

Nationella folkhälsoenkäten 2020

Nationellt urval. Teknisk rapport – En beskrivning av genomförande och metoder.

Om publikationen

Detta är en teknisk rapport som beskriver genomförande och metoder till den nationella folkhälsoenkäten, Hälsa på lika villkor 2021. Rapporten är ursprungligen framtagen av Statistiska centralbyrån i juni 2020 (2020-06-09) med beteckning, 894968-7/20066. Därefter har den tillgänglighetssanpassats av Folkhälsomyndigheten.

Bakgrund

Den här rapporten beskriver hur SCB har genomfört uppdraget ”Hälsa på lika villkor” (även kallad Nationella folkhälsoenkäten) och vilka metoder som har använts.

Om du har frågor är du välkommen att höra av dig till den ansvariga på SCB: Erika Bergentz, telefon 010-4796501 eller e-post erika.bergentz@scb.se.

Uppdragets syfte: Syftet med undersökningen var att ta reda på hur befolkningen i Sverige mår och lever.

Uppdragets parter: Enkätenheten vid SCB genomförde under perioden februari–juni 2020 en enkätundersökning på uppdrag av Folkhälsomyndigheten.

Kontaktpersoner: Undersökningsledare på SCB var Erika Bergentz och Michael Franzén var metodstatistiker. Namnet på Folkhälsomyndighetens kontaktperson gentemot SCB var Malin Kark.

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

Leverans: Resultatet i form av datafiler levererades vecka 26 enligt överenskommelse.

Urval: Det nationella urvalet bestod av 40 000 personer varav 57 personer utgjorde övertäckning. Det innebär att urvalet, där övertäckningen är borträknad, var 39 943 personer. Det var totalt 16 907 personer som besvarade frågeblanketten, vilket var 42,3 procent av urvalet (övertäckning borträknad).

I årets undersökning genomfördes dessutom ett experiment där 1 700 personer i åldern 16–44 år fick en förkortad enkät. Analysen av experimentet redovisas i en separat rapport.

Tilläggsurval: I årets undersökning deltog Region Västmanland, Region Gotland, Region Värmland, Region Östergötland och Västra Götalandsregionen med tilläggsurval. Separata rapporter finns för dessa tilläggsurval samt för den totalfil som innehåller undersökningens alla svar.

Genomförande och metoder

Population och urval

Populationen, det vill säga de objekt som man vill kunna dra slutsatser om, utgjordes av samtliga personer i Sverige i åldern 16–84 år.

För att kunna dra ett urval från populationen skapades en urvalsram som avgränsade och identifierade objekten i populationen. Urvalsramen i

undersökningen skapades med hjälp av data från Registret över totalbefolkningen, version 2019-11-30. Antalet personer i urvalsramen var 8 091 396.

Från urvalsramen drogs ett obundet slumpmässigt urval om 40 000 personer (nationella urvalet).

Ett obundet slumpmässigt urval innebär att alla objekt i urvalsramen har samma sannolikhet att komma med i urvalet.

Urvalsstorleken bestämdes av Folkhälsomyndigheten.

Blankettfrågor och registeruppgifter

Folkhälsomyndigheten utformade frågorna i frågeblanketten. Grundläggande för bra kvalitet i en undersökning är kvaliteten på de data som samlas in. För att minska risken för mätfel har blanketten därför genomgått ett mättekniskt test i form av en mätteknisk standardgranskning. Det innebär att mättekniker på SCB gjort en djupgående granskning av frågeblankettens disposition, frågor och svarsalternativ, för att identifiera problem, analysera konsekvenser och ta fram utvecklade förslag på lösningar. Granskningen finns dokumenterad i en skriftlig rapport.

Blanketten bestod av 58 numrerade frågor, flera av dem hade delfrågor vilket genererade totalt 103 frågor. Frågorna handlade bland annat om hälsa, levnadsvanor, ekonomiska förhållanden, arbete och sysselsättning samt trygghet och sociala relationer.

Angående referensperiod för frågorna, se bilaga 1b. Med referensperiod menas vilken tidpunkt svaren avser, till exempel idag, förra veckan, senaste tre månaderna.

Förutom de variabler som samlades in via frågeblanketten hämtades ett antal variabler från register. Vilka dessa variabler är och aktualitet framgår av bilaga 3.

