



Statistics Sweden

Statistiska centralbyrån

Teknisk Rapport

En beskrivning av genomförande, metoder och bortfallsredovisning

EHIS 2014

SCB, Stockholm
08-506 940 00

SCB, Örebro
019-17 60 00

www.scb.se



Inledning

Enheten för Demokratistatistik vid Statistiska centralbyrån (SCB) genomförde under perioden september 2014 – januari 2015 en enkätundersökning på uppdrag av Folkhälsomyndigheten.

Denna undersökning har genomförts i enlighet med ISO 20252:2012 Marknads-, opinions- och samhällsundersökningar vilket innebär att grundläggande kvalitetskrav uppfyllts.

Syftet med undersökningen var att kartlägga och ge information kring: allmänt hälsotillstånd, hälsodata samt uppgifter om hälsans bestämningsfaktorer bland Sveriges befolkning. Undersökningen genomförs parallellt i ett stort antal EU-länder för att kunna jämföra hälsoläget i olika länder i Europa.

Ett tillräckligt stort urval drogs för att minst 6 200 svar kunde samlas in. Urvalet drogs från personer i åldrarna 15 år och uppåt från SCB:s register över totalbefolkningen (RTB).

Urvalet bestod av 11 006 personer och det var totalt 6 292 personer som besvarade frågorna, vilket var 57 procent av urvalet.

Resultatet i form av svarsdata levererades vecka 8 2015 enligt överenskommelse.

Undersökningsledare på SCB var Anna Malmberg och Pär Karlsson var metodstatistiker. Folkhälsomyndighetens kontaktperson gentemot SCB var Richard Bränström.

Omfattning

Population och urval

Populationen, d.v.s. de objekt som man vill kunna dra slutsatser om, utgjordes av befolkningen från 15 år och uppåt i Sverige.

För att kunna dra ett urval från populationen skapades en urvalsram som avgränsade, identifierade och möjliggjorde koppling till objekten i populationen. Urvalsramen i undersökningen skapades med hjälp av data från registret över totalbefolkningen (RTB) från sista juni 2014. De personer som ingick i urvalet till Folkhälsomyndighetens Nationella Folkhälsoundersökningen under 2014 togs inte med i urvalsramen för denna undersökning. Antalet personer i urvalsramen var 7 901 091.

Urvalsramen stratifierades på kön, ålder och födelseland (Sverige, Övriga länder). Totalt bildades 12 strata. Från urvalsramen drogs ett stratifierat obundet slumpmässigt urval om 11 006 personer (se tabell 1 nedan).

Ett stratifierat obundet slumpmässigt urval innebär att alla objekt inom ett stratum har samma sannolikhet att komma med i urvalet.

Tabell 1 Population och urval fördelat på stratum

Stratum nr	Födelse-land	Kön	Ålder	Urvalsram	Urval
1	Sverige	Män	15-29	761 709	1588
2	Sverige	Män	30-54	1 219 081	1838
3	Sverige	Män	55+	1 242 414	1207
4	Sverige	Kvinnor	15-29	721 130	1020
5	Sverige	Kvinnor	30-54	1 159 734	1343
6	Sverige	Kvinnor	55+	1 364 878	1262
7	Övriga	Män	15-29	149 452	469
8	Övriga	Män	30-54	353 727	777
9	Övriga	Män	55+	195 969	229
10	Övriga	Kvinnor	15-29	143 092	357
11	Övriga	Kvinnor	30-54	358 275	637
12	Övriga	Kvinnor	55+	231 630	279
Totalt				7 901 091	11 006

Frågor/Variabler

Folkhälsomyndigheten översatte frågorna från engelska till svenska från det av Eurostat utformade frågeformuläret. Formuläret anpassades sedan i samarbete mellan Folkhälsomyndigheten och SCB till en enkätversion. Formuläret granskades av mätteknisk expertis i syfte att minska risken för mätfel. Grundläggande för bra kvalitet i en undersökning är kvaliteten på de data som samlas in. För att säkerställa att frågorna fungerar så bra som möjligt och enligt intentionerna har därför frågeformuläret genomgått ett mättekniskt test.

Frågeformuläret bestod av 81 numrerade frågor, flera av dem hade delfrågor vilket genererade totalt 111 frågor. Två versioner av frågeformuläret togs fram, ett för åldern 15-64 år och ett för de som är 65 år eller äldre. Det för de äldre innehöll fler frågor.

