

Stöddokument i SPS

# Anvisningar för uppgifts- lämnargranskning i webblanketter och filer

## Inledning

Syftet med anvisningarna är att ge stöd vid utformning och utvärdering av uppgiftslämnarkontroller (ul-kontroller). Målgruppen för anvisningarna är i huvudsak den grupp kompetenser som samverkar i undersökningen (insamling, ämne och metod). Som stöd för utvärdering finns en checklista att använda i arbetet, se SPS process 2.4.2 "Checklista för utvärdering av uppgiftslämnargranskning i webblanketter och filer – förberedelse och genomförande". Checklistan kan dateras och med fördel sparas för det kontinuerliga arbetet med ul-kontroller över tid.

En viktig utgångspunkt för SCB:s arbete med datafångst och granskning är att det är effektivt samtidigt som vi minskar uppgiftslämnarbördan. Det innebär att vi har ett gott samarbete med våra uppgiftslämnare (ul) - ett samarbete som leder till att viktiga uppgifter i blanketter och filer så långt det är möjligt är rätt redan när de skickas in till SCB. Ett sådant samarbete måste bygga på ömsesidig tillit. Det innebär t.ex. att när ul godkännt en uppgift så ska den hanteras som klar i efterföljande steg i granskningsprocessen. Processteget "Uppgiftslämnare granskar" handlar därför om hur vi kan underlätta för ul att lämna rätt uppgifter så att de inte gör fel som riskerar att påverka statistiken. Ul-kontrollerna ska i det här sammanhanget ses som ett komplement till frågor och instruktioner som ska vara utformade för att vägleda uppgiftslämnaren att svara korrekt.

Det är viktigt att utgå både från statistikens ändamål med tillhörande kvalitetskrav och ul:s perspektiv vid val och utformning av ul-kontroller. Det innebär t.ex. att ett fåtal väl fungerande kontroller för de viktigaste variablerna i en blankett är att föredra framför ett större antal kontroller som faller ut för ofta och riskerar att skapa irritation hos ul. Att ha ett makroperspektiv med selektiva inslag är strävan redan i valet och utformningen av ul-kontroller och fokus ska ligga på de viktigaste variablerna och de viktigaste redovisningsgrupperna. Samma fokus gäller även för de andra stegen i granskningsprocessen.

För att kunna bedöma vilka kontroller som ska finnas i ul-granskningen och vilka som ska finnas i andra steg i granskningsprocessen behöver samtliga kontroller i undersökningen utvärderas som en helhet. Det är ett arbete som behöver göras för varje enskild undersökning. Rekommenderat arbetssätt för hur kontroller ska utformas och utvärderas beskrivs i nedanstående avsnitt. Arbetssättet gäller både pågående och nya undersökningar.

## **1 Utforma ul-kontroller**

### *Pågående och nya undersökningar*

För att kunna lämna in korrekta uppgifter behöver ul förstå vad som efterfrågas i undersökningen, ta fram uppgifterna, och lämna in dem i det format som är tänkt. Stor vikt behöver läggas vid att blanketten är väl genomtänkt med tydliga frågor som går att förstå och svara på. Blanketten ska ha en logisk struktur/frågeordning, tydlig visuell design och hjälpmedel i form av instruktioner och ul-kontroller.

Det är även viktigt att tänka sig in i ul:s situation vid uppgiftslämnandet. Vilka system måste ul använda? Hur mycket tid behöver ul lägga på uppgiftslämnandet, vilka andra arbetsuppgifter brukar ul ha samtidigt t.ex. bokslutsarbete? Finns det andra personer ul måste vända sig till för att få uppgifterna? Utformningen av ul-kontroller och meddelanden är beroende av den förståelsen.

### **1.1. Val av ul-kontroller**

Forskning visar att ul vill bli informerade om fel eller oklarheter i data de lämnar. De förväntar sig att vissa grundläggande kontroller görs i samband med själva besvarandet så att de kan justera uppgifter som inte är korrekta. Kontroller ska hjälpa ul att lämna rätt uppgifter, därför är det viktigt att de kontroller som läggs ut är genomtänkta. Ul-kontroller får inte bli en ytterligare börda i uppgiftslämnandet.

Syftet med kontroller och kommentarer är att upptäcka fel. Avsikten är inte att upptäcka den variation eller förändring som statistiken ska beskriva, det är blankettfrågornas uppgift. Omvärldsbevakning ligger utanför granskningen.