Datainsamling

Innan utskick gjordes en kontroll av personerna i urvalet mot de senaste folkbokföringsuppgifterna, för att ta fram aktuella adressuppgifter. Vid kontrollen och under insamlingsfasen framkom det att 57 personer inte längre tillhörde populationen utan utgjorde känd övertäckning, det vill säga den övertäckning som identifierats. Orsakerna till övertäckning var att personerna emigrerat eller avlidit.

Tabell 1. Känd övertäckning för nationella urvalet.

Population	Urval	Övertäckning	Urval exklusive övertäckning
8 091 396	40 000	57	39 943

Totalt fyra utskick gjordes. Samtliga utskick gjordes med post. I ett informationsbrev ombads urvalspersonerna att besvara frågorna och skicka svaret

till SCB. I informationsbrevet kunde urvalspersonerna läsa om undersökningens bakgrund, syfte och att undersökningen genomfördes i samarbete mellan Folkhälsomyndigheten och SCB. Informationsbrevet informerade också om att uppgifter hämtades från SCBs register och att en avidentifierad datafil levereras till Folkhälsomyndigheten och respektive region. Brevet informerade även om dataskyddslagstiftningen samt offentlighets- och sekretesslagen och att det var frivilligt att medverka i undersökningen.

Det första utskicket i form av ett informationsbrev med inloggningsuppgifter genomfördes den 3 februari 2020. Därefter skickades en frågeblankett ut den 19 februari. Sedan skickades två påminnelser till dem som inte besvarat frågeblanketten. Ett påminnelsebrev skickades den 11 mars och en påminnelse med ny frågeblankett skickades den 30 mars. Insamlingen avslutades den 5 maj 2020.

Vid samtliga utskick fick uppgiftslämnaren information om att frågorna även kunde besvaras på webben samt inloggningsuppgifter.

Totalt svarade 16 907 personer på frågeblanketten, vilket var 42,3 procent av urvalet efter att den kända övertäckningen, det vill säga den övertäckning som identifierats, tagits bort (se tabell 1). I tabell 2 redovisas inflödet efter varje utskick som gjordes.

Tabell 2. Beskrivning av inflödet. Antal och andel svar.

Utskicksomgång	Antal	Andel
Efter första utskicket	5 802	14,5
Efter andra utskicket	6 753	16,9
Efter tredje utskicket	2 124	5,3
Efter fjärde utskicket	2 228	5,6
Totalt antal svar	16 907	42,3
Bortfall	23 036	57,7
Urval exklusive övertäckning	39 943	100,0

Det var 9 289 personer som valde att besvara webblanketten, vilket var 54,9 procent av samtliga svarande. I tabell 3a redovisas inflödet uppdelat på svarssätt och i tabell 3b uppdelat på svarssätt och åldersgrupp.

Tabell 3a. Beskrivning av inflödet uppdelat på svarssätt. Antal och andel svar.

Via papper	Antal	Andel
Efter första utskicket	0	0,0
Efter andra utskicket	5 171	30,6
Efter tredje utskicket	881	5,2
Efter fjärde utskicket	1 566	9,3
Totalt antal svar via papper	7 618	45,1
Via webben	Antal	Andel
Efter första utskicket	5 802	34,3
Efter andra utskicket	1 582	9,4
Efter tredje utskicket	1 243	7,4
Efter fjärde utskicket	662	3,9
Totalt antal svar via webben	9 289	54,9
Svar totalt papper + webben	16 907	100,0

Tabell 3b. Beskrivning av inflödet uppdelat på svarssätt och åldersgrupp. Antal och andel svar.

Åldersgrupper	Antal	Andel
Papper 16-29 år	585	27,2
Webben 16-29 år	1 565	72,8
Totalt 16-29 år	2 150	100,0
Papper 30-44 år	1 024	33,3
Webben 30-44 år	2 047	66,7
Totalt 30-44 år	3 071	100,0
Papper 45-64 år	2 358	42,0
Webben 45-64 år	3 250	58,0
Totalt 45-64 år	5 608	100,0
Papper 65-84 år	3 651	60,1
Webben 65-84 år	2 427	39,9
Totalt 65-84 år	6 078	100,0

De frågeblanketter som besvarats på papper registrerades med hjälp av skanning.

Kontroller

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

De inkomna pappersblanketterna granskades i flera olika skeden av bearbetningen.

Handskrivna tecken kontrollerades efter skanningen och tecken som tolkats felaktigt korrigerades.

Vid frågor där svaret skulle anges i form av skrivna siffror, gjordes en rimlighetsbedömning. Orimliga värden togs ut för kontroll i syfte att säkerställa att rätt värde registrerats.

