

Til

## Varedeklarationer for stikprøveundersøgelser

For en konkret stikprøveundersøgelse er det en opgave for dataleverandøren at beskrive dens samlede usikkerhed og kvalitet. Flere af de afgørende kvalitetskomponenter i en stikprøveundersøgelse kan dog ikke måles med tal og må i stedet beskrives med analyser, redegørelse for dataindsamlingens principper og opfyldelsen af relevante standarder for den konkrete stikprøveundersøgelse.

De afgørende kvalitetskomponenter, der ikke kan måles med tal, vil det for brugeren altid kræve en vis faglig indsigt at kunne fortolke, og det forudsætter, at dataleverandøren har givet en fuld dækkende beskrivelse. Dataleverandøren er den bedste til at give en samlet vurdering af sikkerheden, men vurderingen vil altid indeholde et subjektivt element baseret på erfaringen og specifik viden om den konkrete stikprøveundersøgelse, som andre ikke har. Det er en meget omfattende opgave, at udarbejde konkrete principper for varedeklarationer for undersøgelser med helt forskellige stikprøvedesigns, der kort beskriver den samlede usikkerhed – og ingen har hidtil kunnet løfte en sådan opgave konkret. At der ikke findes en enkelt og let løsning fremgår bl.a. af, at Eurostat de sidste 5-10 år har arbejdet med udviklingen af fælles kvalitetsindikatorer for landenes dataindsamlinger. Det er en proces at udvikle kvalitetsmål, der dækker de forskellige landes problemstillinger. Selv om indikatorerne gradvist er blevet bedre, er det stadig kun indikatorer, der ikke kan stå alene, når man skal beskrive den samlede usikkerhed. Fx vil en indikator, der måler størrelsen af bortfaldet, ikke nødvendigvis være udtryk for en tilsvarende positiv/negativ betydning for den samlede sikkerhed i alle landene. Nogle stikprøver trækkes fra paneler, hvor personerne tidligere har indvilget i at deltage, og andre er helt nye stikprøver. Hertil kommer at skævheden i bortfaldet, og dermed dets betydning for den samlede usikkerhed, ikke altid hænger proportionalt sammen med størrelsen af bortfaldet. Et andet eksempel er den såkaldte ”forskerbeskyttelse”, der blev indført i 2001, og som betyder at 12-13% af den danske befolkning er forhåndsudelukket til at deltage i fx brugerundersøgelser på hospitaler og i kommuner, læseundersøgelser som PISA og arbejdskraftundersøgelser, der udvælges fra CPR. Det har placeret Danmark blandt de tre EU-lande, der har det største bortfald, men vi ved ikke præcist, hvor meget det påvirker undersøgelsesernes repræsentativitet og sammenlignelighed over tid.

En stikprøveundersøgelses samlede usikkerhed afhænger af en lang række komponenter:

- Hvorfra trækkes stikprøven? Er det fra hele population eller en del af population, er populationen aktuel og er der tale om en ”stikprøve i stikprøven”, der trækkes fra en population af fx internetbrugere, der tidligere har deltaget i undersøgelser.
- Hvordan trækkes stikprøven? Er der tale om repræsentativ udtrækning, hvor man kender den enkeltes sandsynlig for at blive valgt eller afhænger udvælgelsen af ukendte forhold, fx hvor mange telefonnumre man kan træffes på.
- Hvorledes gennemføres dataindsamlingen? Kontrakter man kun en udvalgt person én gang eller forsøger man mange gange, før man opgiver. Er det tilladt at lave erstatningsinterview med fx naboer, hvis den udvalgte ikke vil deltage. Hvordan administreres reglerne, når man fx skal vælge den, der næste gang har fødselsdag, og denne ikke er hjemme.
- Hvordan spørges der til holdninger? Er det den samme formulering. Hvad er instruktionen, når den udvalgte spørger om hjælp, og indgår spørgsmålet

i en større undersøgelse, der måske indeholder andre holdningsprægede spørgsmål.

- Hvordan opregnes der for skævt bortfald? Hvor meget kender man til de udvalgte baggrundsvariabler, og hvad skal der sammenlignes imod. Fx er tidligere stemmeafgivelse ved folketingsvalg en effektiv korrektionsfaktor ved meningsundersøgelser, hvis det sidste folketingsvalg ikke ligger flere år tilbage og knap så effektiv, hvis der spørges til EU-parlamentsvalget, der har 5 års interval, og hvor man ikke har så mange erfaringer med effektive korrektioner.
- Hvor stort er bortfaldet? Både størrelsen og skævheden af bortfaldet er udtryk for kvalitet. Men hvordan måles størrelsen af bortfaldet, hvis der vælges blandt de, der på internettet har givet tilsagn om at ville deltage i et vælgerpanel, og hvordan sammenlignes dette bortfald med et telefoninterview, der er trukket fra en kendt database med alle telefonnumre. Og hvordan måles skævheden af bortfaldet. Er det kun alder og køn, der inddrages, eller kender man fx også de udvalgte uddannelse, og er denne uddannelse oplyst af den udvalgte og kan den sammenholdes mod kendte registeroplysninger.
- Hvor stor er stikprøven? Denne usikkerhedskomponent kan måles med tal, hvis man kender den udvalgsandsynlighed, de udvalgte er valgt med. Med en stikprøve på 1.000 vil usikkerheden fx typisk være godt 3%, hvis det er et spørgsmål om, der er flertal i befolkningen for en bestemt holdning, omkring 2½%, hvis det er parti med omkring 25%’s opbakning og knap 2%, hvis det er et parti med 10% opbakning.

Det eneste af usikkerhedskomponenterne, der umiddelbart kan måles med et eksakt tal, er den sidste usikkerhedskomponent - stikprøveusikkerheden. De først nævnte faktorer har ofte langt større betydning og deres betydning afhænger ikke af stikprøvestørrelsen. Hvis de andre faktorer betydning er væsentlig, er det eneste man får ud af at øge stikprøven, at resultatet kommer tættere på den undervurdering eller overvurdering, de andre faktorer giver anledning til. Hvis man ønsker at reducere selve stikprøveusikkerheden til det halve, skal man firedoble sin stikprøvestørrelse. Der skal således meget til at reducere den eneste målelige komponent – stikprøveusikkerheden, og det har derfor ikke stor betydning for den samlede usikkerhed, om stikprøven fx er 800, 1.000 eller 1.200 personer.

Problemet med det målelige og ikke målelige sammenfattes bedst af citatet fra Einstein: *Ikke alt der kan tælles tæller, og ikke alt der tæller kan tælles.* Det gælder også mht. kvaliteten og den samlede usikkerhed ved stikprøveundersøgelser. Der er både i Danmark og andre steder i verden en proces i gang med at give brugerne en bedre forståelse af kvaliteten af de stadig flere stikprøveundersøgelser. Men før man kan beslutte at standardisere de forskellige kvalitetsmål, kræver det en proces, hvor man forsøger at harmonisere og konkretisere de enkelte stikprøvedesigns kvalitetsmål.