I bedömningen av en kontrolls lämplighet som ul-kontroll spelar träffsäkerheten in. Med hög träffsäkerhet menas att en kontroll, om den faller ut, i de flesta fall leder till en ändring av ett eller flera värden. Om kontrollen finns i produktionsgranskningen så kan träffsäkerheten beräknas från tidigare omgångar. I annat fall efter avslutad insamlingsomgång, se avsnitt 3.2. ”Beräkna träffsäkerheten för ul-kontroller i SIV”.

### 1.1.1. Kontroller lämpliga som ul-kontroller

- Kontroller som identifierar uppenbara fel, som inte kunnat åtgärdas fullt ut med instruktioner och layout:
  - uppgiften saknas (partiellt bortfall) för variabler där värden måste finnas
  - ogiltigt värde, t.ex. att det registrerade värdet inte tillhör variabelns värdeförråd, t.ex. yrkeskod
  - inkonsistens (motsägelser) mellan olika uppgifter, t.ex. att ett svar på en fråga motsäger svar på en eller flera andra frågor, eller att summan av två variabler inte är lika med en totalsumma.
- Kontroller som identifierar misstänkta avvikelser med hög träffsäkerhet:
  - rimlighets- och värdemängdskontroller där gränsvärden inte har stora variationer mellan t.ex. olika branscher, eller där det är möjligt att anpassa gränserna för olika grupper.

### 1.1.2. Kontroller olämpliga som ul-kontroller

- Rimlighets- och värdemängdskontroller där gränsvärden kan variera mycket mellan t.ex. olika branscher eller storleksklasser. Om det finns behov av att situationsanpassa kontrollen, och den tillgängliga bakgrundsinformationerna för att styra kontrollen inte resulterar i en tillräckligt bra träffsäkerhet är det ett tecken på att kontrollen inte passar i ul-granskningen.
- Kontroller som är för komplexa eller för svåra att kommunicera till ul för att de ska förstå vad som ska kontrolleras. Den typen av fel bör i stället åtgärdas genom att förtydliga i blanketten eller åtgärdas genom automatändring eller tas om hand i makrogranskningen.
- Kontroller där en uppgift jämförs med ett värde som av någon anledning inte kan förtryckas i blanketten eller i felmeddelandet. För att kunna förstå och ta ställning till kontrollen behöver ul se vad uppgiften jämförs med och var det värdet kommer ifrån, t.ex. en annan undersökning eller datakälla.
- Kontroller som används för att upptäcka förändringar i datamaterialet som inte beror på felrapportering, dvs. den variation eller förändring som statistiken ska beskriva. Det är blankettfrågornas uppgift. Omvärldsbevakning ligger utanför granskningen.

## 1.2. Hårda eller mjuka ul-kontroller

Det finns tre olika varianter av kontroller som ska användas utifrån om det är uppenbara eller misstänkta fel som ul ska hantera.

### 1.2.1. Hårda ul-kontroller

**Hård ul-kontroll** ska användas till att identifiera uppenbara fel. Det kan vara vid partiellt bortfall i fält som måste besvaras, värden som har fel format samt icke-valida värden på variabler av stor betydelse. Inget annat än acceptabelt värde accepteras, annars kan blanketten inte skickas in.

**Hård ul-kontroll med följdfråga eller fritextkommentar** innebär att värdet kan accepteras av ul med hjälp av en förklarande kommentar. Den används för att identifiera misstänkta avvikelser där en flagga i mikrogranskningen annars skulle leda till en återkontakt. Kommentarer från ul ska inte användas för att upptäcka den variation eller förändring som statistiken ska beskriva. Omvärldsbevakning ligger utanför granskningen.

- **Hård kontroll med följdfråga** innebär att ul med hjälp av fasta svarsalternativ kan meddela varför uppgiften stämmer. Följdfrågan kommer upp om ul fortsätter till nästa sida utan att ändra sitt svar. Förutom de fasta alternativen ska ul erbjudas alternativet "Annat" eller "Annat, skriv vad:" med möjlighet att själv få formulera en anledning. Ni väljer själva om det ska vara med eller utan möjlighet att kommentera med fritext. Valet beror på om fritextsvaret är viktigt i de efterföljande granskningsstegen eller inte. Tanken är att de fasta alternativen kommer att minska ul:s behov av att använda svarsalternativet "Annat, skriv vad".

Fritextsvaret används endast om värdet faller ut i den selektiva granskningen, om den finns, annars vid behov i makrogranskningen. Det innebär att inte alla fritextsvar behöver tas om hand med automatik.

