

Bortfallsanalys av SCB-undersökningarna HINK och ULF

Jan Qvist

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.

Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

**Bortfallsanalys av
SCB-undersökningarna
HINK och ULF**

Jan Qvist

R&D Report 2000:4

Research - Methods - Development

Analysis of nonresponse in the SCB–surveys ‘HINK’ and ‘ULF’

Från trycket	November 2000
Producent	Statistiska centralbyrån, <i>Statistics Sweden</i> , metodenheten Box 24300, SE-104 51 STOCKHOLM
Utgivare	Lars Lyberg
Förfrågningar	Jan Qvist jan.qvist@scb.se telefon 08-506 943 73

Förord

Bortfallet i olika enkät- och intervjuundersökningar (särskilt individundersökningar) har under en lång följd av år stadigt ökat och har nu nått en oroväckande hög nivå. I de undersökningar som här kommer att gås igenom, inkomstfördelningsundersökningen (HINK) och undersökningen om levnadsförhållanden (ULF), ligger bortfallet numera på mellan 20 och drygt 25 procent. Att upp till var fjärde intervjuperson inte kommer med påverkar naturligtvis undersökningarnas tillförlitlighet negativt.

En rad åtgärder har vidtagits från SCBs sida för att bringa ner bortfallet. Bl.a. har en handbok, "Minska bortfallet" (Japex L. m.fl. 1997), tagits fram som hjälpmedel för att reducera bortfallet vid genomförandet av surveyer. Även metoder för bortfallsjustering har implementerats varigenom man beräkningsmässigt försöker korrigera för bortfallets effekter på resultaten.

SCBs vetenskapliga råd har också engagerats i arbetet med att motverka bortfallet och rådet har initierat föreliggande undersökning. Rapporten är ett led i arbetet med att kartlägga bortfallets grundstruktur.

Rapporten har utarbetats av undertecknad vid BV-avdelningens metodfunktion. Margareta Eriksson (för HINK) och Sven-Erik Johansson (för ULF) har tagit fram datamaterialen och givit värdefull information kring undersökningarna. Sven-Erik Johansson har dessutom varit rådgivande vid analysen av materialet.

Statistiska centralbyrån i augusti 2000

Jan Qvist

Innehåll

	Sida
Abstract (english summary)	1
1. Inledning	3
2. Något om undersökningarna	5
2.1 Basfakta om HINK och ULF	5
2.2 Bortfallets utveckling	6
3. Hur bestämmer vi bortfallsstrukturen	9
3.1 Material	9
3.2 Metod – logitanalys	11
4. Bortfallsstruktur i HINK och ULF	14
4.1 Bortfallet totalt	14
4.2 Bortfallet uppdelat på huvudkategorier	21
4.3 Proportionalitetsantagandets hållbarhet – interaktioner	31
5. Något om bortfallet efter resultatkoder (detaljnivå)	41
6. Sammanfattning och diskussion	45
7. Referenser	48
8. Bilagor	49
- Material	
- Metod	
- Modell med interaktioner	

Abstract

Non-response in different questionnaire and interview surveys (especially surveys of individuals) has grown steadily for many years and has now reached a disturbingly high level. In the surveys that will be examined here – the Swedish Income Distribution Survey (HINK) and Living Conditions Survey (ULF) – non-response now stands at between 20 and 25 per cent, or even a little higher. The fact that up to one in four of those chosen for interview fails to join in naturally has a negative impact on the reliability of these surveys.

The sample survey HINK is conducted on an annual basis. The sample population consists of individuals aged 18 and over. Using **telephone interviews**, data is gathered on employment, occupation, economic situation, housing and the composition of the household. The size of the sample has varied in recent years. For 1996 the sample consisted of about 13 000 individuals, for 1997 about 19 000 and for the latest survey year (1998, with data collection in 1999), the number was again 13 000.

In ULF information is gathered on a continuous basis about the living conditions of the Swedish people. The statistical information relates to the education, economy, work, housing, health and social relations of a sample of individuals aged 16-84. The size of the sample is about 7 500 persons per year. The survey is conducted mainly by means of **face-to-face interviews**.

The sample surveys are analysed in the present report using multivariate regression analysis (SAS, Logistic). The results for the years 1996/97 show that for both HINK and ULF, the odds ratios for all non-response in the surveys rise with age, are higher for unmarried than married individuals, higher for immigrants than for Swedes and much higher for residents of metropolitan regions than residents of other parts of Sweden. Individuals with no earned income also exhibit very high non-response relative to other categories. The variable “earned income” is of course mainly relevant for individuals under the age of 65. The analysis tested one model including the income variable and one model omitting it. This procedure made it possible to establish a more relevant non-response level for the oldest part of the sample, who in fact then exhibited a high level of non-response.

"
"
"
"
"
"
"

A large part of the difference in odds ratios for total non-response is due to differences in non-response because of non-contact with the individuals sampled (this applies to civil status, nationality, region and earned income). This became clear when the analysis was performed with a breakdown of non-response by main categories. In our analysis, non-response because of "refusal to participate" did not turn out to yield very substantial effects. In itself, the relatively small structural impact of refusal to participate is an advantage, in view of the fact that most non-response falls into this category, both in HINK and in ULF. One important exception concerns the variable "age". In both HINK and ULF non-response because of refusal grows considerably as age increases. This is particularly striking in HINK. On the other hand, it may be noted that non-response because of non-contact is more evenly distributed by age. The oldest sample members in fact display very little risk of non-response on account of unavailability, which moderates the combined effect of age on non-response.

In many cases there is a close similarity between the results for HINK and ULF. However, there is one significant difference: in ULF there is less non-response for women than for men.

Specific causes of the high non-response in certain selected high-risk categories (metropolitan region, immigrants, etc.) were studied more closely using detailed result codes. The principle reasons why the level of non-response was particularly high related to the category "non-contact". The causes in detail were "ex-directory telephone number", "no telephone" (Telia) or "moved/gone away". Language difficulties were also important where immigrants were concerned. Regarding refusal to participate, the reason given was often that participation was voluntary.

We have not analysed the time trend for non-response frequencies in this report. A previous study of ULF found that the effects of variables remained fairly stable over time between 1986 and 1997 (however, nationality and civil status were not included in that study). One change, though, was that in the mid-1980s, men and women exhibited about an equal level of non-response, while non-response has now become markedly higher among men than among women.

Finally, we may add that several changes in the demographic structure of society in recent years have contributed to some extent to the deterioration in the propensity to respond. It may be assumed that migration to metropolitan regions, the aging population, the increasing proportion of immigrants and changes in the distribution of incomes have had a significant effect, particularly in the light of the results presented in this report.

"
"
"
"
"
"
"

1. Inledning

Bortfall kan vara en viktig orsak till systematiska fel i urvalsundersökningar ("bortfallsbias"). Skillnader i bortfall mellan olika delgrupper i ett urval indikerar bortfallsbias. År 1998 gjordes en studie av bortfallets struktur i HINK och ULF för perioden 1992-95. I studien undersöktes också bortfallets möjliga inverkan på skattningar av hushållens inkomster (Ribe 1998). Denna inverkan bedömdes vara likartad i de båda undersökningarna.

För att undersöka strukturen i bortfallet i ovannämnda studie med avseende på kön, ålder, region, civilstånd och inkomst användes en multivariat modell (logitanalys med proportionella oddsfaktorer). Man fann att bortfallet var "oberoende" högre för män än för kvinnor, för yngre/äldre än för medelålders, för boende i storstadsområde än för boende i övriga landet, för ej gifta än för gifta och för dem med låg inkomst i förhållande till dem med hög inkomst. Det kan också tilläggas att bortfallsmönstret som helhet uppvisade stora likheter i HINK och ULF.

Den gjorda analysen visade på bortfallsskillnader för ett relativt begränsat antal kända bakgrundsvariabler hämtade ur den officiella registerstatistiken. Analysen skiljer sig därmed från en evalueringsstudie av bortfallet, dvs. en särskild insamling av uppgifter om undersökningsvariablerna i bortfallspopulationen.

I föreliggande studie tas den ovan nämnda analysen som utgångspunkt och utvidgas med några viktiga variabler¹. Med bortfall menas i rapporten genomgående objektsbortfall. Den nya studien gäller också en senare tidsperiod, 1996-97.

Först och främst har den beroende variabeln, bortfallet, delats upp i huvudkategorierna bortfall på grund av att urvalspersonen ej är anträffbar och bortfall på grund av vägran att delta i undersökningen. Denna uppdelning av bortfallet antas ge ökad kunskap om bortfallets natur. Detta kan ha betydelse för vilka åtgärder som är lämpliga att vidta för att minska bortfallet. Den andra utvidgningen består i att vi har delat upp bortfallet efter nationalitet. Under senare år har ett ganska högt bortfall observerats bland just utlandsfödda i flera undersökningar².

Utöver dessa variabelförändringar har vi kontrollerat hur väl proportionalitetsantagandet i regressionsmodellen stämmer. Har t.ex. män högre relativt bortfall³ jämt fördelat i alla indelningar av övriga variabler? Kunskap om eventuella avvikelser från den

¹ Analysen är uppbyggd kring multivariata anpassningar. Vi kommer dock i de olika avsnitten att följa upp hur den enklare univariata ansatsen fungerar.

² I viss begränsad utsträckning beror detta på övertäckning av invandrare i befolkningsregistren (Qvist 1999).

³ Med relativt bortfall menas här den s.k. oddskvoten, vilken beskrivs närmare i metodkapitlet.

proportionella modellen, s.k. interaktioner (samspel), kan ge oss en mer precis bild av bortfallsstrukturen. Resultaten presenteras i kapitel 4. Någon ”bästa” förklarande modell uppställd från början har inte eftersträvats utan olika modelltyper har redovisats.

För vissa utvalda grupper med högt eller lågt bortfall redovisas i kapitel 5 även bortfallet fördelat efter undersökningarnas resultat-koder enligt detaljlista.

Att motverka bortfallet vid insamlingen av statistikuppgifter är naturligtvis eftersträvansvärt och är det grundläggande syftet med denna rapport. Rapporten kan genom kartläggningen av bortfallsstrukturen ge underlag för insatser i denna riktning. Åtgärder för en ökad datafångst vid insamlingen måste dock vidtas på ett sådant sätt att slutresultatet, dvs. det bortfall som trots allt kvarstår, blir så neutralt som möjligt för undersökningsvariablerna. Det bör också nämnas att det finns komplement till ökad datafångst som går ut på att lindra de skevheter som bortfallet eventuellt ger upphov till t.ex. efterstratifiering eller kalibrering.

2. Något om undersökningarna

2.1 Basfakta om HINK och ULF

Hushållens inkomster (HINK/HEK)

Urvalsundersökningen ”hushållens inkomster” genomförs årligen och har som främsta mål att beskriva hushållens disponibla inkomstfördelning. En annan benämning på undersökningen är ”inkomstfördelningsundersökningen”.

Urvalspopulationen består av ett individurval i åldrarna 18 år och däröver. Genom **telefonintervjuer**, som tar 20-30 minuter, insamlas uppgifter om sysselsättning, yrke, ekonomi⁴, boende och hushållets sammansättning. Intervjupersonerna har i förväg fått en skriftlig beskrivning av innehållet i undersökningen. Intervjuaren tar därefter telefonkontakt med intervjupersonen. Fältarbetet tar 10-12 veckor och görs under våren efter referensåret (årgång).

Urvalsstorleken har under senare år varierat. För årgången 1996 var den ca 13 000, för årgången 1997 ca 19 000 och för senaste undersökningsårgång 1998 återigen ca 13 000 (insamling under 1999). Undersökningen har under 1997 varit samordnad med Bostads- och hyresundersökningen (BHU), vilka tillsammans benämns ”Hushållens ekonomi” (HEK).

Undersökningen om levnadsförhållanden (ULF)

I ULF insamlas fortlöpande uppgifter om svenska folkets levnadsförhållanden. Detta har gjorts sedan hösten 1974. Statistikuppgifterna gäller utbildning, ekonomi, arbete, boende, hälsa och sociala relationer för ett urval av personer i åldern 16-84 år. Urvalsstorleken är ungefär 7500 per år.

Undersökningen genomförs huvudsakligen med **besöksintervjuer**. Före intervjun får varje intervjupersonen ett informationsbrev och en folder där undersökningens innehåll presenteras. Överenskommelse om intervju träffas i regel per telefon. I vissa fall skickas ett andra brev till personer som vägrat vid första kontakt.

Intervjun tar ca 1 timma. Telefonintervjuer görs dels om intervjupersonen uttryckligen begär det (fåtal fall), dels i samband med fältarbetets uppföljningsfas. Ca 120 intervjuare deltar i fältarbetet.

⁴ Ett flertal inkomstuppgifter inhämtas från SCB-register.

Datainsamlingen sträcker sig över hela året och resultatet speglar ett årsgenomsnitt för intervjuåret. Undersökningen genomförs visserligen årligen men resultaten sammanställs för två-årsperioder. Den senaste iordningställda dubbelårgången gäller 96/97, vilken också blir föremål för vår bortfallsundersökning.

2.2 Bortfallets utveckling

Som en introduktion visas i diagram 1 en tidsserie av bortfallet och dess orsakskategorier från 1980 fram till och med 1998 (ur senaste Bortfallsbarometer, nr 14). Bortfallet redovisas som en procentuell andel av urvalet för respektive undersökning.

Diagram 1 Bortfallsutvecklingen i HINK (HEK)* och ULF 1980–1998 (ovägt i procent)

— Ej anträffad
—■ Avböjd medverkan
— Övrigt
—● Totalt

Fel! Ogiltig länk.

* referensårgång. ** före 1989 fanns en åldersbegränsning (70/75 år)

— Ej anträffad
—■ Avböjd medverkan
— Övrigt
—● Totalt

Fel! Ogiltig länk.

Källa: Bortfallsbarometer, nr 14

Bortfallet totalt har ökat märkbart under tiden 1980-1998. I HINK från 13 procent 1980 till 26 procent år 1998, alltså en fördubbling, och i ULF från ungefär samma startnivå som i HINK (14 procent) till den något lägre nivån, 23 procent år 1998.

Bortfallet under senare år är sålunda totalt sett något större i HINK än i ULF⁵. Man kan dock notera att ”övrigt” bortfall i HINK har vuxit till under senare tid (3,5 procent 1998, i ULF 1,2 procent samma år). Det övriga bortfallet i HINK består till viss del av ”definierat” bortfall, dvs. i efterhand har vissa individer utelämnats som inte uppfyllt definitionen av HINK-hushållet (gäller i många fall äldre med särskilt boende). Om man bortser från övrigt bortfall och tar hänsyn till olikheter i åldersspannet är HINK och ULF numera ganska lika.

