



PCA/LEDN

## Metoder att kvantifiera mätfel



## Innehåll

1 Inledning .....	5
1.1 Dokumentets syfte och användningsområde .....	5
1.2 Vad är mätfel? .....	5
1.3 Mått eller indikatorer .....	5
1.4 Dokumentera .....	6
1.5 Metoder .....	7
2 Metoder som oftast kräver extra insamling av data, alternativt specifikt designad datainsamling .....	8
2.1 Randomiserade experiment.....	8
2.1.1 Inledning .....	8
2.1.2 Vilken typ av resultat/information kan metoden ge? .....	9
2.1.3 Att tänka på inför genomförande av ett experiment .....	9
2.1.4 Mixed-mode som kontrollerat och okontrollerat experiment .....	10
2.1.5 Läs mer .....	11
2.2 Återintervjustudier .....	12
2.2.1 Inledning .....	12
2.2.2 Återintervju med gold standard .....	12
2.2.3 Replikering.....	14
2.2.4 Tid mellan intervjuerna.....	14
2.2.5 När kan metoden användas?.....	15
2.2.6 Läs mer .....	15
2.3 MTMM.....	17
2.3.1 Inledning .....	17
2.3.2 Survey Quality Prediction (SQP 2.0) .....	17
2.3.3 När kan metoden användas .....	18
2.3.4 Att tänka på inför genomförande av ett test i SQP .....	18
2.3.5 Läs mer .....	18
3 Metoder som innebär analys av befintlig data .....	20
3.1 Record check (jämförelse med annat material på objektsnivå) ..	20

3.1.1 Inledning .....	20
3.1.2 När kan metoden användas?.....	20
3.1.3 Genomförande.....	21
3.1.4 Att tänka på inför genomförandet av en record check.....	21
3.1.5 Samgranskning.....	21
3.1.6 Läs mer .....	22
3.2 Latent klassanalys (Latent Class Analysis, LCA) .....	23
3.2.1 Inledning .....	23
3.2.2 När kan metoden användas?.....	23
3.2.3 Att tänka på inför genomförande av en Latent klassanalys	23
3.2.4 Läs mer .....	23
3.3 Markov kvasi-simplex-modell (Markov quasi-simplex model)	25
3.3.1 Inledning .....	25
3.3.2 När kan metoden användas?.....	25
3.3.3 Att tänka på inför analys med en Markov kvasi-simplex modell.....	25
3.3.4 Läs mer .....	25
3.4 Medlyssning/ Beteendekodning .....	26
3.4.1 Inledning .....	26
3.4.2 Beteendekodning/ medlyssning på SCB.....	26
3.4.3 När kan metoden användas?.....	27
3.3.4 Att tänka på inför genomförande av beteendekodning för att undersöka mätfel.....	27
3.4.5 Läs mer .....	27
3.5 Intervjuarvarians.....	28
3.5.1 Inledning .....	28
3.5.2 Vilken typ av information kan metoden ge?.....	28
3.5.3 När kan metoden användas?.....	28
3.5.4 Att tänka på inför genomförande av en studie av intervjuarvarians.....	29
3.5.5 Läs mer .....	29
3.6 Paradata.....	30
3.6.1 Inledning .....	30
3.6.2 Svarstider .....	30

3.6.3 Respondenter som backar och ändrar .....	32
3.6.4 Avbrott av intervjuer eller webb-enkät .....	33
3.7 Frekventa åtgärder vid mikrogranskning .....	35
3.7.1 Inledning .....	35
3.7.2 När kan metoden användas?.....	35
3.7.3 Att tänka på inför analys av frekventa åtgärder vid mikrogranskning.....	36
3.7.4 Läs mer .....	36
3.8 Indirekta kontroller av mikro- och makro-data.....	37
3.8.1 Inledning .....	37
3.8.2 Frekvenser och fördelningar .....	37
3.8.3 Intern konsistens .....	38
3.8.4 Kontrollera individuella svarsmönster.....	39
3.8.5 Jämförelser med annan statistik.....	39
3.8.6 Stora andelar partiellt bortfall samt "vet ej"-svar .....	39
3.8.7 Tillägg för kontroll i enkät.....	40
3.8.8 Att tänka på vid indirekta kontroller av mikro- och makrodata .....	40
3.7.9 Läs mer .....	40
Översikt: Metoder för att undersöka och kvantifiera mätfel .....	42

## 1 Inledning

### 1.1 Dokumentets syfte och användningsområde

Det här dokumentet är tänkt att i första hand användas av produktansvariga, och motsvarande personer, när det finns intresse att göra en kvantitativ mätfelsstudie av produkten. Dokumentet innehåller beskrivningar av en rad olika metoder att kvantifiera mätfel – vilken typ av information en studie med metoden kan ge, vad som krävs för att det ska gå att genomföra och vad som är viktigt att tänka på inför och under genomförandet. Beskrivningarna är i de flesta fall inte så detaljerade och konkreta att det är möjligt att genomföra en mätfelsstudie endast med hjälp av det här dokumentet. Tanken är att dokumentet ska fungera som ett hjälpmedel i ett första steg för att bedöma vilka metoder som finns och som lämpar sig bäst att använda på den aktuella produkten.

### 1.2 Vad är mätfel?

Mätfel uppstår när det observerade värdet skiljer sig från det sanna värdet enligt variabelns definition. I teorier om mätning menar man att det alltid förekommer fel i en mätning, men felet kan vara olika stort, bero på olika saker och få olika konsekvenser.

Orsakerna till mätfel kan delas in i fyra grupper: de som beror på respondenten, de som beror på instrumentet (frågeformuläret och frågorna), de som beror på intervjuaren och de som beror på sättet som data samlats in (datainsamlingsmetod – papper, webb, telefon, besöksintervju osv.). I den översiktliga tablan i dokumentets bilaga anges vilka orsaker till mätfelet som man kommer åt med de olika metoderna (se Biemer & Lyberg, 2003 kap. 4 för ytterligare orsaker).

På SCB jobbar vi med att förebygga mätfel genom att testa mätinstrumenten innan de används. Det görs genom bl.a. expertgranskningar och kognitiva test. Dessa kvalitativa förebyggande åtgärder ger oss dock ingen information om hur stort problemet med mätfel är och i vilken utsträckning de föreslagna åtgärderna minskar mätfelet. SCB arbetar också mycket med granskning av mikrodata. Granskningen innebär att uppenbara och misstänkta fel identifieras och i många fall åtgärdas. Granskningen bidrar med all säkerhet till att mätfelet minskar i den färdiga statistiken, men stora delar av mätfelen upptäcks sannolikt inte och kan därmed inte heller åtgärdas.

De metoder som beskrivs i det här dokumentet syftar till att i första hand kvantifiera mätfel i genomförda undersökningar. Till stor del kan de också ge kunskap om hur mätfelet kan förebyggas i kommande undersökningsomgångar.

### 1.3 Mått eller indikatorer

De metoder som presenteras i materialet ger olika typer av information om mätfelet i en undersökning. Några av metoderna kan

ge ett mer eller mindre direkt mått på mätfelets storlek, t.ex. den sammanlagda storleken på mätfelet, variansen och sannolikheten för felklassificeringar. Andra metoder ger istället information i form av kvantitativa indikatorer på att det skulle kunna finnas mätfel och om det är ett stort problem. Indikatorerna kan också hjälpa oss i förståelsen för var i mätningen problemen är störst och t.ex. vilka frågor som behöver arbetas om inför kommande omgångar. Eftersom den här typen av mått inte är lika säkra och i många fall kräver mer tolkning är det ofta värdefullt att kombinera flera metoder.

Att kombinera flera metoder är alltid positivt. Det gör det enklare att få en bredare bild av problematiken och en djupare förståelse för orsaken till det eventuella mätfelet och hur det kan förbyggas i kommande undersökningsomgångar.

Samtliga metoder är också värdefulla att använda för att utvärdera åtgärder som görs för att minska mätfel. Genom att jämföra mått från tidigare med motsvarande mått efter att åtgärden genomförts får vi kunskap om vad åtgärden har haft för effekt.

Vilken typ av information det går att få från vilken metod framgår både i texten nedan i samband med respektive metod och i den översiktliga tablan i bilagan.

Ytterligare en fråga att förhålla sig till, och som delvis skiljer sig mellan de metoder som presenteras i materialet, är när i förhållande till granskningen som en mätfelsstudie bör göras. Utgångspunkten i ett sådant beslut måste vara vad det är man vill få ut av studien och hur man ser på och definierar mätfel. Om man ser det som att det huvudsakliga syftet är att förebygga mätfel, genom att t.ex. förbättra mätinstrumentet, och därigenom också minska arbetsinsatserna vid granskning, bör mätfelsstudien göras på ogranskade data. Om man istället främst vill använda mätfelsstudien för att få ett kvalitetsmått på den färdiga undersökningen kan det vara bättre att använda slutgiltig och eventuellt granskad och korrigerad data.

För vissa metoder blir det också mer eller mindre självklart att mätfelsstudien ska göras före granskningen eller på okorrigerade data eftersom inriktningen på metoden är sådan. Det gäller t.ex. studier av intervjuarvariansen och återintervjustudier.

#### **1.4 Dokumentera**

För att det totala mätfelsarbetet på SCB ska bli så bra och effektivt som möjligt är det viktigt att dokumentera planering och genomförande av eventuella mätfelsstudier. Det är också önskvärt om dokumentationen görs tillgänglig. Erfarenheterna från en genomförd mätfelsstudie kan komma andra produkter till nytta genom att resultat i vissa fall och med viss försiktighet kan generaliseras till andra produkter. Andra produkter kan då ha glädje av resultaten i sitt kvalitetsarbete. Även erfarenheter från själva

genomförandet är viktigt att använda – vilka metoder som har varit meningsfulla, svåra och problematiska är viktig information för dem som planerar att genomföra en mätfelsstudie. Liksom var och när problemen har uppstått och hur man har löst dem.

### **1.5 Metoder**

Nedan följer beskrivningar av de olika metoderna. I dokumentet är de uppdelade utifrån om de kräver extra insamling av data alternativt specifikt designad datainsamling eller om metoderna innebär analyser av befintlig data. Uppdelningen är inte perfekt och flera metoder kan passa under båda kategorierna beroende på hur de läggs upp. Vi tror ändå att det kan vara en meningsfull uppdelning för att ge en övergripande struktur. Till dokumentet bifogas också en tablå som syftar till att ge en överblick genom att sammanfatta varje metod i några centrala punkter.

