

Bruker norske revisorer konsistente og tilstrekkelige stikkprøvestørrelser når de tester ukentlig forekommende interne kontroller? En teoretisk analyse¹

Ellen M. Kulset

Førsteamanuensis, Universitetet i Sørøst-Norge og Norges Handelshøyskole

Sammendrag: Tidligere forskning viser at revisorer gjerne bruker tommelfingerregler foreslått av American Institute of Certified Public Accountants (AICPA) når de skal bestemme seg for stikkprøvestørrelser til bruk i forbindelse med test av ikke-frekvente interne kontroller. I Norge har Den norske Revisorforening (DnR) foreslått bruk av tilsvarende stikkprøvestørrelser. Jacoby og Hitzig (2011) kritiserer de amerikanske tommelfingerreglene for å bidra til utilstrekkelige og inkonsistente stikkprøvestørrelser. I dette kapittelet analyseres stikkprøvestørrelsene som benyttes av norske revisorer når de tester ukentlig forekommende kontroller. Basert på data innsamlet av Finanstilsynet i et tematisyn i 2016, finner vi at praksis med tanke på stikkprøvestørrelser synes svært ulik i de ulike revisjonsselskapene. Videre finner vi at stikkprøvepraksisen bare tidvis tilsier at det isolert sett hentes inn tilstrekkelig bevis fra stikkprøvene, og at stikkprøvestørrelsene synes inkonsistente når de sammenlignes med praksis knyttet til test av frekvente kontroller. Videre finner vi at omstendighetene stikkprøven gjennomføres i, tidvis kan tilsi at dagens praksis innebærer tilstrekkelige stikkprøvestørrelser.

Nøkkelord: stikkprøve, stikkprøvestørrelser, ukentlig forekommende interne kontroller, ISA 530

1 Forfatteren takker Jostein Lillestøl, Aasmund Eilifsen og William Messier for nyttige kommentarer knyttet til tidligere versjoner av dette kapittelet.

8.1 Innledning

American Institute of Certified Public Accountants (AICPA) foreslår i *Audit Sampling* (2014) ulike tommelfingerregler revisor kan benytte for å velge stikkprøvestørrelse ved revisjon av ikke-frekvente interne kontroller (kontroller som utføres ukentlig eller sjeldnere). I forbindelse med ukentlig forekommende kontroller foreslås for eksempel en stikkprøvestørrelse i intervallet 5 til 9. Den norske Revisorforening (DnR) foreslår at det kan benyttes stikkprøvestørrelser i intervallet 5 til 10 (Revisorforeningen, 2010). Elder mfl. (2013) oppsummerer tidligere forskning på området revisors stikkprøvepraksis og foreslår at det bør utføres ytterligere forskning knyttet til stikkprøvepraksis i både store og små selskaper og i ulike kontekster. Christensen mfl. (2015) responderer på dette forslaget ved å gjennomføre en studie på amerikanske data og finner blant annet at stikkprøvestørrelser i intervallet 5 til 9 benyttes av amerikanske revisorer når de reviderer ukentlig forekommende kontroller. Kulset (2019) analyserer norske revisorers stikkprøvepraksis i forbindelse med frekvente kontroller, men vi er ikke kjent med annen forskning som analyserer stikkprøvestørrelser benyttet i mindre selskaper eller i andre land enn USA.

Jacoby og Hitzig (2011) kritiserer tommelfingerreglene som foreslås av AICPA for å føre til stikkprøvestørrelser som er både utilstrekkelige og inkonsistente, og foreslår isteden bruk av stikkprøvestørrelser i intervallet 12 til 19 når det revideres ukentlig forekommende kontroller. Hensikten med dette kapitlet er å gjennomføre en teoretisk analyse av stikkprøvestørrelsene som benyttes av norske revisorer når de reviderer ukentlig forekommende kontroller, for å klargjøre når dagens praksis er passende.