Förutom de variabler som samlades in via frågeformuläret hämtades följande variabler från olika register:

- COUNTRY - Land där man bor
- BIRTHPLACE - Födelse-land (Sverige, annat EU-land, utanför EU)
- CITIZEN - Land där man är medborgare (National, Non-national inom EU, Non-national utanför EU)
- REGION NUTS
- DEG_URB - Urbaniseringsgrad
- MARSTALEGAL - Civilstånd

- HHINCOME - Hushållsinkomst i kvintiler
- HATLEVEL - Högsta utbildningsnivå
- JOBISCO - Yrkeskodning enligt ISCO-08
- LOCNACE - Näringsgren

Referensperiod för frågorna, se bifogad frågeblankett. Med referensperiod menas vilken tidpunkt svaren avser, t.ex. idag, förra veckan, senaste tre månaderna.

Datainsamling

Ett förbrev skickades ut med post den 8 september 2014. Brevet beskrev kort undersökningen och gav möjlighet att kontakta SCB för att meddela att man inte ville vara med. Två veckor senare, den 19 september, skickades frågeformulären ut tillsammans med ett informationsbrev som också gav möjlighet att besvara frågorna via webben. Sedan skickades 2 påminnelser till dem som inte besvarat frågeformuläret. Ett tack- och påminnelsekort (ToP-kort) skickades i början på oktober, den 6 oktober, och en påminnelse tillsammans med ny enkät skickades den 20 oktober. I början på november påbörjades en telefonuppföljning till de som ännu inte skickat in svar. Den 11 november startade telefonintervjuerna. Insamlingen avslutades den 14 januari 2015.

Tabell 2 - Beskrivning av inflödet. Antal och andel

	Antal	Andel
Efter första utskick	2 801	25,4
Efter ToP-kort	1 584	14,4
Efter enkätpåminnelse	845	7,7
Telefonintervjuer	1 062	9,6
Totalt	6 292	57,2
Bortfall	4 714	42,8
Urval	11 006	100,0

I informationsbrevet kunde uppgiftslämnarna läsa om undersökningens bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Folkhälsomyndigheten och SCB. Informationsbrevet informerade också om att uppgifter hämtades från andra register och att en avidentifierad datafil levererades till Folkhälsomyndigheten som även skickar data vidare till Eurostat. Brevet informerade även om personuppgiftslagen samt offentlighets- och sekretesslagen och att det var frivilligt att medverka i undersökningen.

För att SCB ska kunna lämna ut data från en enkätundersökning krävs informerats samtycke av uppgiftslämnarna. Det innebär att de genom att besvara blanketten och skicka in den godkänner att deras svar kompletteras med de bakgrundsvariabler, och behandlas på det sätt, som beskrivs i informationsbrevet.

Datainsamlingen genomfördes av enkätenheten vid SCB och telefonintervjuerna genomfördes av SCBs samarbetspartner MIND Research.

Datainsamlingen genomfördes via webb och skanning av de frågeformulär som kommit in via post samt via telefonintervjuer.

De enkäter som kommit in via post registrerades med hjälp av skanning.

Kontroller har genomförts bland annat för att säkerställa så att endast valida värden förekommer i materialet.

SCB kan inte garantera att den utvalda personen själv besvarat frågeformuläret. Någon kontroll av att rätt person har besvarat frågeformuläret har inte gjorts.

Svarsdatafilen kompletterades med vikter för uppräkningsnivå till populationsnivå.

Urvalet har telefonnummersatts av ett telekombolag. Efter denna telefonnummersättning saknades telefonnummer till 1 023 urvalspersoner.

Om urvalspersonen efter telematch saknade telefonnummer genomfördes en manuell spårning av SCB:s särskilda spårargrupp. Efter avslutad spårning saknades telefonnummer för 223 urvalspersoner, vilket motsvarar 4,3 procent av det totala urvalet.

Spårningskällor som vi har använt är:

- Eniro, för att se om det finns någon med liknande namn på adressen, om det bara finns en abonnent eller om namnet förekommer på annan adress.
- Folkbokföringen, för att se om personerna bor kvar på adressen och, om personen är under 25 år, få uppgifter om vilka föräldrarna är.
- Kontrolluppgiftsregistret, för att se om man haft något arbete under förra året.
- Försäkringskassan, för att se om personen har något registrerat arbete eller andra uppgifter som kan hjälpa oss vidare, t.ex. att man studerar.
- Posten
- Pensionsmyndigheten
- Migrationsverket

SCB kan inte garantera att den utvalda personen själv besvarat frågorna i telefonintervjun. Någon kontroll av att rätt person har besvarat intervjun har inte kunnat genomföras.