Innehållet i de olika avsnitten har i grunden en gemensam struktur med liknande underrubriker. Innehållet täcker in en kort övergripande beskrivning, vilken typ av information metoden kan ge, vad som krävs för att metoden ska kunna genomföras samt några punkter som är viktiga att tänka på inför och under genomförandet. Eftersom metoderna skiljer sig avseende hur de fungerar och hur mycket litteratur och erfarenheter det finns varierar dock upplägget något. Det varierar också något mellan metoderna i hur tekniska beskrivningarna är. I de flesta fall är det tekniska innehållet väldigt begränsat. För några metoder har det dock varit omöjligt att ge en adekvat beskrivning utan att använda vissa tekniska termer, det gäller främst MultiTrait-MultiMethod (MTMM)/Survey Quality Prediction (SQP) och Latent klassanalys (LCA). Vi hoppas att dessa något mer tekniska beskrivningar inte ska avskräcka läsaren från att läsa om övriga metoder.

I texten används båda begreppen respondent och uppgiftslämnare när den som besvarar ett frågeformulär avses. Att båda begreppen används beror endast på att det finns behov av en språklig variation och det finns ingen skillnad i betydelsen. Orden kan alltså läsas helt synonymt.

## **2 Metoder som oftast kräver extra insamling av data, alternativt specifikt designad datainsamling**

### **2.1 Randomiserade experiment**

#### **2.1.1 Inledning**

Randomiserade experiment är en klassisk metod inom vetenskap och forskning och har i många sammanhang fått status som den ideala designen. Grunden i metoden är att samtliga experimentpersoner fördelas slumpmässigt i antingen en kontrollgrupp eller i en eller flera experimentgrupper. Den slumpmässiga fördelningen är nödvändig för att vi ska kunna anta att fördelning och resultat är oberoende.

Experimentgrupperna utsätts för någon form av intervention (manipulation, behandling) medan kontrollgruppen lämnas utan eller får eventuellt placebo. Genom att efter interventionen jämföra de olika grupperna kan man dra slutsatser om interventionens effekt. Eftersom gruppernas sammansättning är slumpmässiga och vi kan kontrollera interventionen är det med experimentell design möjligt att undersöka orsakssamband.

Inom surveyområdet görs randomiserade experiment företrädesvis för att utvärdera olika datainsamlingsmetoder, kontakstrategier och mätinstrument. Det kan t.ex. handla om att jämföra svaren mellan en grupp som har fått en pappersblankett och en grupp som har fått svara på webben och se om det är några skillnader i svaren. Eller att jämföra två grupper som fått formulär med olika ordning på frågor eller svarsalternativ. Genom att fördelningen i grupperna är slumpmässig och det enda som skiljer grupperna åt är just det man vill undersöka, kan man uttala sig om t.ex. datainsamlingsmetoden eller frågeordningen har haft någon inverkan på svaren.

Det mest kostnadseffektiva sättet att genomföra ett experiment är oftast att mer eller mindre integrera det i den vanliga produktionen, s.k. inbäddat experiment. I sådana experiment kan man antingen göra ett extra urval av en eller flera grupper som kompletterar det ordinarie urvalet. Det extra urvalet eller urvalen får fungera som experimentgrupp/er. Eller så kan det ordinarie urvalet slumpmässigt delas in i kontrollgrupp respektive experimentgrupp/er. Om det finns resurser är det bra att alltid göra den här typen av inbäddade experiment när undersökningen genomgår en förändring i form av t.ex. annorlunda frågor eller en ny datainsamlingsmetod. Genom att undvika att införa förändringen till en del av urvalet kan vi kontrollera vad förändringen har för effekt.

Experiment kan även göras med andra syften än att undersöka mätfel, t.ex. kan vi se vad olika interventioner har för effekt på bortfallet eller den upplevda uppgiftslämnarbördan.



### 2.1.2 Vilken typ av resultat/information kan metoden ge?

Ett experiment kan i sig sällan ge ett mått på storleken på undersökningens sammanlagda mätfel. Det vi kan få är ett mått på variansen. Det innebär att om de undersökta grupperna (experimentgrupper och kontrollgrupp) skiljer sig åt så vet vi att interventionen har effekt, men vi vet inte vilken av varianterna som är mest korrekt. Om t.ex. en viss frågeordning ger en högre frekvens på en specifik fråga än en annan ordning, hur ska vi då veta vilken av frågetyperna som innebär minst mätfel? I många fall går det dock att göra tolkningar och analyser utifrån teorier och tidigare forskning, dels om frågors egenskaper dels hur fördelningar borde se ut i jämförelser mellan olika grupper (Krosnick, 2011 s.221). På så sätt kan man avgöra vilken av frågorna som har störst respektive minst mätfel. I många fall räcker det också med att konstatera en skillnad mellan experiment- och kontrollgrupp för att syftet med experimenten ska vara uppfyllt. En skillnad innebär ju att minst någon av versionerna innehåller mätfel och om skillnaden är stor är sannolikt också mätfelet betydande.

### 2.1.3 Att tänka på inför genomförande av ett experiment

Det är mycket man ska tänka på om och när ett experiment ska genomföras. Karlberg m.fl. (2002) listar de viktigaste frågorna att fundera på och ta ställning till:

- **Vad är syftet? Kommer upplägget kunna svara mot syftet?**  
Precis som i alla studier och undersökningar måste det vara klart och tydligt vad som är syftet med undersökningen. Det är ingen bra idé att göra ett experiment bara för att det är kul och intressant. Det är också först om syftet är tydligt som det går att göra en rimlig bedömning om experimentet är genomförbart.
- **Power.**  
Kontrollera hur många experimentdeltagare som behövs för att experimentet ska ha en rimlig chans att lyckas.
- **Hur ska resultatet användas och av vem?**  
Är det i huvudsak den specifika produkten som kommer kunna använda resultatet eller kommer det ge generell kunskap som bör förvaltas och spridas av t.ex. någon av metodheterna. Självklart kan det vara både och, men se till att resultaten dokumenteras och når fram till dem som bör förvalta kunskapen.
- **Vilka urvalsenheter ska inkluderas i experimentet?**  
Kan det ordinarie urvalet användas (vid ett inbäddat experiment) eller behöver man göra ett extra urval? Behöver man ha flera experimentgrupper eller räcker det med en?
- **Hur ska randomiseringen gå till?**

Bör randomiseringen av deltagare till de olika grupperna vara obunden eller behöver det stratifieras? Hur ska det göras rent praktiskt?

- **Utbildning till personal.**  
Behövs det någon speciell utbildning av den personal som ska arbeta med experimentet?
- **Hur ska resultaten analyseras?**  
Vilka statistiska metoder kommer kunna användas och hur ska olika resultat tolkas?
- **Integrering av experimentdata i produktionen vid inbäddade experiment.**  
Vilken grupp ska användas som ordinarie data? Vad finns det för risker och problem med att blanda experiment och kontrollgrupper?

#### **2.1.4 Mixed-mode som kontrollerat och okontrollerat experiment**

Mixed Mode eller kombinerade datainsamlingsmetoder är främst en strategi för att effektivisera datainsamlingen och spara resurser. Detta eftersom t.ex. insamling via webben är mindre resurskrävande än telefonintervjuer. Till viss del har kombinerad insamling också varit en strategi för att försöka möta det ökade bortfallet. Att erbjuda flera alternativa insamlingssätt förväntas fånga upp fler respondenter. I vilken utsträckning kombinerad insamling påverkar det slutliga bortfallet är dock oklart.

En konsekvens av att kombinera olika datainsamlingsmetoder är att frågornas och svarens innehåll påverkas. Den skillnad som uppstår i data mellan de olika insamlingsmetoderna p.g.a. detta kallas för mode-effekter och uppfattas oftast som problem som man vill undvika när det går. Men eftersom de här effekterna existerar är det också möjligt att utnyttja skillnaderna i data som de olika metoderna skapar för att få kunskap om mätningarnas kvalitet. Att jämföra de två datainsamlingsmetoderna kan visa på brister i den ena eller andra mätningen. Om skillnaderna är stora mellan svaren på motsvarande fråga i olika insamlingsmetoder tyder det på att det finns mätfel i minst en av dem. Motsvarande tankesätt kan användas då en datainsamlingsmetod ersätts med en annan.

För att utvärdera eventuella effekter av den blandade datainsamlingen på bästa sätt bör undersökningen designas som ett randomiserat experiment på det sätt som beskrivs ovan. Det innebär att det ska vara slumpmässigt vilka respondenter som besvarar undersökningen med vilken metod. I de situationer där det inte är möjligt att anpassa upplägget och randomisera grupperna bör man dock ändå utnyttja den information som skapas genom att data samlas in med två olika metoder. När respondenterna väljer själva vilken datainsamlingsmetod de vill använda kan vi inte vara säkra på att eventuella skillnader mellan grupperna beror på själva datainsamlingsmetoden, men vi kan ändå få vissa indikationer på om

det finns skillnader i mätningarna. Slutsaser från en undersökning som inte är kontrollerad och randomiserad måste helt enkelt dras med stor försiktighet.

#### **2.1.5 Läs mer**

Karlberg, M., Björnram, A., Boynton, I-M., Göransson, B. och Lundquist, P. (2002). "Embedded Experiments in Sample Surveys at Statistics Sweden: The Development of an Experimentation Manual". Joint Statistical Meeting - Section on Survey Research Methods.

Krosnick, J. (2011) "Experiments for Evaluating Survey Questions" i Madans, J., Miller, K., Maitland, A. och Willis, G. (red.) *Question Evaluation Methods*. New Jersey: Wiley

Statistiska centralbyrån (2005), *Experimentell utvärdering - en handbok*. Handbok 2005:1.

Tourangeau, R. (2004). "Experimental Design Considerations for Testing and Evaluating Questionnaires", i Presser m.fl. (red.) (2004), *Methods for Testing and Evaluating Survey Questionnaires*. New Jersey: Wiley.

## 2.2 Återintervjustudier

### 2.2.1 Inledning

En återintervjustudie är en studie där man använder sig av upprepade mätningar bl.a. för att undersöka mätfelet. Precis som namnet antyder handlar det om att ett urval av respondenterna som blivit intervjuade i en undersökning blir återkontaktade och får svara på samma frågor återigen.

I grunden kan man prata om två typer av återintervjustudier med två olika syften och tillvägagångssätt – Golden standard och Replikering. Golden standard innebär att man i återintervjun, med hjälp av en förbättrad metod, försöker samla information som är av allra högsta kvalitet, dvs. information som motsvarar det sanna värdet. I replikeringar görs inget anspråk på att den ena mätningen är mer sann än den andra, en sådan studie visar endast hur stor variationen är. Replikering kan i sin tur delas in i två typer – Test-retest och generell återintervju. I ett Test-retest görs den upprepade mätningen på precis samma sätt som den ordinarie och man kan på så sätt skatta variationen i undersökningen. I en generell återintervjustudie behöver upplägget inte vara likadant i den ordinarie datainsamlingen och i återintervjun. T.ex. så kan en telefonintervju följas av en webblankett.