Uppgången av bortfallet totalt tog sin egentliga början strax före eller vid mitten av 80-talet. Vissa år var uppgången särskilt stor. För HINKs del förelåg en kraftig ökning för 1983. Någon avgränsad och tydlig förklaring har varit svår att komma fram till men nedgången året efter brukar tillskrivas metodförändringen från postenkät till telefonintervju. Metropolitdebatten 1986, som rörde en undersökning om levnadsförhållanden, påverkade svarsfrekvensen negativt i den snarlika ULF-undersökningen medan den för HINKs del hade minde betydelse. Avböjd medverkan steg sålunda kraftigt i ULF. Ökningen i HINK några år senare för årgången 1988 kan delvis förklaras av hög beläggning på intervjuarkåren, som innebär stor arbetsbelastning. 1990 var ett toppår i bortfall både i HINK och ULF. Antagligen hade debatten och diskussionen kring folkräkningen detta år betydelse. För HINKs del fanns också en hög beläggning på intervjuarkåren. Den sista riktiga toppen är 1997 för HINKs del. Uppgången berodde nästan uteslutande på ett förhöjt antal ej anträffade. En förklaring kan vara att urvalet var särskilt stort i den kombinerade HEK - ca 19 000. För omkringliggande undersökningsår var urvalet betydligt mindre - ca 13 000 och likaså var bortfallet mindre.

En viktig orsak till det höga bortfallet totalt är det successivt ökande antalet ej anträffade i båda undersökningarna. Andelen ej anträffade ligger dock på en betydligt lägre nivå i ULF än i HINK, vilket kan förklaras av den längre insamlingsperioden i ULF, som ökar möjligheten till kontakt. Bortfallet i ULF har dock ökat under senare år trots förlängd insamlingsperiod.

En förklaring till den växande andelen ej anträffade är antagligen fältintervjuarnas höga arbetsbelastning, något som påverkar möjligheterna till effektiv uppföljning. Framför allt blir detta märkbart då urvalet är särskilt stort⁶. En annan förklaring kan vara den ökade användningen av telefonsvarare och mobiltelefoner (något heltäckande mobiltelefonregister finns inte).

⁵ Undersökningspopulationens åldersspann för 1998 (Bortfallsbarometer nr 14) skiljer sig dock åt något mellan HINK och ULF men detta kan påverka skillnaden i bortfall mellan undersökningarna med högst en procentenhet.

⁶ För de år som vi särskilt kommer att analysera 1996/97 finns två exempel. Det stora urvalet i HINK avseende årgång 1997 och utvidgningen av ULF (särskild undersökning) rörande vissa grupper av invandrare vid årsskiftet 96/97.

ULF har under större delen av den undersökta perioden haft ett något högre bortfall på grund av avböjd medverkan än HINK. Detta kan väl i och för sig vara naturligt eftersom ULF är en relativt omfattande undersökning med besök i hemmet. Under de senaste tre åren har dock en anmärkningsvärd ökning skett, som inte bara gäller för ULF. Från att ha legat oförändrad (HINK) och till och med sjunkit (ULF) under flera år på 90-talet har vägrarandelen nu ökat märkbart. Denna utveckling kan hänga samman med ett försämrat "uppgiftslämnarklimat", vilket anges som förklaring av många undersökningsansvariga vid SCB. I ULF är det särskilt bland de yngre som den ökade vägrartendensen kan utläsas.

3. Hur bestämmer vi bortfallsstrukturen?

Syftet med denna studie är att analysera bortfallet under senare år i de båda undersökningarna (HINK och ULF). Uppgiften har varit att för respektive undersökning finna en bortfallsstruktur (relativ) efter olika variabelindelningar av urvalet. Med kunskap om strukturen och därmed en identifikation av vilka grupper som har särskilt stort bortfall är förhoppningen att ansvariga för undersökningarnas utförande lättare skall kunna vidta åtgärder för en bättre datafångst. Exempelvis, ge grundläggande beslutsunderlag för frågor som ”behövs det fler intervjuare i vissa delar av landet?”, ”är bortfallsskillnaderna så stora efter t.ex. nationalitet att olika kontaktformer bör övervägas”, ”väljs rätt tid för kontakt med hänsyn tagen till t.ex. ålder?”, etc. En fortlöpande jämförelse mellan HINK och ULF i rapporten ger förhoppningsvis också information om skillnaden mellan undersökningarna och eventuellt intervjuformerna.

3.1 Material

Vi har valt att undersöka bortfallet i HINK och ULF för åren 1996/97. För dessa år finns senaste tillgängliga statistik för hela individurvalet med bortfallsuppgifter och med olika bakgrundsvariabler påförda (ålder, kön, region etc.). För ULFs del har intervjuerna gjorts just dessa år medan intervjuerna i HINK utförts under våren efter respektive år (benämns referensår/årgång)⁷.

Den beroende variabeln är:

■ Bortfall

Omfattningen av bortfallet, som är en dikotom variabel, analyseras först i sin helhet. Därefter studeras de båda huvudkategorierna i bortfallet, ”ej anträffad” och ”avböjd medverkan”. En tredje huvudkategori ”övrigt bortfall” (”förhindrad medverkan”) redovisas här inte separat (ingår i bortfall totalt)⁸. Nämnade tre s.k. huvudkategorier, som sammanfattar näralliggande detaljkoder av bortfallet, är en lämplig indelning vid bortfallsanalyser av intervjuundersökningar (se även Japac m.fl. (1997)).

Bakgrundsvariablerna är:

- Kön
- Ålder (grupper)
- Civilstånd (gift/ej gift)
- Nationalitet (invandrare/svensk)
- H-region
- Inkomst (inkomst (lön)/ej inkomst (lön))
- Panel⁹ (panel/ej panel)

⁷ Tidsskillnaden mellan undersökningarna antas i detta sammanhang sakna betydelse.

⁸ Övrigt bortfall belyses dock något i kapitel 5.

⁹ Förekommer i rapporten enbart för ULF. Denna faktorindelning har genomgående liten betydelse.

Beträffande bakgrundsvariablerna menas med nationalitet att den första kategorien innehåller utländska medborgare samt tidigare utländska medborgare som naturaliserats från och med 1967 (benämns här invandrare) och den andra övriga svenska medborgare. H-region är en tätortsindelning (1) Stockholmsregionen, (2) Göteborg/Malmöregionen, (3) Större städer, (4) Södra mellanbygden, (5) Norra tätbygden och (6) Norra glesbygden. Det bör noteras att indelningen geografiskt inte är strikt avgränsad. T.ex. innehåller "Norra glesbygden" några glest befolkade kommuner i södra Sverige. Inkomst är en uppdelning av urvalspersonerna efter om de har någon löneinkomst eller ej. För ytterligare information se bilagan om datamaterialet (kapitel 8).

I det följande redovisas urvals fördelningarna (marginal) efter de variabler som ingår i studien. Vi har valt att visa de båda undersökningarna som helhet (18+ år för HINK och 16-84 år för ULF). De relativa urvalsandelarna och bortfallsandelarna (ovägda) är baserade på urvalsstorlekarna 31 389 individer för HINK och 14 950 individer för ULF.

Tabell 1. Urvalsfördelningar och bortfallsandelar 96/97 i HINK och ULF. Procent

Variabel	HINK		ULF	
	Urvalsandel	Bortfallsandel i kategori	Urvalsandel	Bortfallsandel i kategori
Kvinna	51,8	26,6	50,7	20,7
Man	48,2	25,5	49,3	22,9
Totalt	100	26,1	100	21,8
Ålder 18-24 år / 16-24 år	9,5	24,6	13,8	16,3
Ålder 25-34 år	16,8	23,9	18,0	19,3
Ålder 35-44 år	15,2	26,5	17,1	21,2
Ålder 45-54 år	16,9	27,2	18,1	22,1
Ålder 55-64 år	12,1	26,0	12,8	24,7
Ålder 65- år / 65-84 år	29,5	27,1	20,2	26,0
Totalt	100	26,1	100	21,8
Gift	46,2	23,6	46,4	19,7
Ej gift	53,8	28,3	53,6	23,5
Totalt	100	26,1	100	21,8
Invandrare	11,2	35,2	11,9	31,0
Svensk	88,8	24,9	88,1	20,5
Totalt	100	26,1	100	21,8
H-reg1	19,5	33,7	19,3	29,4
H-reg2	15,4	31,9	14,8	24,0
H-reg3	35,9	22,9	35,6	20,0
H-reg4	18,8	23,8	19,6	19,2
H-reg5	5,4	16,6	5,5	14,8
H-reg6	5,0	20,3	5,2	16,1
Totalt	100	26,1	100	21,8
Ej ink (lön)	40,6	30,1	34,9	28,9
Ink (lön)	59,4	23,4	65,1	17,9
Totalt	100	26,1	100	21,8

3.2 Metod

Uppgiften är att finna en lämplig form för analys av bortfallet efter olika variabelindelningar. I och för sig kunde vi studera en variabel i taget (marginalfördelningar). Om de bakomliggande variablerna (kovariater) vore oberoende av varandra skulle ett sådant studium vara tillräckligt informativt. Vi vill dock se närmare på den ”oberoende” bortfallseffekten för en given variabel, då vi tagit hänsyn till variabelns eventuella samvariation med de andra bakgrundsvariablerna. För att få fram sådana bortfallseffekter används logistisk regressionsanpassning (multivariat). I regressionsmodellen görs ett sedvanligt proportionalitetsantagande, d.v.s. att den relativa bortfallsstrukturen för en given bakgrundsvariabel gäller genomgående oavsett urvalspersonernas övriga variabelgenskaper (gäller oddskvoterna, se nedan). Det predicerade bortfallet som andelar kan sedan fås genom ett multiplikativt förfarande.

Ett annat sätt skulle vara, åtminstone i teorin, att direkt analysera observerat bortfall för alla enskilda kombinationer av bakgrundsvariablerna. Antalet observationer i varje ”cell” blir då emellertid mycket litet och vi får dessutom svårt att tyda resultatet utan ett proportionalitetsantagande. För att undvika denna svaghet använder vi alltså en viss typ av regressionsmodell.

3.2.1 Modell

Bortfallssannolikheten är ett grundläggande mått på bortfallet. Bakgrundsvariablerna \mathbf{x} är formellt knutna till bortfallssannolikheten med hjälp av en logistisk funktion.

$$p = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n}}$$

där p = bortfallssannolikhet i en sammansatt grupp, x_i , $i = 1, 2, \dots, n$.

Vid den logistiska regressionsanpassningen skattas parametrarna i funktionen ovan.

Vi kan också se att logaritmen av det s.k. oddset för sannolikheten $p/(1-p)$ är en linjär funktion, som är grunduttrycket för logitanalysen.

$$\log(p/(1-p)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Vi önskar dock redovisa varje variabels relativa inverkan på bortfallet. Detta åstadkoms genom att själva oddset för en given variabelindelning divideras med oddset för samma variabels referensgrupp (Odds Ratio). Oddskvoterna får då mycket enkla uttryck eftersom oddset självt är exponentiellt med en linjär exponent. I kapitel 4 redovisas resultaten av oddskvotberäkningarna.

$$OR = p/(1-p)/q/(1-q)$$

och

p = bortfallssannolikhet i en given grupp

q = bortfallssannolikhet i referensgrupp

En skattad oddskvot som representerar relativt bortfall i en godtycklig variabelkategori i förhållande till en vald referensgrupp för variabeln, krymper till ett enkelt faktoruttryck efter divisionen.

$$OR = \exp(\hat{\beta}_i)$$

Relationerna gäller om allt annat är lika i modellen. Oddset för en utvald referenskategori för varje variabel är basnivå med definierad oddskvot 1. Om uttrycket ovan är över värdet 1 för en godtycklig variabelkategori betyder det ett högre odds och därmed en högre bortfallssannolikhet i jämförelse med referensindelningen av samma variabel. Ett värde under 1 anger ett lägre odds (eller bortfallssannolikhet) jämfört med referensindelningen¹⁰. I den följande redovisningen anges även om en skattad effekt är signifikant högre eller lägre referensgruppens nivå (*)¹¹. Observera att vi i den följande resultatredovisningen i princip kommenterar skillnaderna mellan oddskvoterna. I de fall då själva bortfallssannolikheterna inte är alltför stora (gäller särskilt vid uppdelning av bortfallet i huvudkategorier) kan oddkvoten numeriskt sägas vara approximativt lika med över- eller underrisk (sannolikhet) för bortfall (p/q)¹².

Relativa bortfallet (oddkvot) för en kombination av två dummyvariabler, A och B, med variabelvärdena 0 eller 1, utgörs i den "proportionella" modellen av produkten mellan de relativa effekterna e^{β_A} och e^{β_B} , benämnda a och b.

Proportionell modell (oddkvoter):

Variabel	A		Proportionell effekt
	0	1	
B 0	1	a	1
1	b	a x b	b
Proportionell effekt	1	a	

¹⁰ Om oddskvoterna är ungefär 1 för kategorierna hos en variabel har variabeln ingen effekt på bortfallet.

¹¹ Signifikans avseende intercepten har inte särskilt markerats i det följande. Intercepten är dock genomgående signifikanta (gäller samtliga analyser).

¹² I följande tabeller i kapitel 4 anges också interceptet. Detta kan användas för beräkning av sannolikhetsnivån.

Det proportionella antagandet (se tabellen ovan) prövas i avsnitt 4.3. Frågan gäller om proportionaliteten $a \times b$ i verkligheten gäller. Om inte, måste kombinationen justeras för samspelet mellan variablerna. Då beräknas, vad vi kallar interaktionseffekter, genom en särskild interaktionsfaktor (c), som kan höja eller sänka den proportionella faktorprodukten. Interaktionsfaktorn är enligt definition inte en oddskvot i egentlig mening utan en justeringsfaktor.

Interaktiv modell(oddskvoter):

Variabel	A		Proportionell effekt	
	0	1		
B	0	1	a	1
	1	b	$a \times b \times c$	b
Proportionell effekt		1	a	

Det bör observeras att vi i rapporten endast prövar parvisa kombinationer. Om interaktionseffekter skulle behövas vid alla typer av kombinationer av kategori-variablerna utökas parameteruppsättningen kraftigt och vi skulle åter vara vid utgångspunkten med fullständig matrisredovisning (jfr. med beskrivning på sid 11).

3.2.2 Urvalsdesignens betydelse för analysen

Analyserna i resultatkapitlen är gjorda utan användning av uppräkningsvikter i båda surveyerna. HINK är dock i viss utsträckning ett stratifierat urval, vilket innebär att representativiteten blir sämre. Detta behöver emellertid inte påverka analysen negativt om vi inkorporerar bakgrundsvariabler associerade med design-variablerna i modellen (Nordberg, 1989). Stratifieringen i HINK¹³ balanseras dock inte tillräckligt av kovariat-uppsättningen. Ett test med uppräkningsvikter för att ta hänsyn till den speciella designen i HINK gav dock ett approximativt likartat resultat (parameterskattningar) med en analys utan uppräkningsvikter. Detta har föranlett oss att låta HINK vara oviktad vid den kommande analysen.

¹³ Företagare (1996), jordbrukare (1996), bostadsrätter (1997) samt äldre, 75- år.

