Myndigheten för samhällsskydd och beredskap

GÖTEBORGS UNIVERSITET

STUDIE

# Vad hände med opinionen?

## En översikt om metoder och strategier för att mäta snabba opinionsförändringar och undersöka opinioner/ händelser i sociala medier

# Faktaruta

Projekt: Utveckling av opinionsstudier för ett effektivt psykologiskt försvar

2017

Utförarorganisation: Göteborgs universitet

Författare: Bengt Johansson, Johan Martinsson, Peter Dahlgren, Elina Lindgren, Maria Andreasson och Elias Markstedt

Rapporten tar upp olika metoder och strategier för att mäta snabba opinionsförändringar och undersöka opinioner/händelser i sociala medier. Trots att det finns betydande svårigheter att undersöka opinioner och händelser i sociala medier visar rapportens slutsatser att det går att göra både undersökningar som fångar in förändringar av opinion och händelseförlopp med god kvalitet. En förutsättning är dock att tid och resurser läggs för att planera dessa studier.

MSB:s kontaktperson:
Fredrik Konnander, 010-240 41 32

Publikationsnummer MSB1169 - februari 2018
ISBN 978-91-7383-798-9

MSB har beställt och finansierat genomförandet av denna studierapport. Författarna är ensamma ansvariga för rapportens innehåll.

# Förord

Rapporten består av flera delar med delvis olika författare. Bengt Johansson och Johan Martinsson har haft det övergripande ansvaret för projektet och rapporten. De olika kapitlen har författats självständigt av respektive författare. Litteraturöversikten i kapitel 1, om datainsamlingsmetoder i opinionsundersökningar är författad av Elina Lindgren och Johan Martinsson. Kapitel 2 tar upp olika experiment som gjorts vid LORE, Göteborgs universitet och handlar om opinionsmätningar med korta mätperioder. Den har samförfattats av Johan Martinsson, Maria Andreasson, Elina Lindgren och Elias Markstedt. Rapportens tredje kapitel presenterar olika designer och strategier för att mäta snabba opinionsförändringar. Denna del har skrivits gemensamt av Bengt Johansson och Johan Martinsson. För rapportens sista kapitel står Peter Dahlgren som ensam författare till, vilket tar upp hur man kan undersöka opinioner och händelser i sociala medier. Alla författarna är verksamma vid Göteborgs universitet.

# Innehållsförteckning

1.1 Inledning ..... 6
1.2 Metoder för datainsamling ..... 6
1.2.1 Insamlingsmetoder ..... 6
1.2.2 Urvalsmetoder ..... 7
1.2.3 Val av undersökningsmetod ..... 8
1.2.4 Fördelar och nackdelar med olika insamlingsmetoder ..... 8
1.3 Skattningsfel i undersökningar som bygger på ett urval/stickprov ..... 9
1.3.1 Bortfall och täckningsfel ..... 10
1.3.2 Mätfel ..... 12
1.4 Opinionsmätningar och rörliga opinionsgrupper ..... 13
1.5 Svenska förhållanden och avslutande diskussion ..... 15
2.1 Inledning ..... 17
2.2 Vilken roll spelar många påminnelser och kontaktförsök för svarsfördelningen? ..... 18
2.2.1 Tidigare forskning ..... 19
2.2.2 Metod ..... 20
2.2.3 Data och analys ..... 20
2.2.4 Resultat ..... 21
2.2.5 Slutsatser och implikationer ..... 22
2.3 Empirisk delstudie 1 ..... 23
2.4 Empirisk delstudie 2 ..... 25
2.4.1 Design, data och resultat ..... 25
2.5 Empirisk delstudie 3 ..... 26
2.5.1 Resultat ..... 28
2.6 Sammanfattning ..... 30
3.1 Inledning ..... 31
3.2 Opinionsnivåer eller förändringar ..... 32
3.3 Grundmodell ..... 32
3.4 Befolkningsurval eller respondentpaneler ..... 33
3.5 Tvärsnittsdata eller paneldata (longitudinell data) ..... 35
3.6 Design ..... 36
3.6.1 Design 1. En mätning — efter en händelse ..... 37
3.6.2 Design 2. Två mätningar – under och efter händelse ..... 40
3.6.3 Design 3. Tre mätningar – före, under och efter händelse ..... 43
3.7 Vad ska man välja? ..... 45
4.1 Inledning ..... 48
4.2 Unga överrepresenterade i sociala medier ..... 49
4.3 Opinioner i sociala medier är inte representativa för befolkningen ..... 50

4.3.1 Ofullständig och felaktig information ..... 50
4.3.2 Ett användarkonto är inte detsamma som en person ..... 51
4.3.3 Aktiva användare är överrepresenterade ..... 51
4.3.4 Många normer är specifika för den sociala plattformen ..... 52
4.3.5 Sociala medier som understödjande del i ett övergripande arbete ... 52
4.4 Metoder för att analysera sociala medier ..... 52
4.5 Web scraping och API – samla in data ..... 54
4.6 Sentimentsanalys – klassificera positiv och negativ opinion ..... 54
4.6.1 Identifiera värdeladdning med lexikon ..... 55
4.6.2 Identifiera nyckelord med statistik ..... 56
4.6.3 Identifiera mönster med maskininlärning ..... 56
4.6.4 Analys av nätverkets struktur ..... 58
4.6.5 Utbrottsidentifiering – identifiera händelser automatiskt i realtid .. 58
4.6.6 Mäta opinionsförändringar över tid ..... 59
4.7 Sammanfattning och rekommendation av metoder ..... 61
4.8 Verktyg för analys av sociala medier ..... 61
4.9 Insamling av data från sociala medier ..... 62
4.9.1 Creepy ..... 62
4.9.2 DMI-TCAT ..... 62
4.9.3 Echosec ..... 62
4.9.4 Netvizz ..... 63
4.9.5 Social Feed Manager ..... 63
4.9.6 Observatory on Social Media ..... 63
4.10 Analys av data från sociala medier ..... 63
4.10.1 GATE ..... 63
4.10.2 Gephi ..... 63
4.10.3 NodeXL ..... 63
4.10.4 Opinion Finder ..... 64
4.10.5 OpenNLP ..... 64
4.10.6 Weka ..... 64
4.11 Slutsats och rekommendationer ..... 64
4.11.1 Sociala medier är inte representativa för befolkningen ..... 64
4.11.2 Följ opinionsförändringar för att upptäcka påverkansförsök ..... 64
4.11.3 Välja rätt verktyg kräver metodkunskaper ..... 65
4.11.4 Låt extern part följa händelser och opinionsförändringar ..... 65

# 1. Översikt över metoder för datainsamling i opinionsundersökningar

Elina Lindgren och Johan Martinsson

## 1.1 Inledning

Detta inledande kapitel avser att ge en översikt över olika sätt att genomföra opinionsundersökningar. Översikten omfattar olika metoder och tillvägagångssätt för datainsamling och analys för att skatta opinionslägen och opinionsförändringar. Översikten består av tre avsnitt. Först görs en genomgång av olika metoder för datainsamling och deras för- och nackdelar, sedan diskuteras olika typer av felkällor i opinionsundersökningar, och slutligen i det tredje avsnittet diskuteras de svenska förhållandena mer specifikt och resultaten sammanfattas.

## 1.2 Metoder för datainsamling

Med opinionsundersökningar menas en vetenskapligt grundad metodisk datainsamling, som syftar att kartlägga en definierad populations, till exempel ett lands befolknings, åsikter, attityder, kunskaper eller beteenden. Opinionsundersökningar avser oftast olika typer av samhällsfrågor, medan marknadsundersökningar exempelvis fokuserar attityder till kommersiella produkter eller tjänster. Opinionsundersökningar görs oftast kvantitativt, genom att ställa standardiserade frågor till ett urval av personer. Det finns dock även undersökningar med kvalitativa inslag i det insamlade datamaterialet så som användande av öppna frågor, men dessa ingår då oftast som komplement i en större statistisk undersökning. I följande avsnitt presenteras och diskuteras olika former av metoder för datainsamling, urvalsförvaranden, samt för- och nackdelar med de olika metoderna.

### 1.2.1 Insamlingsmetoder

Datainsamling för opinionsmätningar genomförs vanligen med hjälp av någon av följande tekniker: face-to-faceintervju, telefonintervju, postenkät eller enkät via internet (webbenkäter). Traditionellt har face-to-face, telefon- och post-undersökningar varit vanligast, men användandet av internet som insamlingsteknik har ökat dramatiskt under senare år (Dillman, 2011; Sánchez-Fernández m fl., 2012). Så kallade mixed-mode tekniker har också ökat, vilket innebär en kombination av två eller flera insamlingsmetoder (Dillman m fl., 2009; de Leeuw, 2005). Vidare har användandet av så kallade respondentpaneler, oftast kallade webbpaneler, ökat i popularitet, vilket innebär att undersökningar görs med deltagare som antingen själva aktivt anmält sig, eller som på olika sätt har blivit rekryterade till respondentpanelen. Dessa personer har då i förväg lämnat sina kontaktuppgifter och godkänt att bli inbjudna att delta i framtida undersökningar och därmed att få återkommande inbjudningar till att delta i olika studier (Brüggen & Dholakia, 2010).

### 1.2.2 Urvalsmetoder

Då opinionsundersökningar ofta rör större populationer (som till exempel befolkningen i Sverige) görs dessa vanligen inte på hela populationen, utan på ett så kallat stickprov, ett slumpmässigt urval, av personer. Den totala population, vilken innefattar hela gruppen individer om vilka man vill dra generella slutsatser, kallas ofta ”målpopulation.” Stickprovet, vilket utgör en del av målpopulationen, kallas istället ”undersökningspopulation”,”urvalsgrupp” eller rätt och slätt ”urval”.

Det är fullt möjligt att på en urvalsgrupp göra statistiskt tillförlitliga opinionsmätningar för en större målpopulation (ett urval på några tusen personer anses till exempel ofta vara tillräckligt för att dra slutsatser om svenska befolkningens åsikter). För att man skall kunna dra slutsatser utifrån ett stickprov måste man dock noga beakta vilka individer, eller vilka grupper, det är som man vill dra slutsatser kring – det vill säga vilken målgrupp det är man önskar undersöka. Om undersökningen gäller opinionen hos Sveriges befolkning måste man fundera på vilka man vill inkludera i definitionen ”befolkning.” Vill man till exempel att undersökningen ska innefatta, eller exkludera personer som är medborgare i Sverige men bor utomlands, personer som inte är medborgare i Sverige men som tillfälligtvis bor här, personer som har rösträtt i Sverige men inte är medborgare, och så vidare.

Metoderna för att välja ut personer för att delta i undersökningen, det vill säga vilka som ska ingå i urvalsgruppen, är flera. Den främsta metoden, den som de flesta undersökningar (dock inte alla) utgår ifrån, är ett så kallat slumpmässigt urvalsförfarande. Den grundläggande utgångspunkten i slumpmässigt urvalsförfarande är att alla individer som ingår i målpopulationen ska ha samma kända sannolikhet att komma med i urvalsgruppen (Blair m fl., 2013).

För att ett slumpmässigt urval ska kunna dras behövs ett register eller liknande, en så kallad ram, från vilken kontaktuppgifter till de individer som ingår i målpopulationen kan hämtas. Urvalsgruppen blir då en delmängd av personerna som ingår i urvalsramen. Det är därför viktigt att ramen som används täcker in så många som möjligt av individerna i målpopulationen. Ett folkbokföringsregister eller liknande (till exempel en lista över alla hushåll i en region), fungerar ofta bra som sådan. Det är dock sällan möjligt att skapa en ram som täcker in varje individ som ingår i målpopulationen. I detta föreligger en risk för att ett så kallat täckningsfel uppstår. Ett täckningsfel kan delas in i undertäckning (rampopulationen täcker inte in vissa individer i målpopulationen), eller övertäckning (rampopulationen täcker in individer som inte finns med i målpopulationen). Om till exempel målpopulationen för en undersökning är *boende i Sverige*, skulle undertäckning uppstå om rampopulationen uteslöt personer som inte är svenska medborgare, medan övertäckning skulle ske om rampopulationen innefattade personer som är medborgare i Sverige, men boende utomlands. Se till exempel Blair m fl. (2013) och Brewitz (2015) för mer diskussion kring detta.

Trots att ett slumpmässigt urvalsförfarande allmänt anses utgöra den bästa grunden för att dra statistiska slutsatser om en större målpopulation, kan även andra urvalsförfaranden, beroende på undersökningens syfte, ibland anses acceptabla. Ett exempel är ett så kallat självrekryterade urval där de som är intresserade av att delta i en undersökning själva anmäler sig för att delta vilket bland annat är vanligt i många medicinska studier och i psykologisk forskning. Som andra exempel på urvalsförfaranden som faller utanför kategorin

sannolikhetsbaserade urval, kan nämnas rena så kallade kvoturval. Idag är dock läget mer komplicerat då ibland ett i grunden slumpmässigt urval kombineras med kvoter för att säkerställa en demografisk representativitet eller för att minska behovet av viktning av urvalet. För mer information om olika typer av urvalsförfaranden, se till exempel Blair m fl. (2013), Esaiasson m fl. (2017) eller Johansson (2010).

### 1.2.3 Val av undersökningsmetod

En välgenomtänkt och väldokumenterad urvalsmetod är avgörande för undersökningens kvalitet. Vilken typ av metod som är lämplig att använda kan skilja sig åt beroende på vilken frågeställning undersökningen syftar till att besvara. När syftet med undersökningen är att fastställa generella åsikter och opinionstrender hos en större målpopulation (såsom Sveriges befolkning) används oftast stora urvalsgrupper som väljs ut genom ett slumpmässigt urvalsförfarande. Här används också oftast traditionella metoder för datainsamling, såsom telefonintervjuer och postenkäter, även om som påpekats tidigare internet blivit en alltmer vanlig teknik. Dessa undersökningar och urvalsförfaranden anses fortfarande ha den bästa generaliserbarheten, och därmed utgöra den bästa grunden för officiell statistik om en större befolkning. Se till exempel Blair m fl. (2013), Dillman m fl. (2009) och de Leeuw (2005) för översikter av undersökningsmetoder.

När syftet är att besvara frågor gällande specifika socio-demografiska grupper som är svåra eller kostsamma att identifiera med andra metoder, eller att genomföra experimentstudier inom ramen för en enkätundersökning, så kallade surveyexperiment, används istället allt oftare så kallade webbpaneler. Ibland kan då kraven på urvalets representativa kvalitet av praktiska skäl få ge vika och större fokus läggas istället på datakvalitet hos varje enskilt svar, det vill säga att de som deltar bidrar med så rättvisande svar som möjligt. Frågor som deltagarnas motivation att delta och besvara hela enkäten, även mer krävande frågor, är då av central vikt. Låg motivation ökar risken att personerna väljer att inte svara, bara svarar på delar av enkäten eller svarar slarvigt och oengagerat. Det kan innebära att personer svarar mindre sanningsenligt, och ger ogenomtänkta svar som inte speglar verkligheten.

### 1.2.4 Fördelar och nackdelar med olika insamlingsmetoder

Jämförelser mellan olika typer av datainsamlingstekniker för opinionsmätningar har genomförts i decennier. Sammantaget visar dessa att det finns en betydande skillnad i hur människor svarar på opinionsundersökningar som görs muntligen, via telefon och i direkt möte mellan respondent och intervjuare (face-to-face), och mellan undersökningar som görs skriftligen, via brev eller internet (Dillman m fl., 2009).

När det gäller möjligheten att nå ut till en bred population har internetenkäter och telefonintervjuer flera fördelar gentemot undersökningar som genomförs face-to-face eller via postenkäter. De förra brukar anses som mer kostnadseffektiva (Baca m fl., 2007; Dillman, 2011), i nuläget i Sverige är dock i allmänhet även postenkäter kostnadseffektiva. Huruvida en telefonundersökning eller en postenkät är ett dyrare alternativ beror i nuläget på omfattningen av undersökningen (hur lång den är) och vilken svarsfrekvens man önskar nå. Generellt kan sägas att för korta undersökningar kan telefon vara effektivare, medan postenkäter har en fördel för mer omfattande undersökningar med fler frågor. Särskilt internetundersökningar har

ökat i popularitet, då de erbjuder en både tids- och kostnadseffektiv metod för datainsamling (Dillman, 2011), ger snabbare svar (Ilieva m fl., 2002; Göritz, 2002), och är bekväma för respondenterna som kan vara mer flexibla med när de svarar på undersökningen (Sax m fl., 2003; Keusch, 2015).

Jämförelser mellan olika tekniker för datainsamling visar också att postenkäter och internetenkäter[1] tenderar att ge en lägre svarsfrekvens än telefon- och face-to-face intervjuer (Couper, 2000; Fricker m fl., 2005; Rogelberg & Stanton, 2007), men att de i vissa avseenden kan ge en högre svarskvalitet (Dillman m fl., 2009). Att särskilt internet kan öka svarskvalitén indikeras bland annat genom studier som visar att människor tenderar att ge längre öppna svar (Paolo, m fl., 2000; Smyth m fl., 2009), och även mer ärliga svar (Bachmann m fl., 1999; Mehta & Sivadas, 1995; Kiesler & Sproull, 1986) i internetenkäter. Studier har också visat att variationen i svar verkar vara högre i internetenkäter än i telefon- och face-to-face intervjuer, vilket indikerar att respondenter är mer bekväma att uttala mindre konsekventa åsikter när de genomför undersökningen i en mer anonym svarssituation (Homola m fl., 2016), liksom de är mer benägna att uttala impopulära eller känsliga åsikter.

Vidare har användandet av så kallade ”mixed-mode” tekniker, det vill säga olika kombinationer av datainsamlingsmetoder i samma undersökning, visat sig kunna förbättra svarsfrekvenser något (de Leeuw, 2005). Användning av olika datainsamlingsmetoder på ett och samma urval, men på olika delar av en undersökning (till exempel mellan inledande studie och uppföljningsstudier), eller användning av olika datainsamlingsmetoder för *olika* urval, som till exempel äldre och yngre, och högutbildade och lågutbildade, kan också öka svarsfrekvensen. Samtidigt innebär sådana här typer av förfaranden alltid risker för mätfel och oönskade skillnader mellan svaren, på grund av att utformningen och designen av enkäten skiljer sig mellan de olika insamlingsteknikerna. För mer information kring ”mixed-mode” tekniker, se till exempel Dillman m fl. (2009).

## 1.3 Skattningsfel i undersökningar som bygger på ett urval/stickprov

Opinionsskattningar som görs från ett urval (det vill säga det beräknade värdet på en viss variabel i en målpopulation som görs med hjälp av en undersökning på ett urval individer) antas alltid ha en viss avvikelse från det ”sanna” värdet för målpopulationen, populationsparametern. Hur mycket och på vilka sätt en sådan skattning avviker från det sanna värdet (undersökningens totalfel) påverkas av både så kallade slumpmässiga, och systematiska fel.

Slumpmässiga fel uppstår alltid i urvalsundersökningar. Det är här som det välkända begreppet felmarginal kommer in. Felmarginalen i en undersökning återspeglar bara det slumpmässiga fel som uppstår på grund av att vi inte undersöker alla individer i en population utan just ett slumpmässigt urval. Felmarginalen visar helt enkelt hur säkra vi är

[1] Problem med låg svarsfrekvens är ofta extra stora i webbpaneler, där bortfallet kan ske i tre olika faser. Den första fasen är rekryteringsfasen, där personer kan välja att inte gå med i panelen. Den andra bortfallsfasen uppstår när medlemmarna i panelen får en inbjudan att delta i en särskild undersökning, men väljer att inte delta. Till den tredje bortfallsfasen räknas de personer som väljer att aktivt avsluta sitt medlemskap i panelen (se Brüggen & Dholakia, 2010; Brewitz, 2015).

på att resultaten inte beror på just på slumpens roll när vi drar ett urval. Det systematiska felet, å sin sida, står för systematiska skillnader mellan det skattade variabelvärdet i undersökningar och det sanna populationsvärdet som inte beror på slumpmässigheten. Detta fel kallas ibland ”snedvridning”, eller ”bias”, och kan i många situationer stå för den stora delen av totalfelet. Systematiska fel uppstår ur alla andra källor till osäkerhet i undersökningen. Vanliga orsaker till systematiska skattningsfel är så kallade *täckningsfel* som nämndes tidigare (fel som uppstår till följd av att en del personer inte ingår i rampopulation – de kan inte nås med det kontaktsätt som används i urvalsförfarandet, alternativt att individer som *inte* tillhör målpopulation ingår i rampopulationen), svarsbortfall (fel som uppstår till följd av att det finns systematiska skillnader mellan de individer i urvalsgruppen som fullföljde undersökningen och de som inte gjorde det), och *mätfel* (fel som uppstår till följd av att personer, omedvetet eller medvetet, svarar fel på frågor, och på grund av dålig design av till exempel frågeformuleringar och svarsalternativ). Se till exempel Dillman & Smyth (2007) och Biemer m fl. (2011) för översikter.

### 1.3.1 Bortfall och täckningsfel

Det är i allmänhet praktiskt tagit omöjligt att skapa en urvalsgrupp som helt perfekt reflekterar målpopulationen. Ofta uppstår någon form av täckningsfel på grund av att ramen, eller registret från vilket urvalet dras, saknar täckning för vissa grupper av individer i målpopulationen. Ofta brister representativiteten i det insamlade datamaterialet också genom mer eller mindre stora svarsbortfall, särskilt allvarligt är detta i situationer då bortfallet inte är jämnt fördelat mellan olika grupper. Med detta i åtanke är det viktigt att noga överväga på vilka sätt, och i vilken mån avvikelserna från målpopulationen kan komma att påverka undersökningens tillförlitlighet.[2] Graden av risk för att olika typer av skattningsfel uppstår kan skilja sig mellan olika tekniker för datainsamling, och val av insamlingsteknik ska därför alltid göras med åtanke att minimera undersökningsfel. När sedan olika potentiella brister i datamaterialet har identifierats, finns ett antal metoder som kan användas för att förbättra opinionsmätningarnas generaliserbarhet.

En tumregel sägas vara att det systematiska felet i en opinionsskattning ökar om andelen bortfall i undersökningen är hög. Med bortfall åsyftas de personer i urvalsgruppen som på grund av att de inte kan, eller inte vill, avstår från att svara på en undersökning. Om alla tillfrågade i ett undersökningsurval hade samma sannolikhet att faktiskt svara skulle bortfallet inte vara något problem utan vara helt slumpmässigt (”missing at random”), men tyvärr verkar detta ofta inte vara fallet. Med hjälp av information som finns i befolkningsregistret kan man i Sverige till exempel se att bortfallet tenderar att vara större bland unga, i storstäder och bland utrikes födda (Brewitz, 2015). Detta leder till ökad osäkerhet vad gäller korrektheten i opinionsskattningarna. Om bortfallsgrupperna också beter sig annorlunda med avseende på undersökningsvariabeln jämfört med hela målpopulationen eller jämfört med de som deltar i undersökningen, ökar det systematiska felet och resultaten kan bli missvisande. Skattningarna som görs utifrån urvalsundersökningen blir då, med andra ord, antingen signifikant högre eller signifikant lägre än det ”sanna” värdet, det vill säga värdet

[2] Ett sätt att ta reda på hur tillförlitliga skattningar från ett undersökningsurval är, är att jämföra resultaten från undersökningen med resultat från andra representativa urval på relevanta variabler. Detta kan ge en fingervisning om hur träffsäkra resultaten är för en målpopulation (se Brewitz, 2015).

för målpopulationen. Detta minskar giltigheten och generaliserbarheten av resultaten, se till exempel Blair m fl. (2013) och Biemer m fl. (2011).

Utöver risken för systematiska skattningsfel kan ett stort bortfall medföra att det blir svårare att identifiera signifikanta samband mellan variabler och signifikanta skillnader mellan grupper och att möjligheten att dra statistiskt säkerställda slutsatser minskar då antalet som deltar i studien blir mindre (Cohen, 1992). Det finns sedan många år en generell trend av ökat bortfall i opinionsundersökningar, och denna sträcker sig i princip över alla typer av undersökningstekniker (Fan & Yan, 2010). Eftersom många politiker och andra beslutsfattare använder resultat från opinionsundersökningar som underlag för politiska beslut, ses ett stort bortfall i dessa vara särskilt allvarligt, och metoder för att justera för stora bortfall blir centralt.

En möjlig metod för att minska bortfallet under datainsamlingsfasen (som redan nämnts tidigare under avsnittet ”metoder för datainsamling”) är att använda sig av två eller flera olika metoder för datainsamling i samma undersökning (Dillman m fl., 2009). De positiva effekterna som detta kan ha på svarsfrekvensen kan dock, som också det tidigare nämnts, ske på bekostnad av att svaren hos de tillfrågade skiljer sig mellan de olika typerna av datainsamling, och metoden bör alltså användas med försiktighet. Även användandet av incitament, olika former av ersättning till deltagare, har visat sig ha positiva effekter (för ett aktuellt exempel i Sverige, se Arkhede m fl., 2017).

En annan vanlig metod för att hantera potentiella skattningsfel på grund av bortfall är att man justerar för bortfallet *efter* datainsamlingen (i analysfasen). Tillförlitligheten i de skattningar som görs beror nämligen inte bara på hur urvalet har skapats eller hur stort eller skevt bortfallet är, utan även på hur skattningarna *beräknas*. Statistiska skattningar görs normalt genom att varje observation i en undersökning ges en vikt som multipliceras med variabelvärdet. Produkten för varje observation adderas sedan till den så kallade populationsnivån. När sådana skattningar görs kan justeringar för bortfall göras. En vanlig metod för detta är att, med stöd av så kallade hjälpvariabler göra en omvägning (eller viktning) av datamaterialet, där enkelt uttryckt observationer för underrepresenterade grupper ges mera tyngd när skattningen av målpopulationens inställning i en viss fråga beräknas.