### 2.2.2 Återintervju med gold standard

Det kanske vanligaste sättet att få fram det "sanna värdet" i återintervjun är genom så kallad rekonciliering. Det innebär att respondenten återkontaktas en bestämd tid efter intervjun av en speciellt tränad intervjuare. Intervjuaren upprepar hela eller delar av intervjun och kontrollerar svaren mot vad respondenten svarat i den tidigare intervjun. Om samtliga svar överensstämmer avslutas intervjun. Om ett eller flera svar inte är desamma i de två intervjuerna blir det aktuellt med den så kallade rekoncilieringen. Själva rekoncilieringen går ut på att intervjuaren försöker få fram det sanna svaret genom fördjupande noggranna uppföljningsfrågor. Det sanna värdet kan vara antingen det ursprungliga svaret, svaret vid återintervjun eller ett helt nytt svar. Det är genom att se skillnaderna mellan den skattade proportionen i populationen utifrån den ordinarie intervjun och utifrån de rekoncilierade svaren som man får ett mått på mätfelet.

Återintervjun och rekoncilieringen kan också innehålla en del där den specialtränade intervjuaren försöker ta reda på orsaken till att det inte blev rätt värde från början. En sådan del kräver ännu mer av intervjuaren och gör att intervjun blir längre. Men upplägget innebär också att man utöver ett mått på mätfelets storlek kan få kunskap om vad som bör åtgärdas för att mätfelet ska minska.

Att skatta sanningen utifrån rekoncilieringen är dock inte oproblemiskt. Bland annat så kan fel i den ursprungliga intervjun

förbises eftersom rekonciliering med fördjupade frågor endast blir aktuellt om ursprungsintervjun och återintervjun ger olika värden. Att intervjuerna felaktigt ger samma värde kan bero på att respondenten minns svaren som hon eller han gav vid ursprungsintervjun och försöker ge samma svar igen. Men också att intervjuaren, om denne känner till de ursprungliga svaren, medveten eller omedvetet styr och tolkar de nya svaren på samma sätt.

Fel kan också uppstå när det skiljer sig mellan den ordinarie intervjun och återintervjun och en rekonciliering blir aktuell. Det kan t.ex. bero på tolkningsfel och kommunikationsbrister mellan respondent och intervjuare. Det är alltså möjligt att både återintervjuer och rekoncilieringsintervjuer innehåller samma fel som den ursprungliga intervjun.

Ett alternativ till rekonciliering är att inte göra en regelrätt återintervju med samma frågor som i ursprungsintervjun utan istället göra djupintervjuer med ett urval av objekten. I djupintervjuerna ställs detaljerade och utredande frågor redan vid det första steget i återintervjun. Dessa fördjupade frågor ska ge en ordentlig och heltäckande bild av respondentens situation i förhållande till det som ska mätas och därmed kunna bedöma det sanna värdet på den aktuella frågan. Även här krävs intervjuare med särskilt utbildning och god kunskap för vad som undersökningen syftar till att mäta. Graden av fördjupning kan variera i olika studier, men syftet är att få fram det sanna värdet i återintervjun.

#### *Brutto- eller nettofel?*

Vid återintervjustudier enligt gold standard får man kunskap om både brutto- och nettofelet. Bruttofelet motsvarar både felaktigt medtagna och felaktigt uteslutna från en viss kategori, t.ex. de som felaktigt hamnat i kategorin höginkomsttagare och de som felaktigt hamnat utanför kategorin höginkomsttagare. Nettofelet motsvarar istället vilken effekt felet får på statistikvärdena, dvs. de som felaktigt hamnat i kategorin minus de som felaktigt hamnat utanför. Båda de här måtten kan vara relevanta för att få information om problem i undersökningen, men det är viktigt att tänka på skillnaden.

Att tänka på inför genomförande av en återintervjustudie enligt golden standard:

- Vilken typ av återintervju ska göras för att fånga det "sanna värdet"? Upprepad intervju med rekonciliering eller direkt djupintervju.
- Vilken typ av utbildning behöver intervjuarna?
- Ska rekoncilieringen även innehålla försök att utreda orsakerna till de upptäckta felen?

### 2.2.3 Replikering

Som framgått ovan så kan replikering i sin tur delas in i två typer av återintervjustudier med olika inriktningar och upplägg, test-retest och generell återintervju. Den vanligaste formen av replikering är så kallat test-retest. I ett test-retest försöker man hålla svarsprocessen så konstant som möjligt mellan de två mätningarna. Detta för att kunna uttala sig om att eventuella skillnader mellan mätningarna beror på slumpmässiga variationer. Ett test-retest ger inte kunskap om systematiska fel i mätningen eller om skillnaden mellan observerade värdet och det sanna värdet. Det man får är helt enkelt skillnaden mellan två olika mätningar där ingen är mer sann än den andra. Den uppmätta skillnaden blir då ett mått på graden av slumpmässiga fel, eller annorlunda uttryckt, på stabiliteten (reliabiliteten). En stor variation mellan de två mätningarna innebär att graden av slumpmässiga fel är hög och att stabiliteten är låg. Samstämmiga svar betyder således också att det är en frånvaro av slumpmässiga fel. Samstämmighet betyder dock inte nödvändigtvis en frånvaro systematisk fel (hög validitet). De systematiska felen kan ju finnas i båda mätningarna.

Att test-retest endast fångar de slumpmässiga felen innebär också att det inte ger någon kunskap om orsakerna till de eventuella felen. För att metoden ska kunna leda fram till åtgärder och bidra till ett minskat mätfel krävs alltså att den kombineras med andra metoder.

Grundläggande vid ett test-retest är att insamlingsproceduren går till på samma sätt vid ursprungsintervjun och återintervjun. Det innebär att:

- Återintervjuarna ska ha samma utbildning och väljas på samma sätt som ursprungsintervjuarna.
- Frågeformuläret ska vara detsamma.
- Det ska vara samma respondenter.
- De båda intervjuomgångarna ska vara oberoende, t.ex. så ska svaren från ursprungsintervjun inte presenteras för respondenten vid återintervjun.

I den varianten som vi kallar för generell återintervju är kraven på en identisk procedur i de båda insamlingarna inte lika hårda. För att kunna använda resultaten som mått på mätfel krävs då istället mer avancerade analyser och modellering. Detta till skillnad från återintervju enligt golden standard och test-retest.

### 2.2.4 Tid mellan intervjuerna

En viktig men svår fråga att ta ställning till när man ska genomföra en återintervjustudie och som är aktuell både för replikering och Golden standard är hur lång tid det ska gå mellan ursprungsintervjun och återintervjun. Tiden måste vara tillräckligt

lång för att respondenterna inte ska komma ihåg vad de avgav för svar vid ursprungsintervjun. Men inte så lång så att respondenten inte kan minnas den handling, situation och referensperiod som efterfrågas i frågorna. Den får inte heller vara så lång så att det sker en faktisk förändring mellan ursprungsintervjun och återintervjun.

Vilken tidsperiod som passar in på dessa kriterier varierar mellan olika undersökningar och beror på vilken typ av frågor och vilka referensperioder som används.

#### **2.2.5 När kan metoden användas?**

Återintervjustudier genomförs främst inom undersökningar riktade till individer och hushåll. När metoden beskrivs i litteraturen är beskrivningarna också anpassade för intervjustudier till individer. Det är dock möjligt att göra återintervjustudier även för att testa undersökningar riktade till företag, organisationer och offentlig sektor. Med vissa modifieringar och anpassningar kan man också använda metoden på självadministrerad datainsamling (pappers- och webbenkäter). En replikering i form av ett test-retest av en självadministrerad undersökning innebär inte så stor skillnad från en intervjuundersökning. Respondenterna ombeds helt enkelt att besvara samma frågeformulär en gång till. En återintervju enligt Golden standard av ett självadministrerat formulär kräver dock att man kombinerar den ursprungliga självadministrerade insamlingen med en intervju. Antingen först vid en eventuell rekonciliering eller direkt vid återintervjun.

Den här typen av blandningar, modifieringar och anpassningar kan också göras även när ursprungsmätningen är en intervju, t.ex. så kan återintervjun direkt ha formen av en rekonciliering.

#### **2.2.6 Läs mer**

Biemer, P. & Forsman, G. (1992). "On the Quality of Reinterview Data with Applications to the Current Population Survey", *Journal of the American Statistical Association*, Vol. 87, No. 420.

Biemer, P. & Lyberg, L. (2003). *Introduction to Survey Quality*. New Jersey: Wiley. (kapitel 5 och kapitel 8)

Forsman, G., och Schreiner, I. (1991). "The design and analysis of reinterview: An overview", in *Measurement errors in surveys*, eds. P. Biemer et al., New York: John Wiley & Sons, pp. 280-301.

Statistiska centralbyrån (2013). *Mätfelsstudie AKU – projektrapport*.

The Federal Committee on Statistical Methodology (2001). *Measuring and Reporting Measurement Errors in Surveys*. Statistical Policy Working Paper 31.



## 2.3 MTMM

### 2.3.1 Inledning

MultiTrait-MultiMethod (MTMM) är en metod som används för att utvärdera surveyfrågor genom att skatta frågornas systematiska och slumpmässiga fel. Metoden bygger på att samma fråga ställs flera gånger till samma respondenter, men där frågorna har vissa variationer. Man kan således likna MTMM vid ett experiment. Både metoden och dess beskrivning är relativt tekniskt och lämpar sig bäst för den som är extra insatt och intresserad.

I genomförandet av ett MTMM-experiment har man dels de olika egenskaper (beteenden, attityder, fakta) man vill mäta, dels tre olika sätt att ställa frågorna som ska mäta egenskaperna. Varje egenskap måste mätas på tre olika sätt. Det innebär att om man har tre olika egenskaper som ska ingå i MTMM-experimentet kommer materialet bestå av nio olika frågor – tre sätt att fråga om vardera tre egenskaper. Frågornas kvalitet skattas sedan i en modell, s.k. Strukturell Ekvationsmodellering (SEM).

Eftersom ett MTMM-experiment kräver att varje fråga ställs på tre olika sätt är genomförandet väldigt arbets- och resurskrävande. För att göra det möjligt att skatta frågornas kvalitet utan att behöva genomföra ett MTMM har ett forskarteam vid Research and Expertise Centre for Survey Methodology (RECSM) vid Universitat Pompeu Fabra i Barcelona utvecklat verktyget *Survey Quality Prediction (SQP 2.0)*.