4. Bortfallsstruktur i HINK och ULF

4.1 Bortfallet totalt

Vi har undersökt bortfallet som helhet i undersökningarna HINK(HEK) respektive ULF för 1996-97. Bortfallsstrukturen (oddskvoterna) har beräknats utifrån bakgrundsvariablerna i en multivariat analys. Det bör noteras att det är oddskvoterna som formellt åsyftas vid beskrivningen i det följande¹⁴. För att få bättre jämförbarhet mellan undersökningarna studeras det gemensamma åldersspannet (18-84 år).

I HINK ligger oddskvoten för kvinnor i förhållande till män nära 1 och uttrycker sålunda en minimal skillnad i bortfall mellan könen (tabell 2, huvudmodell). Kvinnor och män är alltså i HINK-undersökningen benägna att delta i ungefär samma utsträckning. I ULF har kvinnor däremot en tydligt lägre bortfallsnivå än män (ca 15 procent lägre¹⁵). Skillnaden mellan HINK och ULF i detta avseende är anmärkningsvärd.

Oddskvoterna varierar en hel del mellan de olika åldersklasserna. Sammanlagt kan sägas att medelålders personer (35-64 år) i HINK har ca 30 procent högre oddskvot än de yngre tillsammans (jämfört med 18-34 år¹⁶). I ULF föreligger också en sådan skillnad. De medelålders ligger dock på en märkbart högre bortfallsnivå än de yngre, med en oddskvot som är ca 60 procent större. I både HINK och ULF har de äldsta en lägre oddskvot än medelålders. Vi har dock tidigare visat att de äldsta har ett observerat högt bortfall (tabell 1). Bakgrunden till den låga nivån i den simultana analysen är att inkomstfaktorn, som i modellen generellt är förhöjd för dem utan löneinkomst, får som återverkning en låg ålderseffekt för de äldsta, eftersom de flesta bland dem saknar löneinkomst (ca 90 procent). Den låga ålderseffekten blir därmed något svårtolkad eller rent av missvisande (jfr tabell 1). Modellen är ändå formellt korrekt men borde inrymt fler relevanta förklaringsvariabler än vad som här står till buds. Ett annat sätt att se på åldersfaktorn är att som i tabell 2 redovisa en förenklad modell med inkomstvariabeln utelämnad ur modellspecifikationen och allt annat lika. Vi kan då lägga märke till att de allra äldsta får betydligt högre oddskvoter. I ULF är bortfallet för denna äldregrupp till och med mycket högt. De äldsta har alltså ett grundläggande högt bortfall, vilket allra helst skulle ha framgått direkt ur modellens ålderseffekt (se diagram 2).

Beträffande civilståndsfaktorn kan observeras att gifta har i runda tal 30 procent lägre oddskvot än ej gifta i både HINK och ULF (nivån något lägre i ULF än i HINK).

¹⁴ Det kan vara bra att veta vid läsningen och värderingen av siffrorna här att i de flesta fall motsvaras oddskvoterna nästan av nivån för relativa risker (sannolikheter) för bortfall (enklare mått, se även sid 12).

¹⁵ I detta fall är bortfallssannolikheten 13 procent lägre för kvinnor än för män (övriga variabler avser referensgrupperna). I tabellerna finns interceptet angivet så att det är möjligt att exakt beräkna de relativa riskerna för varje variabel eller kombinationer av variabler.

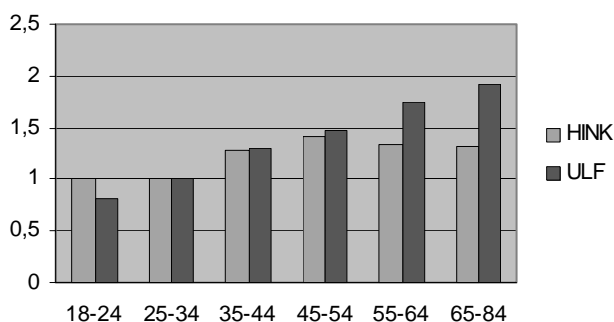
¹⁶ Notera byte av referensgrupp.

Tabell 2. Logitanalys av bortfallet totalt i HINK och ULF 96/97 (oddskvoter)

Bortfall					
Variabel	Odds kvoter – HINK			Odds kvoter - ULF	
	Huvudmodell	Utan inkomst		Huvudmodell	Utan inkomst
Kvinna	0,99	(1,00)		0,85 ***	(0,85***)
Man	1			1	
Ålder 18-24	0,98	(1,00)		0,79 **	(0,82*)
Ålder 25-34	1			1	
Ålder 35-44	1,27 ***	(1,28***)		1,28 ***	(1,30***)
Ålder 45-54	1,39 ***	(1,41***)		1,45 ***	(1,48***)
Ålder 55-64	1,21 ***	(1,34***)		1,53 ***	(1,74***)
Ålder 65-84	0,90 *	(1,32***)		1,15	(1,91***)
Gift	0,75 ***	(0,74***)		0,67 ***	(0,65***)
Ej gift	1			1	
Invandrare	1,47 ***	(1,61***)		1,56 ***	(1,79***)
Svensk	1			1	
Hreg1	1,64 ***	(1,60***)		1,72 ***	(1,67***)
Hreg2	1,48 ***	(1,48***)		1,31 ***	(1,31***)
Hreg3	0,96	(0,96)		1,06	(1,05)
Hreg4	1			1	
Hreg5	0,65 ***	(0,66***)		0,74 **	(0,75**)
Hreg6	0,82 **	(0,83*)		0,81	(0,81*)
Ej ink (lön)	1,63 ***	-		1,94 ***	-
Ink (lön)	1	-		1	-
Panel	-	-		1,02	(1,02)
Ej panel	-	-		1	
Intercept	-1,3905 ⇒ p=0,20			-1,6569 ⇒ p=0,16	

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån. p=bortfallssannolikhet (huvudmodell) baserad enbart på interceptet, dvs gäller i kombinerad referensgrupp. Uppgifter inom () avser modell utan inkomstfaktor

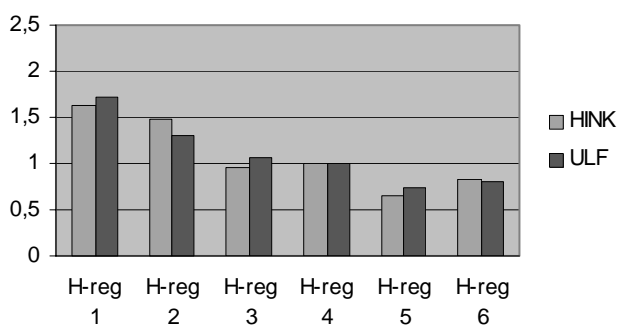
Diagram 2 Oddskvoter för bortfall efter ålder (modell utan inkomst)



Den ”nya” variabeln som avser utländska medborgare samt naturaliserade utländska medborgare (från 1967) visar att denna grupp har ett betydligt högre bortfall än övriga svenska medborgare, i HINK är oddskvoten förhöjd med ca 45 procent och i ULF ca 55 procent¹⁷.

Boende i storstadsområde har högre risk än i övriga Sverige (tabell 2). För stockholmsregionen är nivåvärdena av oddskvoterna så höga som ca 65-70 procent. De regionala skillnaderna är alltså av största betydelse för bortfallssituationen för dessa intervjuundersökningar (se även diagram 3).

Diagram 3 Oddskvoter för bortfall efter H-region (huvudmodell)



Risken för bortfall är också betydligt högre för dem utan löneinkomst i förhållande till dem med inkomst. Oddskvotens nivåvärde för dem utan inkomst i förhållande till inkomsttagare är förhöjt ca 65 procent i HINK och ca 95 procent i ULF¹⁸. Vissa problem uppstår dock vid tolkningen av resultatet för de äldre som tidigare nämnts. Det är ju så få

¹⁷ Om endast utomnordiska medborgare ingick (naturaliserade och nordiska medborgare ingick inte) ändrades bortfallet för denna variabel i begränsad omfattning (i HINK ökade effekten något och i ULF minskade den något).

¹⁸ I ULF, som har den högsta oddskvoten, är även den relativa risken (sannolikheten) för bortfall kraftigt förhöjd, knappt 70 procent (övriga variabler avser referensgrupperna).

som har inkomst men oddskvotnivån gäller även för de allra äldsta (de som har lön har lägre bortfall). Den egentliga bortfallsnivån för den större delen av de äldre framkommer därför istället först genom produkten av oddskvoten för de äldsta och oddskvoten för dem utan inkomst. Detta är något bakvänt men är modellmässigt korrekt.

Bortfallsstrukturen (oddskvoter) är med några undantag ganska likartad mellan HINK och ULF (tabell 2). Detta fann man även i undersökningen för 1992-95 (Ribe 1998). Likheten är kanske något man kan förvänta sig eftersom undersökningarna har många gemensamma drag¹⁹. Allmänt kan dock nämnas en tendens till aningen större variabilitet av bakgrundseffekterna i ULF än i HINK²⁰.

4.1.1 Univariat analys av bortfallet

Den skattade inverkan från en given variabel på bortfallet är i den simultana (multivariata) modellanpassningen rensad från påverkan från sam- eller motvariation bland bakgrundsvariablerna. Vi har alltså fått enskilda eller oberoende effekter från bakgrundsvariablerna, åtminstone med ett teoretiskt betraktelsesätt.

Vid en jämförelse med en univariat skattning kan betydelsen av den simultana tekniken utläsas. Med univariat menas att vi frångår att analysera alla variablerna i ett sammanhang och i stället studerar varje variabel för sig²¹. Notera dock att variablerna kön och ålder genomgående tas med i alla modeller vid den "univariata" skattningen av oddskvoterna (grundläggande standardisering). I en första analys ingår endast kön och ålder (oddskvoter redovisade i tabellerna).

I tabell 3 redovisas resultaten för HINK samtidigt med bägge metoderna (huvudmodell). Motsvarande redovisning för ULF görs i tabell 4. Vid dessa tillfällen låter vi respektive undersöknings åldersspann gälla²².

¹⁹ Stabiliteten i oddskvoterna undersöktes också genom att individer i den lägsta och högsta åldersklassen exkluderades ur analysen. Strukturen bibehölls med ungefärlig överensstämmelse avseende de kvarvarande oddskvotnivåerna.

²⁰ Detta hänger antagligen delvis samman med att ULF-undersökningen baseras på ett mindre antal observationer. ULF innehåller därmed en större slumpmässig variation (konfidensintervall i bilagan).

²¹ Man kan notera att den univariata analysen på sätt och vis motsvaras av marginalfördelningarna vid vanlig bortfallsredovisning.

²² Gäller även i den fortsatta analysen.

Tabell 3. Logitanalys av bortfallet totalt i HINK 96/97 (oddskvoter)

Bortfall		
Variabel	Oddskvoter – multivariat	Oddskvoter – univariat ¹
Kvinna	1,00	1,05
Man	1	1
Ålder 18-24	0,97	1,04
Ålder 25-34	1	1
Ålder 35-44	1,27 ***	1,15 **
Ålder 45-54	1,39 ***	1,19 ***
Ålder 55-64	1,21 ***	1,12 *
Ålder 65-	0,96	1,18 ***
Gift	0,73 ***	0,73 ***
Ej gift	1	1
Invandrare	1,46 ***	1,71 ***
Svensk	1	1
Hreg1	1,61 ***	1,66 ***
Hreg2	1,46 ***	1,52 ***
Hreg3	0,95	0,96
Hreg4	1	1
Hreg5	0,65 ***	0,65 ***
Hreg6	0,80 **	0,82 **
Ej ink (lön)	1,64 ***	1,73 ***
Ink (lön)	1	1
Intercept	-1,3806 ⇒ p=0,20	-1,1842 ⇒ p=0,23

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån

p=bortfallssannolikhet baserad enbart på interceptet, dvs gäller i kombinerad referensgrupp.

1) I de univariata analyserna ingår alltid ålder och kön. Oddskvoterna för ålder och kön ovan gäller en analys med enbart dessa båda faktorer (interceptet från analysen med denna faktoruppsättning).

Tabell 4. Logitanalys av bortfallet totalt i ULF 96/97 (oddskvoter)

Bortfall				
Variabel	Odds kvoter – multivariat		Odds kvoter – univariat ¹	
Kvinna	0,85	***	0,86	***
Man	1		1	
Ålder 16-24	0,71	***	0,82	**
Ålder 25-34	1		1	
Ålder 35-44	1,28	***	1,13	
Ålder 45-54	1,45	***	1,19	*
Ålder 55-64	1,54	***	1,38	***
Ålder 65-84	1,17	*	1,49	***
Gift	0,67	***	0,65	***
Ej gift	1		1	
Invandrare	1,56	***	1,90	***
Svensk	1		1	
Hreg1	1,70	***	1,81	***
Hreg2	1,31	***	1,36	***
Hreg3	1,06		1,08	
Hreg4	1		1	
Hreg5	0,73	**	0,73	**
Hreg6	0,79	*	0,80	*
Ej ink (lön)	1,90	***	2,07	***
Ink (lön)	1		1	
Panel	1,03		1,02	
Ej panel	1		1	
Intercept	-1,6505 ⇒ p=0,16		-1,3634 ⇒ p=0,20	

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån

p=bortfallssannolikhet baserad enbart på interceptet, dvs gäller i kombinerad referensgrupp.

1) I de univariata analyserna ingår alltid ålder och kön. Odds kvoterna för ålder och kön ovan gäller en analys med enbart dessa båda faktorer (interceptet från analysen med denna faktoruppsättning).

Skillnaderna i HINK mellan simultan och univariat analys är inte särskilt stora (tabell 3). För variablerna kön, civilstånd, H-region och inkomst föreligger endast små skillnader. Däremot är ålderseffekten mera uttalad vid simultan än vid univariat analys, med höga risker för medelålders personer. Medelålders har alltså en större benägenhet att inte delta i undersökningen än vad som vanligtvis kan utläsas utifrån vanliga bortfallsandelar efter ålder. Bortfallet bland invandrare ändras också ganska kraftigt. I detta fall är bortfalls-effekten lägre vid simultan analys. Själva invandrareffekten tunnare alltså ut och minskar då man tar hänsyn till andra faktorer.

Beträffande ULF ser ovanstående skillnader mellan simultant/univariat nästan likadana ut som för HINK. Ålderseffekterna och invandrareffekten uppvisar även i ULF störst skillnader mellan analysformerna. I ULF är oddskvoten högre för de allra äldsta än i den simultana analysen. Detta beror på att inkomsteffekten här inte påverkar resultatet²³.

²³ Det bör påpekas att beräkningen av signifikansområdena är något osäker då det gäller univariat analys (se även metodavsnitt 8.3).