Med hjälpvariabler menas variabler för vilka värden finns för respondenterna som svarat på undersökningen, och för vilka totalsummorna finns för hela målpopulationen (eller, om sådana saknas, i alla fall för hela urvalsgruppen). När undersökningen bygger på ett sannolikhetsurval används ofta registervariabler som hjälpvariabler (se till exempel Särndal & Lundström, 2005, för mer information). När undersökningen bygger på ett självrekryterat urval, som till exempel ofta är fallet vid online-paneler, kan hjälpvariablerna också hämtas från den så kallade ”profilundersökningen” där paneldeltagarna svarar på olika frågor om social bakgrund och andra centrala demografiska variabler (Brewitz, 2015).[3] Viktigt att notera här är att bortfallsjustering med hjälp av viktning bara är effektivt om de hjälpvariabler som används för att få fram nya vikter är korrelerade

[3] Ett annat sätt att minska risken för skattningsfel är att använda sig av så kallade kvoturval, vilket enkelt sett innebär att datainsamlingen för olika grupper pågår tills ett visst antal svar från denna grupp har inkommit (se Blair m fl., 2013).

*både* med svarsbenägenheten hos olika respondenter, och med ”målvariablerna” (variablerna som undersökningen avser mäta). För mer information om olika former av viktning av data, se till exempel Kalton & Flores-Cervantes (2003), Betlehem & Keller (1987), och Särndal & Lundström (2005).

Det bör dock observeras att senare tiders metaanalyser har visat att lägre svarsfrekvenser inte på något automatiskt sätt medför mindre träffsäkra skattningar (se exempelvis Groves & Peytcheva, 2008). Omfattande genomgångar av ett stort antal studier har klarlagt att åtminstone när det gäller svarsfrekvensspannet 30 till 60 procent inte finns någon stark koppling mellan träffsäkerheten i skattningarna och andelen bortfall i en undersökning. Groves och Peytcheva (2008) trycker istället på att vi i varje enskild undersökning måste fundera på om de faktorer som är kopplade till bortfallet och svarsbenägenheten också är kopplade till det vi vill undersöka (utfallsvariablerna). Det är i dessa lägen vi har stark anledning att oroa oss för skattningsfel orsakade av låg svarsfrekvens. Ett exempel är till exempel medieanvändning som är starkt kopplat till ålder samtidigt som unga har lägre benägenhet att delta i surveyundersökningar (jämför Markstedt, 2014).

### 1.3.2 Mätfel

Med mätfel menas skillnaden mellan det observerade variabelvärdet, det vill säga det svar som en respondent anger, och variabelns sanna värde (det som respondenten under ideala omständigheter skulle ange). Mätfel kan uppstå genom brister i undersökningsdesign och mätinstrument, såsom formuleringar av frågor och svarsalternativ med mera. Det kan också uppstå genom misstag hos respondenterna, eller på grund av intervjuarens beteende. En väldesignad undersökning och motiverade respondenter är därför centralt för att minimera mätfel.

Något som tyder på att design av undersökningsfrågor och svarsalternativ kan leda till mätfel, är forskning som visar att skillnader i uppbyggnaden av undersökningsfrågor kan leda till skillnader i respondenternas svar. Ett exempel på detta är en studie av Christian m fl. (2008), som visade att skalor som är helt märkta med svarsetiketter (1 = mycket positiv, 2 = ganska positiv… 7 = mycket negativ), så kallade ”fully labelled scales”, ofta producerar mer positiva svar än skalor som bara har utmärkta ändpunktsvärden/ändpunktsetiketter.[4] Frågan är emellertid hur vi kan veta vilket av svaren som är mest rättvisande? Ett annat välkänt fenomen är när en frågetyp används där deltagarna ombeds säga om de ”instämmer” eller ”inte instämmer” med olika former av påståenden. Ibland leder sådana frågor till en förhöjd grad av instämmande jämfört med andra frågekonstruktioner. Särskilt problematiskt är att denna förhöjda tendens till instämmande ibland har visat sig vara högre bland respondenter med lägre utbildning.

Ett vanligt mätfel i opinionsundersökningar uppstår när respondenterna ger så kallade socialt önskvärda svar (på engelska, ”social desirability”). Med detta menas att respondenter tenderar att svara på opinionsundersökningar på ett sådant sätt de uppfattar vara mest socialt accepterat eller ”korrekt.” Det här problemet uppstår framförallt när frågorna gäller känsliga ämnen, som till exempel frågor om moral och politiska ställningstaganden, och det är större när

[4] Helmärkta skalor är vanligare i internetundersökningar och postenkäter, medan utmärkta ändpunktsvärden är vanligare i telefon- och face-to-face intervjuer (Dillman m fl., 2009).

undersökningen genomförs i direkt möte mellan intervjuare och respondent, än anonymt via brev eller internet (se till exempel Krysan m fl., 1994). Fenomenet kan också urskiljas i studier som visar att respondenter i muntliga undersökningar tenderar att ge mer positiva svar än deltagare i skriftliga undersökningar (se till exempel Christian m fl., 2008). Dessa resultat tyder på att respondenter som svarar direkt till en intervjuare helt enkelt inte vill uttrycka missnöje med en produkt eller annat som undersökningen avser, eller svara på annat sätt som går emot intervjuarens förväntningar (Dillman m fl., 2009). Mätfel som uppstår genom socialt önskvärda svar uppstår med andra ord ofta som en kombination av beteendet hos respondenten och en så kallad ”intervjuareffekt.” Undersökningar som genomförs över internet eller via postenkäter, i vilka respondenterna är mer anonyma i förhållande till intervjuaren, antas därför minska risken för mätfel som baseras på socialt önskvärda svar (Baker m fl., 2010).

Ett annat vanligt mätfel uppstår när respondenterna har svårt att förstå, eller helt missförstår, innebörden av en undersökningsfråga. Internetundersökningar och postenkäter kan här ha en viss fördel genom att respondenten kan besvara formuläret i sitt eget tempo, och därmed får större möjlighet att reflektera över frågan (Brewitz, 2015). Å andra sidan kan dessa tekniker för datainsamling också öka risken för den här typen av mätfel, då det inte finns någon intervjuare som kan svara på följdfrågor från respondenterna eller fånga upp uppenbara missförstånd. När opinionsundersökningar genomförs över internet och via postenkäter är det därför extra noga att tänka igenom frågeformuleringar och svarsalternativ för att minimera risken för missförstånd. Det är också viktigt att tänka igenom layout och design, för att minimera risken för att respondenterna missar viktig information. Dessutom är det önskvärt att för-testa frågor på en testgrupp eller ett mindre urval för att på så sätt säkerställa att frågeformuleringen fungerar på det sätt som är avsett.

## 1.4 Opinionsmätningar och rörliga opinionsgrupper

Följande avsnitt resonerar kring frågan huruvida vi fångar upp grupper i befolkningen som kan tänkas vara extra rörliga vad gäller politiska och sociala åsikter. Vi vet sedan tidigare att grupper som är mindre benägna att delta i opinionsmätningar bland annat har lägre utbildning och troligen också lägre intresse för politik och samhällsfrågor. Den samhällsvetenskapliga forskningen har försökt ta fram modeller och teorier för att förstå hur graden av politiskt engagemang, intresse och sofistikering hänger samman med tendensen att ta in politisk information och med att byta åsikt i olika frågor. Vi gör här därför en kort genomgång av några sådana teorier och avslutar med att diskutera hur och om detta kan påverka vår förmåga att fånga upp opinionsförändringar i undersökningar.

Begreppet politisk medvetenhet syftar på en persons grad av politiska kunskaper och intresse, samt förmåga att processa politisk information och forma stabila politiska åsikter. En hög grad av dessa egenskaper anses ofta, av demokratiteoretiker, som eftersträvansvärda, eller till och med nödvändiga, för en välfungerande representativ demokrati. Orsaken är att politisk medvetenhet antas öka individers förmåga att koppla politiska frågor till abstrakta, men grundläggande, politiska värderingar och uppfattningar (Luskin, 1990, 2002; Zaller, 1992, 2002; Converse 1964; DelliCarpini and Keeter 1996), vilket i sin tur antas göra de mer motståndskraftiga mot politisk manipulation.

I sin bok *The Nature and Origins of Mass Opinion* (1992) presenterar Zaller en teori om hur individer former politiska åsikter. Teorin sammanfattas i den så kallade ”RAS-modellen” (Receive-Accept-Sample), vilken kan översättas ungefär med att individer formar sina attityder utifrån de politiska budskap de har 1) *mottagit* (received), 2) *accepterat* och 3) *för tillfället prioriterar* (sampled). Enligt den här teorin ökar individers benägenhet att motta politiska budskap, även motstridiga sådana, med politisk medvetenhet och intresse. Politiskt medvetna blir på detta sätt bättre på att koppla politiska ståndpunkter till abstrakta värden, och kan av denna anledning bättre strukturera och forma stabila och genomtänkta politiska preferenser. Enligt Zaller innebär detta i förlängningen att personer med hög grad av medvetenhet blir mer konsekventa i sina åsikter, och därför också mindre lättpåverkade av politiska budskap och manipulation.

Det finns även andra teorier som bekräftar bilden av att de mest störst politiskt intresse också har relativt stabila åsikter. Teorin om ”motiverat politiskt resonerande” (på engelska ”motivated political reasoning”) är en sådan. Denna teori härstammar från kognitionsvetenskap och socialpsykologi, och används för att beskriva situationer där människor bildar och håller fast vid åsikter och föreställningar även i situationer där det finns mycket information som motsäger dessa föreställningar (se till exempel Festinger, 1957; Kunda, 1990; Nickerson, 1998; Redlawsk, 2002). Processen är känslomässigt motiverad, och har framförallt ansetts vara ett sätt att mildra det obehag människor känner när de upplever så kallad ”kognitiv dissonans” (Festinger, 1957). Med andra ord, snarare än att rationellt söka information som antingen bekräftar eller motbevisar en viss föreställning eller attityd, söker människor information som bekräftar de uppfattningar de redan har för att därmed rättfärdiga sina tidigare beslut och intagna ståndpunkter (Tavris & Aronson, 2007, sid. 3-5; Lodge och Taber, 2013). Givet att individer med hög politisk medvetenhet ofta har mer stabila och starka politiska åsikter, blir de utifrån den här teorin också mer benägna att använda sig av den här typen av psykologiska mekanismer.

Det finns i dagsläget visst stöd för ett positivt samband mellan ”motiverat resonerande” och hög grad av politisk medvetenhet (se till exempel Nickerson, 1998, för en översikt). Dessutom visar annan forskning kring opinionsfenomen som fientliga medieeffekter (hostile-media effect) att personer med starkt engagemang och tydliga åsikter inte tar till sig information som går emot deras egna åsikter. (Strömbäck & Johansson, 2017).

Men alla är inte intresserade av politik. I Zallers resonemang utgår han ifrån att majoriteten av medborgare istället har ett begränsat politiskt intresse, och därmed få strukturerade uppfattningar från vilka de kan härleda politiska preferenser från. Bilden av den allmänna opinion blir då något som drivs, och ändras av hur politiska eliter kommunicerar politiska budskap. Med detta perspektiv blir både allmänna opinionen och resultaten från opinionsundersökningar främst en reflektion av de budskap som för stunden är mest framträdande i den politiska debatten (se också till exempel Iyengar, 1991, och Miller & Krosnick, 1997 för liknande resonemang).

Konsekvensen för opinionsundersökningar av detta resonemang är att de kan ge ett överdrivet stabilt intryck av den allmänna opinionen. Detta eftersom politiskt medvetna och politiskt intresserade är överrepresenterade i många urval. Om undersökningarna inte förmår de ointresserade och mindre politiskt sofistikerade att delta och svara kan opinionssvängningar riskera underskattas. Det som talar mot detta är att de mindre intresserade också är svårare att nå med politisk

information då de inte är lika benägna att ge sådan information någon uppmärksamhet alls. Kanske är det därför framförallt i sammanhang då vi har anledning att misstänka att yttre händelser eller påverkansförsök faktiskt har nått ut till grupper som på ett allmänt plan har låg grad av politiskt intresse och engagemang vi bör vara särskilt uppmärksamma på risken att underskatta opinionsförändringar. I sådana lägen blir det av särskilt vikt att försöka anpassa undersökningsmetoderna för att nå ut till vanligtvis svårnådda grupper som exempelvis unga, utlandsfödda och lågutbildade.

## 1.5 Svenska förhållanden och avslutande diskussion

Inledningsvis bör sägas att det i Sverige föreligger jämförelsevis mycket goda förhållanden för att genomföra opinionsundersökningar av hög kvalitet. Utan att vara uttömmande kan flera faktorer nämnas. En sådan fördel är tillgången till ett befolkningsregister på individnivå av hög kvalitet som får användas för sådana undersökningar. Vi har också i Sverige i ett internationellt perspektiv fortfarande relativt höga svarsfrekvenser trots en tydlig nedåtgående trend de senaste decennierna. Vidare har vi även mycket god täckning av internettillgång både i hushåll och genom utbrett smartphoneanvändande. Många faktorer talar därför för att det är möjligt att genomföra opinionsundersökningar av hög kvalitet i Sverige.

Sverige utmärker sig också genom att många företag som arbetar med webbpaneler har byggt upp stora slumpmässigt rekryterade paneler, medan situationen i många andra länder är den att de självrekryterade webbpanelerna dominerar marknaden, i Sverige ses de fortfarande snarast som undantaget. Det är dock värt att lägga märke till att de flesta av dessa framförallt (och ofta även uteslutande) är rekryterade via telefon, vilket gör att de troligen dras med samma problem som de traditionella telefonintervjuerna vad gäller täckning och svarsbortfall med mera.

I sammanfattning kan sägas opinionsundersökningar med syfte att nå högsta möjliga kvalitet på nivåskattningar som exempelvis insamling av officiell statistik om en befolknings åsikter, attityder eller beteenden med fördel bör utföras på slumpmässiga urval så långt det är möjligt. Mindre undersökningar på specifika grupper, inklusive så kallade surveyexperiment, kan dock ofta även genomföras på andra typer av urval, som kvoturval och självrekryterade urval. När det gäller metod för datainsamling, det vill säga huruvida datamaterialet exempelvis samlas in via telefon eller face-to-face intervjuer, i postenkäter eller över internet, är det svårare att ge en generell rekommendation. Här måste olika fördelar och nackdelar vägas mot syftet med undersökningen; vilka krav på generaliserbarhet och representativitet som föreligger, vilka personer ingår i målpopulationen, hur undersökningens tid- och kostnadsramar ser ut, och så vidare.

Någon aktuell kartläggning över antalet vetenskapliga studier med olika surveymetoder är svårt att finna. Mängden studier med internet som huvudsaklig datainsamlingsmetod (Dillman, 2011; Sánchez-Fernández et al, 2012), och studier som använder en kombination av två eller flera insamlingsmetoder (Dillman m fl., 2009; de Leeuw, 2005) ökar dock stadigt.

Kostnaderna för hembesöksintervjuer (face-to-face) har ökat kraftigt och ligger numera långt över kostnaderna för andra typer av datainsamlingar. Telefonintervjuer å andra sidan dras med kraftiga och

ökande problem med sjunkande svarsfrekvenser och ökande problem att hitta telefonnummer till befolkningsurval. Webbundersökningar å sin sida har fortfarande vissa, men begränsade, problem med internettäckningen i den äldre delen av befolkningen, och svårigheterna att direkt använda slumpmässiga befolkningsurval vilket leder till bortfall i flera steg och medföljande oklarheter angående kvalitén. De postala enkäterna kan ibland vara ett alternativ, men även där sjunker svarsfrekvenserna och de kräver avsevärt längre datainsamlingsperioder. Vi återvänder diskussionen om hastighetens betydelse i kapitel 2 och i kapitel 3.

# 2. Opinionsmätningar med korta mätperioder

Johan Martinsson, Maria Andreasson, Elina Lindgren och Elias Markstedt.

## 2.1 Inledning

Detta kapitel behandlar frågan i vilken mån det är möjligt att genomföra opinionsmätningar via surveyundersökningar under väldigt korta tidsintervall. Tiden då fältarbetet pågår kan idag variera enormt mellan olika undersökningar. En del undersökningar har idag ett fältarbete som pågår i åtskilliga månader medan andra i extremfallet helt genomförs under en och samma dag. Längre fältarbetsperioder ger i allmänhet självfallet bättre möjligheter till en högre svarsfrekvens, och en högre svarsfrekvens brukar i sin tur anses hänga samman med mer träffsäkra resultat, vilket diskuterades i kapitel 1. Men hur stora skillnader är det i kvalitet mellan så kallade snabbmätningar och mer traditionellt upplagda surveyundersökningar? Och finns det några särskilda tekniker eller metoder som brukar tillämpas för att motverka skevheter vid snabbmätningar?

Överlag kan sägas att det inte finns särskilt mycket skrivet om detta ämne. Flera opinions- och marknadsundersökningsföretag i Sverige gör ibland så kallade snabbmätningar, där de genomför fältarbetet under 1-2 dagar och direkt efteråt rapporterar resultaten. De flesta sådana verkar göras online, men de kan också göras per telefon. En misstanke är då att de som är mest intresserade och har högst villighet att delta i den typen av undersökningar kommer bli överrepresenterade i svaren jämfört med de mindre intresserade. Man kan därför tänka sig att olika former av mer sofistikerade justeringsåtgärder skulle vara aktuella. De flesta företag i Sverige som genomför snabbmätningar verkar mena att denna metod i deras fall inte ger sämre kvalitet, utan att de har utvecklat metoder som ger tillförlitliga resultat. Bristande transparens i vissa fall gör detta väldigt svårt att bedöma. En samlad bild utifrån den information vi har är emellertid att de olika företagen gör få extra åtgärder för att säkerställa att snabbmätningarna ger lika hög kvalitet som andra undersökningar med betydligt högre svarsfrekvens. Ibland tillämpas en sorts kvoter efter vilka partier deltagarna röstade på i föregående riksdagsval, vilket även är brukligt i andra sammanhang.

Vi börjar kapitlet med en djupdykning i ett centralt arbete i den begränsade vetenskapliga litteratur som finns inom ämnet. Därefter följer sammanfattningar av tre empiriska fallstudier genomförts inom ramen för uppdraget och sedan en avslutande diskussion.

## 2.2 Vilken roll spelar många påminnelser och kontaktförsök för svarsfördelningen?

Ett välkommet undantag till bristen på litteratur om snabba undersökningar och (extremt) korta fältperioder är en arbetsrapport från National Centre for Research Methods vid University of Southampton från 2016. Följande delavsnitt av kapitlet sammanfattar och diskuterar därför rapporten ”Fieldwork effort, response rate, and the distribution of survey outcomes: a multi-level meta-analysis” författad av Sturgis, Williams, Brunton-Smith och Moore.

I den här forskningsrapporten undersöker Sturgis, Williams, Brunton-Smith och Moore hur utfallet i surveyundersökningar förändras när svarsfrekvensen ökar till följd av att fler kontaktsamtal till potentiella respondenter görs. Studien, som är en metastudie, innefattar sammanlagt 559 variabler från sex större brittiska face-to-face undersökningar. Resultaten visar att punktskattningarna på de flesta variablerna ligger mycket nära den slutliga skattningen efter bara ett till två samtalsförsök, och skattningarna varierar endast marginellt över de olika undersökningarna såväl som frågetyper (till exempel beteende, attityder och uppfattningar). Detta stärker tidigare studier som ifrågasatt styrkan i förhållandet mellan svarsfrekvens och snedvridning av resultat på grund av bortfall i opinionsmätningar. Implikationerna är att flera kontaktförsök till ”ovilliga” respondenter som syftar till att öka svarsfrekvensen inte är ett kostnadseffektivt sätt för att minimera skattningsfel.

Bakgrunden till studien menar Sturgis m fl. är de ökande utmaningarna som surveyundersökningar med slumpmässiga urval står inför idag: höga och ökande utförandekostnader, och låga och sjunkande svarsfrekvenser (Groves, 2011; Peytchev, 2013). Dessa två utmaningar, som varit synliga ända sedan 1990-talet, har blivit allt mer akuta de senaste åren. En viktig orsak antas vara framväxten av självrekryterade internetpaneler, som erbjuder både en snabbare, mer flexibel och billigare metod för att genomföra surveyundersökningar (Callegaro m fl., 2014). Då samtidigt svarsfrekvensen för sannolikhetsbaserade undersökningar blir allt lägre, ifrågasätter allt fler beställare av opinionsundersökningar om de kostnader som krävs för att uppnå höga svarsfrekvenser verkligen är motiverade. En fråga som ställs är om en undersökning med slumpmässigt urval, men låg svarsfrekvens ger bättre kvalitet på datamaterialet i jämförelse med en väldesignad undersökning med kvoturval, t ex från en självrekryterad respondentpanel. En annan fråga är huruvida kvaliteten på datamaterialet faktiskt ökar med högre svarsfrekvens.

En standardmetod för att nå målen om hög svarsfrekvens är att intervjuare gör upprepade kontaktförsök (genom till exempel besök eller telefonsamtal) tills kontakt, och förhoppningsvis också ett deltagande i undersökningen uppnås. För många intervjuundersökningar handlar detta ofta om att genomföra åtminstone sex samtal, vid olika tider på dagen och på olika dagar i veckan, innan en potentiell respondent klassificeras som en så kallad ”icke-kontakt” (Legleye m fl., 2013). Det innebär också ibland att intervjuare återkommer till individer som initialt tackat nej till att delta, i syfte att övertala personen att ändra sig.

Att övertala personer som redan tackat nej till en inbjudan innebär dock flera tveksamheter. Praktiskt sett innefattar sådana här typer av övertalningsprocesser ofta att individen erbjuds (ytterligare)

ekonomiska incitament för att delta, vilket gör att kostnaderna för att öka svarsfrekvensen hos dessa initialt ovilliga individer blir särskilt höga. Metoden skapar också forskningsetiska frågor, där särskilt principen om frivilligt deltagande och rätten till privatliv – grundläggande krav inom forskning på mänskliga subjekt (AAPOR 5, 2014) – ifrågasätts. Vidare kan ifrågasättas huruvida vinsterna i form av ökad representativitet (om sådana ens finns) faktiskt ökar kvaliteten på de avgivna svaren, då svaren kommer från mer motvilliga, och därmed antagligen mer oengagerade, respondenter (Kreuter m fl., 2010; Roberts m fl., 2014). Sammantaget kan kostnaderna för att nå dessa så kallade ”svårnåbara” respondenter, både genom upprepade kontaktförsök och genom övertalningsförsök, vara höga på flera plan.

Trots att metoderna i allmänhet visat sig vara effektiva för att öka svarsfrekvensen i opinionsundersökningar, är det mindre klart hur effektiva de är för att minska snedvridning i svaren som kan uppstå på grund av ett stort bortfall. Det finns ett starkt teoretiskt antagande att ökad svarsfrekvens hos ”svårnåbara” individer minskar snedvridning i datamaterialet, då personer som är mindre tillgängliga och samarbetsvilliga sannolikt också är de som är underrepresenterade i kvoturval och i slumpmässiga urval med låg svarsfrekvens. Det finns dock vissa indikationer på att de vinster i representativitet som intervjuer med svårnåbara individer kan vara mindre än vad som traditionellt antagits (Curtin m fl., 2000). Utöver att svaren från svårnåbara (och mindre samarbetsvilliga) individer kan tänkas ge mindre engagerade och genomtänkta svar som potentiellt kan försämra datakvaliteten, anges två andra faktorer som huvudsakliga anledningar till tveksamhet. För det första utgör gruppen individer som är särskilt svåra att nå i allmänhet en jämförelsevis liten del av det totala urvalet. Därför, för att ha substantiell inverkan på en skattning hos en målpopulation måste gruppen svårnåbara respondenter skilja sig väsentligt på de *relevanta undersökningsvariablerna*, jämfört med den grupp av svarande som är mer lättillgängliga. I tillägg till detta finns det ökande bevis för att benägenheten att svara på inbjudningar till undersökningar endast är svagt korrelerad med många av de typer av egenskaper som rutinmässigt mäts i undersökningar (Groves 2006; Peytchev 2013). Gruppen med svårnåbara respondenter verkar med andra ord vara ganska lik den grupp av svarande som är samarbetsvilliga redan efter ett litet antal kontaktförsök på många observerbara karaktärsdrag. Detta indikerar, enligt de brittiska författarna, att kostnaderna för att genomföra surveyundersökningar kan minskas betydligt genom att man tolererar en lägre svarsfrekvens som norm, utan att man för den skull äventyrar kvaliteten och representativitet på datamaterialet och kvaliteten på svaren.

### 2.2.1 Tidigare forskning

Sammanfattningsvis menar Sturgis med kollegor (2016) att den tidigare forskningen på det här området visar att snedvridning i svar (bortfalls bias) och påföljande skattningsfel endast verkar vara svagt korrelerade med svarsfrekvens (se till exempel Groves, 2006; Groves och Peytcheva, 2008; Schouten, Cobben och Bethlehem, 2009). Ett litet, men växande antal studier har också visat att även om individer som är svårnåbara skiljer sig från mer samarbetsvilliga och kontaktbara individer på olika bakgrundsvariabler, har bortfallet av dessa individer ofta endast marginell inverkan på variationen i den slutgiltiga skattningen (se till Curtin m fl., 2005). Det har till och med, i vissa fall, observerats ett svagt *negativt* samband mellan ökad svarsfrekvens och

representativiteten i svaren hos urvalsgruppen (Bethlehem m fl., 2011; Fuchs, Bossert och Stukowski, 2013).

Trots att forskningen på det här området är relativt stor är enligt de brittiska forskarna en svaghet att de ofta varit relativt begränsade i form av ämnesområde, undersökningsmetod, och undersökningsfrågor. Genom att studera variation i punktskattningar efter olika antal kontaktsamtal på ett stort och varierat antal undersökningsvariabler, bidrar den här studien således till forskningsfältet med en ökad generaliserbarhet av resultaten.

### 2.2.2 Metod

Sturgis m fl. (2016) menar att den principiellt bästa designen för att undersöka hur svarsfrekvens är relaterad till fördelningen på en undersöknings utfall då information om respondenterna på ett särskilt, externt kriterium, är tillgängligt för både de svarande i undersökningen, och för bortfallet. Variationen på det här kriteriet kan då jämföras med den eller de variabler som mäts i undersökningen, för att på så sätt uppskatta hur stor snedvridning av svaren som uppkommer på grund av bortfallsbias (och andra former av fel), allteftersom undersökningens svarsfrekvens ökar.