Vi stiller følgende forsknings spørsmål: (1) Hvor mye bevis hentes inn ved hjelp av de stikkprøvestørrelsene som benyttes av norske revisorer i forbindelse med test av ukentlig forekommende kontroller? (2) Hvilke forutsetninger må være på plass for at denne stikkprøvepraksisen skal resultere i konsistente og tilstrekkelige stikkprøvestørrelser?

For å besvare spørsmålene benytter vi data samlet inn av Finanstilsynet i deres tematilsyn *Stikkprøver i revisjon* (2016). Svar på spørsmålene vil være av interesse for både revisjonspraktikere, forskere og regulerende myndigheter ved at det kan identifiseres behov for ny

revisjonspraksis, behov for bedre beslutningsstøtte og behov for ytterligere forskning.

I resten av kapittelet presenterer vi først relevant bakgrunnsinformasjon i delkapittel 8.2. Vår forskningsmetode presenteres i delkapittel 8.3 før vi presenterer studiens funn i delkapittel 8.4. I delkapittel 8.5 konkluderer vi samt diskuterer studies implikasjoner, bidrag og begrensninger.

8.2 Bakgrunn

8.2.1 Bruk av stikkprøver ved test av frekvente interne kontroller

Stikkprøver defineres i ISA 530 (International Standard on Auditing 530), pkt. 5a som:

Anvendelsen av revisjonshandlinger på mindre enn 100 % av enhetene i en populasjon som er relevant for revisjonen, under forutsetning av at alle enhetene har en mulighet til å bli valgt ut, for å skaffe revisor et betryggende grunnlag for å kunne trekke konklusjoner om hele populasjonen.

Helt siden den industrielle revolusjon har stikkprøver vært et viktig verktøy i revisors verktøykasse (Johansen, 2015). Men selv om moderne IT-basert teknologi i stadig større grad muliggjør testing av hele populasjoner, er det ikke grunn til å tro at revisors behov for å benytte stikkprøver vil forsvinne helt, i det minste ikke i den nærmeste fremtid (se blant annet Eilifsen et al., 2014, s. 259).

En av stikkprøvemethodene som revisor gjerne benytter, er attributtstikkprøver der man «søker å anslå andelen avvik fra en foreskrevet kontrollprosedyre i forhold til de tilfeller der kontrollen skal brukes» (Lillestøl, 1996, s. 17).

Når vi tester en intern kontroll, har vi fra et matematisk ståsted følgende situasjon: Vi har en populasjon som består av N enheter (antall ganger kontrollen skulle ha vært utført). I denne populasjonen er M enheter spesielle (antall ganger kontrollen ikke har blitt utført som intendert). Hvis revisor nå ønsker å bruke en stikkprøvebasert metodikk for å estimere M , kan han trekke en tilfeldig stikkprøve bestående av n enheter. La oss

anta at han finner Y spesielle enheter i stikkprøven. Basert på disse forutsetningene vet vi nå at Y er hypergeometrisk fordelt med parametrene (N, M, n) . Passende stikkprøvestørrelse n avhenger av hvor presist man ønsker at estimatet skal være (ønsket konfidensintervall), og ønsket konfidensnivå, se Kulset og Lillestøl (2020) for en nærmere fremstilling av faktorene som stikkprøvestørrelsen n avhenger av.