### 2.3.2 Survey Quality Prediction (SQP 2.0)

Verktyget är uppbyggt utifrån en stor mängd MTMM-experiment som genomförts inom ramen för the European Social Survey (ESS) på de frågor som ställs i undersökningen. Resultaten från dessa experiment, d.v.s. kvaliteten för en mängd olika frågor, ligger sedan till grund för skattningar av kvalitet även på frågor som inte ingått i ESS eller i något MTMM-experiment. Verktyget finns tillgängligt på [www.sqp.nl](http://www.sqp.nl) och vem som helst kan registrera sig som användare.

När en specifik frågas kvalitet ska utvärderas med hjälp av verktyget ska frågans egenskaper kodas enligt ett 60-tal variabler. Egenskaper som kodas är t.ex. frågeområde; referensperiod; antal ord i frågan; antal svarsalternativ; om svarsalternativen har etiketter osv. Frågans kvalitet beräknas sedan utifrån vilken kvalitet frågor med samma egenskaper visat sig ha i de tidigare genomförda MTMM-experimenten.

De MTMM-experiment som databasen bygger på har gjorts även på den svenska delen av ESS. Det finns därför förutsättningar för att använda verktyget för att utvärdera svenska frågor. Ett problem är att inte alla svenska frågor som ingått i experimentet, och där det alltså finns skattad kvalitet, är kodade enligt de ca 60 variablerna. För

att de ska kunna användas i prediktionen av andra frågors kvalitet krävs att de kodas. Det finns dock tillräckligt många frågor kodade för att verktyget ska gå att använda.

Kodningen av frågor är generellt ett av de stora problemen med SQP. Det finns utförliga instruktioner i programmet i hur kodningen ska göras, men det finns alltid utrymme för tolkning och vissa variabler är svårare att koda än andra. Detta leder till viss osäkerhet i skattningarna. Felaktiga koder kan vara aktuella både i den fråga som ska testas och som kodas av oss som användare och i de frågor som redan ligger kodade i databasen.

### **2.3.3 När kan metoden användas**

Metoden är mest användbar när man vill testa kvaliteten hos enskilda frågor och inte har möjlighet att göra något eget experiment eller annan typ av kvalitetsmätning.

De kvalitetsmått som används inom metoden är reliabilitet och validitet. Hur dessa begrepp definieras varierar något i olika sammanhang. Det kan därför vara bra att dels fundera över hur man själv ser på begreppen, dels läsa på om vad de har för innebörd inom metoden innan metoden används. För en utförligare diskussion om innebörden av reliabilitet och validitet i MTMM se Alwin, 2007. Boken innehåller flera olika avsnitt om MTMM där frågan tas upp.

Eftersom databasen är uppbyggd av experiment genomförda på European Social Survey är verktyget lämpligast att använda på frågor riktade till individer och hushåll.

### **2.3.4 Att tänka på inför genomförande av ett test i SQP**

- Behövs någon hjälp i kodningen av frågorna?
- Hur ska de mått (kvalitetsmått) som testet resulterar i tolkas och användas?

### **2.3.5 Läs mer**

Om MTMM

Saris & Andrews (1991) "Evaluation of measurement instrument using a structural modeling approach". I Biemer, Groves, Lyberg, Mathiowets & Sudman (eds.) (1991) *Measurement errors in surveys*. New York: John Wiley & Sons

Alwin, D.F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. New Jersey

Om SQP 2.0

Verktyget finns här: <http://www.sqp.nl/>

Ljung, B. (2012) *Korrigeringsmetoder för slumpfel och metodeffekter i sambandsanalys av enkätdata med SQP 2.0*. Magisteruppsats HT2012, Statistiska institutionen, Uppsala universitet

Saris, W. och Gallhofer, I. (2007) *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, New Jersey: Wiley.

## 3 Metoder som innebär analys av befintlig data

### 3.1 Record check (jämförelse med annat material på objektsnivå)

#### 3.1.1 Inledning

Record check går ut på att jämföra insamlade data med en annan jämförbar datakälla som bedöms hålla en hög kvalitet. Datakällan kan vara antingen från ett administrativt register eller från en annan undersökning. Jämförelserna görs på objektsnivå, dvs. insamlade data för varje enskilt objekt jämförs med värdet för samma objekt enligt den alternativa källan. Om vi kan anta att den alternativa källan håller så hög kvalitet att data kan jämföras med det sanna värdet kan vi få ett mått på det sammanlagda mätfelet i en undersökning. Om variabelvärdet i den undersökning som kontrolleras skiljer sig från motsvarande variabelvärde i den alternativa källan finns mätfel i undersökningen. Eftersom jämförelser görs på objektsnivå får vi också vetskap om mätfelet är större eller mindre i vissa grupper, t.ex. olika typer av företag. Genom att kombinera informationen från analysen med en genomgång av mätinstrumentet och andra aspekter av mätningen kan vi också få en förståelse för möjliga orsaker till mätfelet.

Det kan tyckas märkligt att man samlar in undersökningsdata i de fall motsvarande data finns tillgängliga genom administrativa register eller andra undersökningar. Men i en del fall där det finns jämförbara data i alternativa källor är insamlingen motiverad. Det kan t.ex. vara så att den andra källan endast har data på årsbasis medan behov finns av månadsdata. Andra skäl till insamling av undersökningsdata kan bero på en tidsmässig fördröjning i ett administrativt register eller behov av fördelade totalvärden på olika underliggande nivåer (t.ex. produkter, varor och länder).

Förutom att skatta mätfelets storlek kan jämförelser med annat material på objektsnivå också ge mått på bl.a. bortfallsfel. I det här informationsmaterialet fokuseras dock på hur man kan använda metoden för att skatta mätfel.

#### 3.1.2 När kan metoden användas?

För att kunna göra en record check krävs att det finns jämförbara data för de värden man vill kontrollera. Det kan uppstå problem då en datakälla som vid en första anblick verkar vara jämförbar skiljer sig från undersökningsdata i centrala detaljer vilket stör jämförbarheten. Det kan t.ex. skilja sig något i hur de centrala variablerna är definierade. Periodiseringen/referensperioden kan också vara annorlunda. För att metoden ska gå att genomföra krävs att dessa aspekter är samstämmiga, eller att de går att korrigera, för att källorna ska bli helt jämförbara.

Det krävs också att den alternativa källan håller tillräckligt hög kvalitet för att kunna användas som bedömningsgrund och motsvara

det sanna värdet. Detta bör kontrolleras inför genomförandet av en record check-studie genom att studera registrets innehåll och tillförlitlighet.

Möjliga datakällor att använda för jämförelser på objektsnivå finns främst för företag, organisationer och offentlig sektor. Det är alltså företrädesvis inom dessa områden som record check-studier kan bli aktuella.

### **3.1.3 Genomförande**

Weideskog (2013) beskriver genomförandet av registerstudier i sju steg:

- (1) Avgränsning av population.
- (2) Val av källa med hög kvalitet
- (3) Val av matchningsvariabler och variabler som ska studeras
- (4) Avgränsning för att få jämförbarhet
- (5) Val av referensperiod
- (6) Val av omfattning för matchning
- (7) Skattningar av mätfel

### **3.1.4 Att tänka på inför genomförandet av en record check**

De centrala punkterna att hänsyn till i planeringen har framkommit i de beskrivna stegen ovan. Vi väljer ändå att sammanfatta det viktigaste här.

- Kontrollera att den källa som ska användas håller tillräcklig hög kvalitet.
- Kontrollera att det är möjligt att matcha källan och undersökningsdata
- Kontrollera att de variabler som ska jämföras är definierade på samma sätt i båda källorna.
- Kontrollera att tidsperioderna är samstämmiga.

### **3.1.5 Samgranskning**

Samgranskning kan uppfattas som en typ av record check eftersom det också handlar om att jämföra insamlad data med andra källor på objektsnivå. Det finns dock en skillnad mellan samgranskning och record check som handlar om att man i samgranskningen inte jämför med en källa som antas motsvara det sanna värdet. Samgranskning görs istället mellan två källor med mer eller mindre okänd kvalitet. Istället för att kunna skatta det sammanlagda mätfelet ger en samgranskning istället ett mått på variationen. Vi kan få information om att värden mellan de två källorna skiljer sig, men vi vet inte vilken av dem som motsvarar det sanna värdet. Med hjälp av kompletterande information bör det dock i de flesta fall gå att utröna vad variationen kan antas bero på.

I praktiken kan man också se förhållandet mellan Record check och samgranskning som en glidande skala, där många jämförelser av olika källor skulle hamna någonstans i mitten mellan record check och samgranskning. Vi har sällan tillgång till en datakälla helt utan fel – även när vi vid en record check gör antagandet att data håller hög kvalitet och att vi kan använda källan som bedömningsgrund så finns det alltid större eller mindre problem med data. På motsvarande sätt gör man vid en samgranskning ofta bedömningen att den ena källan är mer tillförlitlig. Även om den inte används för att motsvara det sanna värdet, ges den kanske en högre status vid jämförelsen.

Se också avsnittet "Jämförelser med annan statistik" under "Indirekta kontroller av mikro- och makrodata" nedan. Skillnaden mellan samgranskning och den typen av jämförelser som beskrivs det senare avsnittet är att samgranskning görs på objektsnivå medan det som avses i "Jämförelser med annan statistik" är jämförelser av aggregerad data.

#### **3.1.6 Läs mer**

Biemer, P. & Lyberg, L. (2003). *Introduction to Survey Quality*. New Jersey: Wiley. (kapitel 8.4.5)

Skinner, C. "Measurement errors" i Davies, P. & Smith, P. (ed.), *Model Quality Report in Business Statistics*.

Weideskog, F. (2013) Arbetsdokument. *Record check – förslag på praktisk tillämpning*.

## 3.2 Latent klassanalys (Latent Class Analysis, LCA)

### 3.2.1 Inledning

LCA är en statistisk analysmetod som kan användas för att uppskatta sannolikheten för fel i mätningen. Grunden i LCA är latent variabler, d.v.s. antagna inte direkt observerbara variabler, som antas utgöra sanningen. Utifrån flera mätningar av observerbara variabler relaterade till samma latent variabel kan man skatta de underliggande latent variablerna.

De upprepade mätningarna kan vara av olika slag:

- Olika frågor, menade att mäta samma sak, som ställs vid samma tillfälle.
- Upprepning av frågan vid ett annat mättillfälle, en återintervju eller en ny självadministrerad enkät. Frågan kan då vara identisk eller förändrad.
- Samma fråga ställd vid olika omgångar i en panelstudie/longitudinell studie.