Tänk på att:

- Det ska framgå av följdfrågan vilken kontroll som avses och vad ul förväntas göra, se avsnitt 1.3. "Granskningsmeddelandet"
- Svarsalternativen ska utformas utifrån den specifika undersökningen, utgå om möjligt från fritextsvar som lämnats i samband med följdfrågan.
- Svarsalternativen ska vara varandra uteslutande
- Använd radioknappar om endast ett svarsalternativ är aktuellt
- Använd checkboxar om flera svarsalternativ kan vara aktuella
- Begränsa antalet svarsalternativ till max sju inkl. fritextsvaret.

Ul-kontroll med följdfråga och fasta svarsalternativ ska alltid användas där det är tekniskt möjligt.

- **Hård kontroll med fritextkommentar** kräver manuell hantering av både ul och SCB och ska därför endast användas där följdfråga med

fasta svarsalternativ inte är tekniskt möjligt, som i automatiska loopar och dynamiska tabeller.

### 1.2.2. Mjuk ul-kontroll

Mjuk ul-kontroll ska användas för att identifiera misstänkta avvikelser med hög träffsäkerhet där inte finns användning av ul:s kommentarer för hantering i efterföljande granskningssteg. De används främst för att signalera misstänkta värden, t.ex. när ett värde ligger utanför ett visst intervall (avvikelsefel). I vissa fall kan mjuka kontroller även användas för uppenbara fel om man vill minska risken för objektsbortfall till följd av för mycket felsignaleringar. Det gäller främst frivilliga undersökningar där det finns risk att ul tröttnar om de inte kan skicka in uppgifterna utan att först ha åtgärdat fel.

Mjuka kontroller ska inte kommenteras med fritext eller följdfråga utan hanteras som klar i efterföljande steg i granskningsprocessen när ul godkänt uppgiften genom att inte ändra den.

### 1.3. Granskningsmeddelandet

Meddelanden ska ses som en del i SCB:s dialog med ul. Syftet är att stödja ul i arbetet med att lämna så korrekta uppgifter som möjligt. Det innebär att meddelandet ska vara lättläst och tydligt, att ul lätt kan förstå vad som ska kontrolleras. Om kontrollen kräver ett långt och invecklat meddelande så är det ett tecken på att kontrollen är för komplex och inte ska läggas ut.

I meddelandet ska ul få svar på följande frågor:

- Varför har jag fått upp kontrollen?
- Vad ska jag göra?
- Hur ska jag göra det?

#### Så här formulerar du granskningsmeddelandet:

- Använd aktiv form och använd samma begrepp som i frågorna. Om olika termer används i meddelanden och frågor finns en stor risk för feltolkning.
- Var tydlig med vad misstanken avser, t.ex. vilka frågenummer eller variabler som det gäller och vad ul förväntas göra, t.ex. kontrollera ett misstänkt värde, eller fylla i ett fält som måste fyllas i.
- När en följdfråga med fasta svarsalternativ ska formuleras är det viktigt att det framgår vilken kontroll som avses och vad ul förväntas göra. Ett exempel:

**Det lämnade värdet på *Handelsvaror* på föregående sida har mer än halverats jämfört mot det lämnade värdet föregående kvartal. Vad är anledningen till det?**

- När en fritextkommentar efterfrågas ska det vara tydligt att ul ska lämna en förklarande kommentar om varför uppgiften avviker från SCB:s förväntan. Risken är annars att ul skriver "Det stämmer" eller "Ok" och en återkontakt måste tas för att reda ut orsaken.

Ett förslag: ”Ändra uppgiften eller lämna en förklarande kommentar om varför uppgiften stämmer.”

#### **1.4. Placera ul-kontroller nära svaren**

Ul-kontrollen ska ske direkt när svar på frågan/frågorna lämnas och granskningsmeddelandet ska visas i så nära anslutning till svaren som möjligt. Studier visar att ul föredrar att få resultat av kontroller direkt efter att frågan är besvarad i stället för att få en lista med alla granskningsmeddelanden i slutet av blanketten. Är det en sambandskontroll mellan variabler, t.ex. en kontroll av konsistens mellan helheten och olika delar, ska granskningsmeddelandet ligga i så nära anslutning till den senast ifyllda uppgiften som möjligt, inte överst eller nederst på sidan.

#### **1.5. Testa ul-kontroller och meddelanden**

Testa nya ul-kontroller och meddelanden innan kontrollerna läggs ut genom att låta någon som inte är så insatt i undersökningen besvara blanketten (ett användbarhetstest). Det är ett effektivt sätt att se om kontroller förstås på ett korrekt sätt.