Det är dock i praktiken ofta inte så meningsfullt att göra sådana undersökningar. En första anledning är att de flesta egenskaper som rutinmässigt mäts i undersökningar ofta saknar en extern referent. Den andra är att om det ändå skulle finnas en sådan referent tillgänglig, skulle det inte finnas någon anledning att göra en skattning på denna i en (ny) undersökning som endast innefattar ett urval av målpopulationen. Sådana här studier är därför sällsynta, och de som finns är sannolikt inte representativa för den mängd olika individuella egenskaper som mäts i surveyundersökningar. Det är därför vanligare att forskare jämför fördelningen för svarande och bortfall på variabler som är tillgängliga på rampopulationen för undersökningsurvalet, eller som finns länkade till rampopulationen från en extern källa. Andra vanliga metoder innefattar uppföljande undersökningar på ett urval av bortfallet, eller jämförelser mellan svarande och bortfall på frågor som ställts i inledande profilundersökningar eller liknande.

### 2.2.3 Data och analys

Studien bygger på data från sex brittiska face-to-face undersökningar med slumpmässigt urval, vilka täcker en variation av ämnesområden, urvalsförfaranden och svarsfrekvens (svarsfrekvensen varierar mellan 54 och 76 procent). Undersökningarna som inkluderas är: the British Crime Survey (2011); the British Election Study (2010); the Skills for Life Survey (2010-11); the Taking Part Survey (2011); the Community Life Survey (2013-14); and the National Survey for Wales (2013-14). Från varje undersökning används alla icke-demografiska variabler som ställts till alla deltagare (559 undersökningsfrågor totalt).

Inför analyserna transformerades alla frågor till binära kategoriska variabler (där respondenterna kunde välja mer än ett svarsalternativ behandlades varje svarsalternativ som ett separat binärt utfall). För varje sådan variabel räknades sedan den absoluta procentuella variationen i varje kategori, för varje antal kontaktsamtal, och jämfördes mot variationen i den slutgiltiga skattningen. För kategoriska frågor med k svarsalternativ härleddes, mer precist, ett ”k-1” absolut

procentuellt skillnadsmått, med kategorin med lägst antal svar utelämnad. För en binär svarsvariabel (k = 2) kodad som 0 eller 1 beräknades med andra ord den procentuella skillnaden i proportionen som svarade ”1” på frågan efter ett enda samtal, och proportionen som svarade ”1” efter alla samtal, och så vidare. Författarna kunde således för varje undersökningsvariabel uppskatta den förväntade skillnaden mellan en punktuppskattning som gjordes vid samtal *n,* och skattningen som gjordes efter alla samtal och därmed på det slutgiltiga antalet svarande i undersökningen.

Författarna använder ett meta-analytiskt flernivå-ramverk (se till exempel Goldstein, 2011) för att analysera de 1 250 absoluta procentuella skillnaderna i skattningar från de 559 frågorna. Detta tillvägagångssätt gör det möjligt att göra en sammanslagen uppskattning av den genomsnittliga absoluta procentuella skillnaden i svarsdistribution för varje kontaktsamtal som genomfördes. Den ”poolade” skattningen är ett viktat medelvärde av den procentuella skillnaden för varje fråga, inom varje undersökning.

### 2.2.4 Resultat

Analyserna visar att svarsfrekvensen ökar stadigt för varje samtalsperiod, med viss variationen mellan de olika undersökningarna,[5] men att skillnaderna i svarsfrekvens endast förklarar en liten del av variationen som uppstått i vissa skattningar till följd av snedvridning på bortfallet. Istället verkar majoriteten av den här variansen kunna förklaras på en ”frågetyps-nivå” (det vill säga om frågan handlar om attityder, beteenden eller uppfattningar).

Den förväntade medelskillnaden från den slutgiltiga skattningen (som alltså inkluderar alla svarande i undersökningen), över alla 559 variabler är efter ett samtal 1.6 procent. Skillnaden sjunker till 0.7 procent efter tre samtal, och till 0.4 procent efter fem samtal. Den genomsnittliga absoluta procentuella skillnaden för post-stratifierade skattningar (skattningar som görs efter att datamaterialet viktats, och svaren från underrepresenterade grupper getts mer tyngd) efter det första samtalet är 1.4 procent, med en minskning till 0.8 procent efter det andra samtalet, och till endast 0.3 procent efter det femte samtalet.

Analyser på frågenivå visar att den absoluta procentuella skillnaden för mer än hälften av frågorna (54 procent) med statistisk säkerhet inte kan särskiljas från noll, ens efter det första samtalet. Vid det femte samtalet har talet stigit till nästan 60 procent av alla frågor. Medan den genomsnittliga skillnaden ligger (anmärkningsvärt) nära den slutliga fördelningen även efter endast ett samtal, noteras vid varje kontaktsamtal en liten minoritet av frågor med signifikant högre absoluta skillnader jämfört med medelvärdet. Detta är särskilt synligt efter ett samtal, där den största skillnaden är 6.6 procent. Alla dessa frågor handlar om beteende (två handlar om användningen av Internet, en om hur många timmar människor vanligtvis tillbringar hemifrån under dagen, och en handlar om antalet pubbesök som gjorts den

[5] The British Crime Survey har högst svarsfrekvens vid varje uppringningsperiod, vilket enligt författarna reflekterar det stora intresset för dessa frågor i Storbritannien, medan till exempel undersökningen ”Skills for Life,” vilken enligt författarna är både tids- och arbetskrävande för respondenterna, har lägre svarsfrekvenser särskilt i tidigare kontaktskeden.

senaste månaden). Efter det första samtalet har flera skillnader minskat, och efter det femte samtalet har även den största skillnaden på 6.6 procent minskat till endast 1.4 procent.

En analys där samvariationen för undersökningsfråga (i vilken undersökning frågan ställdes) och frågetyp (beteende, attityder eller uppfattningar) inkluderas visar att de procentuella skillnaderna mellan olika skattningar är signifikant större i vissa av undersökningarna. Skillnaderna är mest synliga för TP undersökningen, där den absoluta procentuella skillnaden generellt är nästan två gånger så stor som för BCS undersökningen. Trots detta, om man betänker de små initiala skillnaderna mellan undersökningsfrågorna, utgör detta mått fortfarande endast en absolut procentuell skillnad på 1.9 procent i TP undersökningen vid det första samtalet (jämfört med strax under 1 procent för BCS undersökningen). Det är troligt, menar författarna, att effekterna för de olika undersökningarna reflekterar skillnader i undersökningarnas ämnesinnehåll, då vissa ämnen som är mer känsliga för variation i svarsfrekvensen är mer frekvent förekommande i vissa av undersökningarna.

Den här analysen visar också att skillnaderna ofta är större för enkelkodade frågor, med absolut procentuell skillnad upp till 50 procent större än de för kategoriska frågor där respondenterna ombeds att ge flera svar på samma fråga. Såsom är fallet när det gäller ”undersökningseffekter” är det dock också här svårt att säga om dessa effekter är resultat av formen på svarsalternativen, eller om de beror på samvarians mellan ämnesområdet för frågan och det svarsformat som används. De flesta skillnaderna syns också generellt mellan beteendefrågor och frågor som mäter attityder och uppfattningar. Den absoluta procentuella skillnaden för attitydfrågor och frågor som mäter uppfattningar skiljer sig inte signifikant från beteendefrågor. Dock, om interaktionen mellan nummer på samtal och frågetyp inkluderas, finns större skillnader för frågor som mäter uppfattningar efter det första telefonsamtalet. Denna skillnad minskar vid det andra telefonsamtalet och försvinner vid det tredje samtalet. Både attitydfrågor och frågor om uppfattningar har också något större absoluta procentuella skillnader vid det femte samtalet jämfört med beteendefrågor. Det övergripande mönstret för samvariationen är dock att det endast finns en liten marginell variation, och denna finns alltså inom vad som redan är små skillnader.

### 2.2.5 Slutsatser och implikationer

Sammanfattningsvis visar den brittiska studien att punktskattningarna på de flesta variablerna som analyserats ligger mycket nära den slutliga skattningen efter bara ett till två samtalsförsök (även innan någon post-stratifierad viktning tillämpats). Skattningarna varierar också endast marginellt över de sex undersökningarna som studien innefattar, och över de olika frågetyper (beteende, attityd och uppfattningar) som ingår. Sammantaget bekräftar därmed resultaten från studien resultaten från tidigare studier som påvisat att förhållandet mellan svarsfrekvens och snedvridning i urvalsskattningar på grund av bortfall är marginellt.[6]

[6] Det går, enligt Sturgis m fl., inte att med den här studien säkert säga att svarsfrekvens är svagt relaterad till snedvridning till följd av bortfall, eftersom bortfallet för de sex studierna inte observerats. Resultaten kan endast användas som indikator på snedvridning på grund av bortfall, om de slutgiltiga skattningarna

I praktiken indikerar resultaten att ett större antal kontaktförsök (och därmed högre svarsfrekvens) inte är ett kostnadseffektivt sätt att minimera undersökningsfel (skattningsfel). Om fältarbetet hade upphört efter det första samtalet, skulle de resulterande punktskattningarna ha varit mycket lika dem som erhållits efter alla samtal, om än med en kraftigt minskad urvalsstorlek. Att stoppa efter 5 samtal skulle ge skattningar som är praktiskt taget omöjliga att skilja från de slutgiltiga skattningarna. Enligt författarna implicerar detta att betydande kostnader för utförande av surveyundersökningar skulle kunna sparas, eller användas mer effektivt, genom att antalet samtalskontakter begränsas och att försök att övertala ovilliga individer att delta minskades eller stoppades helt.

## 2.3 Empirisk delstudie 1

Följande avsnitt använder sig av en liknande analysstrategi som den rapport av Sturgis m fl. (2016) som diskuterades ingående i kapitlets första del. Vi använder här den stora databasen med svar från SOM-undersökningarna, databasen med svar från webbpanelen Medborgarpanelen, vid Göteborgs universitet samt data från en telefonundersökning från SIFO. På detta sätt täcker vi in tre datainsamlingsmetoder. Datasetet från SIFO är dock mycket mindre då det bara kommer från ett enskilt forskningsprojekt, medan databaserna från SOM och Medborgarpanelen är en sammanslagning av många olika projekts data. Strategin är enkel. Vi jämför svarsfördelningen bland dem som svarar direkt på undersökningen med hur svarsfördelningen ser ut bland dem som svarar under en fältperiod på tre veckor.

**Fel! Hittar inte referenskälla.** visar hur stor skillnaderna är mellan fältdag 1 å ena sidan och fältdag 2 till 21 å andra. De tre första raderna summerar hur stor skillnaderna är i termer av hur medelvärdena skiljer sig åt. Det är beräknat som den absoluta skillnaden i medelvärdet som procent av hela skalan. 10 procents medeldifferens skulle till exempel innebära ett skalsteg av en elvagradig skala mellan 0 och 10 ($0{,}1 * (10 - 0)$). Tabellen visar att det i snitt endast handlar om cirka en femtedel av ett skalsteg på en elvagradig skala om man skulle stoppa datainsamlingen inom den första fältdagen jämfört med om den hade hållits öppen tre veckor till.

Det är dock inte bara hur stor skillnaden är som är relevant, utan också om den skillnaden är statistiskt signifikant. I snitt är p-värdet av skillnaden 0,44 där ett värde under 0,05 räknas som signifikant, men variationen är stor som standardavvikelsen visar. Men i 1 fall av 20 kan skillnaderna slumpmässigt vara signifikanta (om signifikansnivån är definierad 0,05). Andelen signifikanta differenser, 10 procent, visar att det antagligen inte är en slump att det ändå finns skillnader. För att inte riskera att dessa resultat speglar en datainsamlingsmiljö som inte längre finns så visar resultaten i tabeller bara perioden sedan 2010 även om data åtminstone för SOM-undersökningarna finns betydligt längre tillbaka i tiden. I stort sett är skillnaderna desamma, men om möjligt så skiljer sig måtten åt något mer på senare år.

---

själva antas vara fria från snedvridning på grund av bortfall. Detta är ett starkt antagande att göra, men kommer antagligen alltid att i viss mån vara nödvändigt för undersökningar av attityder och beteenden.

**Tabell 2.1. Medelvärden av differenserna mellan fältdag 1 och fältdag 2-21.**

| | *SOM (papper)* | | *Medborgar-panelen (webb)* | | *SIFO (telefon)* | |
|---|---|---|---|---|---|---|
| | *Medel* | *Std.av.* | *Medel* | *Std.av.* | *Medel* | *Std.av.* |
| Absolut relativ medeldifferens | 2,1% | (2,0%) | 1,5% | (1,3%) | 1,6% | (1,7%) |
| p-värde, test av medelvärdesdifferens | 0,445 | (0,303) | 0,287 | (0,314) | 0,541 | (0,308) |
| Andel signifikanta medeldifferenser | 10% | (30%) | 33% | (47%) | 5% | (21%) |
| Absolut proportionsdifferens | 3,0% | (2,7%) | 2,1% | (1,8%) | 3,0% | (2,4%) |
| p-värde, prtest av proportionsdifferens | 0,453 | (0,301) | 0,322 | (0,319) | 0,449 | (0,299) |
| Andel signifikanta proportionsdifferenser | 10% | (30%) | 32% | (47%) | 7% | (26%) |
| Observationer | 2 248 | | 699 | | 37 | |

Kommentar: Riks-SOM-undersökningarna och MP gäller data insamlad under perioden 2010-2015. SIFO-mätningen gäller bara 2007. I praktiken räknas fältdag 1 som den eller de fältdagar som sammanlagt har samlat in 4 procent av de som skulle komma att svara. Anledningen är att postgången skiljer sig åt mellan år och områden. Som oftast är det inom tre dagar när det gäller SOM och en när det kommer till de andra undersökningarna.

För att ge en annan bild så går det även att visa hur mycket andelen som väljer vissa kategorier skiljer sig åt mellan de olika insamlingsperioderna, dag 1 och dag 2-21. Proportionen beräknas här som andelen som väljer alla kategorier under mittenkategorin (om en sådan finns, annars räknas första halvan av skalan). Det kan till exempel vara den andel som har ett mycket eller ganska stort förtroende för regeringen 2015, vilket var 32 procent. Proportionen kan delas upp i de som svarade den första fältdagen, 29 procent, mot de som svarade fältdag 2 till 21 vilket var 32 procent. Den absoluta proportionsdifferensen är i det här fallet 3 procent (32 % - 29 %) och innebära en viss underskattning av förtroendet om vi bara tittar på dag 1, men skillnaden är inte statistiskt signifikant ($p$ = 0,30). I snitt är proportionsdifferensen också 3 procent. I princip visar inte proportionsdifferensen någon annan bild än vad medelvärdesdifferensen visar.

Överlag visar alltså våra resultat att det kan finnas vissa systematiska skillnader mellan de som svarar omedelbart på en undersökning och de som svarar under resten av en fältperiod på tre veckor. Men överlag är skillnaderna substantiellt små. Det tycks alltså återigen inte föreligga några stora risker med korta fältperioder. Med hjälp av efterviktning av materialet kan eventuellt dessa skillnader minskas än mer. Vi vill betona att dessa resultat bygger på ett mycket stort material. För SOM-undersökningarna har över två tusen variabler insamlade under sex år undersökts, och för Medborgarpanelen sju hundra variabler under samma tidsperiod.

## 2.4 Empirisk delstudie 2

Opinionslaboratoriet LORE har tidigare genomfört studier om fältperioder för e-postutskick kan kortas genom att skicka tätare påminnelser. Rapporten undersöker effekter av att korta tidsintervallet mellan att undersökningspåminnelser skickas ut till respondenter, och av olika längd på fältperioden, på undersökningens svarsfrekvens. Studien bygger på ett storskaligt experiment med 9 000 respondenter, slumpmässigt indelade i tre olika experimentgrupper där varje grupp får upp till två påminnelser om de inte besvarar enkäten, men med olika tidsintervall.

Listan över faktorer som påverkar svarsfrekvensen i survey-undersökningar är lång, men inte alla av dessa delar är lika lätta att påverka. Vi fokuserar här på en smal del av de tänkbara faktorerna, nämligen fältarbetsrelaterade faktorer som gäller tidpunkter för inbjudan och påminnelser som exempelvis inkluderar 1) antal påminnelser som skickas till respondenter under fältperioden, 2) densiteten på påminnelserna (tätheten, tiden mellan olika påminnelser skickas ut), 3) längden på fältperioden, 4) vilken veckodag och 5) vilken tid på dagen inbjudan och påminnelser till en undersökning skickas till respondenterna. I tidigare metodrapporter har LORE visat att antalet påminnelser (LORE methodological note 2014:3 and 2014:5) påverkar svarsfrekvensen som förväntat, men att veckodag (LORE methodological note 2014:17) och tidpunkt på dagen (LORE methodological note 2014:20) verkar ha mycket begränsade (om någon) effekter på svarsfrekvensen i undersökningar, och att dessa endast påverkar på väldigt kort sikt. För att korta fältperioder kan tätare påminnelser vara nödvändigt, och detta avsnitt fokuserar på hur en sådan ökad täthet i påminnelserna påverkar svarsfrekvenserna.

### 2.4.1 Design, data och resultat

Datainsamlingen genomfördes med deltagare i Medborgarpanelen där 9 000 panelmedlemmar blev slumpmässigt tilldelad en av tre experimentgrupper. Alla respondenter fick två påminnelsemail var till undersökningen, men dessa fördelades så jämnt som möjligt under de hypotetiska datainsamlingsperioderna (2, 3 respektive 4 veckor långa). De kallas här ”hypotetiska” eftersom undersökningen inte tekniskt stängde förrän sista dagen på den längsta perioden (4 veckor). Denna design gjorde det möjligt att testa om det är längden på fältperioden som påverkar svarsfrekvensen, eller om det är tätheten hos påminnelserna. Som mått på svarsfrekvens används de som svarat på hela, eller delar av enkäten.

Resultaten visar att när respondenterna skickas två påminnelser med olika tidsintervall spelar längden på perioden för datainsamling en positiv, men liten roll. När perioden sträcker sig över 4 veckor jämfört med 2 veckor ökar svarsfrekvensen enbart med 2.5 procent, och effekterna går inte att fastställa med statistisk säkerhet. När effekterna testades på olika subgrupper av individer, kunde inga skillnader ses mellan individer med olika kön, ålder och utbildningsnivå. Resultaten implicerar att de relativa vinsterna med att ha en längre fältperiod är, om än positiva, mycket marginella. Sammanfattningsvis stärker resultaten från den här studien resultat från tidigare studier som indikerar att det är antalet påminnelser, snarare än längden på fältperioden eller intervallet mellan påminnelseutskicken, som påverkar svarsfrekvens i den här typen av panelbaserade

undersökningar. Studien kan dock tyvärr inte svara på hur stor skillnaden blir om perioden kortas ytterligare till en vecka eller 2-3 dagar.

## 2.5 Empirisk delstudie 3

Ett tänkbart sätt att snabbt nå ut med undersökningar som ska genomföras på kort tid är via SMS. Undersökningar via SMS är dock ovanligt i Sverige. Såvitt vi känner till arbetar ingen av de etablerade opinionsinstituten med en sådan metod i betydelsen att de skickar SMS med en inbjudan till en undersökning direkt till människor utan att först kontakta dem på något annat sätt. Anledningarna är säkert flera. Men två uppenbara anledningar är att det betraktas med stor misstänksamhet och inte ses som seriöst av många deltagare. Det är särskilt lätt att mottagarna tror att det är en del i en bluff av något slag om sms:et inkluderar en internetlänk. För det andra får vi ganska stort bortfall då vi inte kan hitta mobilnummer till alla i ett befolkningsurval, precis som för vanliga telefonundersökningar. Men nu tillkommer det faktum att vi även måste hitta ett mobilnummer. Vill vi dessutom använda tekniken med att skicka med en länk till en webbenkät måste mottagarna även ha en smartphone. Summerar vi dessa bortfall kan det bli ganska stort. Å andra sidan har metoden tiden på sin sida och kan troligen bli allt mer användbar med tiden.

Opinionslaboratoriet LORE genomförde därför för denna rapport ett första storskaligt test av en snabbmätning via sms-inbjudan till en webbenkät som rapporteras här. Syftet med studien var flerfaldigt. För det första testas vilken svarsfrekvens som är möjlig att uppnå i en webbundersökning med sms-inbjudan och mycket kort fältperiod. För det andra undersöker vi huruvida det går att påverka personer att svara snabbare genom att påpeka i sms-inbjudan till enkäten att svaren behövs snabbt. För det tredje testas hur svarsfrekvensen påverkas av hur nära inpå första inbjudan en påminnelse skickas ut.

Ett slumpmässigt urval drogs från Skatteverkets folkbokföringsregister. Detta nummersattes sedan av ett kontrakterat företag och vi tog från detta material fram ett bruttourval om 8000 personer som hade minst ett mobiltelefonnummer. Urvalet var representativt utifrån kön, ålder (18-80 år), födelseland (Sverige eller annat land), antal barn och län. Med en faktoriell 2x2 design delades urvalet in i fyra randomiserade grupper, 2000 personer i varje grupp, se Tabell 2.2.

Personerna i urvalet fick en första sms-inbjudan att delta i studien klockan 10.00 den 30 januari 2017 (de olika varianterna av inbjudningarna finns bifogade i appendix). I inbjudan informerades personerna om att de blivit slumpmässigt utvalda att delta i en webbundersökning som genomfördes av Göteborgs universitet i samarbete med Myndigheten för samhällsskydd och beredskap (MSB). SMS-inbjudan innehöll en personlig länk till webbenkäten, och inga ekonomiska incitament för att delta användes. Undersökningen stängdes vid midnatt sex dagar efter första utskick, den 5 februari 2017.

Två av grupperna (1 och 3) fick i sms-inbjudan en uppmaning om att svara så snart som möjligt (”då undersökningen berör läget just nu”). Utöver uppmaningen om att svara snabbt för grupp 1 och 3, fick två av grupperna (1 och 2) en påminnelse om att delta redan samma dag som de fick inbjudan, klockan 16.00, det vill säga sex timmar efter första inbjudan. De andra två grupperna (3 och 4) fick påminnelse om att delta två dagar senare, vid samma tidpunkt, klockan 16.00.

I Tabell 2.2 visas bruttourvalsstorleken för de olika delurvalen (2000 per delgrupp), samt den faktiska nettourvalsstorleken. 3.2% av mobilnumren i bruttourvalet studsade, och levererades därmed inte. Dessa räknas därför bort från nettourvalsstorleken.

**Tabell 2.2. Bruttourval, nettourval samt antal studsar.**

| | Info i inbjudningstext | Påminnelse | Bruttourval | Nettourval | Antal studsar |
|---|---|---|---|---|---|
| Delurval 1 | Snabba svar | Dag 1 | 2000 | 1930 | 70 |
| Delurval 2 | Ej snabba svar | Dag 1 | 2000 | 1927 | 73 |
| Delurval 3 | Snabba svar | Dag 3 | 2000 | 1944 | 56 |
| Delurval 4 | Ej snabba svar | Dag 3 | 2000 | 1942 | 58 |
| Hela urvalet | | | 8000 | 7743 | 257 |

I tabell 2.3 jämförs den demografiska fördelningen i bruttourvalet (i procent) med de svarande i studien. Totalt sett svarade, lite beroende på vilka som räknas in som svarande, 850-950 personer på undersökningen. Vi jämför i tabell 2.3 sammansättningen med avseende på kön, ålder fördelat på tre kategorier (18-29 år, 30-54 år och 55-80 år), samt födelseland. Skillnaderna är små, mellan 1-2 procentenheter, förutom när det gäller respondenter i ålderskategorin 30-54 år som är överrepresenterade bland de svarande med 4.6 procentenheter.

**Tabell 2.3. Demografisk fördelning i hela urvalet respektive bland de svarande.**

| | Hela urvalet (%) | Svarande (%) | Differens mellan hela urvalet och svarande, N-n (procentenheter) |
|---|---|---|---|
| Kön | | | |
| Kvinna | 50.7 | 52.7 | 2.0 |
| Man | 49.3 | 47.3 | -2.0 |
| Ålder | | | |
| 18/29 | 18.5 | 16.4 | -2.1 |
| 30/54 | 43.3 | 47.9 | 4.6 |
| 55/80 | 38.2 | 35.7 | -2.5 |
| Födelseland | | | |
| Sverige | 85.3 | 86.5 | 1.2 |
| Utanför Sverige | 14.7 | 13.5 | -1.2 |

Resultaten i tabell 2.3 är mycket uppmuntrande. De svarande avviker mycket lite från hela befolkningens fördelning. Särskilt är andelen unga ovanligt hög. Denna metod verkar fungera relativt bra för den yngre gruppen. De är fortfarande underrepresenterade med några procentenheter, men i många andra studier brukar denna skillnad vara långt större. I många undersökningar kan vi exempelvis se att svarsfrekvensen kan vara dubbelt så hög bland de äldre som bland de yngsta. Naturligtvis måste vi fortfarande vara mycket försiktiga med

slutsatser kring dessa resultat då det kan finnas stora skevheter bland de som svarar när det gäller innehållet i svaren bortom demografin.

### 2.5.1 Resultat

I tabell 2.4 anges andelen som klickade på länken i sms-inbjudan och startade enkäten (startfrekvens), 13.6%, samt nettosvarsfrekvensen för hela urvalet, 11.2%.