Värdet på de latent variablerna, de sanna värdena, beräknas genom modellering. Modelleringen kräver att de observerade variablerna, liksom de underliggande latent variablerna, är kategoriska variabler. Det mått vi får fram är därmed sannolikheten för att den observerade värdet är samma kategori som det sanna värdet. Fördelen med LCA är att vi genom att använda befintlig data kan få ett tydligt och direkt mått på sannolikheten för fel klassificeringar.

### 3.2.2 När kan metoden användas?

Metoden kan användas när det finns upprepade mätningar av samma variabel, samt mätningarna består av kategoriska data. För kontinuerliga data kan en motsvarande metod med en annan typ av modell användas, s.k. Markov quasi-simplex model (se Alwin, 2007).

### 3.2.3 Att tänka på inför genomförande av en Latent klassanalys

- Finns den data som krävs för att analysen ska vara möjlig?
- Hur ska resultaten användas?
- Ta hjälp från kunnig personal på någon av metodenhetererna.

### 3.2.4 Läs mer

Alwin, D.F. (2007). *Margins of Error: A Study of Reliability in Survey Measurement*. New Jersey

Biemer, P. (2010), *Latent Class Analysis of Survey Error*. New York: Wiley and Sons.

SCB (2010) *Bakgrundsfakta, Befolknings- och välfärdsstatistik 2010:1, Alternativa datainsamlingsmetoder i ULF, fas 2. En jämförelse mellan två olika datainsamlingsmetoder*  
([http://www.scb.se/statistik/\\_publikationer/LE0101\\_2010A01\\_BR\\_BE96BR1001.pdf](http://www.scb.se/statistik/_publikationer/LE0101_2010A01_BR_BE96BR1001.pdf)).



### **3.3 Markov kvasi-simplex-modell (Markov quasi-simplex model)**

#### **3.3.1 Inledning**

Liksom den latenta klassanalysen (LCA) så bygger Markov kvasi-simplex på en modell med latenta variabler där man utifrån flera mätningar av observerbara variabler skattar det sanna värdet. Skillnaden är att kvasi-simplex-modellen analyserar kontinuerliga svarsdata, medan LCA analyserar kategoriska svardata. Även måttet på mätfelet är olika – i LCA skattas sannolikheten för att det observerade värdet är i samma kategori som det sanna värdet, med en kvasi-simplexmodell kvantifierar man den slumpmässiga komponenten i mätfelet, dvs. den rena mätosäkerheten.

I kvasi-simplexmodellen används en reliabilitetskoefficient, som ligger mellan 0 och 1 där 1 betyder avsaknad av mätosäkerhet och 0 betyder att det bara finns mätosäkerhet. Reliabiliteten skattas utifrån korrelationen mellan tre mättillfällen där samma objekt deltagit vid samtliga tillfällen.

#### **3.3.2 När kan metoden användas?**

Metoden kan användas när det finns upprepade mätningar av samma variabel med samma objekt och variabeln är kontinuerlig.

Som mått på den slumpmässiga mätosäkerheten kan Markov kvasi-simplex-modellen endast användas om det är liten variabilitet inom objekt mellan mättillfällena. Använder man den skattade reliabiliteten för att se om insamlingsprocessen är stabil över längre tidsperioder så är inomobjektsvariabiliteten av mindre betydelse.

#### **3.3.3 Att tänka på inför analys med en Markov kvasi-simplex modell**

- Finns de data som krävs för att analysen ska vara möjlig?
- Hur ska resultaten användas?
- Hur stor är inomobjektsvariabiliteten? Finns det säsongmönster?
- Ta hjälp från kunnig personal på någon av metodenheterna.

#### **3.3.4 Läs mer**

Alwin, D.F. (2007). Margins of Error: A Study of Reliability in Survey Measurement. New Jersey

### 3.4 Medlyssning/Beteendekodning

#### 3.4.1 Inledning

I litteraturen beskrivs beteendekodning i första hand som en metod för att förtesta frågeformulär eller utvärdera intervjuares prestationer. Samma metod går dock att använda för att kvantifiera mätfel i en redan genomförd undersökning. Självklart krävs dock inspelade intervjuer. Beroende på hur man genomför beteendekodningen kan man få olika typer av kunskap om undersökningens mätfel.

Grunden i metoden är att man genom att lyssna på och koda genomförda intervjuer kan identifiera problem i mätningen. Kodningen ska göras utifrån ett standardiserat och genomtänkt kodschema där felkoder kan användas för att bedöma kvalitet på enskilda frågor och på undersökningen i stort.

Det är vanligt att koderna fokuserar på intervjuarens beteende. Kodschemat berör då punkter som *om intervjuaren läser frågan exakt efter manus* och *om intervjuaren ställer ledande följdfrågor*. För att kodningen ska ge en heltäckande bild av hur mätningen fungerar och för att få fram meningsfulla indikatorer på eventuella mätfel krävs dock också koder som fokuserar på respondentens reaktioner och beteende. Exempel på koder med fokus på respondenterna är *om respondenterna kommer med följdfrågor* och *om respondenten avbryter intervjuaren för att svara på frågan innan frågan är färdigläst*. Det finns forskning som visar att det främst är avvikelser i respondenters beteende som indikerar mätfel. Att när respondenten t.ex. avbryter intervjuaren så ökar risken för att fel data samlas in (Hess m.fl. 1999). Intervjuarens avvikelser från manus indikerar inte mätfel i samma utsträckning och fungerar därmed sämre som mått på datas kvalitet. Att många intervjuare avviker från manus på vissa frågor kan dock vara ett tecken på att de tycker frågan är svår att läsa vilket i sin tur kan vara en tillräcklig anledning att förändra de problematiska frågorna.

För att beteendekodningen ska kunna användas i mätfelsarbetet på ett effektivt sätt är det också att föredra att kodningen görs på frågenivå. För att kunna få information om problem med enskilda frågor behövs data på avvikelser kopplade till de specifika frågorna.

#### 3.4.2 Beteendekodning/medlyssning på SCB

SCB har inom ramen för Medlyssningsprojektet utvecklat rutiner och tekniker för medlyssning och beteendekodning av intervjuer. Resultaten används i kvalitetsarbetet genom feedback och utbildning till intervjuarna och i rapporter tillbaka till respektive produkt. I den ordinarie medlyssningen görs dock ingen kodning på frågenivå, så när beteendekodning ska användas i mer omfattande mätfelsstudier krävs ett utförligare kodschema.

Det finns också genomförda beteendekodningar på SCB där mer omfattande kodscheman används, se t.ex. Mättekniska utvärdering av ULF – CATI 2006 och slutrapporten för Fallstudie i medlyssning där medlyssningen testats på AKU.

#### **3.4.3 När kan metoden användas?**

Beteendekodning kan göras av telefonintervjuer och besöksintervjuer. Det krävs dock att intervjuerna spelas in, vid besöksintervjuer bör intervjun spelas in på video. För att metoden ska kunna användas krävs därmed teknik för inspelning.

Ytterligare en förutsättning är att aktuella respondenter informeras om och godkänner att bli inspelade.

#### **3.3.4 Att tänka på inför genomförande av beteendekodning för att undersöka mätfel**

- Kommer metoden kunna ge den information som vi är intresserade av?
- Hur ska vi använda resultaten på bästa sätt?
- Finns det tillgång till den teknik vi behöver?
- Vilka koder är relevanta att ha med i kodschemat?
- Behövs det någon ytterligare kompetens för t.ex. konstruktion av kodschemat och tolkning av resultaten?
- Behöver intervjuerna någon extra utbildning?
- Dokumentera genomförandet.

#### **3.4.5 Läs mer**

Dykema, J., Lepkowski, J.M. & Blixt, S. (1997). "The Effect of Interviewer and Respondents Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study". I Lyberg, L., Biemer, P. Collins, P. de Leeuw, E. Dippo, C. Schwarz, N & Trewin, D. (eds.) *Survey Measurement and Process Quality*. New York: John Wiley & Sons.

Ongena, Y.P. & Dijkstra, W. (2006). "Methods of Behavior Coding of Survey Interviews". *Journal of Official Statistics*, Vol. 22, No. 3, Pp. 419-451.

Hess, J., Singer, E. & Bushery, J. (1999) "Predicting Test-Retest Reliability from Behavior Coding". *International Journal of Public Opinion Research*, Vol. 11. No. 4. Pp. 346-359.

SCB-rapport (2006). *Mätteknisk utvärdering av ULF – CATI 2006*.

SCB Projekt rapport (2008). *Tekniskt test och fallstudie i medlyssning*.

## 3.5 Intervjuarvarians

### 3.5.1 Inledning

Intervjuarvarians-studier handlar om att undersöka intervjuarens effekt på svarsdata och alltså den delen av mätfelet som kan orsakas av intervjuaren och interaktionen mellan intervjuaren och respondenten. Målet i en survey är att intervjuaren inte ska styra eller på annat sätt påverka respondentens svar – svaren ska vara helt oberoende av intervjuaren. Det betyder att det inte bör vara någon skillnad mellan olika intervjuare och om det inte finns någon intervjuareffekt skulle två olika intervjuare få samma svar från samma respondenter.

### 3.5.2 Vilken typ av information kan metoden ge?

Genom att undersöka intervjuarvariansen kommer man endast åt fel som har med intervjuaren att göra eller där åtminstone intervjuaren är inblandad. Det går inte att nå fel som främst handlar om att en fråga misstolkas av respondenterna. Genom att tolka och analysera resultatet kan vi dock även använda intervjuarvariansen som en indikator på vilka frågor som är problematiska och som i interaktion med intervjuaren bidrar till mätfelet. Om det t.ex. finns en stor varians mellan intervjuarna på en specifik fråga, men en stor samstämmighet inom varje intervjuare är det sannolikt att även frågan bidrar till felet. Frågan är sannolikt vag och otydlig och leder till problem när respondenten ska tolka frågan vilket i sin tur kan göra att intervjuareffekten blir starkare. Det kan t.ex. handla om att respondenten kommer med följdfrågor som intervjuarna hanterar på olika sätt. Om instruktionerna till intervjuarna är bristfälliga blir den här problematiken extra allvarlig. Analysen i det här fallet ger oss dels information om att det existerar allvarliga mätfel, dels ger den oss information om att intervjufrågan eller intervjuarinstruktionen behöver förbättras för att minska framtida mätfel. Det den inte kan ge oss är information om vilka intervjuare som orsakar mer fel än andra eller vilken intervjuare som samlar in korrekt data.

I undantagsfall kan metoden även användas för att få fram indikationer på att en specifik intervjuare orsakar mätfel. En indikation på detta skulle vara om intervjuarens insamlade värden avviker från en i övrigt samstämmig grupp intervjuare.

### 3.5.3 När kan metoden användas?

För att kunna analysera och få fram intervjuarvariansen krävs att det finns en dokumenterad koppling mellan intervjuare och respondent, man måste veta vilken intervjuare som har intervjuat vem.