4.2 Bortfallet uppdelat på huvudkategorier

Det är viktigt att kunna skilja på olika typer av orsaker för att få ökad förståelse för uppkomsten av bortfall. Med kunskap om bakgrunden kan man bättre vidta adekvata motåtgärder. Vanligen delas bortfallet upp efter de två huvudkategorierna ”ej anträffade” och ”avböjd medverkan”. I både HINK och ULF är avböjd medverkan den största kategorin.

I HINK består det sammanlagda bortfallet till övervägande delen av avböjd medverkan/vägrare, ca 48 procent, och till nästan en lika stor del, ca 40 procent, av ej anträffade samt en kanske inte helt obetydlig restgrupp (ca 12 procent), som bl.a. innehåller ”definerat” bortfall. I ULF är förhållandena något annorlunda. Så mycket som 66 procent av bortfallet totalt utgörs av vägrare och 25 procent av ej anträffade, samt en mindre del (9 procent) av sjuka mm²⁴.

Det kan noteras att den ”bias” som uppstår på grund av strukturskillnader i bortfallet (varierande oddskvoter) är i HINK och ULF tämligen överensstämmande, vilket framgån av föregående avsnitt (tabell 2). I de följande avsnitten redovisas bortfallsmönstret efter kategori. I princip är det mest fördelaktigt om skillnaderna mellan oddskvoterna är som minst för den bortfallskomponent som väger tyngst. För de bägge undersökningarna och särskilt för ULF är det alltså mest fördelaktigt om oddskvoterna för vägrare varierade så lite som möjligt.

4.2.1 Bortfallsmönster för huvudkategorierna i HINK

Resultaten i det följande bygger på den multivariata analysen. I tabell 5 redovisas oddskvoterna i HINK för bortfall på grund av ej anträffade respektive på grund av vägran.

Ej anträffade

Bortfallet på grund av ej anträffade varierar ganska kraftigt för de olika indelningarna av bakgrundsvariablerna. Nivåskillnaderna i oddskvoter är här större än för bortfallet som helhet. Bortfallet på grund av att man ej anträffat urvalspersonerna är betydligt lägre för kvinnor än för män. Detta resultat, nästan 20 procent lägre bortfall, avviker från redovisningen för hela bortfallet, där skillnaden mellan män och kvinnor var marginell i HINK.

Bortfallsnivåerna för olika åldersgrupper är tämligen likartade före 55 års ålder. Skillnaderna är små i förhållande till referensgruppen (25-34 år). I högre åldrar är oddskvoterna betydligt lägre och därmed representerar de ett mindre bortfall. I tabell 5 finns även en modell utan inkomstvariabel. Oddskvoten för de äldsta är även i detta fall betydligt under 1 och därmed understryks att bortfallet på grund av ej anträffad är mycket lågt för de äldsta. Det är alltså i grunden lättare att komma i kontakt med äldre än yngre urvals-

²⁴ Kategoriandelarna avser hela undersökningspopulationerna HINK och ULF.

Tabell 5. Logitanalys av bortfallskategorier i HINK (oddskvoter) 1996-97

Variabler	Bortfall - oddskvot			
	Ej anträffade		Vägrare	
	Huvudmodell	Modell utan inkomst	I hela urvalet	Utan ej anträffade
Kvinna	0,83 ***	(0,83***)	1,14 ***	1,11 **
Man	1		1	1
Ålder 18-24	1,02	(1,04)	0,97	0,98
Ålder 25-34	1		1	1
Ålder 35-44	1,04	(1,05)	1,52 ***	1,54 ***
Ålder 45-54	0,99	(1,01)	1,78 ***	1,79 ***
Ålder 55-64	0,61 ***	(0,69***)	1,97 ***	1,86 ***
Ålder 65-	0,31 ***	(0,48***)	1,46 ***	1,27 ***
Gift	0,51 ***	(0,50***)	1,21 ***	1,13 **
Ej gift	1		1	1
Invandrare	1,82 ***	(2,02***)	0,74 ***	0,82 **
Svensk	1		1	1
Hreg1	2,28 ***	(2,21***)	1,13 *	1,26 ***
Hreg2	1,82 ***	(1,83***)	1,14 *	1,22 ***
Hreg3	1,05	(1,06)	0,91	0,91
Hreg4	1		1	1
Hreg5	0,54 ***	(0,54***)	0,72 ***	0,69 ***
Hreg6	0,95	(0,97)	0,77 **	0,76 **
Ej inomst (lön)	1,71 ***	-	1,12 *	1,22 ***
Ink (lön)	1	-	1	1
Intercept	-2,0652 ⇒ p=0,11		-2,4983⇒ p=0,08	-2,3488⇒ p=0,09

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån. p=bortfallssannolikhet baserad enbart på interceptet, dvs gäller i kombinerad referensgrupp. Uppgifter inom () avser modell utan inkomstfaktor.

personer. Några förklaringar är att de är hemma i större utsträckning, har liten omflyttning och inte så ofta valt hemligt telefonnummer/mobiltelefon.

Det är nästan dubbelt så svårt att komma i kontakt med ej gifta som gifta²⁵. Sannolikheten att någon är hemma vid telefonkontakt är troligtvis större för gifta par än för ogifta som många gånger är ensamstående.

Även invandrare är svårare att komma i kontakt med. Bortfallet är förhöjt med ca 80 procent i förhållande till urvalspersoner som är svenska. Vid en närmare granskning av de mera detaljerade orsakerna visar det sig att invandrare i större utsträckning saknar telefon (Telia) eller har hemligt telefonnummer (se tabell 14).

Den största skillnaden mellan olika grupper gäller var i landet man bor. Man kan särskilt notera att oddskvoten är drygt 4 gånger så stor i Stockholm (H-reg 1) som i norra tätbygden (H-reg 5). Risken att misslyckas med att få kontakt med en urvalsperson i Stockholm är alltså betydligt större jämfört med norra tätbygden (den relativa risken är i själva verket 3,5 gånger så stor²⁶). Slutligen är oddskvoten för att ej anträffa en urvalsperson utan löneinkomst 70 procent högre än för dem med löneinkomst.

Avböjd medverkan (=vägran)

Vid beräkning av oddskvoterna för vägran har vi använt oss av två olika ansatser (tabell 5). I det ena fallet utgör hela urvalet baspopulation och i det andra endast de som anträffats²⁷. Skillnaderna mellan ansatserna är dock ganska små men inte helt obetydliga.

Det finns flera strukturella skillnader i oddskvoternas variation mellan de två bortfalls-kategorierna. Först kan vi lägga märke till att kvinnor i HINK i större utsträckning än män vägrar delta, vilket var det motsatta för komponenten ”ej anträffade”.

Det mest iögonenfallande är att personer i medelåldern och särskilt övre medelåldern är mera ovilliga att delta. För en person som är 55-64 år är oddskvoten nästan dubbelt så stor att avböja en intervju som för en ung person i åldern upp till 35 år. Åldersmönstret för ”avböjd medverkan” är sålunda helt motsatt det för ”ej anträffad”.

Vidare kan nämnas att personer i norra tätbygden har en relativt liten risk för vägran (samma förhållande som för ”ej anträffade”).

²⁵ Formellt är ej gifta referensgrupp. För en dikotom variabel kan man dock enkelt växla referensgrupp (invertering).

²⁶ Den relativa risken har beräknats då alla andra variabler representeras av referensgruppsnivåerna. Vi kan också observera att grundsannolikheten för denna bortfallsdel är relativt liten ($p=0,11$), vilket innebär att överensstämmelsen mellan relativa risker och motsvarande oddskvoter i allmänhet blir större.

²⁷ Endast de som anträffats är en mera adekvat studiepopulation

Det finns två andra företeelser som man särskilt kan lägga märke till. Den första är att för variabeln nationalitet går oddskvoterna i olika riktning för avböjd medverkan och ej anträffade. Trots att språksvårigheter många gånger kan lägga hinder i vägen låter sig invandrare intervjuas i större utsträckning än svenskar. Det måste dock påpekas att rutinerna för bortfallskodning här kan spela en viss roll, vilket kan leda till underskattning av oddskvoten. Eventuellt finns det också ett samband mellan ”språksvårigheter” (ingår i ”övrigt bortfall”) och vägran. Om språksvårigheter helt och hållet förs till vägrargruppen uppstår istället en överrisk för invandrare (oddskvot=1,2). Ovanstående resultat, att invandrare har större benägenhet att intervjuas, är därför i viss mån osäkert. Det stora bortfallet totalt bland invandrare hänger dock i första hand samman med att intervjuarna inte får kontakt med urvalspersonerna.

Den andra företeelsen är att inkomstvariabeln inte har någon större effekt då det gäller vägran. På grund av detta finns det heller ingen anledning att redovisa modellen utan inkomstvariabeln.

Slutligen kan sägas att oddskvoterna för vägran i allmänhet varierar mindre än motsvarande oddskvoter för ej anträffade.

4.2.2 Bortfallsmönster för huvudkategorierna i ULF

I tabell 6 redovisas för ULF oddskvoterna (multivariat) för bortfall på grund av ej anträffade respektive på grund av vägran.

Ej anträffade

Bortfallet på grund av ej anträffade varierar mycket kraftigt för de flesta bakgrundsvariabler. Nivåskillnaderna är här (liksom i HINK) betydligt större än vad som är fallet för bortfallet utan uppdelning i bortfallsorsak. Dessutom gäller att grundstrukturen för ej anträffade är ungefär densamma här som i HINK. Detta är naturligt eftersom undersökningarna så långt i genomförandet är ganska lika.

Bortfall på grund av att urvalspersonen ej anträffas är i ULF mycket lägre för kvinnor än för män, ungefär 30 procent lägre (lägre än i HINK). Även åldersfaktorn har betydelse. För HINK nämndes att det var de allra äldsta som var relativt lätta att komma i kontakt med. I ULF gäller samma sak även i modellen när inkomstvariabeln är utelämnad (se tabell 6). I ULF har även de allra yngsta ett litet bortfall på grund av oanträffbarhet.

Bortfallet på grund av oanträffbarhet är extremt högt för ej gifta i förhållande till gifta samt för invandrare i förhållande till svenskar. Oddskvoterna är i bägge fallen ca 3 gånger högre än respektive referensgruppsnivå²⁸.

²⁸ Formellt är ej gifta referensgrupp. För en dikotom variabel kan man dock enkelt växla referensgrupp (invertering).

Tabell 6. Logitanalys av bortfallskategorier i ULF (oddskvoter) 1996-97

Bortfall - oddskvot					
Variabler	Ej anträffade			Vägrare	
				I hela urvalet	Utan anträffade. ej
Kvinna	0,70	***	(0,69***)	0,94	0,91 *
Man	1			1	1
Ålder 16-24	0,63	***	(0,72**)	0,82 *	0,79 *
Ålder 25-34	1			1	1
Ålder 35-44	1,14		(1,20)	1,28 **	1,31 **
Ålder 45-54	1,30	*	(1,37**)	1,41 ***	1,45 ***
Ålder 55-64	1,04		(1,31*)	1,59 ***	1,61 ***
Ålder 65-84	0,32	***	(0,67**)	1,37 ***	1,26 *
Gift	0,33	***	(0,31***)	0,97	0,90 *
Ej gift	1			1	1
Invandrare	2,94	***	(3,58***)	0,90	1,02
Svensk	1			1	1
Hreg1	3,18	***	(2,97***)	1,30 ***	1,41 ***
Hreg2	2,31	***	(2,29***)	1,12	1,18 *
Hreg3	1,34	*	(1,33*)	1,02	1,03
Hreg4	1			1	1
Hreg5	0,54	*	(0,55*)	0,73 *	0,72 **
Hreg6	0,76		(0,75)	0,80	0,79
Ej ink (lön)	2,77	***	-	1,21 **	1,33 ***
Ink (lön)	1		-	1	1
Panel	0,88		(0,88)	1,10	1,08
Ej panel	1			1	1
Intercept	-3,2353		⇒ p=0,04	-2,1115⇒0,11	-2,0489⇒0,11

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån. p=bortfallssannolikhet baserad enbart på interceptet, dvs gäller i kombinerad referensgrupp. Uppgifter inom () avser modell utan inkomstfaktor

Bostadsort och inkomst är faktorer som har stort inflytande på bortfallet. Boende i storstadsregionerna har enligt oddskvoterna 4-6 gånger så stort bortfall som i norra tätbygden (i HINK var det 3-4 gånger). Personer utan löneinkomst har i ULF nästan 3 gånger så stort bortfall som för dem med inkomst.

Vi kan också lägga märke till att panelgruppen har ett lägre bortfall (nära signifikans). Anträffbarheten är alltså större bland dem som en gång tidigare har intervjuats (8 år tidigare)²⁹.

Avböjd medverkan

För vägran varierar oddskvoterna betydligt mindre. Även i ULF- analysen har vi använt oss av två olika ansatser³⁰. Skillnaderna i oddskvoterna är också här minimala mellan de båda fallen men variationen är något mera uttalad efter korrigering av urvalet.

För vissa av variablerna är mönstret nästan helt jämnt mellan variabelindelningarna. Detta gäller civilstånd och nationalitet.

Oddskvoterna visar att skillnaden mellan mäns och kvinnors benägenhet att medverka i undersökningen är ganska liten. Det man kan notera är att kvinnor i ULF har något mindre bortfall än män, vilket inte gällde HINK, som hade en motsatt tendens. Kanske speglar bortfallsdifferensen kvinnors och mäns olika intresseområden och därmed villighet att delta. Den stora könsskillnaden för bortfallet totalt i ULF beror alltså på anträffbarheten.

Vägran att delta i undersökningen ökar med åldern. Största bortfallet på grund av vägran föreligger hos medelålders personer, 55-64 år (jfr. även HINK). De äldsta har också relativt hög oddskvot. I ULF pressas denna ner (mot 1) i viss utsträckning av inkomstfaktorn (större än i HINK)³¹.

Bostadsort och ekonomi har en viss betydelse för benägenheten att delta i undersökningen. Lägst intresse föreligger i Stockholm och störst i norra tätbygden. Dessutom gäller att de utan löneinkomst har en högre bortfallsrisk. Bägge dessa faktorer har en motsvarighet i HINK.

²⁹ Notera att panelurvalet i viss begränsad utsträckning även innehåller nya urvalspersoner som definieras som panel (särkilt i yngsta åldersklassen). Denna omständighet påverkar dock inte effekterna i någon större utsträckning. Det kan tilläggas att om panelfaktorn exkluderades förblev nivåerna för övriga oddskvoter desamma i samtliga analyser av ULF i rapporten. Se även bilagan.

³⁰ I det ena fallet utgör hela urvalet baspopulation och i det andra endast de som anträffats.

³¹ I en modell utan inkomstvariabel höjdes värdet på denna variabel till nästan 1,6 men når inte upp till värdet för åldersklassen 55-64 år (oavsett modelltyp).

4.2.3 Sammanfattning för huvudkategorierna i HINK – ULF

Sammanfattningsvis kan sägas att skillnaderna mellan variabelindelningarna är stora för ”ej anträffade” och att mönstret i grunden är likartat mellan HINK och ULF. Variationen är dock något större i ULF.