**Tabell 2.4. Startfrekvens och svarsfrekvens (netto) (%)**

| | Startfrekvens (%) | Svarsfrekvens (%) |
|---|---|---|
| 1: Påminnelse dag 1, snabba svar nämnt | 14.2 | 11.6 |
| 2: Påminnelse dag 1, snabba svar ej nämnt | 11.6 | 9.4 |
| 3: Påminnelse dag 3, snabba svar nämnt | 15.7 | 13.2 |
| 4: åminnelse dag 3, snabba svar ej nämnt | 12.8 | 10.7 |
| Hela urvalet | 13.6 | 11.2 |

Skillnaderna i svarsfrekvens som kan ses i tabell 2.4 är inte så stora, men systematiska. Att vänta med påminnelse till dag 3 ger något högre svarsfrekvens när vi stänger efter sex dagar än att påminna redan första dagen (jämför grupp 1 och 3, samt grupp 2 och 4). Att be deltagarna att svara snabbt för att undersökningen handlar om något aktuellt ger också något högre svarsfrekvens i slutändan (jämför grupp 1 och 2, samt grupp 3 och 4). Överlag verkar att be respondenterna svara snabbt ge omkring 2,5 procentenheters högre svarsfrekvens än när en sådan uppmaning inte ingår, och en senarelagd påminnelse ger omkring 1,5 procentenheters högre svarsfrekvens än en mycket snabb påminnelse (samma dag som inbjudan). Skillnaderna är dock som sagt små. Den enda gruppskillnad som blir statistisk signifikant (p=.001) är den mellan grupp 3 och 2. Det vill säga, den grupp som gör båda de positiva åtgärderna (längre tid till påminnelsen och ber som snabba svar) och den grupp som inte gör någotdera (påminner snabbt och inte ber om snabba svar). Skillnaden mellan grupp 4 och 3 är dock signifikant på 90 %-nivån (p=.055), vilket indikerar en självständigt signifikant effekt av att be om snabba svar.

**Tabell 2.5 Svarsfrekvens (netto) fördelat på kön, ålder (tre kategorier) och födelseland (%).**

| | Delurval 1 | Delurval 2 | Delurval 3 | Delurval 4 | Hela urvalet |
|---|---|---|---|---|---|
| Kön | | | | | |
| Kvinna | 10.9 | 10.5 | 13.1 | 11.9 | 11.6 |
| Man | 12.2 | 8.3 | 13.4 | 9.1 | 10.8 |
| Ålder | | | | | |
| 18/29 | 8.5 | 9.1 | 12.5 | 9.3 | 9.9 |
| 30/54 | 13.6 | 10.1 | 14.8 | 11.1 | 12.5 |
| 55/80 | 10.7 | 8.7 | 11.6 | 10.8 | 10.4 |
| Födelseland | | | | | |
| Sverige | 11.6 | 9.2 | 13.3 | 11.2 | 11.3 |
| Utanför Sverige | 11.3 | 10.6 | 12.6 | 7.5 | 10.5 |

I tabell 2.5 jämförs svarsfrekvenser i olika grupper. Som tidigare har nämnt är svarsfrekvenserna efter kön, ålder och födelseland ovanligt jämnt fördelade jämfört med andra metoder. Det är dock tydligt att metoden är mindre populär bland äldre då de annars brukar ha mycket högre svarsfrekvenser än yngre. Sannolikheten att få svar från någon över 55 är dock med denna metod lika stor som för någon under 30.

**Tabell 2.6 Utbildning och politiskt intresse bland de svarande i de olika delurvalen (%)**

| | Delurval 1 | Delurval 2 | Delurval 3 | Delurval 4 | Hela urvalet |
|---|---|---|---|---|---|
| Utbildning | | | | | |
| Låg/medel utbildning | 56.4 | 55.6 | 55.5 | 61.8 | 57.2 |
| Hög utbildning | 43.6 | 44.4 | 44.5 | 38.2 | 42.8 |
| n | 248 | 201 | 280 | 228 | 957 |
| | | | | | |
| Politiskt intresse | | | | | |
| Inte alls intresserad | 6.9 | 2.0 | 5.7 | 3.0 | 4.6 |
| Inte särskilt intresserad | 20.6 | 19.3 | 16.6 | 22.4 | 19.6 |
| Ganska intresserad | 49.2 | 52.5 | 58.0 | 55.6 | 54.0 |
| Mycket intresserad | 23.4 | 26.2 | 19.8 | 19.0 | 21.9 |
| n | 248 | 202 | 283 | 232 | 965 |

*Kommentar*: Hög utbildning motsvarar tre års högskole-/universitetsutbildning eller mer, låg/medel utbildning allt därunder.

Andelen högutbildade och politiskt intresserade, två grupper som ofta är överrepresenterade i surveyundersökningar, skiljer sig inte särskilt åt mellan de fyra grupperna. Vi ser dock att båda är klart överrepresenterade med denna metod (2.6). Det är som sagt vanligt med andra undersökningsmetoder också, men det tycks här vara mer markant än i exempelvis konventionella telefonintervjuer eller postenkäter. Om vi jämför med de årliga SOM-undersökningarna ser vi att SMS-undersökningens svarsgrupp innehåller fler högutbildade och fler politiskt intresserade än i SOM-institutets postenkäter.

Den generella slutsatsen från vårt första storskaliga försök med SMS-inbjudan till slumpmässiga urval där de får en länk direkt till en webbenkät är att metoden faktiskt tycks som ett lovande alternativ inför framtiden, men att mer utvecklingsarbete återstår att göra. Det är troligen ett utmärkt sätt att snabbt få in svar från representativa urval av den yngre delen av befolkningen. Bland äldre måste metoden om god representativitet är ett prioriterat mål troligen kompletteras med andra metoder. Vi fick även känna på den stora misstänksamhet som omgärdar metoden bland allmänheten. Många känner till problemen med olika bluff- och kapningsförsök via internetlänkar i SMS och e-post. Den exakta utformningen är därför av stor vikt. Efter dessa erfarenheter tror vi dock inte att de är oöverstigliga. Trovärdigheten i meddelandet kan höjas genom kontaktuppgifter utanför internetlänken och andra åtgärder som underlättar för deltagarna att bekräfta att det är en seriös undersökning. Svarsfrekvensen i sig, som högst i försöken som redovisats här omkring 13 procent, kan tyckas låg. Men faktum är att efter bara två kontaktförsök och en mycket kort fältperiod kan detta inte sägas vara dåligt då många traditionella telefonundersökningar idag ligger mellan 15-25 procent i svarsfrekvens.

## 2.6 Sammanfattning

Sammantaget ser vi att både den begränsade tidigare forskning som finns på området och våra egna empiriska studier indikerar att det troligen är möjligt att göra datainsamlingar under mycket kortare fältperioder än vanligt utan att kvalitén drabbas alltför mycket. Någon fullständig säkerhet att så inte är fallet i en given undersökning går dock inte att ge. Ibland ser vi att resultaten systematiskt skiljer sig åt, men dessa skillnader är med få undantag små. Det förutsätter dock troligen att vi har erforderlig kompetens och att frågan om hur materialet bör viktas behandlas genomtänkt. Slår vi samman indikationerna som vår genomgång av tidigare forskning och våra egna empiriska studier som rapporteras här ger så har vi resultat både från besöksintervjuer, telefonundersökningar, postenkäter, e-postutskick från webbpaneler, och SMS-utskick till slumpmässiga urval. I stort sett pekar alla resultaten i samma riktning. Det är möjligt att göra snabba opinionsmätningar med korta fältperioder.

# 3. Att mäta snabba opinionsförändringar

Bengt Johansson och Johan Martinsson

## 3.1 Inledning

Opinioner förändras. Ibland handlar det om långsamma värderingsförändringar i samhället som är svåra att hänföra till en enskild händelse. Det kan vara förändringar av det politiska klimatet som beror på ett komplext samspel mellan folkopinion, politisk opinionsbildning och medierapportering. Ibland sker opinionsförändringarna dramatiskt. Ett terrorattentat, ett krig som bryter ut eller någon annan dramatisk händelse förändrar opinionen snabbt och helt plötsligt är opinionsläget helt annorlunda.

Det händer att sådana förändringar blir bestående, att opinionen lägger sig på en helt ny nivå jämfört med tidigare. Ett exempel på detta är oron för situationen i Ryssland som på årsbasis låg still sedan början av det nya millenniet, men som först ökade dramatiskt i samband med Rysslands agerande i Krim och Ukraina för att sedan stabiliseras på en väsentligt högre nivå (Svenska Trender 2015). I andra fall blir det bara ett hack i kurvan och sedan är opinionen tillbaka till ett läge där den var innan händelsen ägde rum. Sedan många år ser vi exempelvis sådana hack i opinionskurvor under valår. Då ökar både politiskt intresse och förtroendet för politiska institutioner och partier för att sedan sjunka igen då valåret är över (Strömbäck & Johansson, 2007). Ett annat exempel är kärnkraftsolyckorna i Tjernobyl och Harrisburg. Dessa resulterade i en kraftig opinionspåverkan, som dock visade sig vara snabbt övergående. Tjernobylkatastrofen orsakade en opinionsförändring i negativ riktning för stödet för kärnkraften med hela mellan 15 och 20 procentenheter. Harrisburgeffekten visade sig vara försvunnen redan efter ett halvår, och Tjernobyleffekten efter max ett och ett halvt år (Holmberg, 2012). Att samma frågor kan ha olika opinionsutveckling visar kärnkraftsolyckan i Fukushima. Efter olyckan 2011 blev den svenska opinionen mer negativ till kärnkraft och denna gång har det inte skett en tillbakagång. Flera år efteråt ligger kärnkraftsmotståndet kvar på ungefär samma nivå (Holmberg, 2016).

Att mäta opinioner och dess förändringar kan också behöva göras i mer nära relation till en händelse (exempelvis en opinionspåverkan, terroristattentat eller annan samhällskris). Mätningar som görs en gång om året fångar de långsiktiga trenderna, men för att få kunskap om hur enskilda händelser påverkar opinionen och om eventuella informationsinsatser som görs har någon effekt måste tätare mätningar genomföras. Här är snabbhet och förberedelse viktigt. Utan planering är det svårt att kunna göra mätningar överhuvudtaget och vill man uttala sig om opinionens förändringar är detta centralt. I samband med kriser har exempelvis tidigare forskningsprojekt vid Göteborgs universitet samarbetat med opinionslaboratoriet vid Göteborgs universitet (LORE) för att ta fram en modell för hur krisreaktioner hos allmänheten kan undersökas. Erfarenheterna var att förberedelse och standardisering var centrala villkor för att det överhuvudtaget skulle

vara möjligt att snabbt genomföra undersökningarna i direkt anslutning till krisförloppet (Odén m fl., 2016).

I detta kapitel ska vi diskutera vilka strategier och val som måste göras för att kunna undersöka opinionsförändringar i direkt anslutning till en händelse. Det handlar med andra ord om att göra det möjligt att spåra snabba opinionsförändringar och besvara frågor som exempelvis: Påverkade det som hänt opinionen? Var förändringen bestående eller opinionen tillbaka till nivåerna före händelsen ganska snart? Hade insatser från myndigheterna någon påverkan på opinionen? I kapitlet presenteras en grundmodell över opinionsförändring i samband med en händelse och tre olika typer av design som kan användas för att undersöka hur opinionen förändras. I samband med dessa tre designer diskuteras ett antal olika aspekter och val som den som vill undersöka opinionsförändringar ställs inför. Det handlar om övergripande strategi ifråga om att arbeta med tvärsnittsdata eller paneldata, om urvalet ska göras från befolkningsregister eller från respondentpaneler, om det är via telefonintervjuer, internet eller kanske via postenkäter data om opinionsläget ska samlas in, om det är opinion på nationell nivå som är intressant eller om olika grupper av medborgare eller boende i vissa geografiska områden står i centrum och sist men inte minst vilka typer av frågor som ska ställas till de svarande.

## 3.2 Opinionsnivåer eller förändringar

När opinioner studeras är det oftast fråga om att antingen studera hur vanliga vissa uppfattningar är i sig, så kallade nivåskattningar (exempelvis ”en majoritet stödjer ett fortsatt svenskt EU-medlemskap”, eller ”X procent tycker att den offentliga sektorn bör minska”) eller att mäta förändringar i opinionen (exempelvis ”andelen som stödjer kameraövervakning på offentliga platser har ökat/minskat med X procentenheter”). Generellt anses det svårare att mäta och precist bestämma nivåer på opinioner än att studera förändringar. Skälen är flera, bland annat är andelen som svarar på ett visst sätt starkt kopplat till exakt hur frågan har ställts och vilka svarsalternativ som gavs. Det finns många exempel där forskning visat att opinionsnivåer kraftigt förändras bara genom att frågors ordval eller svarsalternativ modifieras lätt (se till exempel Schuman & Presser, 1981, samt kapitel 1 i denna rapport). Ett annat skäl är att om vi har ett urval som inte är perfekt representativt för befolkningen, exempelvis på grund av skevt svarsbortfall så kan detta påverka nivån på en viss opinion, medan vi under sådana omständigheter har större chans att ändå fånga in förändringen från en tidpunkt till en annan på ett korrekt sätt.

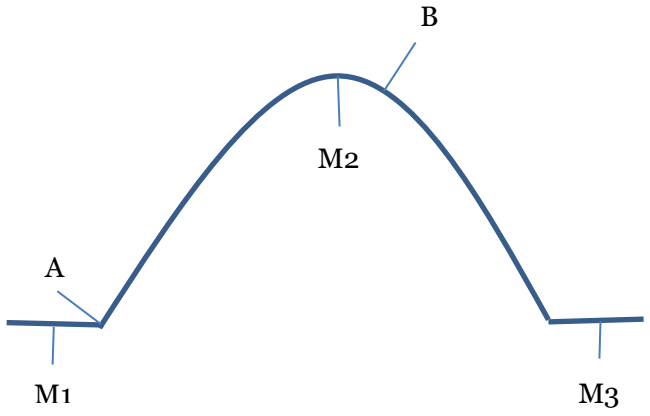
I fokus för kapitlet är inte i första hand nivåskattningar av åsikter och attityder, utan snarare hur man kan studera om det har skett förändringar i ett opinionsläge eller inte. De undersökningsdesigner som presenteras är framförallt framtagna med studier av opinionsförändringar i åtanke, snarare än för att ta fram precisa nivåskattningar. Studier av förändringar av opinioner är mer robusta och inte lika beroende av de exakta frågeformuleringarna eller vilka svarsalternativ som erbjuds. Trots detta konstaterande finns det ändå skäl att diskutera tillförlitligheten även i resultat angående opinionsförändringar.

## 3.3 Grundmodell

När vi diskuterar opinionsförändringar utgår vi från ”normalläge” i opinionen, som sedan förändras på grund av någon sorts yttre påverkan eller händelse (A) och hur opinionen utvecklas efter händelsen. I

uppdraget från MSB ingår att kunna studera vilken effekt informationsinsatser som vidtagits (B) för att motverka den eventuella opinionsförändring som skett efter händelsen. Detta villkor är inte nödvändigt för de olika typer av undersökningsdesign som vi tagit fram, men kan inkluderas om så önskas. Man kan naturligtvis studera hur opinionen förändras i samband med en händelse oavsett om man gör en informationsinsats med anledning av händelsen eller inte.

**Figur 3.1. Grundmodell för att studera opinionsförändringar**




*Kommentar*: M1=mätning före en händelse, M2=mätning under en händelse, M3=mätning efter en händelse. A=händelse, B=åtgärd för att förändra opinionen.

Mätpunkter finns innan händelsen äger rum (M1), i direkt anslutning eller under det att händelsen äger rum (M2) och en tredje mätning ytterligare en tid senare efter att händelsen ägt rum (M3). Vi diskuterar framförallt tre olika avgränsningar av målpopulationen som studeras. Den första varianten är den generella opinionen på nationell nivå, dvs. hela befolkningen (vanligen inom något rimligt åldersintervall). Den andra är att undersöka opinionsförändringar i olika socio-demografiska grupper av befolkningen, och den tredje riktar in sig på hur opinionen skiftar över tid hos befolkningen i geografiskt avgränsade områden. I princip kan naturligtvis även en kombination av en socio-demografisk och en geografisk avgränsning tänkas.

## 3.4 Befolkningsurval eller respondentpaneler

När man väljer urvalsmetod för en surveyundersökning är en av de grundläggande frågorna att ta ställning till om urvalet ska göras från befolkningsregistret eller om urvalet ska dras från någon form av förrekryterad panel av respondenter (se även kapitel 1). Traditionellt har slumpmässiga urval från befolkningsregister varit det självklara valet för surveyundersökningar (åtminstone i länder där sådana finns, andra länder använder andra metoder som exempelvis olika former av adressbaserade urval) eftersom man då under rätt förutsättningar kan uttala sig med en större säkerhet (och i strikt mening, med en känd osäkerhet) om i vilken utsträckning resultaten från en undersökning

speglar verkligheten i den population man vill uttala sig om, och dessutom anses få bättre träffsäkerhet. Idag är det dock i praktiken även vanligt med urval från olika former av respondentpaneler, vanligen så kallade webbpaneler, och mycket tyder på att kvalitetsskillnaden mellan dessa urvalsstrategier minskar (se t ex Sturgis m fl., 2016). Dels beror detta på att det blir allt svårare att få människor att delta i undersökningar som bygger på urval från befolkningsregister (sjunkande svarsfrekvenser), vilket både har att göra med att det är svårare att få kontakt med dem som valts ut och att färre vill delta i surveyundersökningar. Dels har det att göra med att kvaliteten på en del respondentpaneler blir allt bättre. En del respondentpaneler har idag många tio-tusentals deltagare eller mer och rekryterar strategiskt för att panelen ska bli en miniatyr av hela befolkningen. Genom att dessutom göra stratifierade slumpmässiga urval (eller kvoturval) från respondentpanelens medlemmar och sedan olika former av viktning av resultaten (exempelvis poststratifiering) kan kvaliteten på dessa ökas. Även slumpmässiga befolkningsurval är idag i praktiken med få undantag beroende av viktning för tillförlitliga resultat. Det är dock inte alla webbpaneler som kan anses ha en kvalitet som närmar sig de slumpmässiga befolkningsurvalen, utan träffsäkerheten och representativiteten i dessa varierar stort mellan olika företag och paneler (se exempelvis Yeager m fl., 2011). Sammanfattningsvis brukar webbpaneler (eller andra former av förrekryterade respondentpaneler) ännu inte anses lika tillförlitliga som högkvalitativa undersökningar från befolkningsurval (se exempelvis Brewitz 2015). Man ska också vara medveten om att båda urvalsmetoderna kan missa att fånga upp grupper i samhället — speciellt socialt utsatta grupper och yngre — vilket kan medföra brister ifråga om representativitet.

Vid en närmare titt på de två grundmetoderna är det dock så att det finns flera nyansskillnader mellan befolkningsurval och respondentpaneler. En sak att uppmärksamma är att respondentpaneler rekryteras på olika sätt. En del bjuder in paneldeltagare genom slumpmässigt befolkningsurval, medan andra baserar sig på självrekrytering där alla som vill får anmäla sig (dessa kallas ibland även ”öppna paneler”). Frågan är då om detta spelar någon roll för kvaliteten. Generellt sett bör slumpmässigt rekryterade paneler vara bättre då intresse av att delta i framtida studier hos deltagarna rimligen spelar en mindre roll för huruvida en viss person deltar i en studie än för självrekryterade paneler. Samtidigt är det alltid en generell svarsvilja hos individerna som avgör om de accepterar att delta även om de har valts ut slumpmässigt.

Det diskuteras ibland även om inte självrekryterade paneler löper en risk att ”kapas” av organiserade intressegrupper. I praktiken är detta troligen ett mindre problem eftersom paneler ofta använder olika sätt att motverka att panelerna ska användas för att driva kampanjer av olika slag. Förutom att det inte är så enkelt som att den som anmäler sig till en självrekryterad webbpanel automatiskt väljs ut att delta i alla undersökningar så används också av många företag olika former av ”karantänperioder” som innebär att respondenter inte får svara på frågor om exempelvis politik för ofta, utan bara med ett bestämt tidsintervall (till exempel var tionde gång, varannan månad eller liknande). På detta sätt försvåras eventuella försök att driva opinion genom att organiserat anmäla sig för att delta i paneler eftersom arbetsinsatsen höjs markant när de då måste besvara andra typer av undersökningar (exempelvis marknadsundersökningar) många gånger mellan varje undersökning om politik. Men det är viktigt att påpeka att medvetenheten detta problem och rutinerna för att motverka det skiljer

sig åt mellan olika paneler och företag. Sammantaget kan det sägas att vetenskapliga studier oftast menar att slumpmässigt rekryterade paneler ger något högre tillförlitlighet än självrekryterade paneler (se exempelvis Yeager m fl., 2011), men skillnaden kan ofta vara ganska liten.

Inte heller slumpmässiga befolkningsurval är en enhetlig kategori, utan även där kan det finnas skillnader i tillvägagångssätt. I praktiken har idag också många opinionsinstitut och fältföretag som gör telefonintervjuer börjat frångå renläriga befolkningsurval. Traditionellt görs urvalet med en postenkät genom att beställa ett slumpmässigt urval från skatteverkets folkbokföring (närmare bestämt från statens personadressregister), en tjänst som är tillgänglig för både företag och myndigheter. Urval från folkbokföringen innehåller dock inga uppgifter om telefonnummer, utan ska en telefonundersökning göras måste en så kallad nummersättning/nummersökning göras på urvalet. Idag har många börjat istället göra urval direkt från olika kommersiellt tillgängliga databaser med adress och telefonuppgifter, vi kommer här kalla dessa för konsumentdatabaser (ett exempel är PAR konsument). I strikt mening är det tveksamt om sådana urval kan sägas vara slumpmässiga befolkningsurval eftersom urvalen inte görs från ett äkta befolkningsregister. Å andra sidan går det inte att hitta telefonnummer till hela det urval som fås från befolkningsregistret, utan bortfallet kan där vara ganska stort. Det är inte ovanligt att cirka 30 procent av ett äkta befolkningsurval inte kan nummersättas och därför inte ingår i telefonundersökningar. Det är då rimligt att tänka sig att de urval som man får från så kallade konsumentdatabaser ganska väl motsvarar ett nummersatt befolkningsurval. En del som arbetar med urval från sådana konsumentdatabaser använder även någon form av kvoturval baserat på befolkningsdata från Statistiska centralbyrån på de urval som dras från konsumentdatabasen för att säkerställa den demografiska representativiteten ytterligare.

I den privata undersökningsbranschen brukar konsument-databasurvalen anses kunna fungera relativt bra, trots den mer oklara urvalsramen och att de i strikt vetenskaplig mening inte kan sägas vara äkta befolkningsurval. Inom akademin däremot brukar de inte betraktas som lika säkra. Såvitt vi vet har inte några vetenskapligt utformade och publicerade studier som jämför kvalitén mellan de två metoderna gjorts. Däremot har privata företag själva utvärderat frågan innan de har valt metod. Tyvärr publiceras sällan sådana interna studier av opinionsinstitut och marknadsundersökningsföretag. Vi har dock indikationer på att en del i den privata undersökningsbranschen som arbetar med sådana urval anser att de främst fungerar tillfredsställande för svarande från medelåldern och uppåt, medan andra företag använder sådana urval för hela åldersspannet. De praktiska fördelarna är stora med konsumentdatabaserna då det tar cirka 5-10 dagar att få ett urval från statens personadressregister medan urval från kommersiella konsumentdatabaser går att få direkt. Både respondentpaneler och konsumentdatabaser har snabbheten på sin sida då det inte finns någon fördröjning vid själva urvalsförfarandet. Konsumentdatabaserna innehåller också vanligen geografisk information som möjliggör urval för exempelvis specifika kommuner.

## 3.5 Tvärsnittsdata eller paneldata (longitudinell data)

Ett annat grundläggande val när det gäller strategin för att undersöka opinionsförändringar är om de ska studeras med tvärsnittsdata, där olika personer undersöks vid varje undersökningstillfälle eller med

paneldata där samma personer undersöks vid flera olika tillfällen. Det finns fördelar och nackdelar med båda strategierna.

Generellt sett kan man säga att paneldata ger vissa förstärkta möjligheter till säkra analyser av förändringar då vi kan observera förändringar på individnivån direkt utan att behöva nöja oss med att jämföra aggregerad statistik som procent eller medelvärden hos hela urvalet eller undergrupper av urvalet. Paneldata kan även ge fördelar när det gäller analyser av mekanismer som ligger bakom eventuella förändringar då vi kan se vilka uppfattningar eller egenskaper hos individerna vid ett tidigare mättillfälle som korrelerar med förändring av åsikter vid ett senare mättillfälle. Eftersom samma personer undersöks flera gånger är det även lättare att uppnå statistisk signifikans för de opinionsförändringar som hittas eftersom vi i princip slipper det slumpmässiga felet som alltid uppstår när vi jämför två oberoende slumpmässiga urval med varandra

Nackdelen med paneldata är desamma som dess fördel – att samma personer undersöks vid flera tillfällen. Det kan innebära att vi drabbas av bortfall av personer från analysen då alla som deltar i den första undersökningen vanligen inte accepterar att ingå även i de följande undersökningarna, vilket kan ge problem med representativiteten om de som hoppar av undersökningarna systematiskt avviker från de som fortsätter att medverka. Ett annat problem som kan uppstå med paneler är att personers svarsmönster påverkas av att de svarat på samma frågor vid ett tidigare tillfälle. Det är inte helt klarlagt i forskningen hur allvarliga sådana så kallade paneleffekter är, eller när de uppträder. En del tidigare studier indikerar att det finns sådana effekter medan andra finner att de inte utgör ett problem. Vi menar dock att risken får anses större när det är kort tidsrymd mellan undersökningstillfällena.

Tvärsnittsdata löser problemen med bortfall mellan mättillfällen och paneleffekter där tidigare svar kan påverka hur de svarar i efterföljande undersökningar eftersom olika personer ingår vid varje undersökningstillfälle. Nackdelen är dock att tvärsnittsdata från olika urval alltid medför en statistisk osäkerhet i form av en felmarginal då det inte alltid kan säkerställas om relativt små opinionsförändringar återspeglar en verklig förändring hos hela målpopulationen eller om de bara är ett resultat av slumpmässighet i urvalsförfarandet.

## 3.6 Design

I detta kapitel diskuteras tre grundmodeller för att undersöka snabba opinionsförändringar där man antingen kan använda tvärsnitts- eller paneldata.