Under datainsamlingen granskades otillåtna värden löpande, till exempel om det fanns frågor där uppgiftslämnare markerat flera svar trots att frågor skulle besvaras med ett alternativ. Hur otillåtna värden hanterats beskrivs i avsnittet Datafilens innehåll.

Datamaterialet granskades före leverans med hjälp av frekvenstabeller för att upptäcka och korrigera eventuella kvarstående fel.

Kontroll av att rätt person har besvarat frågeblanketten har skett genom att svar på frågan om födelseår har jämförts med motsvarande registeruppgift. SCB kan ändå inte garantera att den utvalda personen själv besvarat frågeblanketten.

Bortfall

Bortfallet består dels av objektsbortfall, som innebär att frågeblanketten inte är besvarad alls, och dels av partiellt bortfall som innebär att vissa frågor i blanketten inte är besvarade. Om bortfallet skiljer sig från de svarande, med avseende på undersökningsvariablerna, kan skattningarna som grundar sig på enbart de svarande vara skeva. För att reducera bortfallsskevheter har vikter beräknats med hjälp av kalibrering (se bilaga 2).

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 4 nedan.

Tabell 4. Beskrivning av objektsbortfall.

Orsak	Antal
Ej avhörda	21 926
Avböjd medverkan	325
Förhindrad medverkan	106
Ej anträffad	532
Insänd blank eller oanvändbar blankett	21
Fel person har svarat	126
Totalt	23 036

Med ”Ej avhörda” menas att ingen uppgift om varför frågeblanketten inte är besvarad har lämnats. Med ”Avböjd medverkan” menas att SCB meddelats att uppgiftslämnaren inte vill medverka i undersökningen. Med ”Förhindrad medverkan” menas att uppgiftslämnaren av olika skäl inte kan medverka i undersökningen (exempelvis pga. sjukdom eller språksvårigheter). ”Ej anträffad”

innehåller personer som saknar adress i Registret över totalbefolkningen (RTB), har hemlig adress, är tillfälligt bortresta samt postreturer. Med "Fel person har svarat" menas att någon annan än den person som frågeblanketten var avsedd för har svarat.

Partiellt bortfall kan bero på att en fråga är svår att förstå, är känslig, att uppgiftslämnaren missar att besvara frågan eller att instruktionerna vid hoppfrågor misstolkas. Det partiella bortfallet för respektive fråga redovisas i bilaga 6.

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 objektbortfall och täckningsfel. Beräkningen gjordes med hjälp av ett av SCB egenutvecklat SAS-makro (ETOS).

Vikterna i denna undersökning kan beskrivas med formeln som redovisas i Figur 1.

Figur 1. Formel för att beräkna vikter i undersökningen.

$$w_k = d_k \cdot v_k$$

där w_k = vikt/uppräkningsstal för objekt k

d_k = designvikt för objekt k

v_k = kalibreringsvikt för objekt k 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. kan kvinnor svara i högre grad än män. Om de grupper som svarat i högre grad har en annan fördelning med avseende 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). Vikterna bygger på antagandet att ramen återspeglar populationen väl och därmed att över- och undertäckningen är försumbar.

För en mer utförlig beskrivning av kalibreringsestimatorn 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 objektbortfallet, men inte för det partiella bortfallet.

För beräkning av skattningar av totaler används följande formel (se Figur 2):

Figur 2. Formel för att beräkna skattningar av totaler i undersökningen.

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

där w_k = vikt/uppräkningsstal för objekt k
 y_k = variabelvärde för objekt k

Summering sker av de svarande r . För beräkning av skattningar av medelvärden används följande formel (Figur 3).

Figur 3. Formel för att beräkna skattningar av medelvärden i undersökningen.

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

där w_k = vikt/uppräkningsstal för objekt k
 y_k = variabelvärde för objekt k

Summering sker av de svarande r . Vikten kalvikt i svarsdatafilen, motsvarande w_k ovan, ska användas vid framtagning av resultat. Summan av vikten över alla svarande motsvarar antalet personer i urvalsramen, 8 091 396.

Datats tillförlitlighet

Tillförlitlighet totalt

Statistiken är behäftad med osäkerhet. Vid bedömning av hur olika osäkerhetskällor påverkar statistiken från en undersökning skiljer man på slumpmässiga och systematiska avvikelser. Slumpmässiga avvikelser orsakar enligt statistisk teori beräkningsbar osäkerhet hos de skattade resultaten. Systematiska avvikelser påverkar resultaten i en viss riktning. Det är svårt att ange hur tillförlitligheten påverkas av eventuella systematiska avvikelser (skevhet) då det ofta kräver resurskrävande utvärderingsinsatser. En sådan utvärdering har inte gjorts för detta uppdrag.