Bortfall

Bortfallet består dels av objektsbortfall, som innebär att frågeformuläret/intervjun inte är besvarad alls, och dels av partiellt bortfall som innebär att vissa frågor i frågeformuläret/intervjun inte är besvarade. Om bortfallet skiljer sig från de svarande, med avseende på undersökningsvariablerna, så kan skattningarna som grundar sig på enbart de svarande vara skeva. För att reducera bortfallsskevheten har vikter beräknats med hjälp av kalibrering (se bilaga).

Objektsbortfall kan bland annat bero på att uppgiftslämnaren inte är villig att delta i undersökningen, att uppgiftslämnaren inte går att nå eller att uppgiftslämnaren är förhindrad att medverka. Objektsbortfallet i denna undersökning redovisas i tabell 3 nedan.

Tabell 3 Beskrivning av objektsbortfall

	Antal
Ej avhörda	3 954
Ej anträffad	253
Avböjd medverkan	406
Förhindrad medverkan	101
Totalt	4 714

Med "Ej avhörda" menas dels att ingen uppgift om varför frågeformuläret/intervjun inte är besvarad har lämnats men även de som skickat in en tom eller delvis ifyllt enkät. Med "Avböjd medverkan" menas att SCB meddelats att uppgiftslämnaren inte vill medverka i undersökningen. "Förhindrad medverkan" innehåller avlidna, flyttat från Sverige, sjukdom och språksvårigheter. "Ej anträffad" innehåller saknad adress i Registret över totalbefolkningen (RTB), hemlig adress, tillfälligt bortrest och skyddad identitet.

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig, att uppgiftslämnaren glömmer att besvara frågan eller att instruktionerna vid hoppfrågor misstolkas. Det partiella bortfallet varierar mellan 1 och 3 procent för de flesta av frågorna.

Viktberäkning och estimation

För varje svarande person (kallas objekt i fortsättningen) har en vikt beräknats. Syftet med detta är att kunna redovisa resultat för hela populationen och inte bara för de svarande. Vikten kallas därför även för uppräkningsstal.

Vikterna har beräknats utifrån urvalsdesignen samt antaganden om objektsbortfall och täckningsfel. Beräkningen gjordes med hjälp av ett av SCB egenutvecklat SAS-makro (ETOS v2).

Vikterna i denna undersökning kan beskrivas med formeln:

$$w_k = d_k \cdot v_k$$

Där w_k = vikt/uppräkningsstal för objekt k

d_k = designvikt

v_k = kalibreringsvikt baserad på hjälpinformation

Designvikten är den del av vikten som beror på urvalsdesignen. Vid bortfall kan det vara så att vissa grupper av urvalet svarar i större utsträckning än övriga, t.ex. svarade kvinnor i högre grad än män. Om de grupper som svarat i högre grad har en annan fördelning på undersökningsvariablerna än övriga kan detta ha en snedvridande effekt på resultatet. För att kompensera för detta har kalibreringsvikter använts (se bilaga för mer detaljer).

För en mer utförligare beskrivning av kalibreringsestimatorens se Lundström och Särndal (2001): *Estimation in the Presence of Nonresponse and Frame Imperfections, Statistics Sweden*.

Vikterna/uppräkningsstalen multipliceras med objektens variabelvärden för att skapa statistikvärden gällande för populationen. Om vikterna inte används så kan resultaten bli helt missvisande. Vikterna kompenserar för objektsbortfallet men inte för det partiella bortfallet.

För beräkning av skattningen av totaler används följande formel:

$$\hat{Y} = \sum_r w_k y_k$$

där w_k = vikt/uppräkningsstal för objekt k

y_k = variabelvärde för objekt k

summering sker av de svarande (r)

Och för beräkning av skattningen av medelvärden används följande for-

$$\hat{Y} = \frac{\sum_r w_k y_k}{\sum_r w_k}$$

där w_k = vikt/uppräkningsstal för objekt k

y_k = variabelvärde för objekt k

summering sker av de svarande (r).