Tre olika designer:

- 1) **En mätning — efter en händelse:** En mätning görs av opinionen efter det att insatser gjorts för att motverka den (eventuellt) förändrade opinionen.
- 2) **Två mätningar — under och efter en händelse:** mätning av opinionen görs direkt efter att en händelse ägt rum samt ytterligare en efter det att insatser gjorts för att motverka den (eventuellt) förändrade opinionen.
- 3) **Tre mätningar — före- under och efter en händelse:** Innan någon händelse äger rum görs en planerad basmätning med allmänna frågor om det område som de tänkta händelserna och opinionsförändringarna gäller. Därefter mäts opinionen direkt efter en händelse har inträffat samt även efter

det att insatser gjorts för att motverka den (eventuellt) opinionen.

### 3.6.1 **Design 1. En mätning — efter en händelse**

Att endast mäta opinionen efter att händelsen ägt rum (M3, se figur 3.1) — och efter att informationsinsatser med anledning av händelsen har genomförts — kan tyckas märkligt eftersom man med en enda mätpunkt knappast kan uttala sig om förändringar. En sådan mätning kan dock likväl ha ett visst värde, även om designen givetvis har stora begränsningar.

Fördelen är att den kan genomföras även om man inte haft möjlighet att göra någon undersökning i direkt anslutning händelsen och de informationsinsatser som gjorts, till exempel på grund av tidsbrist eller bristande förberedelser. Men vad kan då sägas med hjälp av resultaten? För det första kan själva räckvidden hos en informationsinsats bedömas, hur många och till vilka den har nått ut. En möjlig analys är även att jämföra åsikter, attityder eller beteenden hos dem som säger att de har tagit del av informationsinsatser med dem som inte alls exponerats för dessa insatser. På så sätt kan man försöka uppskatta om informationsinsatserna har haft någon påverkan. Problemet här är dock att de som tagit del av informationen kanske skiljer sig systematiskt från dem som inte fått/exponerats för information. Det finns därmed en risk att de som tagit del av informationsinsatserna kanske har ett större engagemang och intresse, vilket kan göra att de kanske hade en annan åsikt redan innan informationsinsatsen ägde rum. De skillnader som identifieras mellan de som exponerats för informationsinsatser och de som inte har det beror då inte på att man tagit del av informationen utan på att grupperna redan från början hade olika åsikter. Ett sätt att försöka lösa detta är att på olika sätt försöka jämföra grupper av respondenter som i allt annat liknar varandra förutom att de uppger att de exponerats för informationsinsatsen (kontrollera för olika faktorer). Om grupperna är identiska i övrigt, men de som har exponerats för information har andra åsikter i frågan det gäller är det rimligt att anta att opinionsskillnaden beror på informationen de exponerats för. Det är dock svårt att veta om vi verkligen tar hänsyn till alla relevanta egenskaper hos grupperna vi jämför och det finns med andra ord en viss risk att vi överskattar informationseffekten. Det säkraste sätter att säkerställa effekter av ett informationsmaterial i sådana fall är genom experiment där vissa i en kontrollerad randomiserad studie utsätts för informationen och andra inte, varefter vi jämför gruppernas åsikter. Men då tappar vi givetvis möjligheten om informationen hade haft denna effekt utanför experimentsituationen.

Är det en geografiskt avgränsad kris kan man eventuellt också jämföra med situationen i andra liknande områden, där man då försöker likställa så mycket som möjligt ifråga socio-demografiska faktorer om de svarande och där den enda skillnaden är vilken ort de bor på och om de därmed utsatts för påverkan och informationsinsatser.

De urvalsmetoder som är aktuella kan båda vara befolkningsurval och respondentpanelurval. Eftersom man bara mäter en enda gång handlar skillnaden mellan urval nästan uteslutande om att kunna säkerställa representativiteten hos urvalet, vilket generellt sett är bättre med befolkningsurval. Eftersom det inte är lika bråttom (ingen mätning görs i omedelbar anslutning till händelsen) kan man lite beroende på tidshorisonten använda urval som dragits ur befolkningsregister. Beroendet av att mäta med snabba metoder såsom via telefon eller internet är inte heller säkert nödvändigt i dessa lägen. Med andra ord kan även postenkät vara aktuell som insamlingsmetod.

Ser vi på olika målgrupper kan den nationella opinionen undersökas med båda urvalsstrategierna. Socio-ekonomiska grupper kan nås med befolkningsurvalen, men om man med en eftermätning vill nå specifika grupper (förutom grundläggande indelningar såsom till exempel ålder och kön) måste screeningfrågor ställas för att nå den målgrupp man vill undersöka, vilket kan medför betydligt större kostnader. En del respondentpaneler kan däremot erbjuda möjligheter där olika grupper kan identifieras innan utskick görs genom att rikare bakgrundsinformation redan finns i panelen. En screeningfråga fungerar så att den tillfrågade får svara på om de tillhör den efterfrågade gruppen (exempelvis personer med barn i kommunala skolor, personer som arbetar i vårdsektorn eller liknande). Om man svarar ja på frågan får man fortsätta att svara på hela enkäten, vid ett nej på frågan avslutas enkäten. Vid behov att ställa screeningfrågor ökar kostnaden för undersökningen. Urvalet måste ökas för att få med tillräckligt många svarande i de grupper som ska undersökas och då ökar även kostnaderna för kontakter med de svarande då många telefonintervjuer måste göras eller enkäter skickas ut till personer som sedan inte ska ingå i själva undersökningen. För telefonintervjuer kan det även innebära att det tar längre tid då fler samtal måste ringas innan tillräckligt många i målgruppen har hittats.

Om en respondentpanel används har man större möjligheter att undersöka olika socio-demografiska grupper. Är panelen omfattande med många svarspersoner kan man enkelt göra urval från de grupper man vill undersöka, förutsatt att den efterfrågade bakgrundsinformationen om deltagarna redan finns insamlad. Eftersom de svarande besvarat ett omfattande frågebatteri när de rekryterades till panelen finns det oftast goda förutsättningar att identifiera socio-demografiska grupper, men även respondenter med olika intressen, vanor eller åsikter. Vid långsiktiga samarbeten med ett företag/organisation som har en respondentpanel kan rikare bakgrundsinformation möjligen även beställas förutsatt att vi vet ungefär vilken typ av målgrupper som kan bli aktuella i framtida undersökningar.

Ibland kan det vara av intresse att rekrytera specialpaneler med grupper som man vet är svåra att komma åt med traditionella urval, eller som vanligen är för få i vanliga slumpmässiga urval eller befintliga paneler. Det pågår exempelvis försök med att rekrytera specialpaneler med asylsökande eller boende i socialt utsatta områden eftersom dessa grupper generellt sett är starkt underrepresenterade i urval (exempelvis på grund av större bortfall i dessa grupper) från befolkningsregister samt i befintliga respondentpaneler.

Ibland har dock befolkningsurvalet fördelen jämfört med urval från respondentpaneler i att det är lättare att nå relativt små geografiskt avgränsade områden. Ofta blir det för få personer som ingår i en panel på mindre orter. Även i konsumentdatabaser går det att nå mindre geografiska områden med relativt god precision. Möjligheten att undersöka mindre geografiska områden med panelurval avgörs helt av hur stor panelen är. Om man som exempel skulle vilja undersöka en mellanstor svensk kommun med 80 000 invånare och vi utgår från Sverige befolkning över 15 år om knappt åtta miljoner invånare så utgör en sådan kommun en procent av befolkningen. Det innebär att man i en respondentpanel om 50 000 deltagare som är representativ för befolkningen således har cirka 500 personer. Svarsfrekvenserna varierar även stort mellan respondentpaneler från några få procent i vissa till uppemot 70-80 procent i andra. Om inte panelerna är väldigt stora och har en pålitligt hög svarsfrekvens kan alltså möjligheterna till

att undersöka mindre geografiska områden vara begränsade med den metoden.

När det gäller vilka frågor som ska ingå i undersökningen är det här uppenbart av högsta vikt att ställa specifika frågor kring det område undersökningen gäller. Men eftersom vi inte kan studera förändringar är det i eftermätningen även centralt att ställa precisa räckviddsfrågor om informationsinsatser så att man i detalj kan jämföra respondenter som exponerats för informationsinsatserna med dem som inte nåtts av informationen. Dessutom bör man för att öka möjligheterna till att ändå i någon mån göra analyser av informationseffekter i en sådan situation också ställa frågor kring allt man tror kan predicera huruvida någon har tagit del av informationen eller inte. I bristen på möjligheter att studera förändringar finns det också en poäng att försöka i detalj identifiera de uppfattningar/åsikter/beteenden som man vill motverka och dessutom ta ställning till vilka nivåer som kan vara acceptabla. Även om vi inte kan veta exakt hur opinioner förändrats kan vi i alla fall mäta och uppskatta om de ligger på nivåer som inte uppfattas som skadliga eller hotande.

Utöver ovanstående bör man också ställa mer allmänna frågor om området (till exempel synen på Sveriges försvarsberedskap). Förutom att nivåskattningarna i sig kan vara intressanta kan resultaten också jämföras med andra undersökningar för att validera resultaten och åtminstone ge ett underlag för att diskutera om händelsen och insatser haft någon påverkan. Dessutom kan man ställa frågor om andra områden (exempelvis förtroende för olika samhällsinstitutioner, bedömning av Sveriges ekonomi, syn på mellanmänsklig tillit, vilka samhällsproblem man oroar sig för) för att säkerställa representativiteten hos det urval som gjorts. Dessa frågor kan i kombination med socio-demografiska data även vid behov användas för att poststratifiera (vikta) urvalet om undersökningen har stort bortfall eller osäker representativet. På så sätt kan urvalets kvalitet förbättras.

**Tabell 3.1. Sammanfattning design 1. En mätning — efter en händelse**

| | | |
|---|---|---|
| **Urval och metod** | | |
| *Urval* | Befolkningsurval<br>I första hand urval ur befolkningsregister. Finns även möjlighet att göra urval direkt ur så kallade konsumentdatabaser, dock med något mer oklar träffsäkerhet. | Respondentpaneler<br>Urval av paneldeltagare. Paneler kan variera mycket i representativet, men kan vid behov stratifieras för att spegla olika populationer bättre. |
| Metod | Främst telefon- eller postenkät. Även webbenkät är möjlig via brev- eller sms-kontakt. | Oftast webbenkät men kan också göras på andra sätt t ex via telefon eller sms. |
| **Målpopulation** | | |
| *Nationell opinion* | Representativt på nationell nivå som en traditionell opinionsundersökning. | Urval kan göras som speglar populationen (möjlighet till stort n-tal). |
| *Socio-demografiska grupper* | Viss information finns i urvalsram, men kräver screeningfrågor för att identifiera särskilda grupper. | Större möjlighet att undersöka särskilda grupper t ex yrkesgrupper, utsatta grupper, opinionsledare. Antingen genom mer detaljerad bakgrundsinformation om panelmedlemmar eller via specialpaneler. Eventuellt måste sådana grupper preciseras i förväg för att öka bakgrundsinformationen i panelen. |
| *Geografiskt område* | Kan undersökas på kommunnivå (info finns i även konsument-databaser). | Även mindre orter, starkt beroende på panelstorlek/urval. |
| **Enkätinnehåll** | | |
| *Specifika frågor* | Specifika frågor om det som hänt<br><br>Eftersom man inte kan säkert studera effekter är det centralt med räckviddsfrågor om insatser.<br><br>Anstränga sig mer för att identifiera de uppfattning/åsikter man vill motverka. Detta eftersom vi inte kan studera förändring måste nivåskattningar tas i beaktande (och bedöma om dessa är acceptabla). | |
| *Generella frågor* | Allmänna frågor om området som undersöks (som kan jämföras med andra undersökningar).<br><br>Frågor om andra områden med syfte att kunna poststratifiera (vikta) vid stort bortfall eller urval med osäker representativitet. | |

### 3.6.2 Design 2. Två mätningar – under och efter händelse

I denna design görs en mätning i direkt anslutning till att en händelse har ägt rum (M2) och en tid senare mäts om de insatser som gjorts för att motverka opinionspåverkan av händelsen haft någon effekt (M3). Med två mätningar kan vi uttala oss om förändringar och om de insatser som gjorts haft någon påverkan på opinionen. Eftersom det rimligen är viktigt att kunna mäta opinionen direkt efter att en händelse skett och blivit känd är snabbheten en viktig faktor, vilket i sig begränsar valet av undersökningsmetod.

Den övergripande strategin kan då antingen vara att använda tvärsnittsdata eller paneldata. Den förra fungerar så som traditionella opinionsundersökningar när man till exempel jämför fördelningen av partisympatier mellan två månader för att undersöka om någon förändring av den aggregerade opinionen har skett, som exempelvis i de väljarbarometrar som rapporteras i de flesta nyhetsmedier.

När paneldata används är det samma personer som ingår vid båda mättillfällena. Den största fördelen med paneldata är att vi med större säkerhet kan fastslå att opinionsförändringarna sker på individnivå och också med större säkerhet kan kontrollera att det inte är andra faktorer som påverkat opinionsförändringen. Nackdelen är, som nämndes tidigare, att det kan finnas s.k. paneleffekter om man ställer samma frågor till samma personer med för täta intervaller. Svaren kan då påverkas av hur man svarade vid det tidigare tillfället. Om bortfallet mellan panelstegen inte är slumpmässigt fördelat så kan det också påverka resultaten.

Vill man göra återkommande täta mätningar av läget efter en händelse för att i detalj följa opinionsutvecklingen är det knappast att rekommendera att använda paneldata. Skulle samma person svara varje dygn på samma fråga är risken för paneleffekter uppenbara. Att jobba med tvärsnittsdata är därför i så fall att föredra. Här får dock inte paneldata som strategi blandas ihop med urval från respondentpaneler. Även respondentpaneler kan givetvis användas för att göra upprepade tvärsnitt vid flera tidpunkter, förutsatt att panelerna inte är alldeles för begränsade till sin storlek.

Har man ont om tid är i princip de enda alternativen att använda konsumentdatabaser eller respondentpaneler, då urval från befolkningsregistret ofta tar ungefär en vecka. Med rätt förberedelser kan då i princip en undersökning lanseras inom några timmar, exempelvis via e-postutskick eller via telefon. Alternativt kan också respondenter kontaktas med sms där de får en inbjudan till att besvara en enkät via webben. Eftersom snabbheten är en faktor faller postenkät bort med denna design. Det är inte möjligt att använda postenkät som datainsamlingsmetod vid M2 som behöver göras i omedelbar anslutning till en händelse.

När det gäller målpopulationen finns det inga skillnader mot vad som gäller för design 1. De svagheter och fördelar som finns med slumpmässiga urval ur befolkningsregister jämfört med respondentpaneler gäller även här. För urval ur specifika socio-demografiska grupper eller andra grupper är det mindre kostsamt och enklare att använda respondentpaneler. Samma argument som tidigare gäller för geografiska områden, men där är även urval ur så kallade konsumentdatabaser ett alternativ. Vi bör även hållas i åtanke att även om slumpmässiga befolkningsurval tar för lång tid att beställa i detta läge då snabbheten är en faktor så kan det om det är prioriterat eventuellt lösas om vi har planerat i förväg vilken typ av urval som kan komma att behövas och införskaffat det i förväg, eller om vi använder en prenumerationstjänst där man med vissa tidsintervall får ett aktuellt urval levererat regelbundet.

För de frågor som bör ställas är det också mycket gemensamt med design 1. Dock blir det inte riktigt lika centralt att ställa frågor om räckvidd av informationsinsatserna, även om det ändå kan vara av stort intresse att veta vilka som faktiskt exponerats, dels av intresse i sig självt, men dels för att kunna mer precist utvärdera om informationsinsatser har haft avsedd effekt. Den stora vinsten med design 2 jämfört med design 1 är att vi nu kan jämföra opinionsläget vid

två mätpunkter och därmed kan uttala oss om eventuella förändringar mellan tidpunkterna, vilket är den stora vinsten.

**Tabell 3.2. Sammanfattning design 2. Två mätningar — under och efter en händelse**

| | | |
|---|---|---|
| **Urval och metod** | | |
| *Strategi* | Tvärsnittsdata<br>Undersökningar där resultat från två separata urval jämförs (likt traditionella opinionsundersökningar). | Paneldata<br>Undersökningar av samma individer vid flera tillfällen. |
| *Urval* | Befolkningsregister<br>I första hand urval ur befolkningsregister. Finns även möjlighet att göra urval direkt ur så kallade konsumentdatabaser, dock med något mer oklar träffsäkerhet. | Respondentpaneler<br>Urval av paneldeltagare. Paneler kan variera mycket i representativet, men kan vid behov stratifieras för att spegla olika populationer bättre. |
| *Metod* | Främst telefon. Även webbenkät är möjlig via sms-kontakt. Postenkät troligen för långsam. | Oftast webbenkät men kan också göras på andra sätt t ex via telefon eller sms. |
| **Målpopulation** | | |
| *Nationell opinion* | Representativt på nationell nivå som en traditionell opinionsundersökning | Urval kan göras som speglar populationen (möjlighet till stort n-tal). |
| *Socio-demografiska grupper* | Viss information finns i urvalsram, men kräver screeningfrågor för att identifiera särskilda grupper. | Större möjlighet att undersöka särskilda grupper t ex yrkesgrupper, utsatta grupper, opinionsledare. Antingen genom mer detaljerad bakgrundsinformation om panelmedlemmar eller via specialpaneler. Eventuellt måste sådana grupper preciseras i förväg för att öka bakgrundsinformationen i panelen. |
| *Geografiskt område* | Kan undersökas på kommunnivå (info finns i även konsumentdatabaser). | Även mindre orter, starkt beroende på panelstorlek/urval. |
| **Enkätinnehåll** | | |
| *Specifika frågor* | Specifika frågor om det som hänt<br><br>Räckviddsfrågor om insatser blir inte lika centrala som för design 1 eftersom effekter av insatser kan mätas direkt.<br><br>Nivåskattningar av uppfattningar/åsikter blir heller inte lika centrala som i design 1 eftersom förändringar av åsikter kan mätas. | |
| Generella frågor | Allmänna frågor om området som undersöks (kan jämföras med andra undersökningar).<br><br>Frågor om andra områden med syfte att kunna poststratifiera (vikta) vid stort bortfall eller urval med osäker representativitet. | |

### 3.6.3 Design 3. Tre mätningar – före, under och efter händelse

Med mätningar före-, under och efter en händelse finns det goda möjligheter att med relativt stor säkerhet slå fast att en opinionsförändring har ägt rum (jämföra M1 och M2) och om de insatser som gjorts för att motverka denna förändring också haft effekt (jämföra M3 med M2). Nackdelen som design 2 har jämfört med denna är att vi i den förra inte säkert kan veta att opinionen förändrades av händelsen. Då design 2 först mäter opinionen efter det att händelsen ägt rum vet vi ju inte om människors åsikter faktiskt förändrades av det som skett. Även om vi kan mäta förändringar av opinionen efter att händelsen ägt rum och vilka effekter informationsinsatser har, vet vi ju faktiskt inte om det egentligen behövdes. Vi kanske försöker återställa ett opinionsläge som egentligen inte hade ändrats.

Med design 3 slipper vi detta problem. Eftersom vi mäter opinionen innan något har hänt (en basmätning) kan vi med större säkerhet säga om opinionen troligen hade påverkats av händelsen och om det därmed finns behov av att göra några insatser. Sedan kan vi med mätningar under (M2) och efter (M3) händelsen också avgöra om de insatser som satts in haft någon effekt.

Detta är det mest omfattande om ambitiösa upplägget. Den största fördelen är att vi kan följa opinionsutvecklingen med flera mätpunkter. Svårigheten ligger främst i att kunna ställa relevanta frågor om en händelse som ska undersökas innan vi vet exakt vad som kommer hända. Vi återkommer till detta längre fram i detta avsnitt.

Även om tre mätningar ska göras kommer kraven på snabbhet vid M2 vara avgörande för ett antal av de val som görs. Vad gäller strategi fungerar både tvärsnittsdata och paneldata med de olika för- och nackdelar som nämnts tidigare.

Även val av urvalsmetod sker enligt samma principer som för design 2. Eftersom man av jämförbarhetsskäl vill göra urval på samma sätt vid alla tre mätpunkterna blir M2 styrande. Om man t ex har använt befolkningsurval vid basmätningen (M1) kan det bli problematiskt vid M2 då kraven på snabbhet plötsligt kan bli stora. Gör man valet att använda konsumentdatabaser (telefonnummerlistor) eller respondentpaneler för basmätningen är detta i allmänhet dock inget problem. I vissa undantagsfall kan det dock tänkas att det går att utnyttja olika former av rullande prenumerationer på befolkningsurval som kan lösa problemet med den långsamma leveransen av slumpmässiga befolkningsurval. Detta gäller dock främst då det är den nationella opinionen som ska undersökas. Om det gäller små geografiska enheter innan vi vet vilket geografiskt område vi kommer behöva urval från känns inte prenumerationer på befolkningsurval som en realistisk lösning.

Vi vill dock nämna att det i princip kan gå att arbeta med olika typer av urval vid olika mättillfällen även om det innebär en del metodologiska utmaningar och ökad osäkerhet. Genom att man vid M1 kalibrerar olika urvalsmetoder kan olika sätt att mäta jämföras. Om man vill använda konsumentdatabaser eller respondentpaneler vid mätningen i samband med händelsen (M2) kan man vid M1 göra ett metodtest där både urval från befolkningsregister och ett från konsumentdatabas eller respondentpanel ingår och sedan jämföra utfallet. Dessa resultat kan sedan i princip användas för att framöver kalibrera de olika metoderna och göra dem så jämförbara som möjligt. I viss mån kommer dock alltid

kvarstående osäkerhet angående jämförbarheten att finnas och det är inget vi i första hand rekommenderar.

När det gäller metod för att samla in data och målpopulation finns det ingen principiell skillnad mellan design 2 och 3. Här gäller samma för- och nackdelar.

När det gäller vilka frågor som ska ställas finns det dock en del att tänka på för design 3. Mätningen innan händelsen sker (M1) är en planerad basmätning och den uppenbara svagheten är ju att vi inte säkert kan veta att de frågor vi ställer kommer att vara relevanta för den händelse som (kanske) sker. Därför är det av största vikt att ta fram allmänna frågor om det tänkta området som kan komma att bli aktuellt. Dessa mer allmänna frågor kan sedan ställas vid alla tre mättillfällena och jämföras. Utöver detta ska givetvis även mer specifika frågor kring händelsen ställas och jämföras mellan M2 och M3. Det är dock värt att ytterligare betona att hela idén med design 3 står och faller med att vi lyckas identifiera en uppsättning centrala frågor som kan ställas redan innan en händelse har ägt rum och följas både före, under och efter händelsen. Har vi en basmätning som inte fångar in centrala frågor som är tillräckligt väl relaterade till det som sedan kommit att hända så har hela fördelen med tre mätpunkter förfelats.

Vad gäller basmätningen kan vi även tillägga att vi hittills endast berört svårigheten med att formulera relevanta frågor i förväg. Detta tror vi emellertid inte är helt oöverstigligt. Några exempel kan t ex vara förtroende för olika samhällsaktörer eller institutioner, stödet för olika politiska förslag, eller förtroende och attityder till olika grupper i samhället eller andra länder. Om det däremot inte är den nationella opinionen vi är intresserade av utan vill förbereda oss för att studera effekter av en geografiskt avgränsad händelse eller om vi är intresserade av specifika delpopulationer så tillkommer utmaningen att utöver frågorna utse vilka områden eller grupper som är de relevanta.

**Tabell 3.3. Sammanfattning design 3. Tre mätningar – före, under och efter händelse**

| | | |
|---|---|---|
| **Urval och metod** | | |
| *Strategi* | Tvärsnittsdata<br>Undersökningar där resultat från tre separata urval jämförs (likt traditionella opinionsundersökningar). | Paneldata<br>Undersökningar av samma individer vid flera tillfällen. |
| *Urval* | Befolkningsregister<br>I första hand urval ur befolkningsregister. Finns även möjlighet att göra urval direkt ur så kallade konsumentdatabaser, dock med mer oklar representativitet. | Respondentpaneler<br>Urval av paneldeltagare. Paneler kan variera mycket i representativet, men kan vid behov stratifieras för att spegla olika populationer bättre. |
| *Metod* | Främst telefon. Även webbenkät är möjlig via sms-kontakt. Postenkät troligen för långsam. | Oftast webbenkät men kan också göras på andra sätt t ex via telefon eller sms. |
| **Målpopulation** | | |
| *Nationell opinion* | Representativt på nationell nivå som en traditionell opinionsundersökning. | Urval kan göras som speglar populationen (möjlighet till stort n-tal).<br>. |
| *Socio-demografiska grupper* | Viss information finns i urvalsram, men kräver screeningfrågor för att identifiera särskilda grupper. | Större möjlighet att undersöka särskilda grupper t ex yrkesgrupper, utsatta grupper, opinionsledare. Antingen genom mer detaljerad bakgrundsinformation om panelmedlemmar eller via specialpaneler. Eventuellt måste sådana grupper preciseras i förväg för att öka bakgrundsinformationen i panelen. |
| *Geografiskt område* | Kan undersökas på kommunnivå (info finns i även konsumentdatabaser). | Även mindre orter, starkt beroende på panelstorlek/urval. |
| **Enkätinnehåll** | | |
| *Specifika frågor* | Specifika frågor om det som hänt, kan dock givetvis inte ställas i för-mätning.<br><br>Räckviddsfrågor om insatser blir inte lika centrala som för design 1 eftersom effekter av insatser kan mätas direkt.<br><br>Nivåskattningar av uppfattningar/åsikter blir inte lika centrala som i design 1 eftersom förändringar av åsikter kan mätas. | |
| *Generella frågor* | Allmänna frågor om området som undersöks är av avgörande betydelse. Sådana frågor möjliggör jämförelser av opinionslägets förändring när en händelse inträffat.<br><br>Frågor om andra områden med syfte att kunna poststratifiera (vikta) vid stort bortfall eller urval med osäker representativitet. | |

## 3.7 Vad ska man välja?

Ska man dra ett antal slutsatser av de tre olika typerna av design ger de alla möjligheter till att erhålla värdefull information om hur opinionen

ser ut och framförallt förändras. Design 1, där endast eftermätning är möjlig ger minst möjlighet att uttala sig om förändringar, men om inget annat står till buds går det ändå att ge viss kunskap om opinionen kring en händelse. Har man tid och resurser är de andra typerna av design (2 och 3) att föredra. De ger mer kunskap och större möjlighet att faktiskt följa hur opinionen förändras och kontrollera hur informationsinsatser påverkar opinionen.