Det krävs också att fördelningen av respondenter till respektive intervjuare är slumpmässig. Respondenternas egenskaper ska inte skilja sig på ett systematiskt sätt mellan intervjuarna, t.ex. att en intervjuare endast intervjuar kvinnliga respondenter eller respondenter från storstadsområden. Om det från början finns en

skevhet i fördelningen respondenter-intervjuare kan inte eventuell intervjuarvarians förklaras av intervjuareffekt och det handlar således inte om ett mätfel.

#### **3.5.4 Att tänka på inför genomförande av en studie av intervjuarvarians**

- Kommer metoden kunna svara mot det vi är intresserade av?
- Går det att knyta intervjuare till data på objektsnivå?
- Vilka statistiska analysmetoder blir relevanta? Finns den kompetens som behövs för att genomföra analyserna?
- Fundera igenom hur resultaten ska kunna användas.
- Dokumentera genomförandet.

#### **3.5.5 Läs mer**

Biemer, P. & Lyberg, L. (2003). *Introduction to Survey Quality*. New Jersey: Wiley. (kapitel 5)

Groves, R.M. (1989). *Survey Errors and Survey Costs*, New Jersey: Wiley

Japac, L. (2005). *Quality issues in interview surveys*. Akademisk avhandling, Statistiska institutionen, Stockholms universitet.

Japac, L. & Lundquist, P. (2000). *Bortfallet – påverkas det av intervjuarnas attityder och strategier?* Rapport. Statistiska centralbyrån.

Isaksson, A. Lundquist, P. och Thoburn, D. (2008). *Optimalt antal kontaktförsök i en telefonundersökning*. Research and Development. Methodology reports from Statistics Sweden 2008:1. Stockholm: SCB

Lundquist, P. (2006). *Estimating interviewer effects in sample surveys. Some contributions*. Akademisk avhandling, Statistiska institutionen, Stockholms universitet.

### 3.6 Paradata

#### 3.6.1 Inledning

Paradata (även kallat processdata) är data om statistikproduktionsprocessen. Det kan vara data som automatiskt uppkommer och lagras under pågående statistikproduktion och som tas om hand och används i kvalitetskontrollen. Det kan också vara data som särskilt samlas in för att få mer kunskap om hur processen sett ut och som sedan kan användas i kvalitetsarbetet.

Paradata används mycket i arbete och analyser med bortfall, dels för att samla in hjälpinformation inför bortfallsanalyser och imputering, dels för att utveckla effektiva kontaktstrategier. Det finns dock exempel på hur man också kan använda paradata i mätfelsarbetet.

I texten nedan tar vi upp några exempel på hur paradata kan användas i mätfelsarbetet. De exempel som ges är inte på något sätt heltäckande, samtliga exempel är inte heller genomförbara på SCB som det ser ut idag med de verktyg och system som används. Framst handlar det om att lagring av data på det sätt som anges i exemplen är prestandatungt och svårt att göra i SIV och alltså i SCB:s webbenkäter. Det skulle bl.a. kunna leda till ett försvårat uppgiftslämnande och en ökad uppgiftslämnarvärda. Vilket i sin tur skulle leda till försämrade datakvalitet. Här presenteras några metoder inom paradata som skulle kunna vara mest meningsfulla utifrån ett mätfelsperspektiv även om de inte alltid går att genomföra i dagsläget.

#### 3.6.2 Svarstider

Att mäta tiden det tar för en respondent att svara på specifika frågor kan ge information om potentiella mätfel. T.ex. så kan en väldigt kort svarstid betyda att en respondent inte läst frågan ordentligt, inte tänkt efter innan hon eller han gav sitt svar, eller inte tagit sig tid att plocka fram korrekta uppgifter till det som efterfrågats. Men även långa svarstider kan indikera mätfel då en respondent som tar lång tid på sig att avge ett svar sannolikt är mer osäker på vad hon eller han ska svara, har svårt att förstå frågan eller har svårt att få fram de uppgifter som efterfrågas.

Om man ska mäta svarstider för att undersöka förekomsten av mätfel bör man alltså både titta på svarstider som är misstänkt långa och misstänkt korta.

*När kan analys av svarstider användas*

Svarstider kan mätas både vid intervjuundersökningar och för webbenkäter. För webbenkäter krävs processdata som är svårt att få fram på SCB. Vid besök- och telefonintervjuer kan intervjuaren administrera mätningen av svarstider manuellt och på det sättet få fram de mått man vill ha. Med hjälp av viss manuell hantering är det också möjligt att använda processdata som finns registrerad i

WinDATI (det datorstöd SCB använder för intervjuer). Det handlar då om tiden mellan två registrerade svar. Vi kan alltså inte vara säkra på att tiden motsvarar endast uppläsning av frågan och avgivande av svar. Perioden kan också innehålla kommentarer eller annat som tillfälligt avbryter intervjun (se nedan: När börjar och slutar svarsprocessen).

Att tänka på och ta ställning till inför mätning av svarstider

- **Baslinje.**  
Vad är en orimligt kort respektive lång svarstid? För att få fungerande mått på avvikande svarstider krävs en relevant baslinje att förhålla svarstiderna till. Baslinjen kan konstrueras utifrån svarstiden på enkla okomplicerade frågor. Baslinjen konstrueras individuellt för varje respondent.
- **När börjar och slutar svarsprocessen?**  
För att kunna mäta svarstiden måste man ta ställning till när svarsprocessen börjar och slutar. En del ser det som att processen pågår mellan det att respondenten hört eller läst hela frågan och det att hon eller han påbörjar sitt svar. Medan andra menar att processen startar redan när respondenten börjar läsa eller höra frågan och avslutas när besvarandet är helt färdigt. Man kallar de olika perspektiven på svarstider för aktiv och latent svarstid.
- **Outliers.**  
Det är vanligt med outliers i mätningar av svarstid för självadministrerade webbenkäter. Framför allt outliers med väldigt lång svarstid. Ofta handlar det om respondenter som blir avbrutna i sitt besvarande och gör någonting annat. Svartiden blir då inte ett mått på hur lång tid respondenter spenderar på svaret. Dessa respondenter bör sorteras bort från analysen, först måste man dock bestämma var gränsen ska dras för att ett svar ska definieras som en outlier.

Ett alternativ till att mäta svarstider på specifika frågor är att mäta tiden för besvarandet som helhet. Svarstiden kan då inte användas som en indikator på mätfel på enskilda frågor utan fungerar snarare som ett mått på uppgiftslämnarbördan som i sin tur antas påverka mätfelet. För vissa undersökningar som hanteras genom WinDATI finns det också data på den totala svarstiden för enskilda frågeblock. Sådan information kan användas för att bedöma vilka frågeblock som främst påverkar den totala bördan.

*Läs mer om svarstider:*

Bassili, J.N. (1996). "The How and Why of Response Latency Measurement in telephone Surveys" i Schwarz & Sudman (ed.) *Answering Questions. Methodology for determining Cognitive and Communicative Process in Survey Research*. San Francisco: Jossey-Bass Publishers.

Couper, M. & Kreuter, F. (2013) "Using paradata to explore item level response times in surveys", *Journal of the Royal Statistical Society*.

Draisma, S. & Dijkstra, W. (2004). "Response Latency and (Para) Linguistic Expressions as Indicators of Response Error" i Presser, S. (et.al.) (ed.) *Methods for Testing and Evaluating Survey Questionnaires*. Wiley.

Kreuter, F. (red.) (2013). *Improving surveys with paradata. Analytic use of process information*. New Jersey: Wiley.

Kreuter, F. & Casas-Cordero, C. (2010). *Paradata*. Working Paper No. 136. German Council for Social and Economic Data.

Kreuter, F., Couper, M. & Lyberg, L. (2010) "The use of paradata to monitor and manage survey data collection". Section on Survey Research Methods – JSM.

Yan, T. & Tourangeau, R. (2007) "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology*.

### **3.6.3 Respondenter som backar och ändrar**

På samma sätt som svarstider kan data om respondenters beteende i form av i vilken grad de backar i formuläret (webb-enkät) och eventuellt ändrar i sina svar ge information om problem i mätningen.

Forskningen på området är inte särskild omfattande men det finns studier som visar på samband mellan ändrade svar och försämrade svars kvalitet. De som backar och ändrar avger felaktiga svar i större utsträckning.

*När kan metoden användas (Respondenter som backar och ändrar)*



Om en mätning av ändringar ska göras på en webb-enkät krävs verktyg (s.k. skriptprogram/skriptspråk) som för närvarande inte används i SCB:s datainsamling. Om mätningen ska göras av en undersökning där datainsamlingen görs via telefonintervjuer krävs det att antingen intervjuarna kodar respondentens eventuella ändringar eller att kodning görs vid så kallad beteendekodning/medlyssning (se avsnittet om beteendekodning ovan).

*Läs mer om respondenter som backar och ändrar:*

Kreuter, F. (red.) (2013). *Improving surveys with paradata. Analytic use of process information*. New Jersey: Wiley.

Stieger, S. & Reips, U.D. (2010). "What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study." *Computers in Human Behavior*.

#### **3.6.4 Avbrott av intervjuer eller webb-enkät**

Avbrott i besvarandet av en enkät är ett problem främst för att det innebär bortfall. Men data om avbrott och var i ett frågeformulär de sker kan också fungera som en indikator på vilka frågor som är problematiska. Det finns forskning som visar att avbrott ofta hänger samman med komplexa, multipla och öppna frågor, d.v.s. frågor där uppgiftslämnarbördan är stor (Pyetchev, 2009). Var avbrotten sker ger oss därför information om vilka frågor som uppfattas som extra krävande. Detta innebär att vi kan få kunskap om i vilka frågor risken för mätfel är stor. Avbrotten används således som en indikator och ett indirekt mått på mätfel.

Uppgifter om avbrott kan bl.a. tas fram utifrån data på partiellt bortfall, den sista besvarade frågan blir då en markör för var besvarandet avbröts. För intervjuer är det också möjligt att använda intervjuarens resultatkod för avbrott för att vara säkra på att det partiella bortfallet verkligen handlar om ett avbrott. Den senaste besvarade frågan markerar då var i intervjun avbrottet skedde. I de fall data samlas in genom SIV är det möjligt att få information om var avbrottet skedde även om blanketten aldrig skickas in. Detta fungerar eftersom data om den senaste visade sidan lagras i systemet.

Att använda processdata på det här sättet överlappar till stor del med de metoder som beskrivs i avsnitten *Indirekta kontroller av uppgiftslämnandet* nedan. Läs även det avsnittet för att få tips på indikatorer på mätfel som är relativt enkla att mäta.