## **2 Inventera och utvärdera befintliga kontroller**

### *Pågående undersökningar*

Som ett led i SCB:s arbete med ständiga förbättringar ska samtliga kontroller i en undersöknings granskning inventeras och utvärderas regelbundet. Förslagsvis sker detta efter genomförd insamlingsomgång i årliga och intermittenta undersökningar och en gång per år i månads- och kvartalsundersökningar. Inventering och utvärdering är också viktigt vid designförändringar som påverkar granskningen.

Arbetet sker ur ett helhetsperspektiv så att ul-kontrollerna kan sättas i relation till andra kontroller i granskningsprocessen (ul-kontroller ska till exempel inte återkomma i eventuell produktionsgranskning). För att få en översikt är ett förslag att lista och ordna kontrollerna i granskningsprocessen utifrån de mål- och observationsvariabler de berör.

Processdata, ämneskunskaper och erfarenheter från uppgiftslämnarkontakter (till exempel via uppgiftslämnarservice) utgör underlag för utvärderingen. Se avsnitt 3 om processdata i SIV. För översyn av eventuella kontroller i produktionsgranskningen finns anvisningarna ”Översyn av produktionsgranskning” som finns i Statistikproduktionsstödet process 5.2 ”Granska mikrodata”. Se även ”Handledning för outputgranskning” som finns i Statistikproduktionsstödet process 6.2 ”Granska makrodata”

## 2.1. Att tänka på vid utvärdering av ul-kontroller i granskningen

- Är kontrollerna utformade utifrån undersökningens statistikändamål och tillhörande kvalitetskrav?
- Har undersökningens statistikändamål och tillhörande kvalitetskrav ändrats sedan föregående utvärdering?
- Är det tydligt vad som ska besvaras och hur? En ul-kontroll får inte ersätta en instruktion. Hantera t.ex. tusenfel genom att lägga till "000 kr" i anslutning till svarscellen i blanketten.
- Om ul-kontrollen är alltför komplex, se om det i stället går att förtydliga i blanketten eller åtgärda genom automatändring eller ta hand om i makrogranskningen. Se över vilka ul-kontroller som eventuellt är komplexa och utvärdera layout och instruktioner i blankett och fil tillsammans med en mättekniker.
- Om ul-kontrollen faller ut ofta kan det vara ett tecken på att ul har svårt för att besvara frågan. Går det att ändra frågan eller förtydliga med instruktioner för att underlätta för ul?
- Gruppera ul (objekt) och anpassa kontroller per grupp. Eventuellt även till enskilda objekt, t.ex. företag som har lämnat uppgifter tidigare.
- Är acceptansgränserna i ul-kontrollerna aktuella?
- Har ul-kontrollerna hög träffsäkerhet? Med hög träffsäkerhet menas att en kontroll, om den faller ut, i de flesta fall leder till en ändring av ett eller flera värden. Rekommendationen är en träffsäkerhet på minst 80% (8 av 10 ul som får upp kontrollen rättar värdet). Gäller både hårda och mjuka ul-kontroller. Om kontrollen finns i produktionsgranskningen så kan träffsäkerheten beräknas från tidigare omgångar. Utgå annars från avsnitt 3.2. "Beräkna träffsäkerheten för ul-kontroller i SIV".
- Finns det kontroller där det mer handlar om omvärldsbevakning, dvs. att upptäcka den variation eller förändring som statistiken är tänkt att beskriva. Ta i så fall bort kontrollen.
- Se över svarsalternativen i hårda ul-kontroller med följdfråga så att de är aktuella. Saknas det något svarsalternativ? t.ex. ett svar som många ul har skrivit i fritextrutan.

## 3 Processdata i SIV

Vid utvärdering av ul-kontroller kan processdata från SIV användas. Det går att se om ett granskningsmeddelande har aktiverats under uppgiftslämnandet. I dagens system går det dock inte att med säkerhet fastställa hur ul har reagerat på meddelandet. Men en kontrollers träffsäkerhet kan approximeras genom en analys av flaggningsdata från SIV och de svar som ul skickar in till SCB. Den analysen behöver ske utanför SIV.

### 3.1. Flaggingar i SIV

I dagens system går det endast att lagra de granskningsflaggingar som sker i webblancketten, inte de som sker vid inlämning av fil. Lagringen innebär att en räknare summerar antalet gånger granskningsmeddelandet aktiverats hos ul. Informationen om hur många gånger som kontrollen aktiverats i kombination med inkomna data, gör det möjligt att utvärdera kontrollernas effektivitet. Det är en procedur som aktiveras genom att beställa tjänsten i självbetjäningssportalen (Topdesk). Det behövs inte i nya SIV. En förutsättning för utvärderingen är att kontrollen både finns ute hos ul och för utvärdering internt, med tillgång till inkomna data finns då möjlighet att utvärdera ul-kontrollerna. Kontrollerna i ul-granskningen ska inte återkomma i mikrogranskningen. I nya SIV kan kontrollerna lagras separat från granskningsflödet och användas för utvärdering. I gamla SIV finns ingen generell lösning för detta. För att beräkna träffsäkerheten i en kontroll som bara finns i ul-granskningen måste den återskapas i t.ex. SAS och tillämpas på ogranskat data.