Jämförs design 2 och 3 är också den senare att föredra. Detta för att vi med större säkerhet kan uttala oss om händelsen faktiskt har påverkat opinionen genom att opinionsförändringar kan jämföras med basmätningen. I design 2 kan vi bara se om en informationsinsats leder till en opinionsförändring i förväntad riktning, men vi kan inte veta om vi har nått tillbaka till utgångsläget innan händelsen. Detta kan i viss mån vara möjligt med design 3 trots svårigheten med att formulera relevanta frågor innan vi vet vad mer precist som kommer att hända.

Vi vill därmed avsluta med att dra några slutsatser om möjligheterna att undersöka snabba opinionsförändringar som skisserats i kapitlet. I både den allmänna debatten och inom forskningslitteraturen har (den bristande) kvalitén på opinionsundersökningar uppmärksammats, inte minst efter att opinionsundersökningarna inte lyckades pricka in att medborgarna i Storbritannien skulle rösta för ett utträde ur EU och Donald Trumps valseger i presidentvalet i USA[7]

Vi menar dock att det är fullt möjligt att undersöka den allmänna opinionens förändringar med god precision och tillförlitlighet. Både metoder och kompetens att genomföra sådana undersökningar finns både inom den kommersiella sektorn och inom akademin. Men som genomgången visar finns det många olika saker att ta hänsyn till och alla de val som görs innan en undersökning genomförs kommer få konsekvenser för kvaliteten på resultaten och de slutsatser som kan dras. Som vi flera gånger påpekar finns det också osäkerheter i de metoder som används. Även om utveckling av undersökningsverksamhet pågår lever alla som undersöker opinioner med stora utmaningar, inte minst ifråga om fallande svarsfrekvenser och att det är svårt att nå vissa grupper i samhället.

Vi vill här trycka på *planering* och *långsiktighet* för att minimera risken för att få bristfälliga beslutsunderlag ifråga om opinionsförändringar. I planeringen ligger flera saker. En viktig faktor är att vara en bra beställare, att veta vilken typ av beslutsunderlag som efterfrågas. Om denna fas i planeringen görs noggrant är det större chans att man också kan beställa undersökningar av god kvalitet. I planeringen ligger också att tid och resurser läggs på att för-testa frågor och design så att man innan skarpt läge är förberedd och har undersökningsinstrument som håller god kvalitet. Ju mer detaljerade och färdiga mallar som finns för de frågor som ska ställas, desto större möjlighet har man också att snabbt kunna agera när en händelse ägt rum. Kan man dessutom inte bara för-testa frågor utan också genomföra ”brandövningar” av de

[7] Det bör dock påpekas att opinionsundersökningarna i det amerikanska presidentvalet inte var så missvisande som det först verkade. De allra flesta mätte opinionen på nationell nivå och visade att Hillary Clinton ledde med några procentenheter före Donald Trump. Valresultatet på nationell nivå visade också en seger för Clinton över Trump på drygt två procentenheter. I de s.k. swingstaterna gjordes det inte lika många opinionsundersökningar och de pekade åt lite olika håll. Vad som var förvånande var möjligen att alla dessa stater gick till Donald Trump trots att flera mätningar visade att Clinton ledde.

planerade undersökningarna har man också större chans att få ett lyckat slutresultat.

I planeringen ligger också att i god tid byggt upp ett samarbete med den organisation som ska genomföra opinionsundersökningarna så att resurser och kompetens finns den dag som en händelse inträffar att kunna genomföra de undersökningar som efterfrågas. Hur stora är till exempel krav på flexibilitet och justeringar av design och frågebatterier efter hand? Vill man snabbt kunna göra undersökningar av nya målgrupper mm?

För att kunna bedriva verksamheten på ett sätt som gör mätningarna kan ge ett relevant beslutsunderlag är det också av vikt att det finns en långsiktighet, både inom den organisation som beställer och i relationen till den organisationen som utför mätningarna. Kommer den egna organisationen och den som utför att fortsätta att kunna/ha intresse av att göra undersökningar så att beställarorganisationen kommer att kunna agera optimalt vid framtida händelser. Finns inte denna uthållighet är risken att kunskap går förlorad och inte blir integrerad i beställarorganisationens verksamhet.

Centralt är här också att vara klar över rollerna. Förväntas utförarorganisationen att leverera färdiga analyser (vilken nivå ska dessa ligga på?). Eller ska analyser skötas in-house och organisationen som gör mätningar ska bara leverera rådata som sedan analyseras av beställarorganisationen.

I rapportens sista kapitel ska vi skifta fokus från att mäta människors åsikter genom att göra olika opinionsundersökningar till att rikta intresset mot sociala medier. För tio år sedan hade knappt svenskarna hört talas om varken Facebook eller Twitter. Idag tar de upp en stor del av våra liv och frågan är om vår aktivitet på de sociala nätverkssajterna kan användas för att fånga in skeenden i samhället och kanske även hur opinionen förändras.

# 4. Metoder och verktyg för att analysera opinioner och händelser i sociala medier

Peter Dahlgren

## 4.1 Inledning

Facebook, Twitter, Google+, Instagram, Snapchat, Flickr, Youtube, Reddit, Flashback – det är namnet på några av alla de sociala medier där användare kan skapa en profil och dela med sig av innehåll de själva har skapat. Mängden texter, bilder och videor som delas på sociala medier ökar varje dag och dessa medier blir allt oftare plattformar för samhällsdiskussion. Samtidigt minskar svarsfrekvenserna för traditionella surveyundersökningar med slumpmässiga urval av befolkningen. Det blir därför allt intressantare att vända blicken mot sociala medier för att ta reda på medborgarnas åsikter, inte minst av ekonomiska skäl.

Data kan dessutom samlas in passivt från sociala medier, vilket inte kräver att användarna själva gör något. Det är också möjligt att samla in data aktivt genom att rekrytera användare på sociala medier, vilket är något dyrare men kan ändå vara enklare och billigare jämfört med telefonsamtal eller skicka ut enkäter med post. Därefter kan en mängd typer av analyser genomföras, såsom att undersöka användarnas sociala nätverk eller andelen positiva åsikter. Kort sagt verkar sociala medier ge en mängd fördelar för insamling och analys av opinioner.

Att utvinna information från sociala medier är därför en lukrativ bransch, inte minst bland företag som vill veta vad deras kunder säger om dem. Men för att analys av sociala medier ska ha något att bidra med är det viktigt att granska hur representativa åsikterna är, eller om det är ett fåtal personers åsikter som får oproportionerligt stort utrymme.

Litteraturöversikten i detta kapitel har därför tre syften: (1) undersöka hur pass väl sociala medier speglar opinioner i samhället, (2) beskriva vilka metoder som finns för att analysera opinioner och händelser i sociala medier, samt (3) identifiera praktiska verktyg för sådana analyser.

Som synes avser vi i kapitlet både diskutera hur opinioner och händelser av olika slag som är intressanta ur ett samhällsperspektiv kan undersökas. Det kan handla om mindre händelser som olyckor, påverkansförsök från andra länder till större händelser som sjukdomsepidemier, naturkatastrofer, terrorism eller krig.

Då det digitala medielandskapet snabbt förändras finns det en hög risk att de verktyg som listas här i dag blir omoderna i morgon. Av det skälet läggs större fokus på metoderna, såsom sentimentsanalys för att identifiera opinioner. Av utrymmesskäl kommer inte heller något större

fokus läggas på de etiska aspekterna av att samla in information från sociala medier, i synnerhet om man passivt samlar in och sparar stora mängder data som låter identifiera känslig information om människor. Det är på inga sätt oviktiga frågor, utan måste konstant övervägas i varje fall.

## 4.2 Unga överrepresenterade i sociala medier

Användningen av sociala medier fortsätter att öka för varje år som går, liksom andelen som söker nyheter via internet. Användningen av internet via mobiltelefon är nu uppe i 9,2 timmar per vecka i genomsnitt, en siffra som stadigt ökar (Davidsson & Findahl, 2016). Drygt 58 procent av befolkningen använder sociala medier dagligen och hälften av befolkningen använder sociala medier via mobiltelefonen dagligen. När mobiltelefonen alltid är närvarande är det därför möjligt att delge sin egen åsikt och ta del av andras åsikter under dygnets alla timmar, liksom att berätta när en extraordinär händelse inträffar.

Även om användningen av sociala medier ökar varje år – totalt sett – varierar användningen av specifika sociala medier. Facebook är i nuläget den överlägset största sociala plattformen, med drygt tre fjärdedelar av svenska befolkningen som medlemmar samtidigt som hälften av befolkningen använder Facebook varje dag (Davidsson & Findahl, 2016). Ungdomar i nedre tonåren (12 till 15 år), däremot, använder Snapchat och Instagram i högre utsträckning, medan de som är äldre (över 15 år) framför allt använder Facebook. Sociala medier domineras generellt sett av unga, och användningen av sociala medier avtar gradvis med stigande ålder. Pensionärer är därför kraftigt underrepresenterade.

Eftersom Facebook används av en så stor andel av befolkningen har information möjligheten att spridas väldigt långt. Men informationsspridningen begränsas delvis av Facebooks slutna struktur, där meddelanden sällan går att läsa utan vare sig registrering eller gemensamma vänner som informationen sprids emellan. Den slutna strukturen gör det också svårt för utomstående att utvinna information från Facebooks plattform, eftersom det i regel kräver godkännande från varje individ. Twitter, å andra sidan, har en öppen struktur vilket innebär att alla meddelanden och användarprofiler går att se utan registrering. Men även om det bara är drygt 6 procent av svenska befolkningen som använder Twitter dagligen (Davidsson & Findahl, 2016), finns där en stor andel offentliga personer och beslutsfattare, exempelvis journalister, politiker och kändisar. Twitter är därför en viktig källa för information och nyheter. Det är därför inte konstigt att Twitter är kraftigt överrepresenterat i medierna jämfört med andra sociala medier (Andersson Schwarz m fl., 2015). På grund av sin öppna struktur är Twitter också fördelaktigt för att sprida information som når beslutsfattare relativt snabbt (Andersson Schwarz m fl., 2015; Wu m fl., 2011). Av det skälet har tusentals forskningsstudier fokuserat på Twitter (Williams m fl., 2013), snarare än andra sociala medier som trots allt används av fler människor. Det är därför viktigt att ha i åtanke att användarna skiljer sig åt beroende på vilken social plattform man talar om.

Det är med andra ord viktigt att inte bara fokusera på hur många som använder en specifik tjänst, utan också vilka som använder den, hur de använder den och vilken roll tjänsten tjänar för informationsspridning. Flest svenskar använder Facebook men Twitter är i nuläget viktigare för att nå och påverka beslutsfattare. Eftersom Twitter har en öppen struktur är det också enklare att utvinna information från Twitter.

## 4.3 Opinioner i sociala medier är inte representativa för befolkningen

Det finns i dagsläget inte mycket forskning som har jämfört opinionerna i sociala medier med opinionen i samhället genom traditionella metoder, såsom surveyundersökningar med ett slumpmässigt och representativt urval av befolkningen (Murphy m fl., 2014). De få studier som finns har i regel modellerat opinionen i sociala medier med surveyundersökningar som facit, för att på så vis försöka matcha dem med varandra. De undersökningar som har försökt förutse valresultat utifrån Twitterdata har nästan uteslutande förutsett valresultatet *efter* valet har ägt rum (Gayo-Avello, 2013), vilket kraftigt överdriver träffsäkerheten. I en undersökning jämfördes politisk missnöjdhet i sociala medier med politisk missnöjdhet i traditionella representativa surveyundersökningar och resultatet visade att de stämmer relativt väl överens (Monti m fl., 2013). Det skulle innebära att det är möjligt att undersöka hur missnöjda människor är med demokratin genom att analysera sociala medier. Men det är i dagsläget inte helt klart huruvida detta faktiskt fungerar eftersom samma problem kvarstår: när modellen görs efter facit riskerar träffsäkerheten att kraftigt överdrivas. Det är därför väldigt osäkert hur väl dessa modeller kan stå på egna ben, utan hjälp av traditionella opinionsundersökningar.

Det finns emellertid väldigt få övertygande skäl till att sociala medier skulle vara representativa för opinioner i samhället. Varje medborgare är för det första inte representerad på sociala medier, och även om de var representerade skulle deras aktivitet inte nödvändigtvis vara jämbördiga.

Att utvinna information från sociala medier innebär ofta stora datamängder. Det kan ge en falsk känsla av precision, i synnerhet om man lyckas analysera *alla* meddelanden som existerar. Traditionella signifikanstest som används inom opinionsundersökningar och andra statistiska analyser blir också i princip meningslösa eftersom tillräckligt stora datamängder gör att variansen i princip blir noll, och därmed alltid leder till statistiskt signifikanta resultat (McFarland & McFarland, 2015).

Ska man karakterisera sociala medier i relation till det som diskuteras i tidigare kapitel kan de betraktas som självrekryterade icke-sannolikhetsurval. Annorlunda uttryckt kan vem som helst kan använda sociala medier för vilket ändamål de än behagar. Nedan kommer vi närmare att belysa vilka konsekvenser det får. Dels att informationen inte alltid är pålitlig, att ett användarkonto inte är detsamma som en person, samt, kanske viktigast av allt, nämligen själva aktiviteten på det sociala mediet.

### 4.3.1 Ofullständig och felaktig information

Till skillnad från traditionella surveyundersökningar är passiv insamling av data från sociala medier nästan alltid förknippat med ofullständig information. Inte minst kan användare välja att inte fylla i uppgifter om sig själva i sina profiler, men de kan också välja att göra eventuell information dold genom den sociala plattformens olika inställningar. Meddelanden som postas kan ibland också redigeras i efterhand eller till och med raderas, och eventuella länkar kan sluta

fungera. Förutom ofullständig information finns det också tillfällen då användarna anger felaktig information. I en amerikansk undersökning hade exempelvis 34 procent av användarna på Twitter i undersökningen felaktigt angett var de befann sig (Murphy m fl., 2014). Detta leder till slutsatsen att man måste vara försiktig om vilka slutsatser man kan dra från informationen som är insamlad från sociala medier, och helst av allt kontrollera informationen mot ytterligare en källa.

### 4.3.2 Ett användarkonto är inte detsamma som en person

Det krävs sällan mer än en e-postadress för att skapa ett användarkonto på sociala medier. Därför finns det heller inte ett en-till-en-förhållande mellan antalet konton och antalet fysiska personer. I stället kan en person sköta flera konton, eller så kan ett konto skötas av flera personer. Det är vanligt förekommande bland företag och organisationer som bedriver kundtjänst i de sociala medierna. Det finns dessutom konton som är helt automatiserade, så kallade *botar*, vars enda syfte är att exempelvis dela de senaste länkarna från en nyhetssida eller i värsta fall bara sprida spam. På Twitter har exempelvis ett nätverk med över 350 000 spambotar identifierats, som alla tycks vara styrda av en enskild individ eller organisation (Echeverría & Zhou, 2017). Antalet botar har uppskattats till mellan 9 och 15 procent bland de aktiva kontona på Twitter (Varol m fl., 2017).

Detta skapar huvudbry för den som undersöker sociala medier: är det personer, konton eller meddelanden som ska analyseras? Det vanligaste sättet är att analysera meddelanden (Williams m fl., 2013), eftersom de är rika på textinnehåll och till viss mån undviker problemet med att dra ett likhetstecken mellan konton och personer. Att analysera meddelanden riskerar därmed att kraftigt överrepresentera en specifik åsikt om det skrivs tillräckligt många meddelanden. Med andra ord kan ett fåtal personer i sociala medier ge sken av att en åsikt är mer vanligt förekommande än vad den faktiskt är. Trots dessa brister är det till skillnad från traditionella surveyundersökningar möjligt att samla in en kontinuerlig ström av meddelanden i realtid, vilket kan vara en fördel av andra skäl än att få en representativ bild av sociala medier eller befolkningen.

### 4.3.3 Aktiva användare är överrepresenterade

En traditionell surveyundersökning med ett representativt urval av befolkningen gör det möjligt att med relativ god säkerhet ta reda på hur opinionen i samhället ser ut vid en viss tidpunkt. På sociala medier är det dock annorlunda. Det vore kanske intuitivt att mäta antalet positiva respektive negativa meddelanden som skickas under en given tidsperiod för att ta reda på opinionen i sociala medier kring en given fråga. Men det är problematiskt eftersom antalet meddelanden influeras av aktiviteten på det sociala nätverket (Hargittai, 2015; Lewis, 2015; McFarland & McFarland, 2015). Under det svenska valåret 2010 var det exempelvis ett fåtal personer som stod för en majoritet av de politiska inläggen på Twitter (Larsson & Moe, 2012), och den tendensen har även hållit i sig några år senare (Andersson Schwarz m fl., 2015). Samma trend har observerats på Twitter som helhet, där ett fåtal personer står för en majoritet av de mest spridda meddelandena (Wu m fl., 2011). Det finns också en risk att man som individ överskattar hur vanligt förekommande något är, bara för att man exponeras för det många gånger (Tversky & Kahneman, 1973). Det vore därför ett misstag att försöka likställa antalet meddelanden i sociala medier med opinionen i sociala medier, och ett ännu större misstag att försöka

likställa antalet meddelanden med opinionen i samhället överlag. Vissa individer är helt enkelt mer aktiva än andra.

### 4.3.4 Många normer är specifika för den sociala plattformen

Hur man beter sig på sociala medier beror ofta på vilka sociala medier vi diskuterar. Det har exempelvis utvecklats en norm i digitala sammanhang som innebär att ett uteblivet svar på en direkt förfrågan av något slag betyder detsamma som ”inte intresserad” (Lewis, 2015). Hur pass utbredd denna norm egentligen är, är en svår fråga att svara på. Poängen är dock att meddelanden i sociala medier inte alltid kan tolkas oberoende av de plattformar där meddelandena skapades eftersom det många gånger uppstår implicita normer för hur man ska bete sig och kommunicera med varandra. Dessa normer framgår dock sällan för en utomstående, utan blir först synliga efter en viss tids användning. Med andra ord krävs det åtminstone en grundläggande förståelse för den specifika kulturella kontext där diskussioner äger rum innan man kan dra slutsatser om hur människor kommunicerar (eller inte kommunicerar). Ett uteblivet svar kan med andra ord vara ett sorts svar.

### 4.3.5 Sociala medier som understödjande del i ett övergripande arbete

Sociala medier kanske allra bäst fungerar som ett stöd till traditionella metoder, som exempelvis en surveyundersökning. Det kan ske på flera sätt. För det första är det möjligt att låta användare i sociala medier influera designen av surveyundersökningar, som frågeformuleringar och svarsalternativ, genom att snabbt och effektivt testa enkäter på de grupper av människor man är intresserad av.

För det andra är det möjligt att observera beteende och språkbruk bland användare på sociala medier under lång tid för att på så vis få insikter som traditionellt sett inhämtas via fokusgrupper eller etnografi. Detta kan minska kostnaderna jämfört med att rekrytera och möta deltagare fysiskt.

För det tredje är det möjligt att också aktivt rekrytera deltagare via annonser på sociala medier, och möjligheten finns att styra annonserna så att de endast visas hos specifika demografiska grupper eller personer med vissa intressen. I vissa fall kan det vara snabbare och billigare att rekrytera deltagare via sociala medier (Murphy m fl., 2014), i andra fall långsammare och dyrare (Moreno m fl., 2016). Det är därför viktigt att på förhand klargöra vilka behov som finns för eventuell rekrytering av deltagare, till exempel hur pass viktigt det är att deltagare från målpopulationen rekryteras (snarare än någon annan population). Dessa metoder kräver emellertid ett aktivt handlande från både den som undersöker och de som deltar i undersökningen. Härnäst vänder vi därför blicken mot passiva metoder som kan automatiseras.

## 4.4 Metoder för att analysera sociala medier

Passiva metoder som automatiskt samlar in och analyserar data från sociala medier har en mängd fördelar. Det kräver minimalt med arbete när det väl är i drift, det minskar kostnaderna och kräver ingen som helst ansträngning från de personer som undersöks. Det är också möjligt att minimera effekten av deltagare som agerar (och svarar) annorlunda när de vet att de blir observerade.

Den information som kan utvinnas från sociala medier består i huvudsak av två typer. Dels själva innehållet (text, bild, ljud, videoklipp etc.) och dels information om innehållet, så kallad metadata. I textinnehållet förekommer förutom texten med åsikten vanligtvis länkar och hashtaggar. Hashtaggar är ord som fungerar som sökbara nyckelord genom att tecknet # placeras framför ordet (såsom #svpol för svensk politik), vilket innebär att meddelandet syns för alla personer som väljer att ta del av hashtaggen. Hashtaggar kan därför vara ett sätt för individer att försöka sprida ett budskap till många.

Genom länkar i ett meddelande går det också att identifiera vilka medier och bloggar som är dominerande och länkas till mest, för att på så vis undersöka vad för slags information som delas och vilka användarkonton som delar det. På så vis kan man exempelvis undersöka mängden av konspirationsteorier (Starbird, 2017).

Tillsammans med ett textmeddelande på Twitter tillkommer dessutom en mängd andra metadata, exempelvis användarnamn, tid och datum, tidszon, språk, geografisk platsinformation (antingen självrapporterad eller i form av longitud och latitud från mobiltelefonens GPS), namnet på appen som användaren postade meddelandet ifrån, användarens sociala nätverk (vänner och följare), antal delningar och gillningar, vilka andra användare som användaren interagerar med, samt i förekommande fall även en mängd olika preferenser som personen har skrivit i sin profil (politik, filmer, böcker, mat, sport med mera). Det är dock viktigt att återigen påpeka att all information inte nödvändigtvis behöver vara korrekt bara för att den är allmänt tillgänglig, och mängden information varierar dessutom mellan sociala plattformar. Facebook tillåter användare att fylla i betydligt mer information, men är å andra sidan svårare för utomstående att få tillgång till.

Denna information kan många gånger verka harmlös, men har exempelvis använts för att identifiera vem som skrivit meddelanden på USA:s president Donald Trumps Twitterkonto under valkampanjen (Robinson, 2016). Meddelanden som innehöll mestadels fakta skrevs ofta från en mobiltelefon av märket iPhone, medan meddelanden som innehöll starka åsikter ofta skrevs från en Android, vilket indikerar att olika personer använde kontot för olika syften.

Tidpunkten när ett meddelande publiceras kan också ses som harmlös information. Men faktum är att det är möjligt att använda tidpunkten för att identifiera om det är en människa eller en bot som skriver meddelanden (Radziwill & Benton, 2016). I regel är människor mer slumpmässiga i sitt beteende jämfört med botar. Om tidsskillnaden mellan två meddelanden är förhållandevis lika över tid, kan det därför vara en indikation på att det inte är en människa som publicerar meddelanden.

Metadata av olika slag ger därför en mängd information utöver innehållet som dessutom kombineras på olika sätt. Men många etiska frågor om personlig integritet uppstår. När det är möjligt att systematiskt se att en individ använder en dator på förmiddagen och en mobiltelefon på eftermiddagen, finns det också en stor möjlighet att identifiera var individen kan förväntas vara vid olika tidpunkter på dagen.

De sociala plattformarna skapar också aggregerad information. Både Facebook och Twitter visar ”trender”, vilket är ämnen som diskuteras flitigt för stunden på en given geografisk plats. Det är därför möjligt, och betydligt mindre resurskrävande, att samla in och analysera denna information. Det är dock inte alltid uppenbart hur de sociala

plattformarna går tillväga när de sammanställer aggregerad information. Offentlig information lider också av självförstärkande popularitetseffekter, det vill säga att det som är populärt tenderar att omtalas mer och därmed bli ännu mer populärt. Aggregerad information är därför öppet för manipulation, åtminstone om det presenteras så att alla kan se den.

## 4.5 Web scraping och API – samla in data

Det är förstås möjligt att manuellt samla in all den data som beskrivits ovan, men överstiger det en viss gräns är det enklare att automatisera insamlingen. Det kan göras på två sätt. Det första är genom så kallad *web scraping*, vilket är en relativt okomplicerad metod som innebär att ett datorprogram automatiskt går in på en webbsida och extraherar relevant information. Datorprogrammet måste med andra ord instrueras vad som är relevant information, och det är inte en helt enkel uppgift. Extraheringen havererar omgående om den sociala plattformen organiserar informationen på ett annat sätt. Web scraping är därför ingen metod som rekommenderas, annat än vid brist på alternativ.

Det andra sättet att samla in data via de sociala plattformarna är via så kallade *application programming interface* (API). Det är ett gränssnitt för att utbyta data. Principen är dock densamma som vid web scraping: ett datorprogram extraherar information från en webbsida, men skillnaden är att webbsidan denna gång endast innehåller data som är uttryckt i standardiserad form. Ett datorprogram som läser information från ett API vet med andra ord på förhand vilken struktur informationen har (vilket står beskrivet i en dokumentation). API är alltså ett standardiserat sätt att utbyta data med de sociala plattformarna och det är också så mobiltelefonens appar kommunicerar med plattformarna över internet. Det är därför förhållandevis enkelt att med grundläggande programmeringskunskaper bygga ett program som hämtar data via dessa API, till exempel alla meddelanden som innehåller ett visst sökord.

Nackdelen är dock att API i regel har en rad begränsningar, bland annat hur mycket data som kan hämtas (utan att betala). API-tjänsterna kan dessutom ändras eller i värsta fall försvinna utan större förvarningar, och det är inte ovanligt med avtal som begränsar vad man kan göra med insamlad data. Trots begränsningarna är API alltid att föredra över web scraping.