For å bestemme hva som er en passende stikkprøvestørrelse, kan revisor bruke statistiske tabeller, Excel eller ulike former for revisjonsprogramvare, for eksempel ACL eller IDEA. AICPAs tabeller A-1 og A-2 (2014) er eksempler på slike tabeller, og i disse tabellene finner vi stikkprøvestørrelsen som en funksjon av valgt konfidensnivå og presisjon, tabulert ved hjelp av tolererbar avviksrate (den høyeste populasjonsavviksraten revisor finner akseptabel hvis han skal konkludere med at kontrollen fungerer) og forventet avviksrate (avviksraten revisor forventer i populasjonen). I disse tabellene skal vi forstå begrepet konfidensnivå på følgende måte: Et konfidensnivå på for eksempel 95 prosent betyr at ved gjentatt bruk av metoden vil den faktiske avviksraten i populasjonen 95 av 100 ganger være lavere enn beregnet øvre grense (øvre konfidensgrense). Risikoen for feilaktig å akseptere en kontroll som fungerende når den faktisk ikke er det, kaller vi gjerne for beta-risiko. AICPAs tabeller A1 og A2 er konstruert slik at det valgte konfidensnivået i tabellene er lik 1 minus beta-risikoen. I resten av dette kapitlet vil vi definere konfidensnivå på denne måten. Vi tar altså ikke eksplisitt hensyn til alfa-risikoen ved valg av stikkprøvestørrelse (risikoen for feilaktig å forkaste en kontroll som ikke-fungerende når den faktisk fungerer). Implisitt tar tabellverket allikevel noe hensyn til alfa-risikoen via faktoren forventet avviksrate, se Kulset og Lillestøl (2020).

Vi ser av tabell 8.1 at dersom revisor forventer en avviksrate på null, tolererer en avviksrate på 10 prosent og ønsker et konfidensnivå på 90 prosent, så bør han/hun benytte en stikkprøvestørrelse på 22. Med de samme forutsetningene om forventet og tolererbar avviksrate, men med et ønsket konfidensnivå på 95 prosent, ser vi i tabell 8.2 at en passende stikkprøvestørrelse vil være 29. Hvis revisor tolererer en avviksrate på 20 prosent, vil en passende stikkprøvestørrelse være 11 med et konfidensnivå på 90 prosent og 14 med et konfidensnivå på 95 prosent. Tabell 8.1 og

8.2 er utarbeidet med hjelp av binomisk fordeling, og stikkprøvestørrelsene avhenger ikke av populasjonsstørrelsen. For små populasjoner ville eksakt beregning av stikkprøvestørrelsene ved hjelp av den hypergeometriske fordelingen gi en noe mindre stikkprøvestørrelse enn de som beregnes med utgangspunkt i den binomiske fordelingen.

Tabell 8.1 Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 90 prosent (hentet fra AICPA tabell A-2, 2014)

Forventet avvikshyppighet i populasjonen	Tolererbar avvikshyppighet									
	3 %	4 %	5 %	6 %	7 %	8 %	9 %	10 %	15 %	20 %
0,00 %	76	57	45	38	32	28	25	22	15	11
0,25 %	129	96	77	64	55	48	42	38	25	18
0,50 %	129	96	77	64	55	48	42	38	25	18
0,75 %	129	96	77	64	55	48	42	38	25	18
1,00 %	176	96	77	64	55	48	42	38	25	18
1,25 %	221	132	77	64	55	48	42	38	25	18

Tabell 8.2 Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 95 prosent (hentet fra AICPA tabell A-1, 2014)

Forventet avvikshyppighet i populasjonen	Tolererbar avvikshyppighet									
	3 %	4 %	5 %	6 %	7 %	8 %	9 %	10 %	15 %	20 %
0,00 %	99	74	59	49	42	36	32	29	19	14
0,25 %	157	117	93	78	66	58	51	46	30	22
0,50 %	157	117	93	78	66	58	51	46	30	22
0,75 %	208	117	93	78	66	58	51	46	30	22
1,00 %	*	156	93	78	66	58	51	46	30	22
1,25 %	*	156	77	78	66	58	51	46	30	22

* For stor til å være effektiv

Alternativet til at revisor benytter en stikkprøvestørrelse basert på statistisk teori, er at han/hun benytter sitt profesjonelle skjønn for å bestemme stikkprøvestørrelsen, jf. ISA 530, pkt. A4, men stikkprøvestørrelsen skal ikke avhenge av hvorvidt revisor benytter statistikk eller skjønn, jf. ISA 530, pkt. A11.