Bortfallet som en följd av ”avböjd medverkan” har visat sig vara relativt oberoende av bakgrundsfaktorerna, d.v.s. en tendens till att vara tämligen generellt (lika över här redovisade bakgrundsvariabler), vilket i och för sig är positivt beträffande bortfallsbias. De skillnader som trots allt finns får dock stor betydelse på grund av att bortfallet i denna huvudkategori numera är mycket stort.

Efter genomgången i avsnitten ovan kan det vara svårt att hålla isär alla bortfalls-skillnader. Vi tänker inte upprepa dem här men vill visa en ”sämsta” kombination ur bortfallssynpunkt för kategorin ”ej anträffade”. Den gäller män, 45-54 år gamla, ej gifta, invandrare, boende i Stockholm och utan löneinkomst. I ULF blir bortfallssannolikheten enligt modellen för denna speciella grupp 57 procent och i HINK 47 procent. De observerade sannolikheterna blir mycket osäkra eftersom angivna gruppindelning endast innehåller 5 personer i ULF och 9 personer i HINK (motsvarande sannolikheter är $4/5=0,80$ respektive $5/9=0,56$).

4.2.4 Univariat analys av bortfall efter kategori

Även det på kategorier uppdelade bortfallet redovisas univariat (tabell 7 och 8). Skillnaderna mellan analysformerna framgår för HINK:s del vid jämförelse mellan tabellerna 5 och 7 och för ULF:s del tabellerna 6 och 8.

Åldersfaktorn för ”ej anträffade” skiljer sig något åt mellan analysformerna. I den simultana analysen är risken att ej anträffas högre i medelåldern än i univariat analys³². Då hänsyn tas till samtliga kovariater i den simultana analysen ger alltså åldersfaktorn förhöjd effekt med åldern med undantag för de allra äldsta, vilket gäller både HINK och ULF.

Beträffande ”ej anträffade” föreligger den största skillnaden mellan analysformerna för nationalitetsvariabeln. Nivån är betydligt högre för invandrare i den univariata analysen. Detta gäller i både HINK och ULF men främst i ULF. Värdet av oddskvoten i ULF med univariat analys är 4 gånger så stort som referensgruppen (tabell 8)³³. Det förhöjda värdet vid univariat analys beror i denna analysform på att flera med nationaliteten samverkande faktorer ligger bakom invandrarnas höga risk att ej anträffas (bostadsort, inkomst, m.m.).

Beträffande ”ej anträffade” kan för övrigt sägas att bortfallsnivån i ULF för H-region 1 (Stockholm) och inkomst har märkbart högre nivå i den univariata analysen.

För ”vägrare” är skillnaderna mellan analysformerna betydligt mindre både i HINK och ULF. Man dock kan notera beträffande HINK att den signifikanta underrisken för invandrare inte kvarstår vid univariat analys. Den stora benägenheten att vilja medverka som vi tidigare sett skymms i den univariata analysen av andra variabler.

³² Modell baserad enbart på kön och ålder.

³³ Detta innebär att den beräknade bortfallsandelen i ULF är ungefär 20 procent bland invandrare och 6 procent bland svenskar i den univariata modellen (för övriga variabler gäller referensgrupperna, dvs män i åldersklassen 25-34 år).

Tabell 7. Logitanalys av bortfallskategorier 96/97 i HINK (oddskvoter). Univariat ¹

Variabel	Bortfall - oddskvot			
	Ej anträffad		Vägrare (urval utan ej anträffade)	
Kvinna	0,87	***	1,12	**
Man	1		1	
Ålder 18-24	1,14	*	0,95	
Ålder 25-34	1		1	
Ålder 35-44	0,87	*	1,57	***
Ålder 45-54	0,75	***	1,86	***
Ålder 55-64	0,50	***	2,02	***
Ålder 65-	0,35	***	1,52	***
Gift	0,51	***	1,11	**
Ej gift	1		1	
Invandrare	2,16	***	0,91	
Svensk	1		1	
Hreg1	2,42	***	1,23	***
Hreg2	1,96	***	1,21	**
Hreg3	1,07		0,91	
Hreg4	1		1	
Hreg5	0,53	***	0,69	***
Hreg6	0,98		0,76	**
Ej ink (lön)	1,88	***	1,17	**
Ink (lön)	1		1	
Intercept	-1,6951 ⇒ p=0,16		-2,2904 ⇒ p=0,09	

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån.

p= bortfallssannolikhet baserad enbart på interceptet (modell – ålder, kön (referensgrupper)).

1) I de univariata analyserna ingår alltid ålder och kön. Oddskvoterna för ålder och kön ovan gäller en analys med enbart dessa båda faktorer (interceptet från analysen med denna faktoruppsättning).

Tabell 8. Logitanalys av bortfallskategorier 96/97 i ULF (oddskvoter). Univariat ¹

Bortfall - oddskvot			
Variabel	Ej anträffad		Vägrare (urval utan ej anträffade)
Kvinna	0,70	***	0,92
Man	1		1
Ålder 16-24	0,85		0,83 *
Ålder 25-34	1		1
Ålder 35-44	0,88		1,25 **
Ålder 45-54	0,84		1,37 ***
Ålder 55-64	0,79		1,58 ***
Ålder 65-84	0,38	***	1,47 ***
Gift	0,33	***	0,89 *
Ej gift	1		1
Invandrare	4,00	***	1,15
Svensk	1		1
Hreg1	3,72	***	1,42 ***
Hreg2	2,57	***	1,18 *
Hreg3	1,41	**	1,03
Hreg4	1		1
Hreg5	0,52	*	0,72 **
Hreg6	0,77		0,79
Ej ink (lön)	3,47	***	1,34 ***
Ink (lön)	1		1
Panel	0,86	*	1,09
Ej panel	1		1
Intercept	-2,4471 ⇒ p=0,08		-1,8954 ⇒ p=0,13

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån.

p= bortfallssannolikhet baserad enbart på interceptet (modell – ålder, kön (referensgrupper)).

1) I de univariata analyserna ingår alltid ålder och kön. Oddskvoterna för ålder och kön ovan gäller en analys med enbart dessa båda faktorer (interceptet från analysen med denna faktoruppsättning).

4.3 Proportionalitetsantagandets hållbarhet - interaktioner

Hittills har vi förutsatt att faktorerna bakom oddsen/oddskvoterna varit proportionella. Detta antagande kan naturligtvis ha brister. Vi har därför undersökt om de gjorda analyserna innehåller några interaktioner (samspel mellan faktorer). Vi har valt att se på interaktionerna med uppdelning av bortfallet på huvudkategorier.

Som exempel på en interaktion kan nämnas att invandrare enligt den proportionella modellen har ca 50 procent större oddskvot (totalt, sid. 15), vilken gäller oförändrad oavsett övriga faktorindelningar. En interaktion kan innebära att oddskvoten varierar t.ex. antar olika nivåvärden i olika åldrar.

För att bestämma om en interaktion föreligger eller inte har vi valt kriteriet att interaktionseffekten "c", definierad på sid. 13, är signifikant skild från 1 på 1-procentnivån³⁴. Tidigare nämnda faktorer a och b finns med som proportionella faktorer³⁵.

Vid fastställandet av interaktionerna har vi testat signifikansen hos parvisa faktorkombinationer i separata test. De kombinationer som uppvisat signifikanta samspel låter vi sedan ingå i en modell med samtliga interaktioner. Förhoppningsvis kvarstår signifikansnivåerna för interaktionerna i det simultana sammanhanget (åtminstone på 5-procentnivån) men det är, för enkelhets skull, inget absolut krav för avgörandet om en interaktion föreligger³⁶.

De proportionella (generella) effekterna kan i den sammanslagna modellen naturligtvis påverkas mer eller mindre. Om det sker en stor förändring påpekas detta i texten (inte så ofta förekommande). I bilagan (Modell med interaktionstermer, tabellerna B3 och B4) redovisas även de proportionella värdena när alla parvisa samspel ingår i modellen. Det är dessa värden som åsyftas då de proportionella effekterna kommenteras i det följande om inte annat anges³⁷.

I framställningen nedan jämförs respektive huvudkategori av bortfallet, dels för HINK, dels för ULF³⁸.

³⁴ Kriteriet är ganska högt ställt. Vi diskuterar senare även andra signifikansnivåer.

³⁵ Vi tillåter oss att beteckna dessa faktorer i det följande som generella faktorer.

³⁶ Om någon kombination blir insignifikant i den sammanslagna modellen finns nivåvärdet trots detta redovisat i tabellerna.

³⁷ Överensstämmer ungefär med proportionella effekter i föregående avsnitt.

³⁸ Vi undersöker här samma populationer (alla åldrar) för HINK och ULF som i tidigare avsnitt med uppdelning av bortfallet på huvudkategori.

4.3.1 Bortfall på grund av ej anträffade - interaktioner

Då det gäller bortfall på grund av att urvalspersonen ej anträffats visar tabellerna 9 och 10 att det saknas interaktioner i modellen för de flesta variabelkombinationer (dvs. c inte statistiskt skilt från 1). Detta innebär att den proportionella modellen för de flesta variabler fungerar tillfredsställande.

Redovisningen av förekommande interaktioner i den följande texten är ordnad efter faktorerna angivna i tabellernas kolumnhuvuden. Åldersfaktorn särredovisas inte som enskild variabel utan den ingår i förekommande fall i redovisningen under de övriga faktorrubrikerna. Allmänt kan sägas att åldersfaktorn ofta är involverad i interaktioner både i HINK och ULF.

4.3.1.1 Samspelskombinationer i HINK – ej anträffade

Könsfaktorn.

Kvinnor har en 17-procentigt lägre oddskvot än män i den tidigare redovisade proportionella modellen. Den generella nivån minskar dock något då vi infört samtliga ”signifikanta” interaktioner (oddskvot =0,89, se tabell B3).

Av tabell 9 framgår att det föreligger en interaktion mellan könsfaktorn och *åldersfaktorn* avseende pensionsåldrarna (som också har låg generell oddskvot). Kvinnor har utöver den redan låga kombinationsnivån (p.g.a. av både köns- och åldersfaktor) ytterligare minskning. I denna åldersklass sjunker nivån enligt tabellen med 36 procent. I förhållande till män har kvinnor sammantaget drygt 40 procent lägre oddskvotsnivå (0,89x0,64) i den högsta åldersklassen³⁹. I detta sammanhang kan det vara värt att uppmärksamma att det breda åldersintervallet 65+ år kan delas upp i mindre intervall och då kan vi se att för de riktigt gamla, i åldrar över 85 år, försvinner denna interaktionseffekt.

Civilståndsfaktorn

Gifta har generellt en nästan 50-procentigt, lägre oddskvot än ej gifta i ursprungsmodellen (ca 40 procent lägre i modellen med interaktioner).

En interaktion mellan civilståndsfaktorn och *inkomstfaktorn* visar att gifta i förhållande till ogifta har en ytterligare förminskad nivå (35 procent) bland dem utan löneinkomst (proportionellt hög oddskvot).

³⁹ Notera att en interaktion har 3 faktorer i produkten enligt sid. 13 (a x b x c). Eftersom jämförelsen mellan män och kvinnor ovan görs för ett enda åldersintervall för båda könen är det tillräckligt med två faktorer i produkten (a x c).

Tabell 9. Ej anträffade. Interaktioner ("c") i HINK 96/97 (utvalda på 1-procentnivån)

	Kvinna	Gift	Invand-rare	Ej lön	H-reg 1	H-reg 2	H-reg 3	H-reg 5	H-reg 6
Gift	1	-	-	-	-	-	-	-	-
Invandrare	1	1	-	-	-	-	-	-	-
Ej lön	1	0,65**	1	-	-	-	-	-	-
H-reg1	1	1	1	1	-	-	-	-	-
H-reg2	1	1	1	1	-	-	-	-	-
H-reg3	1	1	1	1	-	-	-	-	-
H-reg5	1	1	1	1	-	-	-	-	-
H-reg6	1	1	1	1	-	-	-	-	-
Ålder									
18-24	1	1	1	0,64**	1	1	1	1	1
35-44	1	1	1	1	1	1	1	1	1
45-54	1	1	1	1,42**	1	1	1	1	1
55-64	1	1	1	1	1	1	1	1	1
65-	0,64**	1	1,84**	1	1	1	1	1	2,10**

* 5-procentnivån, ** 1-procentnivån

Tabell 10. Ej anträffade. Interaktioner ("c") i ULF 96/97 (utvalda på 1-procentnivån)

	Kvinna	Gift	Invand-rare	Ej lön	H-reg 1	H-reg 2	H-reg 3	H-reg 5	H-reg 6
Gift	1	-	-	-	-	-	-	-	-
Invandrare	1	1	-	-	-	-	-	-	-
Ej lön	1	1	1	-	-	-	-	-	-
H-reg1	1,35	1	1	1	-	-	-	-	-
H-reg2	0,54**	1	1	1	-	-	-	-	-
H-reg3	1	1	1	1	-	-	-	-	-
H-reg5	1	1	1	1	-	-	-	-	-
H-reg6	1	1	1	1	-	-	-	-	-
Ålder									
16-24	1	8,09**	1	0,42**	1	1	1	1	1
35-44	1	1	1	1	1	1	1	1	1
45-54	1	1	1	1	1	1	1	1	1
55-64	1	1	1	1	1	1	1	1	1
65-84	1	1,92*	3,48**	1	1	1	1	1	1

* 5-procentnivån, ** 1-procentnivån

Nationalitetsfaktorn

Invandrare har i allmänhet en hög risk för bortfall. I åldersklassen 65+ år, med i och för sig låg generell ålderseffekt, är oddskvoten för invandrare nästan fördubblad genom interaktionseffekten. Detta pekar på att det relativt sett är särskilt svårt att komma i kontakt med de äldsta invandrarna.

Inkomstfaktorn

I unga år (**16-24 år**) förminskas den höga allmänna bortfallsnivån (fördubblad nivå, tabell B3) bland dem utan löneinkomst genom en interaktionseffekt. Detta innebär för unga utan löneinkomst att inkomstfaktorn tappar mycket av sin betydelse ($2*0,64$). Detta är i och för sig naturligt eftersom inkomsten för unga är oregelbunden under studietiden och inte direkt indikerar anträffbarheten.

Den proportionella modellen visade att bortfallet avseende ej anträffade är relativt högt i **åldern 35-44 år**. Genom en interaktion mellan inkomstfaktorn och åldersfaktorn för nämnda åldersgrupp kan vi se att den relativt höga nivån i medelåldern bland annat beror på särskilt högt bortfall bland dem som saknar inkomst. Interaktionseffekten innebär i åldersgruppen nämligen en förhöjning för dem utan löneinkomst med drygt 40 procent⁴⁰.