Med total osäkerhet avses den sammanlagda osäkerheten = slumpmässiga + systematiska avvikelser. Felmarginalen för andelsskattningar av målstorheter kring 50 procent (inklusive urvalsfel och bortfallsjustering) för hela populationen ligger runt 0,8 procentenheter, förutsatt att det systematiska felet från bortfallet (skevheten) är liten eller försumbar.

Osäkerhetskällor

Bortfallet bedöms vara den största källan till osäkerhet då dess påverkan på skevheten är okänd. Användningen av kalibrering bedöms dock minska skevheten för aktuella redovisningsgrupper. Urvalsfelet är begränsat åtminstone på totalnivå,

även med bortfalljusterade vikter. Täckningsfel, bearbetningsfel och mätfel bedöms vara litet till försumbart i sammanhanget.

Urval

Denna kvalitetskomponent avser osäkerhet som uppkommer på grund av att endast ett urval av populationen undersöks. Urvalsosäkerheten är således den avvikelse mellan ett skattat värde och det faktiska värdet som beror på att man inte undersöker alla objekt i populationen. Urvalsosäkerheten minskar med en ökad urvalsstorlek. Urvalsfelet bör beaktas när man drar slutsatser från undersökningen. Felmarginalen (oaktat bortfallsfelet) är 0,5 procentenheter för andelsskattningar kring 50 procent.

Ramtäckning

Under- och övertäckning innebär att ram- och målpopulation inte helt stämmer överens. Undertäckning innebär att vissa objekt som ingår i målpopulationen saknas i ramen. Övertäckning innebär att objekt som inte ingår i målpopulationen ändå finns i ramen. Ett sätt att minska täckningsbrister är att använda bra och uppdaterade register för att ta fram ramar.

Registret över totalbefolkningen (RTB) uppdateras dagligen via aviseringar från Skatteverket om födöfall, dödsfall, flyttningar inom landet, in- och utvandring. Täckningsbrister beror dels på den undertäckning som uppstår då personer bosatta i Sverige saknas i ramen och dels på övertäckning som t.ex. föreligger då personer i ramen inte längre ingår i landets befolkning. Den undertäckning som finns beror främst på att information om immigranter (invandrare och hemvändare) uppdateras med viss eftersläpning. Effekten på statistiken bedöms dock vara högst obetydlig. Övertäckningen beror framför allt på eftersläpning i rapporteringen av dödsfall och utvandring. Dessa personer skulle helst inte ingå i ramen och bristen ifråga upptäcks oftast inte heller eftersom de inte nås. Därigenom blir de felaktigt klassade som bortfall trots att kategorin övertäckning är den lämpliga. RTB:s uppdateringsrutiner medför att ramen vid urvalsdragningen bedöms ha obetydlig övertäckning. Bristerna leder sammantaget till en liten och oönskad skevhet i statistiken. Sammanfattningsvis är SCB:s bedömning att bristerna i ramen, både med avseende på under- och övertäckning, inte snedvrider statistiken i någon nämnvärd omfattning.

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. För att försöka minska mätfelen har frågorna granskats och/eller testats av mättekniker vid

SCB. Ingen ytterligare bedömning om förekomst av eventuella mätfel eller deras storlek har gjorts.

Bortfall

Skevheter i skattningar på grund av bortfall inträffar om objekten i bortfallet och de svarande skiljer sig åt avseende frågorna/variablerna i undersökningen.

Bortfallet är relativt stort och bedöms kunna ge systematiska fel för skattningar om inte bortfallsjustering görs. Eventuella bortfallsfel kan reduceras genom att på ett effektivt sätt justera undersökningens vikter. En beskrivning av förfarandet finns i den bortfallsanalys som gjorts i samband med valet av hjälpinformation till kalibreringen (se bilaga 2).

Bearbetning

Vid den manuella och maskinella bearbetningen av data kan osäkerhet uppstå. Exempel på bearbetningsbrister är registreringsfel och kodningsfel. Dessa kan förhindras och upptäckas i de kontroller som genomförs vid dataregistreringen. I den här undersökningen bedöms osäkerheten på grund av registrering vara litet eftersom frågeblanketten endast hade fasta svarsalternativ.