8.2.2 Bruk av stikkprøver i små populasjoner

Når revisor benytter stikkprøver for å teste kontroller i små populasjoner, dvs. kontroller som utføres ukentlig eller sjeldnere, har AICPA utviklet en egen tabell (AICPA 2014, tabell 3-5), se tabell 8.3, som revisor kan benytte. Ved testing av ukentlig forekommende kontroller foreslås for eksempel bruk av en stikkprøvestørrelse fra 5 til 9. DnR foreslår i sin *Veiledning: Dokumentasjon ved revisjon av små foretak* (2010) en stikkprøvestørrelse fra 5 til 10 for ukentlig forekommende kontroller.

Tabell 8.3 AICPA tabell 3-5 «Stikkprøvestørrelser i små populasjoner (Testing operating effectiveness of small populations)»

Kontrollfrekvens og populasjonsstørrelse	Stikkprøvestørrelse
Kvartalsvis (4)	2
Månedlig (12)	2-4
To ganger i måneden (24)	3-8
Ukentlig (52)	5-9

Christensen mfl. (2015) undersøker stikkprøvepraksisen i seks internasjonale amerikanske revisjonsselskaper (de fire store samt ytterligere to internasjonale selskaper) og finner blant annet at halvparten av disse selskapene oppgir at de bruker tabellen over, se tabell 8.3, for å fastsette stikkprøvestørrelser i forbindelse med test av ikke-frekvente kontroller. Stikkprøvestørrelsene som benyttes av de resterende selskapene, er også i samsvar med denne tabellen, selv om det ikke eksplisitt oppgis at tabellen benyttes.

Hvis vi benytter den hypergeometriske fordelingen til å beregne passende stikkprøvestørrelse for en populasjon på 52, finner vi for eksempel at med en tolererbar avviksrate på 10 prosent ($5/52$), ingen forventede avvik og en stikkprøvestørrelse på 5, gir dette et implisitt konfidensnivå på 41 prosent,² mens en stikkprøvestørrelse på 9 gitt de samme forutsetningene gir et implisitt konfidensnivå på 63 prosent. Sagt på en annen måte: Hvis vi planla med et konfidensnivå på 63 prosent, forutsatte ingen forventede

² Implisitt konfidensnivå er beregnet ved hjelp av Excel-formelen HYPGEOM.DIST. Implisitt konfidensnivå, x , er gitt som den største x slik at $x \leq 1 - \text{CHyp}(Y, n, M, N)$, hvor CHyp er kumulativ hypergeometrisk sannsynlighet. For $Y = 0$, $n = 0$, $M = 5$, $N = 52$ har vi da et implisitt konfidensnivå på 0,41.

feil og satte den tolererbare avviksraten til 10 prosent, ville 9 være en passende stikkprøvestørrelse, se tabell 8.4.³ Vi ser altså at stikkprøvestørrelser i intervallet 5 til 9 isolert sett gir liten sikkerhet. Hvis vi tolererer høyere avviksrater, for eksempel 19 prosent (10/52), vil vi oppnå høyere sikkerhet med et implisitt konfidensnivå fra 67 til 88 prosent, se tabell 8.4, men vi har fremdeles ikke oppnådd «høy grad av sikkerhet» slik dette defineres av AICPA som et konfidensnivå på 90 prosent eller høyere (AICPA, 2014, pkt. 3.42). Med utgangspunkt i forutsetninger om stikkprøvestørrelser i intervallet 5 til 9 og ingen forventede avvik vil vi ikke oppnå et så høyt implisitt konfidensnivå før vi har en tolererbar avviksrate på 20 prosent eller høyere, se tabell 8.4. En stikkprøvestørrelse på 10 vil akkurat gi et implisitt konfidensnivå på 90 prosent gitt nevnte forutsetninger.