## 4.6 Sentimentsanalys – klassificera positiv och negativ opinion

*Sentimentsanalys*, eller *opinion mining*, innebär att man analyserar texter och bedömer om de uttrycker positiva, negativa eller neutrala åsikter (Pang & Lee, 2008). Detta forskningsområde har utvecklats kraftigt tack vare internets utbredning och den ökade mängden användargenererat innehåll. Mycket av forskningen har fokuserat på produkt- och filmrecensioner, vilket är praktiskt eftersom recensionerna ofta innehåller numeriska betyg som kan användas som ett slags facit för att bedöma tillförlitligheten i metoden. Om analysmetoden uppskattar att texten är överväldigande positiv, men recensenten har gett filmen betyget 1, kan man därför ifrågasätta metoden. Ju mer forskningen utvecklas, desto mer används metoderna också på andra områden än bara produkt- och filmrecensioner. Det är väsentligt eftersom analys av åsikter inom ett ämnesområde inte enkelt låter sig överföras till ett nytt ämnesområde, utan kräver anpassning (Maragoudakis m fl., 2011).

Det finns ett antal metoder för att analysera opinionen i sociala medier som ska redogöras för här, från enklare analyser av nyckelord baserade på lexikon till betydligt mer avancerad maskininlärning. Det är viktigt att tänka på att metoderna som beskrivs här inte enbart gäller sociala medier. Metoderna kan lika väl användas på nyhetstexter eller texter av andra slag. Det vore till och med bättre, eftersom längre texter ger högre tillförlitlighet än korta. Men man kan åtminstone samla in större mängder meddelanden från sociala medier och på så vis kompensera respektive meddelandes korta längd med ett större antal.

### 4.6.1 Identifiera värdeladdning med lexikon

Det mest uppenbara sättet att uppskatta opinionen är att samla in meddelanden med text och därefter använda ett lexikon med positiva respektive negativa ord för att värdera texten. En mening som ”det är bra” skulle därför klassificeras som en positiv mening på grund av ordet ”bra”. Därmed är det möjligt att räkna ut förhållandet mellan andelen positiva respektive negativa meddelanden för att få ett slags mått på opinionen.

Förutom positiv respektive negativ opinion är det måhända mer intressant att klassificera opinionen utifrån vilka känslor som kommer till uttryck, såsom glädje, ängslighet, sorg, rädsla, ilska, chock etc. Antagandet är dock densamma, att känslor representeras av en viss typ av ord – ”äntligen” som uttryck för glädje, ”slå” som uttryck för ilska, och så vidare. Går man ytterligare ett steg längre är det möjligt att identifiera ämnesområdet och det objekt dessa känslor riktar sig mot (Fukuhara m fl., 2007). På så vis är det möjligt att undersöka vilka känslor som yttras om en viss person eller ett specifikt objekt.

Denna intuitiva lösning med att klassificera enskilda ord har dock flera problem, bland annat svårigheten att ta hänsyn till ironi, sarkasm, negationer (”det är inte bra”), innebörd (”jag ska slå honom” respektive ”jag ska slå honom med häpnad”), liksom värdeladdade meningar som trots det saknar värdeladdade ord (”i dag tog jag min examen!”), och inte minst jämförelser av typen ”A är bättre än B” (se Varathan m fl., 2016). Dessutom kan det vara svårt att klassificera enskilda meningar utan hänsyn till sammanhang, i synnerhet som det här exemplet visar där sista meningen negerar de föregående tre meningarna:

> På pappret hade filmen en bra regissör. Utmärkta skådespelare. En prisbelönad manusförfattare. Men ändå blir allt fel.

Därutöver är det inte alltid möjligt att fastställa vad personen ifråga syftar på, det vill säga objektet åsikten handlar om. I exemplet ovan handlar det om filmen – även om regissör, skådespelare och manusförfattare nämns.

Lexikonbaserade analyser har alltså fördelen med att de är relativt enkla att konstruera och genomföra men saknar ofta tillförlitlighet, i synnerhet om man rör sig utanför ett specifikt ämnesområde. De flesta lexikon som redan finns är dessutom författade på engelska eller kinesiska, vilket ytterligare begränsar möjligheten att utgå från befintliga lexikon. Språkbanken på Göteborgs universitet har dock fritt tillgängliga lexikon för svenska språket.[8] Lexikonbaserade analyser bör av dessa skäl tolkas med försiktighet vid identifiering av opinioner (Gayo-Avello, 2013).

[8] https://spraakbanken.gu.se/swe/resurs/sentimentlex

### 4.6.2 Identifiera nyckelord med statistik

Det finns också en rad statistiska metoder för att analysera texter som i huvudsak bygger på statistik av frekvenser över en mängd texter (så kallade *korpus*). En vanlig metod, som framför allt är relevant när man har längre texter, är att räkna förekomsten av ord. Tanken är enkel, ju fler förekomster av ett visst ord, desto sannolikare är det att texten ifråga handlar om detta ord. Först måste man dock sortera bort vanliga ord som förekommer naturligt i alla texter (”och”, ”att”, ”så” etc.). Denna metod har visat sig vara tillförlitlig med att klassificera ämnet i en text. Om orden ”bil”, ”fordon”, ”Volvo” och ”Saab” förekommer i texten tillräckligt många gånger (i förhållande till texten som helhet) kan man vara relativt säker på att textens ämne är bilar.

Denna metod kan också automatiseras genom att man skapar ämnesordlistor på förhand. Nackdelen är att metoden ofta kräver längre texter än de meddelanden som finns i sociala medier, och det behövs ämnesordlistor för varje ämnesområde. Det finns dock ytterligare metoder som syftar till att undersöka relationen mellan ord i en enskild text (*co-occurrence analysis*), liksom mellan ord över flera texter (*term frequency–inverse document frequency*, *TF-IDF*, samt *topic modelling*). Dessa metoder är användbara för att bedöma ett ords närhet med andra ord, och är kanske mest användbart när man behöver finna relevanta sökord för att söka efter meddelanden i sociala medier (se Andersson Schwarz & Hammarlund, 2016). De ord som förekommer ofta tillsammans har rimligen någon relation med varandra.

### 4.6.3 Identifiera mönster med maskininlärning

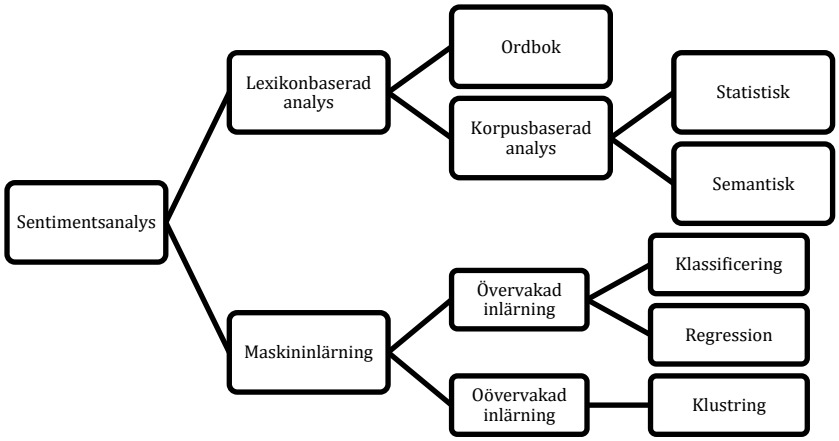
En helt annan metod, som inte använder lexikon med nyckelord men väl avancerad statistik, är maskininlärning. Det är en sammanfattande benämning på metoder som på egen hand hittar mönster i en stor mängd data. Det kan ske på två sätt. Det första sättet kallas *övervakad maskininlärning* och innebär att en algoritm exempelvis instrueras att lära från en delmängd av en stor samling texter som på förhand är klassificerade som positiva, negativa och neutrala av människor (så kallad träningsdata). Man testar sedan hur väl inlärningen har gått genom att algoritmen får försöka klassificera återstående del av textsamlingen som algoritmen inte sett tidigare (så kallad testdata). Detta tillvägagångssätt låter alltså användaren avgöra hur väl algoritmen lyckades klassificera de nya texterna genom att jämföra med ”facit”, och det blir också möjligt att se hur många texter som felaktigt blev klassificerade som exempelvis positiva, när de i själva verket var neutrala eller negativa.

Övervakad maskininlärning av texter utgår med andra ord från antagandet att nya texter kommer att se ut ungefär som tidigare texter. Därför kan det många gånger vara nödvändigt att ”träna om”, det vill säga ge algoritmen nya exempel att klassificera, vilket återigen kräver ytterligare manuell arbetsinsats. Hur många gånger detta behöver göras beror på hur omfattande förändringar som sker i texterna ifråga. Ju mer lika nya texter förväntas vara med gamla, desto mindre resurskrävande är det att träna om, och desto mer träffsäkra resultat kan man förvänta sig över tid. Att klassificera innehållet i sociala medier kan därmed innebära en stor arbetsinsats eftersom språket många gånger utvecklas i takt med användningen. Forskare delar dock ofta med sig av träningsdata, vilket underlättar bedömningar av tillförlitligheten i nya algoritmer. Men när det gäller text sker detta genomgående på engelska, vilket innebär att det behövs resurser för att skapa träningsdata på svenska. Annorlunda uttryckt måste alltså en

människa manuellt klassificera hundratals eller tusentals meddelanden som positiva eller negativa, för att därefter se vilken algoritm som närmast kan återskapa samma resultat.

Bilder, däremot, lider inte av språkproblem. Google har exempelvis offentliggjort drygt nio miljoner träningsbilder som på förhand är uppmärkta med etiketter om sex tusen kategorier (Krasin & Duerig, 2016). En bild på kniv och gaffel har exempelvis märkts upp med etiketterna ”cutlery”, ”tableware”, ”metal”, ”tool”, ”spoon”, och ”fork”, vilket sedan kan användas som facit för en algoritm som försöker gissa vad bilden innehåller. Det är även möjligt för algoritmer att exempelvis klassificera känslouttryck i människors ansikten. Ur ett opinionsperspektiv är dock inte alltid bilden relevant, utan snarare sammanhanget den placeras i, vilket ofta består av text som beskriver bilden.

Fördelen med övervakad maskininlärning är att man alltid har ett facit att jämföra med, vilket också gör det möjligt att bedöma tillförlitligheten i den algoritm man använder, det vill säga antal korrekta klassificeringar i förhållande till antal felaktiga klassificeringar. Nackdelen är att det ofta kräver en stor manuell arbetsinsats eftersom en människa på förhand måste klassificera texterna. Det är också svårt att bedöma tillförlitligheten när texterna förändras över tid, som de så ofta tenderar att göra på sociala medier.




**Figur 4.1. Metoder för analys av åsikter och opinioner**
Att analysera åsikter kan göras på en mängd olika sätt med varierande grad av automatisering. Vilket man väljer beror till stor del på syftet, men det är vanligt att många av metoderna används i olika delar av processen (bild baserad på Serrano-Guerrero m fl., 2015).

Den andra typen av maskininlärning är *oövervakad maskininlärning*. Med denna metod har man inget facit att jämföra med, utan algoritmen analyserar texterna helt på egen hand. I stället för att klassificera texter som exempelvis positiva eller negativa kan en algoritm för oövervakad maskininlärning försöka hitta kluster av texter, det vill säga texter som har något gemensamt med varandra. En människa får därefter göra en bedömning av hur tillförlitligt resultatet blev, liksom hur klusteranalysen ska tolkas. Användningsområdet för oövervakad maskininlärning är därmed något annorlunda.

Övervakad och oövervakad maskininlärning utesluter inte varandra. Det är ofta fördelaktigt att låta båda influera olika delar av processen. Klusteranalys är exempelvis användbart för att hitta nyckelord som kan användas vid sökningar, och algoritmer kan många gånger vara bättre än människor på att hitta relaterade och relevanta sökord (Pang & Lee, 2008).

Efter att en algoritm för maskininlärning har bearbetat meddelanden från sociala medier kan resultatet presenteras som textsammanfattningar som summerar det viktiga. Det kan dock vara opraktiskt om det rör sig om mycket text. Då kan kvantitativa sammanfattningar vara att föredra, vilket i synnerhet är användbart för att visa förändringar över tid.

### 4.6.4 Analys av nätverkets struktur

Genom så kallad *social nätverksanalys* är det också möjligt att se vilka individer som har kontakt med varandra i ett socialt nätverk, vilka som är inflytelserika opinionsledare, och var samt hur information sprids. Det är inte alltid den som har flest följare, fans eller vänner som är mest inflytelserik när det kommer till spridning av information, utan kan i stället vara den individ som överlappar flest lokala nätverk. Detta är en mycket mer tidskrävande analys (i synnerhet om man lägger till tiden som det tar att extrahera data från de sociala plattformarna) som också kräver mycket datorkraft, men som i gengäld är mer stabil över tid och kan förberedas innan en händelse uppstår. Däremot uppstår problem med integritetsaspekten, i synnerhet om en myndighet kartlägger en större del av nätverket i sociala medier i förebyggande syfte.

### 4.6.5 Utbrottsidentifiering – identifiera händelser automatiskt i realtid

Utbrottsidentifiering (*burst detection*) innebär att man identifierar så kallade *utbrott*, snabba ökningar av likartade meddelanden under en kort och intensiv tidsperiod. Utbrotten kan fungera som en indikator för extraordinära händelser, såsom jordbävningar, flygplanskrascher, terrorism, vissa typer av pandemier och händelser som sker plötsligt och oväntat. Då förekommer det en mängd meddelanden vid samma tidpunkt som syftar till att uppmärksamma andra på händelsen. Identifiering av utbrott sker, förenklat uttryckt, genom att en viss typ av meddelanden kontinuerligt bevakas i sociala medier. När dessa meddelanden överstiger deras normala volym indikerar det ett utbrott. Det har utvecklats metoder för att automatiskt identifiera och varna i realtid när sådana utbrott sker (Yin m fl., 2015, Al-garadi m fl., 2016), och det finns fritt tillgängliga programbibliotek för utbrottsidentifiering (Kejariwal, 2014).

Det finns dock ett problem. Varje analys av sociala medier kräver en utgångspunkt, det vill säga sökord för de meddelanden man vill fokusera på att hitta. Det fungerar bra när man undersöker något som redan har hänt, men under krishändelser uppstår därför en paradox: Hur ska man veta vilka sökord man ska använda innan händelsen faktiskt har ägt rum? En lösning på denna paradox är att automatiskt söka efter nya sökord baserat på redan befintliga sökord. Det kan vara sökord som är relaterade till en specifik geografisk plats eller mer allmänna sökord som är relaterade till en rad fenomen, såsom ”upplopp”, ”flygplanskrasch”, ”jordbävning” eller dylikt. Därefter noterar man alla andra ord som förekommer tillsammans med sökordet (genom att identifiera nyckelord med hjälp av statistik, se rubrik 4.2.2). Det vill säga, de ord som ofta förekommer tillsammans med sökordet, under en specifik tidsperiod, kan därför antas vara relaterade med

varandra. Därmed bildar de nyfunna orden nya sökord som sedan används för att utvidga inhämtningen av nya meddelanden. De grupper av ord som ökar snabbt i storlek kan sedan identifieras (det vill säga utbrottsidentifiering) och därefter undersökas närmare. Denna adaptiva metod har föreslagits under namnet *focused crawling* (Zielinski m fl, 2013). Metoden kan antas vara behjälplig tillsammans med utbrottsidentifiering och fungera som en tidig varningssignal på något som bör analyseras manuellt av en människa (Yin m fl., 2015, Algaradi m fl., 2016).

Utbrottsidentifiering är dock i framkanten av forskningen och dessutom något som svårligen låter sig utvärderas under kontrollerade former. Därför är det i mångt och mycket spekulativt i nuläget, och det är tveksamt om den investering i infrastruktur och kompetens som behövs för att driva ett sådant projekt verkligen ger lika stor avkastning. Möjligheten finns dock, och ju större del av befolkningen som använder sociala medier, desto större sannolikhet att utbrott sker.

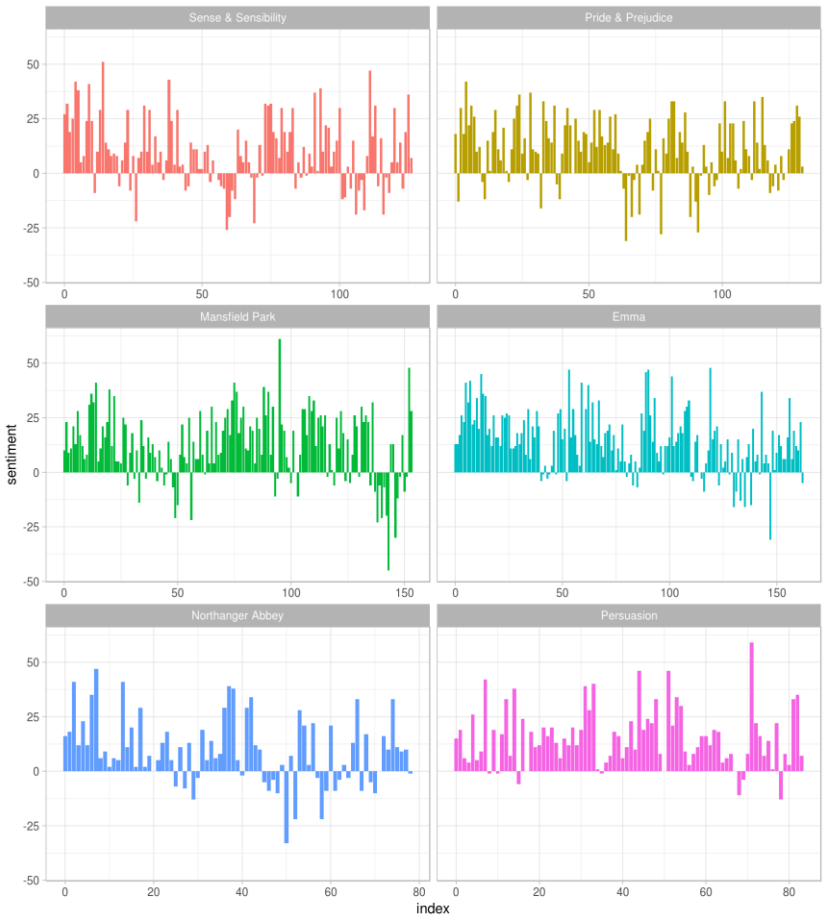
De möjliga användningsområdena är dock begränsade. Vid strömavbrott fyller sociala medier över internet ingen funktion när både trådlösa nätverk inomhus och mobiltelemaster utomhus blir strömlösa. Men å andra sidan bör orsaken till ett sådant strömavbrott bli uppmärksammat på andra sätt.

### 4.6.6 Mäta opinionsförändringar över tid

Sociala medier lämpar sig inte för att mäta nivåer av opinionen vare sig i samhället eller i sociala medier. Men däremot kan det vara fördelaktigt att mäta förändringar över tid. Nackdelen med att aktiva användare producerar fler meddelanden kvarstår förstås. Men denna nackdel kan dock vändas till en fördel om man använder aktiviteten som en indikator på påverkansförsök. Med andra ord, ju fler negativa meddelanden som skrivs om ett objekt över tid, desto större sannolikhet att det är ett medvetet påverkansförsök på opinionen. Därmed kan man följa förändringar över tid för att upptäcka, och i bästa fall *förutse*, påverkansförsök på opinionen i sociala medier.

Eftersom data från sociala medier går ned på sekundnivå är det också möjligt att se hur specifika händelser påverkar den efterföljande opinionen. I synnerhet om opinionen i sociala medier går stick i stäv med opinionen som den ser ut genom traditionella metoder, såsom surveyundersökningar.

Däremot är det svårt att veta vilka individer (i sociala medier) som har påverkats av en eventuell händelse och vilka som inte har gjort det blott genom att analysera känslouttryck efter händelsen ägt rum. Risken finns då att man antingen överskattar eller underskattar orsaken till opinionsförändringar. I en amerikansk undersökning av presidentvalet 2012 visade det sig exempelvis att opinionen på Twitter många gånger gick stick i stäv med opinionen i befolkningen, samtidigt som Twitteropinionen svängde kraftigt fram och tillbaka (Mitchell & Hitlin, 2013). Men frågan om orsak och verkan i sociala system i vid bemärkelse är en naturlig fråga för samhällsvetenskapen, och data från sociala medier kan mycket väl betraktas som ett naturligt experiment (Shadish m fl, 2001). Med andra ord, metoder inom samhällsvetenskapen kommer väl till pass för att analysera tidsserier och bedöma om en specifik händelse har orsakat eller influerat opinionsförändringen, i synnerhet om det kombineras med surveyundersökningar av opinionen före, under eller efter en specifik händelse.




**Figur 4.2. Förändring i värdeladdning över tid**
Jane Austens sex böcker rad för rad, där staplarna representerar förhållandet mellan positivt respektive negativt laddade ord. Staplar som går över noll indikerar mestadels positiva ord, medan staplar som går under noll indikerar mestadels negativa ord (bild från Silge & Robinson, 2017).

Det enklaste sättet att visa förändringarna över tid är genom att använda lexikon för att mäta förhållandet mellan positiva respektive negativa ord. Ett diagram kan sedan konstrueras där värdeladdningen på orden placeras på ena axeln medan tidpunkten placeras på den andra axeln (figur 4.2). Därmed kan man visualisera opinionen, och huruvida den blir mer eller mindre negativ över tid. Det är även möjligt att se vilka ämnen opinionen handlar om, liksom vilken typ av känslor som kommer till uttryck (Ku m fl, 2006; Fukuhara m fl., 2007; Pang & Lee, 2008).

Dock får man vara försiktig med att tolka förändringarna i opinionen i sociala medier. Problemen för en dator att exempelvis skilja ironi från allvar kvarstår. Men genom att i varierande grad kombinera de metoder som redogjorts för tidigare (sentimentanalys med lexikon, statistik och maskininlärning), är det möjligt att minimera de problem som varje enskild metod har.

Man bör dock inte fästa allt för stor tilltro till opinionsförändringar *enbart* i sociala medier. Under 2008 startade Google tjänsten Flu Trends som visade influensaaktiviteten i USA, men också andra länder. Genom att mäta antalet sökningar på influensasymptom kunde Google förutsäga förekomsten av influensa bland befolkningen. Resultaten jämfördes sedan med officiell statistik med bekräftade fall från amerikanska smittskyddsmyndigheter. Inledningsvis följde Flu Trends

den officiella statistiken relativt väl, efter ett fåtal justeringar, men i slutet överskattade Flu Trends influensan som dubbelt så vanligt förekommande jämfört med den officiella statistiken (Lazer m.fl., 2014). Google lade kort därefter ned tjänsten. En av många anledningar till att Google överskattade antalet influensafall kan vara att mediebevakningen påverkade människor att söka efter influensarelaterade symptom.

Automatisk analys låter sig därmed sällan göras helt självständigt, utan måste alltid kontrolleras (och kompletteras) av en människa som tar hänsyn till faktorer utanför analysen. Men automatisk analys kan ge värdefulla förslag på var man bör rikta sina resurser för en mer ingående manuell analys som genomförs av människor. Som ett sätt att upptäcka påverkansförsök på opinionen kan automatisk sentimentsanalys av opinionsförändringar därför vara ett intressant alternativ.

## 4.7 Sammanfattning och rekommendation av metoder

Så långt har ett antal mer eller mindre automatiska analysmetoder för sociala medier beskrivits: *sentimentsanalys* för klassificering av text som positivt eller negativt, eller textens känslouttryck; *utbrottsidentifiering* för att hitta utbrott av meddelanden under en intensiv tidsperiod, såsom vid naturkatastrofer eller terrorhändelser; *nätverksanalys* för undersökning av vilka som är opinionsledare och opinionsföljare i ett socialt nätverk; samt det mer spekulativa *focused crawling* för automatisk identifiering av extraordinära händelser. Det är inte ovanligt att metoderna kombineras (exempelvis Yin m fl., 2015).

Att analysera sociala medier ställer ofta höga krav på teknisk kunskap och anpasslighet eftersom företagen (Facebook, Twitter med flera) ändrar sina tjänster med jämna mellanrum. Det är inte alltid säkert att det går att analysera gamla meddelanden, utan kräver många gånger att man specificerar sina sökord på förhand för att sedan kontinuerligt bevaka och samla in sökresultat. Det är dock en god idé att vara specifik eftersom mängden innehåll i de sociala medierna gör att varje avgränsning sparar både tid och pengar.

Den största behållningen kommer förmodligen inte från analyser av enstaka meddelanden, utan hur dessa meddelanden förändras över tid. Däremot är det svårt att avgöra hur pass tillförlitlig resultatet är utan något att jämföra med. Att förutse valresultat eller opinioner genom sociala medier låter sig i dagsläget inte göras, utan kräver dels mer forskning och dels att metoderna kalibreras med metoder som vi i nuläget vet fungerar (såsom traditionella surveyundersökningar). Däremot är det möjligt att använda sociala medier som en källa för att upptäcka påverkansförsök på svensk opinion.

Information som utvinns från sociala medier fungerar därför bäst i kombination med både traditionella metoder och med andra metoder för sociala medier där de kan komplettera varandra. Rekommendationen är således att använda och utveckla nya metoder tillsammans med redan beprövade traditionella metoder för att utvärdera kvaliteten på de nya metoderna.

## 4.8 Verktyg för analys av sociala medier

Det finns ett överflöd av mer eller mindre tekniskt komplicerade verktyg för att analysera innehållet och nätverken i sociala medier. Då

dessa verktyg med all sannolikhet kommer att förändras i framtiden, beskrivs även de organisationer som utvecklat verktygen om det finns en sådan.

Det finns förhållandevis få kompletta program som gör det mesta av arbetet med att analysera opinioner eller händelser i sociala medier. I stället består utbudet framför allt av programbibliotek som går att ladda ned och integrera i ett program, vilket bäst kan beskrivas som legoblock som kan fogas tillsammans för att utföra en uppgift. Dessa kräver dock mer programmeringskunskaper för att kunna sammanfogas.[9]

I listan nedan utesluts därför programbibliotek och i stället ligger fokus på mer kompletta programvaror för att analysera sociala medier på olika sätt, från ax till limpa. Dessutom utelämnas kommersiella aktörer och i stället fokuseras på fritt tillgängliga alternativ på internet. Det beror i huvudsak på att forskningen är mer transparent i förhållande till de kommersiella alternativen, vilket gör de kommersiella alternativen svårare att utvärdera. Många kommersiella aktörer bygger dessutom tillämpningar på fria alternativ.