För frågor där urvalspersonen ska svara med siffror eller bokstäver i särskilda fält (exempelvis 23 och 41) har en så kallad massverifiering (rättning) gjorts. Det innebär att alla tolkade tecken (bokstäver/siffror) klipps ut och visas uppgradade från 0-9 respektive A-Ö. Om några tecken ser ut att vara feltolkade kontrolleras de särskilt av den person som gör rättningen och kan ändras så att rätt bokstav/siffra registreras. Detta är ett manuellt moment och risken för registreringsfel för den typen av svar bedöms vara något större än för frågor med fasta svarsalternativ.

Modellantagande

Inga andra modellantaganden har gjorts utöver de som redovisats ovan.

Beskrivning av datafil

Datafilens innehåll

Svarsdatafilen kompletterades med vikter. De bifogade vikterna är avsedda att användas vid skattning av målstorheter i populationen, exempelvis totaler och medelvärden.

Vid analyser av samband med hjälp av exempelvis regressionsanalys bör det uppmärksammas att det inte är givet hur vikterna ska hanteras. Urvalets design samt bortfall är komplikationer som kan störa analyser av samband.

Om det är partiellt bortfall (uppgift saknas) i en fråga är värdet blankt. Vi har inte tagit ställning till hur eventuella motsägelsefulla svar ska behandlas. Sådana fel kan

uppstå vid t.ex. följdfrågor och måste undersökas innan den typen av frågor analyseras. I rättningsinstruktionerna (bilaga 4) finns en beskrivning av vad som korrigerats redan vid registreringen och i dokumentet "Upprättningar vid avslut" (bilaga 5) framgår vad som korrigerats då den slutliga datafilen framställdes.

Statistiska mått

Materialet lämpar sig att redovisas som totaler eller procentuella andelar för hela målpopulationen eller fördelat på olika redovisningsgrupper.

Redovisningsgrupper

Ofta redovisas statistik inte bara för hela populationen utan också för delgrupper (redovisningsgrupper). Redovisningsgrupper i den här undersökningen kan avgränsas med hjälp av registervariabler eller bakgrundsfrågor i frågeblanketten.

Jämförbarhet över tid

SCB har genomfört den nationella undersökningen åt Folkhälsomyndigheten (tidigare Statens folkhälsoinstitut) årligen sedan 2004.

Undersökningarna genomfördes på ungefär liknande sätt med avseende på urvalsmetod, urvalsstorlek, insamlingsmetod och antal påminnelser. Dock genomfördes en del experiment 2011 och 2012, vilket innebar att olika kontaktstrategier användes för olika grupper dessa år. 2013 gjordes en förändring jämfört med tidigare år då ett förbrev skickades ut och den sista (tredje) påminnelsen togs bort. Dock skickades en tredje enkätpåminnelse ut till enbart en del av urvalet 2013. Därefter har enbart två påminnelser skickats ut. 2016 gjordes ytterligare ett experiment där ordinarie strategi med förbrev jämfördes med webboffensiv strategi (dvs. istället för förbrev skickades ett missiv med inloggningsuppgifter). Ingen signifikant skillnad gällande svarsfrekvens fanns mellan de båda strategierna och eftersom webboffensiv strategi är ekonomiskt lönsam valdes denna strategi i 2018 års undersökning.

2015 och 2016 gjordes relativt många förändringar i frågeblanketten då flera frågor togs bort och en del frågor lades till. Tidigare år och inför 2018 har enbart ett fåtal frågor ändrats mellan de olika åren. Folkhälsomyndigheten för dokumentation över vilka frågor som återkommit respektive tagits bort de olika åren.

Sekretess och utlämnande

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

Bilagor

- Bilaga 1a: Frågeblankett
- Bilaga 1b: Missiv med inloggningsuppgifter
- Bilaga 1c: Missiv som följde med pappersenkät
- Bilaga 1d: Påminnelsebrev
- Bilaga 1e: Påminnelsemissiv
- Bilaga 2: Kalibreringsrapport
- Bilaga 3a: Postbeskrivning
- Bilaga 3b: Kod och klartexter för vissa registervariabler
- Bilaga 3c: Förteckning över län- och kommunkoder
- Bilaga 4: Rättningsinstruktion
- Bilaga 5: Upprättningar vid avslut
- Bilaga 6: Partiellt bortfall
- Bilaga 7: Urvalsdokumentation