Tabell 8.4 Implisitt konfidensnivå i en populasjon på 52 (1000 i de to kolonnene til høyre med en forventet avviksrate på null)*

Tolererbar avviksrate	Stikkprøvestørrelse							
	(Populasjon = 52)				(Populasjon = 1000)			
	1	2	5	8	10	20	15	60
6 %	6 %	11 %	27 %	40 %	48 %	78 %	60 %	98 %
10 %	12 %	18 %	41 %	58 %	67 %	92 %	79 %	≈ 100 %
19 %	21 %	35 %	67 %	84 %	90 %	≈ 100 %	96 %	≈ 100 %

* Beregningene er utført ved hjelp av den hypergeometriske fordelingen, se fotnote 2. På grunn av denne fordelings egenskaper er det rapportert tilnærmede konfidensnivåer.

Jacoby og Hitzig (2011) kritiserer stikkprøvestørrelsene i tabell 8.3, dvs. AICPAs tommelfingerregler, for å være utilstrekkelige og inkonsistente («deficient and inconsistent») (se s. 34) ettersom så små stikkprøvestørrelser medfører høy risiko for å ikke avdekke avvik samtidig som stikkprøvepraksisen blir inkonsistent ettersom sikkerhetsnivået man benytter i forbindelse med stikkprøver for frekvente og ikke-frekvente kontroller, avhenger av kontrollfrekvensen.

Amerikanske revisorer bruker gjerne et konfidensnivå på 90 prosent når de tester frekvente kontroller (Christensen et al., 2015), og Jacoby og

3 Implisitt konfidensnivå = $1 - \text{CHyp}(0, 9, 5, 52) = 0,63$, hvor CHyp er kumulativ hypergeometrisk sannsynlighet (0,37).

Hitzig (2011) foreslår at revisorer bør benytte de samme forutsetningene om ønsket konfidensnivå og forventet og tolererbar avviksrate uavhengig av frekvensen på kontrollene som testes, det vil si for eksempel planlegge med en tolererbar avviksrate på 10 prosent, en forventet avviksrate på null og et konfidensnivå på 90 prosent uavhengig av kontrollfrekvens. Dette tilsier at revisor bør benytte stikkprøvestørrelser i intervallet 12 til 19 i forbindelse med test av ukentlig forekommende kontroller. Stikkprøvestørrelser i dette intervallet innebærer at revisors beta-risiko («the auditors' risk of failure to detect any deviation») vil være fra 25 til 9 prosent, som gir implisitte konfidensnivåer i intervallet fra 75 til 91 prosent.

Avhengig av situasjon kan imidlertid en stikkprøve som er lavere enn den som foreslås av Jacoby og Hitzig (2011), være passende.

8.3 Metode og data

Finanstilsynet gjennomfører ulike typer tilsyn hos norske revisjonsselskaper, blant annet «kontroll av at revisjonsselskapene virker på hensiktsmessig og betryggende måte i samsvar med lovgivningen» (Finanstilsynet, 2017a). I forbindelse med tilsyn er revisjonsselskapene pliktige å gi alle de opplysninger tilsynet krever. Hvert år utfører Finanstilsynet blant annet tematisyn der de «kartlegger hvordan en problemstilling faktisk blir håndtert av revisjonsbransjen, for deretter å vurdere om Finanstilsynet mener det er behov for endringer» (Finanstilsynet, 2017b). Basert på avdekkede svakheter i tidligere tilsyn knyttet til revisors etterlevelse av ISA 530, gjennomførte Finanstilsynet i 2016 et tematisyn om bruk av stikkprøver for å skaffe «bredere bilde av hvordan ISA 530 er innarbeidet i revisjonsselskapenes metodikk og hvordan den etterleves» (Finanstilsynet, 2016). For å svare på forskningsspørsmålene bruker vi data knyttet til valg av stikkprøvestørrelser som Finanstilsynet samlet inn som en del av dette tematisynet. Data om stikkprøvepraksisen til 65 norske revisjonsselskaper ble samlet inn, dette inkluderer de fem største revisjonsselskapene (Big 4 pluss BDO). De 60 resterende revisjonsselskapene ble tilfeldig valgt blant de små og mellomstore selskapene.