Faktorn för H-region

Norra glesbygden, H-reg 6, har ungefär samma proportionella nivå för bortfall som referensgruppen H-reg 4 (men högre än norra tätbygden H-reg 5). Interaktionseffektens värde för äldre (**65 + år**) i norra glesbygden är anmärkningsvärt högt – en fördubbling. Det bör dock noteras att efter interaktionen minskade den generella effekten för H-reg 6 betydligt och kommer till och med i nivå med H-reg 5⁴¹. Det är alltså särskilt de äldsta som kan vara svåra att få kontakt med i denna region. Det kan också tilläggas att det i första hand är de över 75 år som har ett särskilt stort bortfall.

⁴⁰ Beträffande inkomstfaktorn kan också konstateras att det inte föreligger någon interaktion med faktorn för högsta ålder. De flesta över 65 år saknar löneinkomst men trots detta gäller den proportionella inkomstfaktorn oförändrad även i det högsta ålderintervallet. Även om modellen är korrekt är detta resultat svårtolkat. Eftersom de flesta äldre saknar inkomst bör egentligen endast åldersfaktorn tas i beaktande (se tidigare modell utan inkomstfaktor).

⁴¹ Den generella ålderseffekten var oförändrad.

4.3.1.2 Samspelskombinationer i ULF – ej anträffade

I tabell 10 redovisas interaktionerna i ULF med avseende på bortfallet för ”ej anträffade” (1-procentnivån).

Könsfaktorn

I ULF har kvinnor generellt ett mycket lägre bortfall då det gäller anträffbarheten än män, med ca 30 procent lägre oddskvot. Denna effekt är betydligt lägre än i HINK. Kvinnor i ULF har dock i Stockholm (*H-reg 1*) en särskilt förhöjd risk⁴². En motsatt interaktion gäller för Göteborgs- och Malmö regionen (*H-reg 2*), där interaktionen innebär en minskning. I Göteborgs- och Malmö regionen leder detta till att den relativa bortfallsskillnaden mellan kvinnor och män ökar. Oddskvoten för kvinnor vs män i regionen blir extremt låg, $0,70 \times 0,54 = 0,38$.

Civilståndsfaktorn

Faktorerna för civilstånd och *ålder 16-24 år* interagerar extremt kraftigt. Generellt sett har gifta lägre bortfall än ej gifta – oddskvoten är ca 0,3. Yngre gifta har dock relativt sett en kraftig tendens till förhöjt bortfall genom interaktionen (8 gånger). I dessa unga åldrar får gifta i ULF t.o.m. dubbelt så stort bortfall (oddskvot) som bland ogifta, vilket framgår om vi multiplicerar med den proportionella civilståndseffekten. Observera dock att materialet här är tunt och skattningen därför innehåller en hel del slumpmässig variation.

Det finns även en betydande förhöjning av bortfallet bland gifta i den högsta åldersklassen (**65-84 år**) men i detta fall är förhöjningen inte så extrem (ungefär fördubblad).

Nationalitetsfaktorn

Invandrare har i allmänhet en hög risk för bortfall. I åldersklassen **65-84 år** är oddskvoten kraftigt förhöjd i ULF (drygt 3 gånger). Notera att i HINK var oddskvoten nästan fördubblad. Äldre invandrare är alltså en högriskgrupp, då det gäller svårigheten att få kontakt.

Inkomstfaktorn

I unga år (*16-24 år*) förminsкас den allmänt höga bortfallstendensen för dem utan inkomst genom en interaktion ($c=0,42$). Inkomsteffekten var enligt den proportionella analysen 2,8 i ULF. I den med interaktionerna sammanslagna modellen ökar den till 3,3.

⁴² Vid urvalet var interaktionen signifikant men var insignifikant i analysen då samtliga interaktioner ingick samtidigt.

Interaktionen innebär att låg inkomst mister mycket av sin betydelse bland unga ($3,3 * 0,42 = 1,4$).⁴³

Skillnader och likheter mellan HINK och ULF

Vi har tidigare sett i de proportionella modellerna för HINK och ULF att strukturen i bortfallet av ej anträffade var ungefär densamma. Detta gäller inte för interaktionerna. Två var dock gemensamma.

Man kunde kanske ha förväntat sig att flera interaktioner skulle överensstämma i denna del av undersökningarna som består av att komma i kontakt med urvalspersonerna. Kravet för signifikans (1-procentnivån) är dock högt ställt. Om man släpper något på detta och istället väljer signifikanser på 5-procentnivån, ökar antalet interaktioner i vardera undersökningen avsevärt⁴⁴. Vi redogör dock inte närmare för dessa interaktioner främst på grund av att risken är stor att de mister sin signifikans i den sammanslagna modellen. Det bör dock noteras att de gemensamma interaktionerna ökar från 2 till 6 st.

En av dessa gemensamma ”interaktioner” var signifikant på 1-procentnivån i ULF men bara på 5-procentnivån i HINK, vid de separata testen. Interaktionen gäller faktorn för civilstånd och ålder 16-24 år. I ULF såg vi att denna variabelkombination hade en extremt hög interaktionseffekt ($c=8,1$). I HINK pekar denna interaktionseffekt i samma riktning men ligger på en lägre nivå (ungefär en fördubbling).

Det kan också nämnas att en interaktion var nära signifikans ($p < 0,02$) i både HINK och ULF enligt vårt kriterium. Interaktionen gäller faktorn för invandrare och H-reg5. Invandrare har generellt ca 2-3 gånger högre risk (oddskvot) att ej anträffas än svenskar. I **H-region 5** (norra tätbygden) är bortfallet bland invandrare betydligt större.

⁴³ Jfr med HINK

⁴⁴ HINK går från 6 st till 13 st och ULF från 6 st till 17st .

4.3.2 Bortfall på grund av avböjd medverkan - interaktioner

I tabellerna 11 och 12 redovisas interaktionerna för avböjd medverkan i undersökningarna HINK respektive ULF. Vid beräkningen av bortfallsinteraktionerna för vägran har de ej anträffade utelämnats.

4.3.2.1 Samspelskombinationer i HINK – avböjd medverkan

Antalet interaktioner är ungefär detsamma för denna bortfallshuvudgrupp som för ej anträffade. Två av interaktionerna är insignifikanta och har låga värden i den samlade analysen och kommenteras därför inte vidare. För övriga interaktionseffekter är signifikansnivån som helhet något svagare än för föregående huvudkategori av bortfall.

Könsfaktorn

Interaktioner saknas för denna faktor. Kvinnor har alltså i förhållande till män genomgående i alla variabelindelningar drygt 10 procents större bortfall (oddskvot) på grund av avböjd intervju.

Civilståndsfaktorn

Gifta har i HINK en generellt något högre oddskvot än ogifta i den proportionella modellen (ca 10 procent). Gifta i HINK har alltså enligt den modellen en mindre benägenhet att vilja delta än ogifta har. Det finns emellertid två interaktioner mellan civilståndsfaktorn och **åldersfaktorn**. I ålder 35-44 år består interaktionen av en förminskad effekt för gifta i förhållande till ogifta och i åldersklassen för de äldre (65+ år) det motsatta, en förhöjd sådan effekt. Efter det att interaktionsfaktorerna införts försvinner nästan den proportionella effekten. Den något högre effekten generellt för gifta som vi tidigare sett är till stor del ett resultat av de gifta äldres höga nivå.

Nationalitetsfaktorn

Invandrare har generellt omkring 20 procent lägre bortfallsrisk än svenskar i den proportionella modellen. I **åldern 45-54** föreligger en interaktion. Bortfallsnivån är förhöjd med ca 40 procent. I denna ålder har alltså invandrare en mindre benägenhet att medverka ($0,8 \times 1,4 = 1,1$).

Inkomstfaktorn

För inkomstfaktorn föreligger inte någon interaktion.

Faktorn H-region

Generellt har Stockholm en relativt stor risk för vägran. Tabell 11 visar att det föreligger en tydlig interaktion med hög **ålder (65+år)**. De äldre i Stockholm har enligt interaktionen en förminskad bortfallsrisk på grund av avböjd medverkan. Dessutom har unga medelålders (**35-44 år**) i H-reg 5 en kraftigt förhöjd interaktionseffekt.

Tabell 11. Avböjd medverkan. Interaktioner ("c") i HINK 96/97 (utvalda på 1-procentnivån). Urval utan ej anträffade

	Kvinna	Gift	Invand- rare	Ej lön	H-reg 1	H-reg 2	H-reg 3	H-reg 5	H-reg 6
Gift	1	-	-	-	-	-	-	-	-
Invandrare	1	1	-	-	-	-	-	-	-
Ej lön	1	1,03	1	-	-	-	-	-	-
H-reg1	1	1	1	1	-	-	-	-	-
H-reg2	1	1	1	1	-	-	-	-	-
H-reg3	1	1	1	1	-	-	-	-	-
H-reg5	1	1	1	1	-	-	-	-	-
H-reg6	1	1	1	1	-	-	-	-	-
Ålder									
18-24	1	1	1	1	1	1	1	1	1
35-44	1	0,81*	1	1	1	1	1	1,74**	1
45-54	1	1	1,43*	1	1	1	1	1	1
55-64	1	1	1	1	1	1	1	1	1
65-	1	1,29*	1	1	0,76**	1	1,12	1	1

* 5-procentnivån, ** 1-procentnivån

Tabell 12. Avböjd medverkan. Interaktioner ("c") i ULF 96/97 (utvalda på 1-procentnivån). Urval utan ej anträffade

	Kvinna	Gift	Invand- rare	Ej lön	H-reg 1	H-reg 2	H-reg 3	H-reg 5	H-reg 6
Gift	1	-	-	-	-	-	-	-	-
Invandrare	1	1	-	-	-	-	-	-	-
Ej lön	1	1	1	-	-	-	-	-	-
H-reg1	1	1	1,52**	1	-	-	-	-	-
H-reg2	0,80	1	1	1	-	-	-	-	-
H-reg3	1,36*	1	1	1	-	-	-	-	-
H-reg5	1	1	1	1	-	-	-	-	-
H-reg6	1	1	1	1	-	-	-	-	-
Ålder									
16-24	1	2,70*	1	1	1,42	1	1	1	1
35-44	1	1	1	1	1	1	1	1	2,28**
45-54	0,69**	1	1	1	1	1	1	1	1
55-64	1	1	1	1	1	1	1	1	1
65-	1	1	1	1	0,60**	1	1	1	1

* 5-procentnivån, ** 1-procentnivån

4.3.2.2 Samspelskombinationer i ULF – avböjd medverkan

I tabell 12 anges interaktionerna för bortfall i ULF på grund av avböjd medverkan. Två av interaktionerna är insignifikanta i den samlade analysen och kommenteras därför inte vidare. Antalet interaktioner är ungefär detsamma som för HINK men det finns bara en som är gemensam (äldre i Stockholm).

Könsfaktorn

Kvinnor har i ULF, till skillnad från HINK, generellt ett något mindre bortfall på grund av vägran än män.

Kvinnor i Stockholm (**H-reg 1**) har en interaktion som visar ett förminskat bortfall. I Göteborgs- och Malmö regionen (**H-reg 2**) gäller motsatsen, d.v.s. ett förhöjt bortfall. Vid jämförelse med tabell 11 (bortfall på grund ej anträffad) framgår dock att de två korresponderande interaktionerna har helt motsatt effekt. En förklaring kan helt enkelt vara att man kodar bortfallskategorierna olika i de båda regionerna. Om man ser på bortfallet totalt föreligger faktiskt ingen som helst interaktion.

Den andra interaktionen ($c=0,69$) gäller kön och **åldern 45-54 år**. I denna åldersgrupp har kvinnor särskilt låg bortfallsnivå relativt männen.

Civilståndsfaktorn

Gifta har generellt ett något lägre bortfall än ogifta (det motsatta i HINK). Detta gäller dock ej de som är gifta i åldersgruppen **16-24 år** som har en kraftig förhöjning av risken för vägran ($c=2,7$). Även bortfallet på grund av sämre anträffbarhet var särskilt förhöjt (interaktion) för denna grupp. Den sammanlagda effekten för hela bortfallet (totalt) har en oddskvot på 3,8.

Nationalitetsfaktorn

Den proportionella modellen pekar på att det inte föreligger någon nämnvärd nationalitetsskillnad då det gäller att medverka i ULF. Det finns dock en interaktion nämligen mellan invandrarfaktorn och regionfaktorn för Stockholm (**H-reg 1**). Risken för vägran i denna grupp är förhöjd med ca 50 procent.

Inkomstfaktorn

Här föreligger inte några interaktioner i ULF.

Faktorn H-region

Stockholmsregionen har generellt sett en relativt stor risk för bortfall på grund av vägran. Detta bortfall är, liksom i HINK, ganska kraftigt förminskat i de högsta åldrarna (**65-84 år**). Dessutom föreligger en extrem förhöjning av bortfallet bland medelålders personer (**35-44 år**) i H-reg 6 (norrlands glesbygd).

Skillnader och likheter mellan HINK och ULF

Även för denna bortfallskategori var interaktionerna med ett undantag olika för HINK och ULF.

Signifikansnivåerna i den samlade analysen (med samtliga interaktioner) uppvisade för denna bortfallskategori svagare signifikansnivåer. En signifikansnivå på 5-procent vid fastställandet av en interaktion har här därför inte varit aktuell.

5. Något om bortfallet efter resultatkoder (detaljnivå)

I kapitel 4 redovisades bortfallets struktur efter olika bakgrundsvariabler, både totalt och i huvudkategorier. För vissa variabelindelningar var bortfallet mycket högt. Vi skall i det följande visa hur bortfallet på mer detaljerad nivå är fördelat för några högriskgrupper i populationsurvalen⁴⁵.

Inledningsvis redovisas en fördelning av bortfallet för hela urvalet i HINK respektive ULF (tabell 13). Resultatkoderna för ej anträffade är nästan samtliga högre i HINK än i ULF.

Tabell 13. Fördelning av svar och bortfall i HINK och ULF efter resultatkoder i detaljlista 96/97. Procent

Resultatkod	HINK	ULF
Svar från respondenten	73,9	78,2
Ej anträffad	10,5	5,4
Flyttat /Bortrest (51/52)	1,7	1,4
Kodlås/portlås	0,1	0,1
Hemligt telefonnummer	2,9	1,5
Saknar telefon (Telia)	2,1	1,2
Telefonsvarare	0,2	0,1
Övrigt	3,5	1,1
Avböjd medverkan	12,4	14,4
Ej tid	1,2	2,1
Ställer aldrig upp	1,8	1,9
Frivilligheten	5,7	6,0
Övrigt	3,7	4,4
Förhindrad medverkan	3,2	2,0
Sjukdom/annat hinder (41/43)	1,7	1,6
Språksvårigheter	0,5	0,2
Övrigt	1,0	0,2
Totalt	100	100
Antal personer	31 389	14 950

⁴⁵ Huvudkategorierna ”ej anträffad”, ”avböjd medverkan” och ”övrigt bortfall” (förhindrad medverkan) kan fördelas efter en mer detaljerade resultatkodninglista (vissa av dessa koder har aggregerats).