## 4.9 Insamling av data från sociala medier

Det har utvecklats en mängd verktyg för insamling av data från sociala medier genom åren som sedermera lagts ned eller som helt eller delvis slutat fungera, ofta på grund av ändringar hos de sociala plattformarna. Här följer ett dagsaktuellt urval av programvaror som är långt ifrån komplett.

### 4.9.1 Creepy

http://www.geocreepy.com/

Creepy är ett program för insamling av meddelanden och bilder från en geografisk plats som sedan kan visas på en karta. Creepy har stöd för Twitter, Instagram, Flickr och Google+. Programmet är byggd på öppen källkod.

### 4.9.2 DMI-TCAT

https://wiki.digitalmethods.net/Dmi/ToolDatabase

DMI-TCAT, eller Digital Methods Initiative Twitter Capturing and Analysis Toolset, är en webbapplikation som automatiskt samlar in meddelanden från Twitter baserat på sökord, person eller geografisk plats. Det är därefter möjligt att exportera data för en mängd analyser, däribland nätverksstruktur. Utvecklat av nederländska Digital Methods Initiative, som dessutom har en rad andra verktyg. DMI-TCAT är öppen källkod.

### 4.9.3 Echosec

https://app.echosec.net/

[9] Exempel på programbibliotek är Breakout Detection i R som utvecklats av Twitter för utbrottsidentifiering, samt LingPipe och NLTK för natural language processing i Java respektive Python. Stanford University har också en mängd programbibliotek för både natural language processing (http://nlp.stanford.edu/) och nätverksanalys (http://snap.stanford.edu/).

Echosec är en proprietär sökmotor som utgår från en geografisk plats. Man kan ange en plats på en karta och sedan söka efter offentlig digital kommunikation i realtid som sker på den platsen. Echosec kan därmed vara användbart för att snabbt följa krishändelser. Echosec finns i en gratisversion samt i en kommersiell version.

### 4.9.4 Netvizz

https://wiki.digitalmethods.net/Dmi/ToolNetvizz

Netvizz utvecklas av Bernhard Rieder vid Amsterdams universitet och kan användas för att samla in data från Facebook-grupper och Facebook-sidor. Netvizz fungerar som en webbapplikation inom Facebook och installeras därmed med endast ett klick genom den egna Facebook-profilen. Applikationen kan också exportera data för en mängd typer av analyser.

### 4.9.5 Social Feed Manager

http://gwu-libraries.github.io/sfm-ui/

Social Feed Manager är en webbapplikation som kan samla in, utforska och analysera data från Twitter, Tumblr, Flickr, och kinesiska Sina Weibo. Social Feed Manager bygger på öppen källkod och utvecklas av ett team vid George Washington University.

### 4.9.6 Observatory on Social Media

http://truthy.indiana.edu/tools/

Observatory on Social Media, tidigare kallat Truthy, är ett projekt hos Indiana University som utvecklar webbtjänster för analys av sociala medier, bland annat automatisk klassificering av Twitterkonton som människa eller bot (robot som automatiskt skriver meddelanden). Ursprungligen utvecklade universitetet ett verktyg för att spåra spridandet av så kallade internetmem.

## 4.10 Analys av data från sociala medier

När data väl är insamlad är det möjligt att genomföra en mängd analyser. Nedan listas ett antal fria program för analys av sociala nätverk och texter (så kallad *natural language processing*).

### 4.10.1 GATE

https://gate.ac.uk/

Populärt program för textanalyser av olika slag som dessutom kan byggas ut med nya moduler. Utvecklat av University of Sheffield.

### 4.10.2 Gephi

https://gephi.org/

Program för att analysera och visualisera nätverk upp till hundra tusen noder (såsom personer). Gephi är öppen källkod.

### 4.10.3 NodeXL

http://nodexl.codeplex.com/

Program för analys av nätverk. Till skillnad från Gephi är det även möjligt att direkt via programmet samla in data från Facebook, Twitter,

YouTube och Flickr. NodeXL är öppen källkod och används tillsammans med Microsoft Excel. Utvecklat av Social Media Research Foundation i Kalifornien.

### 4.10.4 Opinion Finder

http://mpqa.cs.pitt.edu/opinionfinder/

Program för sentimentsanalys av texter.

### 4.10.5 OpenNLP

https://opennlp.apache.org/

Program för textanalys. Utvecklat av Apache Software Foundation.

### 4.10.6 Weka

http://www.cs.waikato.ac.nz/ml/weka/

Weka (Waikato Environemnt for Knowledge Analysis) är ett program som kan utföra flera typer av maskininlärningsalgoritmer för att hitta mönster i texter och därefter visualisera resultatet. Utvecklat av University of Waikato i Nya Zeeland.

## 4.11 Slutsats och rekommendationer

Syftet med litteraturöversikten var att (1) undersöka hur pass väl sociala medier speglar opinioner i samhället, (2) beskriva vilka metoder som finns för att analysera opinioner och händelser i sociala medier, samt (3) identifiera praktiska verktyg för sådana analyser.

### 4.11.1 Sociala medier är inte representativa för befolkningen

Slutsatsen är att sociala medier inte alls är representativa för befolkningen och speglar inte opinioner i befolkningen. Eventuella likheter är snarare att betrakta som tillfälliga sammanträffanden. Det beror framför allt på att sociala medier används av olika demografiska grupper, och att dessa grupper kan vara mer eller mindre aktiva. Analyser av sociala medier bör därför betraktas som analyser av vilka intressegrupper som kan skrika högst, åtminstone i politiska sammanhang. Antalet konton i sociala medier behöver heller inte stämma överens med antalet människor, eftersom en person kan ha flera konton liksom ett konto kan skötas av flera personer (såsom kundtjänsten på ett företag). Vi vet i dagsläget inte heller om sociala medier ens är representativa för internetanvändare.

### 4.11.2 Följ opinionsförändringar för att upptäcka påverkansförsök

De metoder som kan användas för att analysera sociala medier bygger antingen på analys av själva innehållet (exempelvis positiva eller negativa textmeddelanden) eller analys av metadata (exempelvis hur användarnas sociala nätverk är uppbyggda). För att undersöka opinioner är meddelandena huvudsakligen av intresse, och det kan även kompletteras med analys av användarnas sociala nätverk för att förstå vilka intressegrupper det är som för fram en specifik opinion.

Att analysera innehållet, och i synnerhet opinioner, i sociala medier kräver en mängd färdigheter och kunskaper. Dels teknisk färdighet med att hantera ett medielandskap i ständig förändring, dels metodkunskap för att värdera och hantera analysverktyg, och dels ämnesspecifik kunskap om hur människor tenderar att agera när det kommer till att

dela och sprida information. Att mäta opinionsförändringar (snarare än specifika nivåer av opinionen) är av huvudsakligt intresse och kan i bästa fall användas för att förutse påverkansförsök på opinionen.

### 4.11.3 Välja rätt verktyg kräver metodkunskaper

Vilket verktyg ska man då välja? De enda programmen på listan som är närmast komplett och samlar in meddelanden och exporterar data för en mängd typer av analyser är DMI-TCAT (för enbart Twitter), samt Social Feed Manager (för en mängd sociala medier), vilket aktivt underhålls av skaparna i nuläget. Resterande program handlar huvudsakligen om analys av redan insamlade data, och kräver också mer metodkunskaper.

Av dessa skäl är det därför viktigt att personer med erforderlig kompetens övervakar och kontinuerligt kan justera de program och system som används för att mäta opinionsförändringar eller identifiera kriser i sociala medier. De verktyg som finns på internet är många gånger gratis men resultatet av analyserna är i bästa fall kryptiska om man saknar kunskaper i hur informationen ska tolkas, eller om man saknar kunskap om teknikens möjligheter och begränsningar.

### 4.11.4 Låt extern part följa händelser och opinionsförändringar

Rekommendationen är således att låta en extern part med både teknisk och ämnesspecifik kompetens kontinuerligt bevaka och följa opinioner och händelser i sociala medier. Analysmetoderna och analysverktygen för sociala medier är i dagsläget inte tillräckliga för att ersätta traditionella metoder (såsom surveyundersökningar), utan bör snarare komplettera och understödja traditionella metoder. Det behövs också mer forskning för att säkerställa att de nya digitala metoderna är tillförlitliga och inte bara producerar tursamma lyckoträffar, och det är i dagsläget en forskningsuppgift som delvis redan bedrivs i stora delar av världen. Däremot kan metoderna och verktygen behöva anpassas efter svenska förhållanden.

# 5. Referenser

Al-garadi, M. A., Khan, M. S., Varathan, K. D., Mujtaba, G., & Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. Journal of Biomedical Informatics, 62, 1–11. https://doi.org/10.1016/j.jbi.2016.05.005

Andersson Schwarz, J., & Hammarlund, J. (2016). Kontextförlust och kontextkollaps: Metodproblem vid innehållsanalys av sociala medier. Nordicom-Information, 38(3), 41–55.

Andersson Schwarz, J., Hammarlund, J., di Grado, S., & Kjellberg, M. (2015). Opinioner och offentligheter online. Stockholm: Internetfonden.

Andreasson, M. (2014). "LORE methodological note 2014:20 Day of the week for survey dispatch: effects on participation rates". Göteborg: Göteborgs universitet, LORE.

Arkhede, S., Oscarsson, H. & Vernersdotter, F. (2017) Incentivizing the SOM surveys: estimating the effects on questionnaire response rates. LORE Methodological Note 2017:1. Göteborg: Göteborgs universitet, LORE.

Baca, C. T., Alverson, D. C., Manuel, J. K., & Blackwell, G. L. (2007). Telecounseling in rural areas for alcohol problems. Alcoholism Treatment Quarterly, 25(4), 31-45.

Bachmann, D. P., Elfrink, J., & Vazzana, G. (1999). E-mail and snail mail face off in rematch. Marketing Research, 11(4), 10.

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Groves, R. M. (2010). Research synthesis AAPOR report on online panels. Public Opinion Quarterly, 74(4), 711-781.

Bethlehem, J, Cobben, F and Schouten, B (2011). Handbook of Nonresponse in Household Surveys. Hoboken, NJ: Wiley.

Bethlehem, J. G., & Keller, W. J. (1987). Linear weighting of sample survey data. Journal of official statistics, 3(2), 141.

Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A., & Sudman, S. (2011). Measurement errors in surveys (Vol. 173): John Wiley & Sons.

Blair, J., Czaja, R. F., & Blair, E. A. (2013). Designing surveys: A guide to decisions and procedures: Sage Publications.

Brewitz, J., (2015). Utredning om metodik för webbpanelundersökningar. Retrieved from Örebro:

Brüggen, E., & Dholakia, U. M. (2010). Determinants of Participation and response effort in web panel surveys. Journal of Interactive Marketing, 24(3), 239-250.

Callegaro, M, Baker, R.P, Bethlehem, J, Goertiz, A, Krosnick, J, & Lavrakas, P (2014). Online Panel Research: A Data Quality Perspective. Wiley.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's Mturk, social media, and face-to-face behavioral testing. Computers in Human Behavior, 29(6), 2156-2160.

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2008). The effects of mode and format on answers to scalar questions in telephone and web surveys. I

Cohen, J. (1992). Statistical power analysis. Current directions in psychological science, 1(3), 98-101.

Curtin, R, Presser, S and Singer, E (2000). The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly 64: 413-428.

Curtin, R, Presser, S and Singer, E (2005). Changes in telephone survey nonresponse over the past quarter century. Public Opinion Quarterly 69 (1): 87-98.

Davidsson, P., & Findahl, O. (2016). Svenskarna och internet 2016. Stockholm: .SE (Stiftelsen för internetinfrastruktur). Hämtad från http://www.soi2016.se/

De Leeuw, D. (2005). To mix or not to mix data collection modes in surveys. Journal of official statistics, 21(2), 233.

Delli Carpini, M., & Keeter, S. (1996). What Americans know about politics and why it matters. New Haven: Yale University Press.

Dillman, D. A. (2011). Mail and Internet surveys: The tailored design method—2007 Update with new Internet, visual, and mixed-mode guide. Hoboken, New Jersey: John Wiley & Sons.

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. Field methods, 17(1), 30-52.

Dillman, D. A., & Smyth, J. D. (2007). Design Effects in the Transition to Web-Based Surveys. American Journal of Preventive Medicine, 32(5, Supplement), S90-S96.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys

using mail, telephone, interactive voice response (IVR) and the Internet. Social Science Research, 38(1), 1-18.

Echeverría, J., & Zhou, S. (2017). The `Star Wars' botnet with >350k Twitter bots. arXiv:1701.02405 [Cs]. Hämtad från http://arxiv.org/abs/1701.02405

Esaiasson, P., Gilljam, M., Oscarsson, H., Towns, A. & Wängnerud, L. (2017) Metodpraktikan: konsten att studera samhälle, individ och marknad. Wolters Klouwer

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. Computers in Human Behavior, 26(2), 132-139.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford: Stanford university press.

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An Experimental Comparison of Web and Telephone Surveys. Public Opinion Quarterly, 69(3), 370-392.

Fuchs, M, Bossert, D and Stukowski, S (2013). Response Rate and Nonresponse Bias - Impact of the Number of Contact Attempts on Data Quality in the European Social Survey. Bulletin de Me´thodologie Sociologique 117: 26-45.

Fukuhara, T., Nakagawa, H., & Nishida, T. (2007). Understanding Sentiment of People from News Articles: Temporal Sentiment Analysis of Social Events. Presenterad vid ICWSM 2007, Boulder, Colorado, USA.

Gayo-Avello, D. (2013). A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. Social Science Computer Review, 31(6), 649–679. https://doi.org/10.1177/0894439313493979

Groves, R. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly 70 (5): 646-75.

Groves, M. R. & Peytcheva, E. (2008) The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly 72(2), pp. 167-189.

Groves, R. (2011). Three Eras of Survey Research. Public Opinion Quarterly 75: 861-71.

Göritz, A. S., Reinhold, Nicole, Batinic, Bernad. (2002). Online panels. In B. Batinic, Ulf-Dietrich Reips, and Michael Bosnjak (Ed.), Online social sciences (pp. 27-47). Göttingen, Ger: Hogrefe & Huber.

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. The ANNALS of the American Academy of Political and Social Science, 659(1), 63–76. https://doi.org/10.1177/0002716215570866

Holmberg, S. (2016). Fortsatt opinionsstöd för att avveckla kärnkraften. I Ohlsson, J, Ekengren Oscarsson, H., & Solevid, M. (red.), Ekvilibrium (s. 267-276). Göteborg: Göteborgs universitet, SOM-institutet.

Holmberg, S. (2012). Fukushimaeffekten. I Bergström, A., Weibull, L., & Oscarsson, H. (red), I Framtidens skugga (s. 263-274). Göteborg: Göteborgs universitet, SOM-institutet.

Homola, J., Jackson, N., & Gill, J. (2016). A measure of survey mode differences. Electoral Studies, 44, 255-274.

Ilieva, J., Baron, Steve, Healey, Nigel M. (2002). Online surveys in marketing research: Pros and cons. International Journal of Market Research, 44(3), 361.

Iyengar, S. (1991). Is anyone responsible? How television frames political issues. Chicago: University of Chicago Press.

Johansson, B. (2010). Surveyundersökningar. I Ekström, M. och Larsson, L (red.), Metoder i kommunikationsvetenskap. Lund: Studentlitteratur.

Kalton, G., & Flores-Cervantes, I. (2003). Weighting methods. Journal of official statistics, 19(2), 81.

Kejariwal, A. (2014, oktober 24). Breakout detection in the wild. Hämtad från https://blog.twitter.com/2014/breakout-detection-in-the-wild

Keusch, F. (2015). Why do people participate in Web surveys? Applying survey participation theory to Internet survey data collection. Management Review Quarterly, 65(3), 183-216.

Kiesler, S., Sproull, Lee S. (1986). Response effects in the electronic survey. Public Opinion Quarterly, 50(3), 402-413.

Krasin, I., & Duerig, T. (2016, september 30). Introducing the Open Images Dataset. Hämtad från https://research.googleblog.com/2016/09/introducing-open-images-dataset.html

Kreuter, F, Muller, G and Trappmann, M (2010). Nonresponse and Measurement Error in Employment Research. Making Use of Administrative Data. Public Opinion Quarterly 74 (5): 880-906.

Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. Public Opinion Quarterly, 58(3), 381-399.

Ku, L.-W., Lee, L.-Y., & Chen, H.-H. (2006). Opinion extraction, summarization and tracking in news and blog corpora. I Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs.

Kunda, Z. (1990). The case for motivated reasoning. Psychological bulletin, 108(3), 480.

Larsson, A. O., & Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. New Media & Society, 14(5), 729–747. https://doi.org/10.1177/1461444811422894

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. Science, 343(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Legleye, S, Charrance, G, Razafindratsima, N, Bohet, A, Bajos, N and Moreau, C (2013). Improving Survey Participation: Cost Effectiveness of Callbacks to Refusals and Increased Call Attempts in a National Telephone Survey in France. Public Opinion Quarterly 77 (3).

Lepkowski & T. J., C., Brick, M., de Leeuw, E., Japec, L., Lavrakas, P., Link, M., Sangster, R. (red.), Advances in telephone survey methodology (pp. 250-275). New York: Wiley-Interscience.

Lewis, I., Watson, B., & White, K. M. (2009). Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participant responses to health-related messages. Australian Journal of Psychology, 61(2), 107-116.

Lewis, K. (2015). Three fallacies of digital footprints. Big Data & Society, 2(2). https://doi.org/10.1177/2053951715602496

Lodge, M., & Taber, C. S. (2013). The rationalizing voter: Cambridge University Press.

Luskin, R. C. (1990). Explaining political sophistication. Political Behavior, 12(4), 331-361.

Luskin, R. C. (2002). From denial to extenuation (and finally beyond): Political sophistication and citizen performance. Thinking about political psychology, 281, 305.

McFarland, D. A., & McFarland, H. R. (2015). Big Data and the danger of being precisely inaccurate. Big Data & Society, 2(2), 1–4.

Maragoudakis, M., Loukis, E., & Charalabidis, Y. (2011). A Review of Opinion Mining Methods for Analyzing Citizens' Contributions in Public Policy Debate. I Electronic Participation (s. 298–313). Springer, Berlin, Heidelberg. Hämtad från http://link.springer.com/chapter/10.1007/978-3-642-23333-3_26

Markstedt, E. (2014). "LORE methodological note 2014:3: The effect of reminders – first case". Göteborg: Göteborgs universitet, LORE.

Markstedt, E. (2014) Representativitet och viktning. Riks-SOM som spegel av det svenska samhället 1986-2013. SOM-rapport 2014:20. Göteborg: Göteborgs universitet, SOM-institutet.

Markstedt, E. & Martinsson, J. (2015) "Survey reminder timing – the effect of time between contact attempts", LORE methodological note 2015:6. Göteborg: Göteborgs universitet, LORE.

Markstedt, E. (2014). "LORE methodological note 2014:5 The effect of reminders – second case". Göteborg: Göteborgs universitet, LORE.

Markstedt, E. (2014). "LORE methodological note 2014:17 When to send e-mail survey invitations during the day."Mehta, R., Sivadas, Eugene. (1995). Comparing responses rates and response content in mail versus electronic mail surveys. Journal of the Market Research society, 37(4), 429-440.

Miller, J. M., & Krosnick, J. A. (1997). Anatomy of news media priming. In S. I. R. Reeves (Ed.), Do the Media Govern? Politicians, Voters, and Reporters in America (pp. 258-275). Thousand Oaks, CA: Sage.

Mitchell, A., & Hitlin, P. (2013, mars 4). Twitter Reaction to Events Often at Odds with Overall Public Opinion. Hämtad från http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/

Monti, C., Rozza, A., Zappella, G., Zignani, M., Arvidsson, A., & Colleoni, E. (2013). Modelling Political Disaffection from Twitter Data. I Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (s. 3:1–3:9). New York, NY: ACM. https://doi.org/10.1145/2502069.2502072

Moreno, M. A., Waite, A., Pumper, M., Colburn, T., Holm, M., & Mendoza, J. (2016). Recruiting Adolescent Research Participants: In-Person Compared to Social Media Approaches. Cyberpsychology, Behavior, and Social Networking. https://doi.org/10.1089/cyber.2016.0319

Murphy, J., Link, M. W., Childs, J. H., Tesfaye, C. L., Dean, E., Stern, M., … Harwood, P. (2014). Social Media in Public Opinion Research: Report of the AAPOR Task Force on Emerging Technologies in Public Opinion Research. American Association for Public Opinion Research. Hämtad från https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOR_Social_Media_Report_FNL.pdf

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2), 175.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1–135. https://doi.org/10.1561/1500000011

Odén, T., Djerf-Pierre, M., Ghersetti, M., & Johansson, B. (2016). Kriskommunikation 2.0. Allmänhet, medier och myndigheter i det digitala medielandskapet. Stockholm: MSB.

Paolo, A. M., Bonaminio, G. A., Gibson, C., Partridge, T., & Kallail, K. (2000). Response rate comparisons of e-mail-and mail-distributed student evaluations. Teaching and learning in medicine, 12(2), 81-84.

Peytchev, A (2013). Consquences of Survey Nonresponse. Annals of the American Academy of Political and Social Science 645 (1): 88-111.

Radziwill, N. M., & Benton, M. C. (2016). Bot or Not? Deciphering Time Maps for Tweet Interarrivals. arXiv:1605.06555 [Cs]. Hämtad från http://arxiv.org/abs/1605.06555

Roberts, C, Allum, N and Sturgis, P (2014). "Non-response and measurement error in online panels based on probability samples - are efforts to recruit reluctant panelists worth it?" In Online panel research: a data quality perspective, eds. Baker, M, Callegaro, R, Bethlehem, J, Göritz, A, Krosnick, J and Lavrakas, P. Wiley.

Robinson, D. (2016, augusti 9). Text analysis of Trump's tweets confirms he writes only the (angrier) Android half. Hämtad från http://varianceexplained.org/r/trump-tweets/

Rogelberg, S. G., & Stanton, J. M. (2007). Introduction: Understanding and Dealing With Organizational Survey Nonresponse. Organizational Research Methods, 10(2), 195-209.

Ryan, R. S., Wilde, M., & Crist, S. (2013). Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. Computers in Human Behavior, 29(4), 1295-1301.

Sánchez-Fernández, J., Muñoz-Leiva, F., & Montoro-Ríos, F. J. (2012). Improving retention rate and response quality in Web-based surveys. Computers in Human Behavior, 28(2), 507-514.

Sax, L. J., Gilmartin, Shannon K, Bryant, Alyssa N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. Research in higher education, 44(4), 409-432.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. Information Sciences, 311, 18–38. https://doi.org/10.1016/j.ins.2015.03.040

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.

Schouten, B, Cobben, F and Bethlehem, J (2009). Indicators for the representativeness of survey response. Survey Methodology 35 (1): 101-13.

Schuman, H., & Presser, S. (1996). Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. London: Sage.

Silge, J., & Robinson, D. (2017). Text Mining with R: A Tidy Approach. S.l.: O'Reilly Media. Hämtad från http://tidytextmining.com/

Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-Ended Questions in Web Surveys Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? Public Opinion Quarterly, 73(2), 325-337.

Starbird, K. (2017). Examining the Alternative News Ecosystem through Tweets about Alternative Narratives of Mass Shooting Events. Presenterad vid The International AAAI Conference on Web and Social Media (ICWSM-17), Montreal, Canada. Hämtad från http://faculty.washington.edu/kstarbi/Alt_Narratives_ICWSM17-CameraReady.pdf

Sturgis, P., Williams, J., Brunton Smith, I., & Moore, J. (2016) Fieldwork effort, responserate, and the distribution of survey outcomes: a multi-level meta-analysis. National Research Centre for Working Papers, 1/16. University of Southampton.

Strömbäck, J., & Johansson, B. (2017). När bjälken sitter i betraktarens ögon. Om fientliga medier-fenomenet. I Lars Truedson (red.) Misstron mot medier (s. 100-116). Stockholm. Institutet för mediestudier.

Strömbäck, J., & Johansson, B. (2007). Electoral Cycles and the Mobilizing Effects of Elections. The Swedish Case. Journal of Elections, Public Opinions and Parties. 17(1): 79-99.

Svenska Trender 2015. SOM-instiutet: Göteborgs universitet.

Särndal, C.E., & Lundström, S. (2005). Estimation in surveys with nonresponse: John Wiley & Sons.

Tavris, C., & Aronson, E. (2008). Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts. Houghton Mifflin Harcourt.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Varathan, K. D., Giachanou, A., & Crestani, F. (2016). Comparative opinion mining: A review. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.23716

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online Human-Bot Interactions: Detection, Estimation, and Characterization. arXiv:1703.03107 [Cs]. Hämtad från http://arxiv.org/abs/1703.03107

Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. Journal of Documentation, 69(3), 384–410. https://doi.org/10.1108/JD-03-2012-0027

Wu, S., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Who Says What to Whom on Twitter. I Proceedings of the 20th International Conference on World Wide Web (s. 705–714). New York, NY, USA: ACM. https://doi.org/10.1145/1963405.1963504

Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2015). Using Social Media to Enhance Emergency Situation Awareness. I Proceedings of the 24th International Conference on Artificial Intelligence (s. 4234–4238). Buenos Aires, Argentina: AAAI Press. Hämtad från http://dl.acm.org/citation.cfm?id=2832747.2832847

Yeager, D, Krosnick, J., Chang, L., Javitz, H., Levendusky, M., Simpser, A., & Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability. Public Opinion Quarterly, 75(4): 709–747.

Zaller, J. (1992). The nature and origins of mass opinion: Cambridge university press.

Zaller, J. (2002). The statistical power of election studies to detect media exposure effects in political campaigns. Electoral Studies, 21(2), 297-329.

Zielinski, A., Middleton, S. E., Tokarchuk, L. N., & Wang, X. (2013). Social-media text mining and network analysis to support decision support for natural crisis management. Presenterad vid ISCRAM 2013, Baden-Baden. Hämtad från http://eprints.soton.ac.uk/359364/

Myndigheten för samhällsskydd och beredskap
651 81 Karlstad Tel 0771-240 240 www.msb.se
Publ.nr MSB1169 - februari 2018 ISBN 978-91-7383-798-9