Det är viktigt att vid tolkningen av resultaten vara medveten om hur granskningsflaggingar aktiveras i webblancketten. Går ul framåt i blanketten utan att fylla i celler som har kontroller, aktiveras dessa och läggs till summan av aktiverade flaggingar.

### 3.2. Beräkna träffsäkerheten för ul-kontroller i SIV

För att beräkna träffsäkerheten i granskningskontrollen ute hos ul med hjälp av SIV:s räknare måste det till viss databearbetning. Det kan ibland vara svårt att med säkerhet veta vad som är första version av varje svar, vilket kan påverka beräkningen av träffsäkerhet. De beräknade indikatorerna kan dock användas som en grov indikation vid utvärdering av kontroller även om värdena är förknippade med viss osäkerhet.

Ul kan aktivera samma kontroll flera gånger för samma cell och då räknas granskningsflaggan flera gånger i räknaren. För att det ska vara möjligt att beräkna träffsäkerhet får antalet granskningsflaggingar ute i SIV maximalt uppgå till antalet celler som kontrollen avser. Ett exempel är att om ett företag ska lämna löner för respektive anställd hos företaget ska antalet granskningsflaggor kopplat till lönen maximalt uppgå till antalet anställda. I tabell 1 nedan så finns ett förslag till hur det går att korrigera för antalet flaggor som räknats ute i SIV.

Tabell 1. Exempel på korrigering av antalet aktiverade granskningsflaggor för en ul-kontroll

Kontroll k avser att kontrollera om lönen på något sätt är misstänkt för att innehålla mätfel.
--



Företag	Antalet anställda som lön rapporteras för	Antalet aktiverade granskningsflaggor för kontroll k	Antalet korrigerade granskningsflaggor för kontroll k = $F_{ul,k}$	Antalet granskningsflaggor internt efter ul-granskning för kontroll k = $F_{p,k}$
X	5	7	5	1
Y	3	2	2	0
Z	8	10	8	1

Träffsäkerheten för en ul-kontroll kan då estimeras med hjälp av formeln nedan:

$$T_{ul,k} = 1 - \frac{F_{p,k}}{F_{ul,k}}$$

$T_{ul,k}$  = träffsäkerhet för ul-kontroll k

$F_{p,k}$  = antal flaggor för kontroll k för första inkomna version

$F_{ul,k}$  = korrigerat antal flaggor för ul-kontroll k

Ett exempel är att träffsäkerheten för kontroll k i tabell 1 beräknas som nedan:

$$T_{ul,k} = 1 - \frac{2}{15} \approx 87 \% \text{ träffsäkerhet}$$

Här har vi antagit att antalet flaggor för kontroll k för första inkomna version är lika med 2.

Ett annat sätt att räkna ut träffsäkerhet för ul-kontrollerna, som kan användas i både webblankett, textfil och Excelfil, är att dela upp insamlingen i två grupper där den ena får ul-kontroller och den andra inte. Då går det att räkna ut en approximativ träffsäkerhet för ul-kontrollerna genom att använda följande formel:

$$T_{ul,k} = 1 - \frac{\frac{n_{tot,k}}{n_{exp,k}} * F_{exp,k}}{\frac{n_{tot,k}}{n_{kon,k}} * F_{kon,k}}$$

I uttrycket ovan betecknar:

$T_{ul,k}$  = träffsäkerhet för ul-kontroll k

$n_{exp,k}$  = antal inkomna objekt som fått ul-kontroller i experimentgruppen som kan ha fått kontroll k

$n_{kon,k}$  = betecknar antal inkomna objekt som inte fått ul-kontroller i kontrollgruppen som kan ha fått kontroll k

$n_{kon,k}$  = betecknar totalt antal inkomna objekt som kan ha fått kontroll k

$F_{exp,k}$  = betecknar antal flaggor för första inkomna version till mikrogranskningen som ingår i experimentgruppen som kan ha fått kontroll k

$F_{kon,k}$  = betecknar antal flaggor för första inkomna version till mikrogranskningen som ingår i kontrollgruppen som kan ha fått kontroll k.