I tabell 14 nedan redovisas bortfall för vissa utvalda grupper i HINK respektive i ULF. Variabelindelningarna har i den tidigare analysen skilt sig kraftigt från respektive referensgrupp. De fyra första grupperna i tabell 14 (storstad, invandrare, ej löneinkomst och ogift) har uppvisat höga oddskvoter i den multivariata analysen. Även fördelningen för de äldsta redovisas (högt bortfall i ”övrigt bortfall”).

Tabell 14. Fördelning av svar och bortfall i HINK efter resultat-koder i detaljlista 96/97. Procent

Resultatkod	Totalt	Storstad, H-reg 1-2	Invand- rare	Ej löne- inkomst, < 65 år	Ej gift	Äldre, >64 år
Svar från respondenten	73,9	67,1	64,8	65,8	71,7	73,0
Ej anträffad	10,5	15,9	20,6	18,1	13,6	5,6
Flyttat /Bortrest (51/52)	1,7	2,8	3,5	2,7	2,4	0,7
Kodlås/portlås	0,1	0,2	0,1	0,1	0,1	0,1
Hemligt telefonnummer	2,9	5,0	7,0	5,3	3,6	1,4
Saknar telefon (Telia)	2,1	2,3	5,3	5,1	3,0	1,2
Telefonsvarare	0,2	0,5	0,3	0,1	0,3	0,0
Övrigt	3,5	5,1	4,4	4,8	4,2	2,2
Avböjd medverkan	12,4	13,7	9,9	12,9	10,8	13,4
Ej tid	1,2	1,5	0,9	1,2	1,1	0,5
Ställer aldrig upp	1,8	1,7	1,2	2,3	1,5	2,1
Frivilligheten	5,7	6,6	4,6	5,6	4,8	6,7
Övrigt	3,7	3,9	3,2	3,8	3,4	4,1
Förhindrad medverkan	3,2	3,3	4,7	3,2	3,9	8,0
Sjukdom/annat hinder (41/43)	1,7	1,9	0,4	1,1	2,2	4,8
Språksvårigheter	0,5	0,6	3,9	1,7	0,3	0,5
Övrigt	1,0	0,8	0,4	0,4	1,4	2,7
Totalt	100	100	100	100	100	100
Antal personer	31 389	10 964	3 519	4 257	16 878	9 273

Det mest utmärkande för högriskgrupperna är att bortfallet på grund av ej anträffad är särskilt högt och gäller främst detaljkoderna ”flyttat/bortrest”, ”hemligt telefonnummer” och ”saknar telefon” (enligt Telias register). Bland invandrare föreligger också en hög

bortfallsandel på grund av språksvårigheter. De äldre har motsatt förhållande med låga bortfallsandelar för just nämnda koder.

Då det gäller bortfall på grund av avböjd medverkan skiljer sig inte högriskgrupperna särskilt mycket från genomsnittet för bortfallet i denna huvudkategori. Genomgående finns en mycket hög andel som anger frivilligheten som skäl till att inte vilja medverka. De äldsta har en något förhöjd nivå för denna resultatkod men framförallt har de en hög bortfallsandel i ”sjukdom/annat hinder”.

Tabell 15. Fördelning av svar och bortfall i ULF efter resultatkoder i detaljlista 96/97. Procent

Resultatkod	Totalt	Storstad H-reg 1-2	Invand- rare	Ej löne- inkomst, < 65 år	Ej gift	Äldre, >64 år
Svar från respondenten	78,2	73,0	69,0	69,1	76,5	74,0
Ej anträffad	5,4	9,2	15,3	12,9	7,6	2,6
Flyttat /Bortrest (51/52)	1,4	2,4	4,1	3,7	2,0	0,7
Kodlås/portlås	0,1	0,3	0,4	0,4	0,2	0,0
Hemligt telefonnummer	1,5	2,7	4,0	3,1	2,0	0,7
Saknar telefon (Telia)	1,2	1,9	3,5	3,2	1,8	0,5
Telefonsvarare	0,1	0,1	0,1	0,2	0,1	0,0
Övrigt	1,1	1,8	3,2	2,3	1,5	0,7
Avböjd medverkan	14,4	16,1	13,6	15,4	13,7	16,9
Ej tid	2,1	2,5	1,3	1,7	2,2	0,7
Ställer aldrig upp	1,9	1,5	1,5	2,4	1,8	2,7
Frivilligheten	6,0	7,4	5,7	5,7	5,4	8,4
Övrigt	4,4	4,7	5,1	5,6	4,3	5,1
Förhindrad medverkan	2,0	1,7	2,1	2,6	2,2	6,5
Sjukdom/annat hinder (41/43)	1,6	1,4	0,7	1,9	1,9	5,6
Språksvårigheter	0,2	0,2	1,3	0,4	0,1	0,2
Övrigt	0,2	0,1	0,1	0,3	0,2	0,7
Totalt	100	100	100	100	100	100
Antal personer	14 950	5 088	1 785	2 482	8 021	3 022

I tabell 15 finns motsvarande fördelning för ULF. Resultaten avseende bortfallsfördelningarna för ULF överensstämmer i stor utsträckning med dem för HINK⁴⁶. Flyttad, avsaknad av telefonnummer är viktiga skäl för de fyra högriskgrupperna. Bortfall på grund av språksvårigheter är dock i ULF betydligt lägre än i HINK.

⁴⁶ Observera att andelarna i tabellerna inte är standardiserade på något sätt (marginalfördelningar). Undersökningarna (HINK, ULF) är därför inte helt jämförbara (gäller främst skillnader i åldersfördelningen).

6. Sammanfattning och diskussion

Resultaten i kapitel 4 (tabell 2) för åren 1996/97 visar för både HINK och ULF att bortfallet totalt i undersökningarna stiger med åldern, är högre för ej gifta än för gifta och högre för invandrare än för svenskar samt mycket högt för dem som bor i storstadsregion i förhållande till dem i övriga Sverige. Även de utan löneinkomst har ett mycket högt bortfall i förhållande till övriga. Variabeln löneinkomst har naturligtvis främst relevans för personer under 65 år⁴⁷. De nämnda variablerna har genom den multivariata modellen visat sig ha signifikanta, ”oberoende” effekter på bortfallet⁴⁸. Det bör noteras att de ”oberoende” effekterna ibland avviker från de skillnader som kan iakttas hos vanliga bortfallsandelar (marginalfördelningar). Ett sätt att bedöma storleken av sådana skillnader är att också göra univariata analyser (faktorerna analyseras en i taget)⁴⁹.

Skillnaderna mellan multivariat och univariat analys var dock i de flesta fall inte särskilt stora. De faktorer som påverkades mest var åldersfaktorn och nationalitetsfaktorn. I den multivariata modellen ökade bortfallsrisken med åldern betydligt kraftigare än i den univariata modellen. Detta resultat pekar på att det föreligger en ålderstendens som är större än vad man är van att iakttä utifrån enkla åldersindelade bortfallsandelar. Invandrarnas högre bortfallsrisk är istället lägre i multivariat analys än univariat. Påverkan från andra korrelerade bakgrundsfaktorer av ekonomisk och social natur ligger bakom den lägre nationalitetseffekten vid univariat analys.

Mycket av variationen av oddskvoterna för det totala bortfallet mellan olika variabelindelningar beror på bortfallet på grund av ej anträffade urvalspersoner (gäller civilstånd, nationalitet, region samt löneinkomst). Detta framgick då analysen gjordes med uppdelning av bortfallet på huvudkategorier (tabellerna 5-6). Bortfallet genom ”avböjd medverkan” har i vår analys inte visat sig ge så stora variabeleffekter. Att de strukturella effekterna är relativt små för avböjd medverkan är i och för sig gynnsamt med tanke på att största delen av bortfallet gäller denna huvudkategori, både i HINK och ULF.

Ett viktigt undantag avseende den senare huvudkategorin av bortfall är åldersvariabeln. I både HINK och ULF växer bortfallet på grund av vägran betydligt med åldern⁵⁰. Detta är särskilt uttalat i HINK. Det kan å andra sidan noteras att anträffbarheten är mer jämt fördelad efter ålder. De äldsta har t.o.m. mycket liten bortfallsrisk på grund av oanträffbarhet. Den sammanlagda ålderseffekten på bortfallet blir därmed mildrad (tabell 2).

⁴⁷ I analysen prövades en modell med inkomstvariabeln och en modell där inkomsten uteslöts. Genom detta förfarande kunde en mera relevant bortfallsnivå fastställas för de äldsta (högt bortfall).

⁴⁸ Med ”oberoende” effekt menas effekten av en faktor då vi kontrollerar för övriga bakgrundsvariabler.

⁴⁹ Faktorerna för ålder och kön ingår i alla modellerna.

⁵⁰ Detta gäller både multivariat och univariat analys.

Resultaten för HINK och ULF är i många fall mycket lika. En väsentlig skillnad finns dock. Bortfallet är mindre för kvinnor än för män i ULF. Variationen i bortfallsstrukturen är i ULF dessutom genomgående mera accentuerad (kan delvis bero på att ULF, som är en mindre undersökning, innehåller en större slumpmässig variation än HINK).

Proportionalitetsantagandet i modellen har visat sig vara i samklang med empiriska data⁵¹. Interaktioner eller samspel mellan bakgrundsvariablerna är enligt den analys vi gjort få till antalet. De som förekommer gäller främst åldersindelningen. Den lägsta och den högsta åldersklassen ger de flesta interaktionerna, dvs. bortfalleffekter för bakgrundsvariablerna är annorlunda i dessa åldersklasser än för övriga åldrar, vilket i och för sig kan anses vara naturligt. En interaktion, som är iögonfallande, är den som gäller för unga gifta (idag en mycket liten grupp), vilka har en särskilt kraftigt förhöjd bortfallsrisk på grund av ej anträffade. Gruppen består till stor del av invandrare.

För att minska bortfallet i undersökningarna är de uppräknade faktorerna som pekar på högt bortfall väsentliga att beakta vid insamlingsförfarandet. Detta av två skäl. Dels identifieras högriskgrupper, som vid en ökad insats vid insamlingen kan ge relativt stort positivt svarsresultat⁵², dels undviks vid en förbättring, snedvridning av resultaten. Visserligen används en viss form av efterstratifiering⁵³ och bortfallskompensation vid statistikproduktionen för att undvika bortfallsbias men trots detta blir representativiteten och tillförlitligheten lidande då bortfallet är stort.

I detta sammanhang kan påpekas att vissa utvalda grupper (variabelkombinationer) kan få ett mycket högt bortfall på grund av bortfallets strukturskillnader. Notera att om invandrare och de utan löneinkomst (i medelåldern), vilka i sig är relativt små grupper, ingår i någon kombination, utgör kombinationen en liten del av urvalspopulationen. Bortfallet totalt för kombinationen (man, ej gift, 45-54 år, invandrare, stockholmsregionen och utan löneinkomst) blir 58 procent i HINK och 59 procent i ULF beräknat på uppgifterna ur tabell 2. För en svensk med samma indelning för övrigt blir bortfallsandelen 48 procent i HINK och densamma i ULF. Om vi i det senare fallet även utväxlar ej löneinkomst mot löneinkomst sjunker bortfallsandelen till 36 procent i HINK och 32 procent i ULF.

⁵¹ Faktoriserat odds.

⁵² Förutsatt lika gruppstorlekar.

⁵³ Gäller ULF. Vid efterstratifieringen ingår de bakgrundsvariabler som använts i föreliggande rapport (utom nationalitetsvariabeln). Två viktsystem används, ett system utan inkomst och ett med inkomst.

De specifika orsakerna till bortfallet angavs i kapitel 5 för högriskgrupperna. De främsta skälen till att bortfallet var särskilt högt gällde främst kategorin "ej anträffad". Detaljorsakerna var "hemligt telefonnummer", "avsaknad av telefon" (Telia) eller "flyttat /bortrest". Även språksvårigheter hade betydelse för invandrare⁵⁴.

Då det gäller avböjd medverkan anges ofta som skäl att det är frivilligt att delta. Denna orsak är dock i större utsträckning jämnt fördelad hos de valda högriskgrupperna.

I denna rapport har vi inte särskilt analyserat tidsutvecklingen av bortfallsfrekvenserna. Enligt en tidigare studie av bortfallsstrukturen i ULF (Johansson 1998) befanns variabel-effekterna vara ganska stabila över tiden 1986-97 (nationalitet och civilstånd ingick dock inte i studien)⁵⁵. En förändring var dock att vid mitten av 80-talet hade män och kvinnor ungefär samma storlek av bortfallet medan män numera har fått ett markant större bortfall än kvinnor.

Inom SCB har under senare år diskuterats vad den växande bortfallsnivån kan bero på. En förklaring som ofta har framförts är att det numera är vanligt med hemligt telefonnummer eller mobiltelefon. Detta har vi kunnat utläsa ur tabellerna i kapitel 5 (1996/97). En annan förklaring är den ökade beläggningen av uppdrag på intervjuarkåren. Denna belastning kan ge tidsbrist vid eftersökningen av intervjupersonerna. En särskilt försvårande omständighet vid tidsbrist är problemet med hemligt telefonnummer och mobiltelefoner. I ULF som har en längre insamlingsperiod är just bortfallet på grund av ej anträffbar mindre. Slutligen anförs ofta att uppgiftslämnarklimatet försämrats. Denna förklaring är dock svårtolkad. Vad betyder egentligen uppgiftslämnarklimat? I detta sammanhang är det anmärkningsvärt att så många pekar på *frivilligheten* som skäl för att inte medverka i undersökningarna. Betydelsen av väl beskrivande introduktionsbrev som för fram värdet av urvalspersonens deltagande kan inte överskattas.

Till sist skall göras ett påpekande som inte alltid förs fram i diskussionen. Flera demografiska strukturförändringar, som inträffat under senare år, har i viss mån bidragit till försämringen av svarsbenägenheten. Inflyttningen till storstadsregionerna, en åldrande befolkning, en allt större andel invandrare samt ändrad inkomstfördelningen har antagligen haft betydelse särskilt i belysning av de resultat som presenterats i rapporten.

⁵⁴ Utländska medborgare utom Norden (ej naturaliserade) hade en ytterligare förhöjning av oddskvoten avseende bortfall på grund av ej anträffad (särskild analys). Förhöjningen var dock måttlig i förhållande till nivån den för hela invandrargruppen redovisad i kapitel 4.

⁵⁵ Analysen utfördes på komplementsannolikheten, dvs. svarsfrekvensen.

7. Referenser

Bortfallsbarometer nr 14. R&D Report 1999:2. SCB

Johansson S-E. (1998) Bortfallsmönstret i ULF 1986-1997. PM 1998-08-30. BV/VÄL. SCB

Japac L, Ahtiainen A, Hörngren J, Lindén H, Lyberg L och Nilsson P (1997). Minska bortfallet. SCB

Nordberg L. (1989) Generalized linear modelling of sample survey data. JOS, 5:3, pp 223-240.