I denna undersökning har följande hjälpvariabler använts i kalibreringen:

X1: Stratum, dvs Ålder (15-29,30-54,55+)*Kön*Födelseland (Sverige/Övrigt)

X2: Kön*Ålder (10-års intervaller)

X3: Kön*Region (NUTS2)

X4: Civilstånd

X5: Utbildningsnivå (Tabell 5)

En fullständig beskrivning av dessa variabler finns i kalibreringsrapporten. Uppräkning har skett till Sveriges befolkning 2014-12-31, enligt registret över totalbefolkningen (RTB)

Statistikens tillförlitlighet

Ramtäckning

Täckningsfel, under- och övertäckning, innebär att urvalsram och population inte helt stämmer överens. Undertäckning innebär att vissa objekt som ingår i populationen saknas i urvalsramen. Övertäckning innebär att objekt som inte ingår i populationen ändå finns i urvalsramen. Ett sätt att minska täckningsfelen är att ha bra och uppdaterade register.

Registret över totalbefolkningen anses återspegla populationen väl, då det bygger på folkbokföringen. I denna undersökning så förekommer en viss undertäckning, då urvalet har haft en negativ sammordning med undersökningen Nationella Folkhälsan. Detta innebär att individer som ingår i urvalet till Nationella folkhälsan inte ingick i urvalsramen till denna undersökning. Genom kalibreringen så har eventuella effekter av undertäckningen minskats, då uppräkning har skett till hela populationen. Eventuella förändringar i befolkningen som skett efter urvalsdragningen har beaktats genom att kalibreringen har skett mot den senaste versionen av RTB, dvs den avspeglar populationen som den såg ut sista december 2014.

Urval

Denna kvalitetskomponent avser fel som uppkommer på grund av att endast ett urval av populationen undersöks. Urvalsfel är således den avvikelse mellan ett skattat värde och det faktiska värdet som beror på att man inte undersöker alla objekt i populationen. Urvalsfelets storlek minskar med en ökad urvalsstorlek.

Mätning

Ett fel som kan uppstå vid mätning är att lämnade uppgifter skiljer sig från faktiska uppgifter. Felet kallas mätfel och kan uppkomma då uppgiftslämnaren inte minns de faktiska uppgifterna, missförstår frågan eller medvetet svarar felaktigt.

Frågorna följer formuleringarna i metodmanualen från Eurostat, och har genomgått mätteknisk granskning av mättekniker på SCB.

Bearbetning

Vid den manuella och maskinella bearbetningen av datamaterialet kan bearbetningsfel uppstå. Exempel på bearbetningsfel är registreringsfel och kodningsfel. Dessa fel kan förhindras och upptäckas i de kontroller som genomförs vid dataregistreringen. I den här undersökningen bedöms registreringsfelet vara litet eftersom frågeblanketten endast hade fasta svarsalternativ.

Bortfall

Bortfallsfel inträffar om objekten i bortfallet och de svarande skiljer sig åt avseende frågorna/variablerna i undersökningen. Kalibrering har använts för att minska eventuella snedvridande effekter av bortfallet. En beskrivning av bortfallet finns i den bortfallsanalys som gjorts i samband med valet av hjälpinformation till kalibreringen (se bilaga).

Datafil

Observera att de bifogade vikterna (variabeln WGT) ska användas vid analys och resultatframställning. I ett stratifierat urval kan datamaterialet inte ses som ett urval av oberoende lika fördelade observationer, som ofta antas i traditionell statistisk teori. Det är därför viktigt att tänka på att de flesta förprogrammerade dataprogram inte klarar av att analysera datamaterialet ifrån en urvalsundersökning på ett korrekt sätt. Var noga med att kontrollera att de dataprogram som används för analys behandlar vikterna på ett riktigt sätt. Att ta fram frekvenser med beaktande av vikter klarar däremot de flesta statistikprogram.

Om det är partiellt bortfall (uppgift saknas) i en fråga har värdet -1 angetts. Vid hoppfrågor har värdet -2 angetts för dem som inte ska besvara följdfrågorna.

Sekretess och utlämnande

SCB har gjort en intern sekretessprövning för utlämnandet av data på mikronivå.

Dessutom har en sekretessöverenskommelse upprättats mellan Folkhälso-myndigheten och SCB med avseende på hur det avidentifierade data-materialet får hanteras.

Bilagor

Frågeformulär

Kalibrerings- bortfallsrapport