*Läs mer om avbrott i intervjuer eller webbenkäter:*

Kreuter, F. (red.) (2013). *Improving surveys with paradata. Analytic use of process information*. New Jersey: Wiley.

Peytchev, A. (2009). "Survey Breakoff". I *Public Opinion Quarterly*, Vol 73, No. 1, 2009.

### 3.7 Frekventa åtgärder vid mikrogranskning

#### 3.7.1 Inledning

Granskning av mikrodata är en etablerad del av SCB:s kvalitetsarbete. Granskningen ska upptäcka och åtgärda fel i det data som ska användas i statistikframställningen. Ett av granskningens huvudsyften är också att identifiera felkällor för att utifrån dessa åtgärda problem i produktionsprocessen och förhindra att fel uppkommer i kommande produktionsomgångar. I CBM (Current best methods) 2002:1 *Guide till granskning* står t.ex. "Ett upptäckt fel ska alltid ses som en representant för en felkälla eller ett problem med undersökningen."

Granskningen ska således, utöver att upptäcka och åtgärda fel i data, vara en del av mätfelsarbetet och det är viktigt att den kunskap och de data som skapas genom granskningen tillvaratas i produkternas förbättringsarbete, t.ex. förändringar i mätinstrumentet.

Granskardebriefings, där PCA/MÅT samlar granskarens erfarenheter och kunskap kring en undersökning, är en viktig del i hur granskningen tillvaratas. Men det finns också möjlighet att kvantitativt analysera problem i undersökningen och mätinstrumentet med hjälp av processdata från granskningen. Vi kan få information om vilka variabler som ofta fastnar i kontroller och som ofta blir föremål för ändringar i mikrodata. Den här informationen kan då också fungera som indikation för vilka variabler och vilka delar av frågeformuläret som är problematiska och bör förändras för att förbättra kvaliteten och undvika fel i fortsättningen. Det är rimligt att anta att de variabler som ofta fastnar i kontroller och som ofta ändras i granskningen innehåller betydande mätfel även efter det att granskningen har genomförts. De enkätfrågor som samlar in det data som ofta fastnar i kontrollerna är problematiska och sannolikt skapar de även fel som inte upptäcks i kontrollerna.

Att använda processdata på det här sättet ger alltså ett indirekt mått på mätfelet. Eftersom de upptäckta felen justeras i och med granskningen fungerar det inte som något direkt mått på mätfelets storlek i slutdata. De upptäckta felen fungerar istället som indikatorer på att frågorna är problematiska och att det sannolikt existerar ytterligare fel än de som upptäckts i granskningen.

#### 3.7.2 När kan metoden användas?

Användning av processdata från granskningen för att upptäcka, åtgärda och förebygga fel i mätningen kan göras i alla de undersökningar där man arbetar med mikrogranskning och där granskningsprocessen har en tillräcklig dokumentation, dvs. det finns uppgifter om kontroller, åtgärder och värden före och efter åtgärd.

Eftersom mikrogranskning av det här slaget endast görs för undersökningar som riktar sig till företag, organisationer och offentlig sektor blir också metoden endast aktuell för dessa undersökningar. I den mån även andra undersökningar anser att de har processdata som kan analyseras och användas är det dock självklart någonting positivt.

En viktig aspekt för att de upptäckta felen och åtgärderna ska kunna fungera som indikatorer för problem i mätningen och användas för att förebygga mätfel i kommande omgångar, är att kontrollerna (och åtgärderna) håller en hög kvalitet. Om ett stort antal fel förbises av kontrollerna, och några åtgärder inte kan ske, leder det inte bara till att ett stort antal fel finns kvar i data utan också till att indikatorerna blir ineffektiva för det fortsatta förbättringsarbetet. Kontrollernas kvalitet bör därmed utvärderas regelbundet.

Utöver processdata från kontroller kan en systematisk genomgång av granskningens loggböcker ge kvantitativa indikatorer på mätfel.

### **3.7.3 Att tänka på inför analys av frekventa åtgärder vid mikrogranskning**

- Är den information som metoden kan ge relevant?
- Finns den processdata som behövs, dvs. finns uppgifter om kontroller, åtgärder och värden före och efter åtgärd.
- Håller kontrollerna en tillräcklig kvalitet för att de ska vara fungerande indikatorer på problem i mätningen?

### **3.7.4 Läs mer**

Granquist, L. (1997). "The New View on Editing" i *International Statistical Review*, Vol. 65, No. 3.

Statistiska centralbyrån, CBM 2002:1. *Guide till granskning*.

Verksamhetsstödet, Statistikproduktionsprocessen process 5.2.

### 3.8 Indirekta kontroller av mikro- och makro-data

#### 3.8.1 Inledning

Nedan presenteras några förslag på indirekta kontroller av uppgiftslämnandet som bör kunna genomföras utan att det krävs allt för omfattande resurser. Förslagen överlappar till viss del både med hur kontroller konstrueras vid mikrogranskning och några av de metoder som beskrivs ovan under rubriken Paradata. Eftersom mikrogranskning sällan görs på data från individ- och hushållsundersökningar kan man se texten som tips på hur kontroller kan göras även vid den typen av undersökningar. Eftersom några återkontakter inte görs vid individ- och hushållsundersökningar så blir identifiering av mätfel och dess orsaker och förbättringar av bl.a. mätinstrumentet huvudfunktionen här. Detta till skillnad från vid mikrogranskningen, där den huvudsakliga funktionen är att åtgärda fel i aktuella data.

De indirekta kontrollerna som beskrivs här överlappar också med vad som görs i delprocessen Granska makrodata 6.2 (se beskrivningar i Verksamhetsstödet). En skillnad är dock att fokus ligger på att identifiera mätfel och på att eventuellt förebygga mätfel i kommande omgångar. Medan den makrogranskning som beskrivs i Verksamhetsstödet i första hand syftar till att kontrollera och eventuellt åtgärda fel i aktuella data.

#### 3.8.2 Frekvenser och fördelningar

Ett sätt att upptäcka mätfel är att använda den kunskap som finns om hur data kan förväntas se ut och utifrån det avgöra om det finns några misstänkta problem. Det kan handla om att jämföra grupper där vi teoretiskt och utifrån andra studier kan göra rimliga antaganden om fördelningar. Förhåller sig grupperna till varandra på förväntat sätt eller finns det en risk att grupperna uppfattar en fråga, dess nyckelbegrepp eller svarsalternativ på olika sätt? Vi kan t.ex. förvänta oss att gruppen äldre har sämre hälsa än en yngre grupp, eller att högutbildade har högre inkomst än lågutbildade. Om data visar på någonting annat är det sannolikt att mätningen innehåller någon form av fel, kanske tolkar grupperna begreppen i frågan på olika sätt.

Att kontrollera samband mellan variabler kan vara meningsfullt även när det inte handlar om att jämföra grupper utifrån tydliga bakgrundsvariabler. Det kan även finnas andra samband mellan variabler som teoretiskt bör respektive inte bör ha samband. Vi bör t.ex. hitta ett negativt samband mellan tid som spenderas på tv-tittande och tid som spenderas ute i skog och mark. Och vi bör hitta ett positivt samband mellan positiv inställning till miljöarbete och beteendet att sopsortera i hemmet. Om de förväntade sambanden inte återfinns i materialet är inte det någon garanti för att det existerar

allvarliga mätfel, men det är tydliga indikatorer som innebär att man bör utreda närmare om det är något som har gått fel.

Man kan också göra kontrollen ännu mindre komplicerad genom att helt enkelt studera enkla frekvenser. Är frekvensen för något värde så högt eller lågt att det är sannolikt att det blivit något fel i mätningen?

Samtliga av dessa tre tillvägagångssätt, att jämföra grupper, kontrollera samband och att studera enkla frekvenser kräver god ämneskunskap. I tolkning och användning av de resultat som fås fram kan det också krävas mätteknisk kompetens.

Det finns också stora risker med den här typen av analyser. De avvikelser som återfinns i data kan mycket väl överensstämma med faktiska förändringar. Eventuella slutsatser bör således dras med stor försiktighet.

### **3.8.3 Intern konsistens**

Att undersöka den interna konsistensen i data är ett annat sätt att få fram indikatorer på eventuella mätfel i en undersökning. Metoden liknar de som beskrivits ovan, men är tillräckligt annorlunda för att förtjäna en egen rubrik. En viktig skillnad är att det vi här avser är kontroller där man kontrollerar data mellan frågor som avser mäta samma sak. Detta alltså till skillnad från ovan där det handlade om samband mellan helt olika variabler. Det kan t.ex. vara att en variabel mäts både genom flera frågor som tillsammans ska fånga konceptet och genom en fråga som ensam förväntas mäta variabeln. Vanligt är också att man i samma frågeformulär både efterfrågar en summa och olika delar, dessa kan då kontrolleras mot varandra. Man kan t.ex. fråga om hur mycket pengar en respondent spenderar på nöjen varje månad, och även fråga hur mycket pengar respondenten lägger på olika preciserade nöjesaktiviteter. Svaret på den första övergripande frågan bör då ge samma svar som summan av de övriga. Om det inte gör det är det en tydlig indikator på att det finns fel i mätningen.

Den här typen av interna kontroller fungerar bäst när de är planerade och kan byggas in i frågeformuläret. Omfattningen bör då begränsas så att de inte ökar uppgiftslämnarbördan på ett omotiverat sätt. I vissa fall finns det också möjlighet att använda redan existerande frågor och alltså genomföra kontroller som inte har planerats i förväg och inte byggts in i instrumentet.

Panelundersökningar passar extra bra för studera den interna logiken. Det som kontrolleras då är de bakgrundsuppgifter som inte bör ha förändrats mellan de olika undersökningsomgångarna. Om svaren på dessa frågor ändå skiljer sig finns det uppenbara fel i någon av eller båda mätningarna.

#### **3.8.4 Kontrollera individuella svarsmönster**

En risk i undersökningar och en möjlig orsak till mätfel är att respondenterna skapar vissa mönster för hur de avger sina svar och fortsätter följa dessa mönster. Istället för att tänka efter och avge det mest lämpade svaret styrs respondentens svar av mönstret. Det kan vara både medvetet och omedvetet. Störst är risken när frågor med likadana svarsalternativ är uppställda i rad i t.ex. en tabell. Att kontrollera om respondenter följer tydliga mönster, t.ex. avger samma svar på samtliga frågor i en tabell kan ge oss information om problem i data.

Det händer att man i konstruktionen av frågeformulär bygger in kontroller genom att varva frågor i olika riktning. De flesta av frågorna konstrueras t.ex. så att svarsalternativet längst till vänster innebär en positiv attityd till olika aspekter av ett område, och några få formuleras så att samma svar innebär en motsatt inställning. Om respondenterna avger samma svar på samtliga frågor fast att svaren till stor del motsäger varandra är det en indikation på att det finns fel i data.