Qvist J. (1999) Täckningsproblem i Registret över totalbefolkningen RTB. R&D Report 1999:1. SCB

Ribe M.(1998) En titt på bortfall i HINK resp. ULF. PM 1998-10-05. SCB

8. Bilagor

8.1 Material

8.1.1 Inkomstfördelningsundersökningen (HINK)

En SAS-datafil sammanställdes utifrån ett basregister (nettourval) för HINK för referensårgångarna 1996-97. Intervjuerna för årgångarna har gjorts under våren påföljande år, dvs. 1997 och 1998⁵⁶. Urvalet är ett stratifierat OSU.

För nämnda årgångar finns senaste statistik för individurvalen med bortfallsuppgifter och med olika bakgrundsvariabler ur befolkningsregistren påförda (ålder, kön, region etc.)⁵⁷.

Den beroende variabeln är:

- Bortfall (svar/bortfall)

Bortfallet är en dikotom variabel och omfattar allt objektsbortfall i undersökningen. Bortfallet bestäms av resultatkoder enligt en detaljlista. Vi kan därför indela bortfallet i huvudkategorier. Dessa är "ej anträffad", "avböjd medverkan" och "övrigt bortfall" (se även "Minska bortfallet", 1997, SCB).

Intervju=10-18

Ej anträffad=50-57

Avböjd medverkan = 60-70, 90

Övrigt bortfall = 40-41, 43-44, 46, 80

Bortfall= Ej anträffad + Avböjd medverkan + Övrigt bortfall

Svar= Intervju

Andra indelningar avseende den beroende variabel (dikotom) som analyseras:

1. Ej anträffade (bortfall)
 - Övriga
2. Avböjd medverkan (bortfall)
 - Övriga (utom ej anträffade)

I det sistnämnda fallet undersöktes även oddskvoterna för det alternativ då gruppen "övriga" innehöll det resterande urvalet.

⁵⁶ I viss utsträckning har indirekta intervjuer utförts då intervjupersonen inte varit tillgänglig. Intervjuformen innebär att någon annan familjemedlem kan tillåtas svara.

⁵⁷ Uppgifter främst från Registret över totalbefolkningen (RTB). Gäller för referensåren.

Bakgrundsvariablerna är:

■ Kön (man/kvinna)

■ Ålder (ålder vid årets slut, åldersspann i HINK 18- år med åldersgruppering)

18-24 år

25-34 år

35-44 år

45-54 år

55-64 år

65- år

■ Civilstånd (gift/ej gift)

Gift=2,3,7 enligt RTB-kod

Ej gift =övriga

■ Nationalitet (invandrare/svensk)

Invandrare =utländskt medborgarskap eller naturaliserad från och med 1967

Svensk= övriga

■ H-region (6 regioner) enligt kommunal indelning 1998-01-01

H-region 1 =Stockholmsregionen

H-region 2= Göteborgs- och Malmöregionen

H-region 3= Större städer

H-region 4=Södra mellanbygden

H-region 5 =Norra tätbygden

H-region 6 =Norra glesbygden

■ Inkomst (inkomst (lön)/ej inkomst (lön))

Inkomst (lön)= kontant bruttolön enligt kontrolluppgift

Ej inkomst (lön)= Övriga

H-region är en tätortsindelning. Det bör noteras att indelningen geografiskt inte är strikt avgränsad.

8.1.2 Undersökningen om levnadsförhållanden (ULF)

En SAS-datafil sammanställdes med hjälp av basregistret (nettourval) för ULF för åren 1996-97. För ULFs del har intervjuerna gjorts just dessa år. Indirekta intervjuer förekommer i mycket begränsad omfattning. Urvalet är ett systematiskt urval (appr. OSU).

För 1996/97 finns statistik för hela individurvalet med bortfallsuppgifter och med olika registeruppgift om bakgrundsvariabler påförda.

Den beroende variabeln är:

■ Bortfall

Bortfallet är en dikotom variabel och omfattar allt objektsbortfall i undersökningen. Bortfallet bestäms även här av resultatkoder enligt en detaljlista.

Intervju=10-22

Ej anträffad=50-56

Avböjd medverkan = 60-69

Övrigt bortfall = 40-44

Bakgrundsvariablerna är:

■ Kön (man/kvinna)

■ Ålder (ålder vid årets slut, åldersspann i ULF 16-84 år med åldersgruppering)

16-24 år

25-34 år

35-44 år

45-54 år

55-64 år

65-84 år

■ Civilstånd (gift/ej gift)

Se HINK

■ Nationalitet (invandrare/svensk)

Se HINK

■ H-region (6 regioner)

Se HINK

■ Inkomst (inkomst (lön)/ej inkomst (lön))

Se HINK

■ Panel (panel/ej panel)

Panelen består av ett s.k. panelurval. De som ingått i ULF 8 år tidigare ingår i panelurvalet men också en andel av ungdomarna som definieras tillhöra panelen.

8.1.3 Urvalsstorlekar

De båda undersökningarna som helhet (18+ år för HINK och 16-84 år för ULF) innehåller populationsstorlekarna 31 389 individer för HINK och 14 950 individer för ULF.

I analyserna som presenterats i texten har urvalsstorlekarna varit de följande,

Tabell B1. Urvalsstorlekar 96/97 i HINK och ULF

	HINK	ULF
Tabell 2	30160	14526
Tabell 3	31389	-
Tabell 4	-	14950
Tabell 5 (Hela urvalet/utan ej anträffade)	31389 / 28083	-
Tabell 6 (Hela urvalet/utan ej anträffade)	-	14950 / 14144
Tabell 7	Se tabell 5	-
Tabell 8	-	Se tabell 6
Tabell 9	31389	-
Tabell 10	-	14950
Tabell 11	28083	-
Tabell 12	-	14144

8.2 Metod

Beräkningsmetod för analysen är logistisk regression eller logit-analys. Metoden som använts är SAS-proceduren (Logistic).

Denna procedur har använts för att skatta parametrarna i logit-modellen.

$$\log(p/(1-p)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

I texttabellerna har vi redovisat oddskvoterna som motsvaras av exponentialfunktionen av parametrarna, $\exp(\beta)$. Metodiken för skattningen har vi tidigare redovisat i avsnitt 3.2. Det kan här tilläggas att det vid signifikanstesten i Logistic inte vid något tillfälle tagits hänsyn till s.k. över- eller underspridning i förhållande till modellantaganden (AGGREGATE (deviance))⁵⁸.

Vid framtagningen av signifikanser för interaktioner har SAS-proceduren CATMOD använts. Denna metod förenklar bearbetningen då en stor mängd interaktioner skall prövas. En interaktion i CATMOD kan nämligen enkelt beskrivas som en variabelprodukt i programmet.

⁵⁸ Det bör noteras att vid univariat skattning ("standardiserad för ålder och kön) avvek i vissa fall "deviance" märkbart från 1 (oftast förhöjt värde). Variansen som är modellbaserad kan därmed underskattas.

8.3 Konfidensintervall (för oddskvoter i tabell 2)

Tabell B2. Konfidensintervall (95%) för oddskvoterna avseende bortfallet totalt i HINK och ULF 96/97

Konfidensintervall (95%) för oddskvoterna (tabell 2, sid.13)				
Variabel	HINK		ULF	
	Huvudmodell	Utan inkomst	Huvudmodell	Utan inkomst
Kvinna	0,94-1,04	(0,94-1,05)	0,79-0,92	(0,79-0,92)
Man	1		1	
Ålder 18-24	0,88-1,09	(0,90-1,11)	0,67-0,93	(0,69-0,96)
Ålder 25-34	1		1	
Ålder 35-44	1,16-1,39	(1,16-1,40)	1,11-1,47	(1,13-1,49)
Ålder 45-54	1,27-1,52	(1,28-1,54)	1,26-1,66	(1,29-1,69)
Ålder 55-64	1,09-1,34	(1,21-1,48)	1,32-1,78	(1,50-2,02)
Ålder 65-84	0,82-1,00	(1,21-1,43)	0,99-1,34	(1,67-2,18)
Gift	0,71-0,79	(0,70-0,78)	0,62-0,74	(0,59-0,71)
Ej gift	1		1	
Invandrare	1,36-1,59	(1,49-1,74)	1,39-1,76	(1,60-2,01)
Svensk	1		1	
Hreg1	1,51-1,78	(1,47-1,74)	1,52-1,95	(1,47-1,89)
Hreg2	1,35-1,61	(1,36-1,62)	1,14-1,51	(1,14-1,50)
Hreg3	0,89-1,04	(0,89-1,04)	0,94-1,19	(0,94-1,18)
Hreg4	1		1	
Hreg5	0,56-0,76	(0,57-0,76)	0,59-0,91	(0,60-0,93)
Hreg6	0,71-0,95	(0,72-0,96)	0,65-1,01	(0,65-1,00)
Ej ink (lön)	1,52-1,75	-	1,75-2,15	-
Ink (lön)	1	-	1	-
Panel	-	-	0,94-1,10	(0,94-1,10)
Ej panel	-	-	1	

Uppgifter inom () avser modell utan inkomstfaktor

8.4 Oddskvoter (proportionella) i modell med interaktionstermer

Tabell B.3. Logitanalys av bortfallet på grund av ej anträffade i HINK och ULF 96/97 (oddskvoter). Proportionella effekter

Bortfall		
Variabel	Oddskvoter – HINK	Oddskvoter – ULF
Kvinna	0,89 **	0,70 **
Man	1	1
Ålder 16-24	1,13	0,83
Ålder 25-34	1	1
Ålder 35-44	1,02	1,15
Ålder 45-54	0,88	1,32 *
Ålder 55-64	0,59 ***	1,03
Ålder 65-84	0,36 ***	0,18 ***
Gift	0,59 ***	0,29 ***
Ej gift	1	1
Invandrare	1,76 ***	2,60 ***
Svensk	1	1
Hreg1	2,28 ***	2,72 ***
Hreg2	1,83 ***	2,87 ***
Hreg3	1,05	1,32 *
Hreg4	1	1
Hreg5	0,54 ***	0,53 *
Hreg6	0,78	0,76
Ej ink (lön)	2,00 ***	3,28 ***
Ink (lön)	1	1
Panel	-	0,88
Ej panel	-	1
Interaktioner ("c")	Se tabell 9	Se tabell 10
Intercept	-2,1246 ⇒ p=0,11	-3,2271 ⇒ p=0,04

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån
 p= bortfallssannolikhet baserad enbart på interceptet, dvs. gäller i kombinerad referensgrupp.

Tabell B.4. Logitanalys av bortfallet på grund av avböjd medverkan i HINK och ULF 96/97 (oddskvoter). Proportionella effekter. Urvalet utan ej anträffade

Bortfall		
Variabel	Oddskvoter – HINK	Oddskvoter – ULF
Kvinna	1,14 ***	0,98
Man	1	1
Ålder 16-24	0,96	0,71 **
Ålder 25-34	1	1
Ålder 35-44	1,71 ***	1,27 **
Ålder 45-54	1,76 ***	1,74 ***
Ålder 55-64	1,90 ***	1,63 ***
Ålder 65-84	1,13	1,38 ***
Gift	1,05	0,89 *
Ej gift	1	1
Invandrare	0,75 ***	0,87
Svensk	1	1
Hreg1	1,35 ***	1,59 ***
Hreg2	1,22 ***	1,02
Hreg3	0,88 *	1,04
Hreg4	1	1
Hreg5	0,62 ***	0,71 **
Hreg6	0,76 **	0,67 **
Ej ink (lön)	1,20 **	1,33 ***
Ink (lön)	1	1
Panel	-	1,08
Ej panel	-	1
Interaktioner ("c")	Se tabell 11	Se tabell 12
Intercept	-2,3293 ⇒ p=0,09	-2,0794 ⇒ p=0,11

Signifikansnivåer: * avser 5%-nivån, **1%-nivån och *** 0,1%-nivån
p= bortfallssannolikhet baserad enbart på interceptet, dvs. gäller i kombinerad referensgrupp.

Förteckning över utkomna R&D Reports

R&D Reports är en för IT-enheten och Metodenheten gemensam publikationsserie, som 1988-01-01 ersatte de tidigare ”gula” och ”gröna” serierna. I serien ingick fram till årsskiftet 1992-93 även **Abstracts** (sammanfattning av metodrapporter från SCB).

Reports published during 1998 and onwards:

- 1998:1 Preliminär statistik: Nybyggnadskostnader - en simuleringstudie
(grön) (*Catarina Elffors*)
- 1998:2 On Inclusion Probabilities for Order Sampling (*Bengt Rosén*)
(grön)
- 1998:3 On the Stratification of Highly Skewed Populations (*Dan Hedlin*)
(grön)
- 1998:4 Bortfallsbarometer nr 13 (*Per Nilsson, Antti Ahtiainen, Stefan Berg, Mats Bergdahl, Monica Rennermalm och Marcus Vingren*)
(grön)
- 1998:5 Estimation from Order π ps Samples with Non - Response (*Bengt Rosén and Pär Lundqvist*)
(grön)
- 1998:6 On variance estimation for measures of change when samples are coordinated by a permanent random numbers technique (*Lennart Nordberg*)
- 1999:1 Täckningsproblem i Registret över totalbefolkning RTB. Skattning av övertäckning med en indirekt metod (*Jan Qvist*)
(grön)
- 1999:2 Bortfallsbarometer nr 14 (*Per Nilsson, Antti Ahtiainen, Mats Bergdahl, Tomas Garås, Jan Qvist och Charlotte Strömstedt*)
(grön)
- 1999:3 Att mäta statistikens kvalitet (*Claes Andersson, Håkan L. Lindström och Thomas Polfeldt*)
(grön)
- 2000:1 Kalibrering av vikter – beskrivning av tekniken och de SCB-fall den prövats i
(grön) (*Sixten Lundström et al*)
- 2000:2 On Inclusion Probabilities and Estimator Bias for Pareto π ps Sampling
(grön) (*Nibia Aires and Bengt Rosén*)
- 2000:3 Bortfallsbarometer nr 15 (*Per Nilsson, Ann-Louise Engstrand, Sara Tångdahl, Stefan Berg, Tomas Garås och Arne Holmqvist*)
(grön)
- 2000:4 Bortfallsanalys av SCB-undersökningarna HINK och ULF (*Jan Qvist*)
(grön)

ISSN 0283-8680

Tidigare utgivna **R&D Reports** kan beställas genom Katarina Klingberg, SCB, MET, Box 24 300, 104 51 STOCKHOLM (telefon 08-506 942 82, fax 08-506 945 99, e-post katarina.klingberg@scb.se). **R&D Reports** from 1988-1997 can - in case they are still in stock - be ordered from Statistics Sweden, attn. Katarina Klingberg, MET, Box 24 300, SE-104 51 STOCKHOLM (telephone +46 8 506 942 82, fax +46 8 506 945 99, e-mail katarina.klingberg@scb.se).