På SCB rekommenderar vi dock inte det här tillvägagångssättet eftersom man riskerar att istället för att kontrollera mätfel faktiskt skapar dem. Strukturen kan också skapa irritation hos respondenter eftersom de upplever sig kontrollerade.

#### **3.8.5 Jämförelser med annan statistik**

Ofta finns det statistik inom samma område som samlats in inom ramen för olika undersökningar. Att jämföra data på individnivå kan då ge rik information om kvaliteten på data på individnivå, se avsnittet om Record check ovan. Men även genom att enbart jämföra aggregerade data kan vi få värdefulla indikatorer på eventuella mätfel. Stora skillnader mellan statistik från olika undersökningar som ska beskriva samma fenomen tyder på att det finns fel i någon eller båda mätningarna.

Att jämföra en undersöknings resultat med aggregerad data låter väldigt enkelt och oproblematiskt. Men samtidigt finns det en rad problem med att göra sådana här jämförelser. Framför allt handlar det om att definitioner och periodiseringar kan skilja sig mellan de olika undersökningarna.

#### **3.8.6 Stora andelar partiellt bortfall samt "vet ej"-svar**

Stora andelar partiellt bortfall fungerar som en kvalitetsindikator dels på ett direkt sätt genom att det leder till mindre data. Det partiella bortfallet kan vara systematiskt genom att det främst är vissa grupper med specifika egenskaper som väljer att avstå från att avge svar på vissa frågor. Bortfallet leder då till felaktiga resultat. Dels kan det partiella bortfallet indikera lägre kvalitet på svaren även från de respondenter som besvarat frågan. Att en betydande andel har hoppat över frågan tyder på att den är problematisk på något sätt.

Det är rimligt att anta att även de som ändå har svarat har haft svårigheter, de kan ha tolkat frågan fel eller haft svårt att avgöra vad de ska svara.

Samma sak gäller för de frågor där andelen "vet ej"-svar är högt. Att många respondenter ansett sig inte kunna svara antyder att många av de som faktiskt avgivit svar är osäkra vilket i sin tur innebär att risken är stor att svaren inte stämmer.

Oavsett om det partiella bortfallet blir en kvalitetsindikator för att data saknas eller om det används för att indikera eventuella ytterligare fel, bör informationen användas i kvalitetsarbete. Frågorna med stora andelar bortfall bör förändras för att minska problematiken till kommande produktionsomgångar.

### **3.8.7 Tillägg för kontroll i enkät**

Ett sätt att få fram indikatorer på mätfel i en undersökning är att ta hjälp av uppgiftslämnarna. Man kan t.ex. be uppgiftslämnarna ange ett mått om hur säkra de är på de sina svar. En sådan fråga kan både kopplas till specifika frågor och till enkäten som helhet. Om man vill ha information om hur en fråga har tolkats och vad svaret innehåller kan man lägga till en öppen följdfråga och be uppgiftslämnaren ge en kort beskrivning av vad svaret avser eller kanske hur det har beräknats. Om uppgiftslämnarna anger en hög grad av osäkerhet och/eller beskriver sina svar på ett sätt som inte överensstämmer med frågeställarens definition finns det starka indikationer på fel i data.

Beroende på hur mycket hjälp man ber uppgiftslämnarna om kan tilläggsfrågorna ställas till hela urvalet eller till en begränsad del av det. Det är viktigt att tänka på den ökade uppgiftslämnarbördan som extrafrågorna innebär och aldrig ställa frågor som inte är motiverade.

### **3.8.8 Att tänka på vid indirekta kontroller av mikro- och makrodata**

- Fundera igenom vilken eller vilka metoder som kan ge relevant information och som går att genomföra.
- Hur ska resultaten användas?
- Ta hjälp av den ämneskunskap och mättekniska kompetens som behövs. Mätteknisk kompetens kan behövas i både planering och analys.
- Kom ihåg att metoderna ovan endast kan ge indikationer på mätfel. Resultaten måste tolkas och bör användas med försiktighet. En fördel är om det går att kombinera flera metoder.

### **3.7.9 Läs mer**

Det finns inte några väldokumenterade erfarenheter av att arbeta med den här typen av indirekta kontroller med syfte att kvantifiera



och förebygga mätfel. Någon omfattande litteratur på området har vi inte heller lyckas hitta. Följande källor kan dock ge visst stöd i arbetet med metoderna:

Dalén, J. (1981). *Metoder för evalvering av noggrannheten i SCB:s statistik. En översikt*. Promemorior från P/STM, nr. 6. Statistiska centralbyrån.

Biemer, P. & Lyberg, L. (2003). *Introduction to Survey Quality*. New Jersey: Wiley. (kapitel 5, avsnitt 8.4.3 och 8.4.4)

Verksamhetsstödet 6.2, *Granska makrodata*

Verksamhetsstödet 6.2, *Checklista för output-granskning*

---

**Översikt: Metoder för att undersöka och kvantifiera mätfel**

---

	<i>Direkt eller indirekt mått?(måttan gäller endast de delar av en undersökning som testats i mätfelsstudien)</i>	<i>Vilken typ av fel kan metoden uppskatta? (fel orsakade av intervjuare, mätinstrumentet, respondenten och/eller datainsamlingsmetoden)</i>	<i>Kan metoden ge kunskaper om orsaken till mätfel?</i>	<i>Vilken kompetens behövs?<sup>1</sup></i>	<i>Engelska begrepp att använda i litteratur- och informationssökning</i>
<b>Analys av befintliga data</b>					
<b>Intervjuarvarians</b>	Resultaten kan fungera som indikatorer på fel och problem i mätningen och är därmed ett indirekt mått. Samtidigt kan metoden ge ett direkt mått på variansen.	Metoden ger i första hand information om problem orsakade av intervjuaren, men kan till viss del även indikera problem i mätinstrumentet.	Till viss del. Metoden kan identifiera problematiska intervjuare och frågor.	Insamling Statistisk metod	Interviewer variability Interviewer variance Interviewer effect Interviewer bias

---

<sup>1</sup> Att viss ämneskunskap liksom grundläggande kunskap i statistisk metod ingår förutsätts här. När dessa listas som nödvändig kompetens är det för att markera att detta är en förutsättningen för en studies genomförande.

<b>Paradata</b>	Indirekt mått. Resultaten kan fungera som indikatorer på fel och problem i mätningen.	Paradata är ett samlingsbegrepp som består av en rad olika metoder och typer av data. I första hand ger dock de olika metoderna information om problem med mätinstrumentet och fel orsakade av mätinstrumentet.	Till viss del. Vissa typer av paradata kan identifiera vilka frågor som är mest problematiska.	Nödvändig kompetens varierar beroende på typ av paradata, ofta behövs dock kompetens i: IT Insamling Mätteknik	Paradata Process data Response latency Response times Survey breakoff
<b>Record check</b>	Beroende på hur metoden tillämpas kan den ge ett direkt mått på mätfelet. D.v.s. vi kan med hjälp av metoden skatta mätfelets storlek.	Record check ger information om det sammanlagda mätfelet, dvs. mätinstrument, respondent och eventuell intervjuare sammantaget. Det är dock inte möjligt att se hur stor del av felet som orsakats av respektive källa.	Endast begränsat. Metoden kan identifiera vilka delar av data som är mest problematiska.	Statistisk metod Ämneskunskap	Record check Administrative record check Administrative register check
<b>Beteendekodning/Medlyssning</b>	Indirekt mått. Resultaten kan fungera som indikatorer på fel och problem i mätningen.	Beroende på hur beteendekodningen genomförs kan den ge information om fel orsakade av både respondenten, mätinstrumentet och intervjuare.	Ja. Men det beror på vilka koder som används och hur avancerad analys och tolkning som görs.	IT Insamling Mätteknik	Behavioral coding Interviewer behavior Respondent behavior Interaction coding

<b>LCA</b>	Metoden kan ge direkta mått på mätfelets storlek i form av sannolikheter för felklassificeringar.	LCA ger information om det sammanlagda mätfelet, dvs. mätinstrument, respondent och eventuell intervjuare sammantaget. Det är dock inte möjligt att se hur stor del av felet som orsakats av respektive källa.	Endast begränsat. Metoden kan identifiera vilka delar av data som är mest problematiska.	Statistisk metod Ämneskunskap	Latent Class Analysis Markov Latent Class Model
<b>Frekventa justeringar vid mikrogranskning</b>	Indirekt mått. Resultaten kan fungera som indikatorer på fel och problem i mätningen.	Kan ge information om fel orsakade av respondent och mätinstrument (intervjuare är sällan aktuellt).	Till viss del. Metoden kan identifiera vilka frågor som är problematiska.	Mätteknik Statistisk metod Insamling/granskning	Statistical data editing Role of editing
<b>Indirekta kontroller av uppgiftslämnandet</b>	Indirekt mått. Resultaten kan fungera som indikatorer på fel och problem i mätningen.	Kategorin innefattar ett flertal metoder med delvis olika fokus. I första hand indikerar dock metoden fel orsakade av mätinstrumentet/frågorna.	Till viss del. Metoderna kan identifiera vilka frågor som är problematiska.	Mätteknik Ämneskunskap	Internal consistency Indirect measures of error

<b>Nya data eller anpassad design</b>					
<b>Randomiserade experiment</b>	Beroende på hur experimentet görs kan de ge ett direkt mått på mätfelets storlek genom att ge mått på variansen.	Vilken källa som fokuseras beror på typ av experiment.	Hur detaljerad information man kan få beror på typ av experiment.	Statistisk metod Mätteknik Insamling	Randomized experiments Embedded experiments Split-sample experiments Split-ballot experiments
<b>Återintervju</b>	Beroende på typ av återintervju kan metoden ge direkta mått på mätfelets storlek.	Metoden kan ge information om fel orsakade av mätinstrumentet, respondenten och intervjuare.	Ja, om studien designas för att ta reda på orsakerna.	Mätteknik Ämneskunskap Statistisk metod Insamling	Reinterview studies Golden standard Test-retest Replication
<b>MTMM/SQP</b>	Metoden kan ge direkta mått på mätfelet genom reliabilitets- och validitetskoefficienter på de testade frågorna.	Metoden kan ge information om mätfel orsakade av frågorna.	I skattningar som görs i SQP anges specifika förklaringar till bristande kvalitet i frågorna och förslag på hur frågan kan förbättras.	MTMM: Statistisk metod Mätteknik  SQP: Mätteknik	MultiTrait-Multimethod Survey Quality Prediction