I undersøkelsen inkluderer Finanstilsynet blant annet et eksempel der revisjonsselskapene blir bedt om å angi hvilken stikkprøvestørrelse de

vil benytte dersom de har «konstatert at en nøkkelkontroll i det reviderte foretaket er fornuftig utformet, at kontrollen er implementert, at revisor planlegger å bygge på kontrollen og derfor skal teste om den har fungert effektivt» (Finanstilsynet, 2016, s. 5). Nøkkelkontrollen blir utført henholdsvis (a) flere ganger daglig (det vil si ca. 2000 ganger i året), (b) én gang per dag (ca. 250 ganger i året) eller (c) én gang i uken (ca. 50 ganger i året). Finanstilsynet legger videre til grunn at det er ønskelig med moderat risiko/sikkerhet, uten at dette begrepet tallfestes.

For å svare på forskningsspørsmålene våre foretar vi en teoretisk analyse av de stikkprøvestørrelsene norske revisorer oppgir til Finanstilsynet at de benytter. Med teoretisk analyse vil vi i dette kapittelet forstå en analyse der vi benytter statistisk teori for å foreta nødvendige beregninger, samtidig som vi benytter to begreper (konsistent og tilstrekkelig) som sammenligningsgrunnlag i analysene vi utfører.

For å finne ut hvor mye bevis som hentes inn ved hjelp av stikkprøvestørrelsene som benyttes, bruker vi statistisk teori for å beregne hvor mye bevis (målt ved hjelp av implisitt konfidensnivå) norske revisorer henter inn med de stikkprøvestørrelsene som benyttes ved revisjon av ukentlig forekommende kontroller. Beregningene legger til grunn at det er ønskelig med en tolererbar avviksrate i intervallet 6 til 19 prosent og en forventet avviksrate på null.⁴

Når vi vurderer om norske revisorers stikkprøvepraksis knyttet til ukentlig forekommende kontroller er konsistent og tilstrekkelig, benytter vi samme definisjon på konsistent stikkprøvepraksis som Hitzig og Jacoby (2011). Et revisjonsselskaps stikkprøvepraksis er med dette konsistent dersom selskapet lar sine valg av forutsetninger om tolererbar og forventet avviksrate samt ønsket konfidensnivå være uavhengig av kontrollfrekvens. For å vurdere om stikkprøvepraksisen er konsistent, sammenligner vi den sikkerheten som norske revisorer implisitt planlegger stikkprøvene med for henholdsvis store og små populasjoner. For at en stikkprøvestørrelse skal gi tilstrekkelig bevis, må den enten være av en størrelse som gjør at det hentes inn tilstrekkelig bevis fra stikkprøven

4 Anekdotiske bevis tilsier at norske revisorer stort sett planlegger test av kontroll med en forventning om ingen avvik i populasjonen.

isolert sett, eller det må finnes omstendigheter som gjør at en stikkprøve som i seg selv ikke er tilstrekkelig til at revisor kan konkludere på en påstand, allikevel er å regne som passende. Dersom en stikkprøvestørrelse for passende forutsetninger om forventet og tolererbar avviksrate tilsier et implisitt konfidensnivå på 90 prosent eller høyere, vil vi legge til grunn at dette isolert sett er en tilstrekkelig stikkprøvestørrelse. For å svare på siste del av forskningsspørsmålet vil vi derfor se på hvilke selskaper som har en slik praksis, og videre diskutere i hvilke situasjoner norske revisorer på tross av eventuell lav stikkprøvestørrelse allikevel kan sies å ha benyttet et utvalg som er tilstrekkelig stort.

8.4 Funn og diskusjon

8.4.1 Valg av stikkprøvestørrelse med tilhørende sikkerhetsnivåer

Funnene i Finanstilsynets tematisyn knyttet til stikkprøvestørrelser fremkommer i tabell 8.5 under. Vi ser at norske revisorer oppgir at de benytter svært ulik stikkprøvestørrelse når de reviderer ukentlig forekommende kontroller, med angitte stikkprøvestørrelser i intervallet fra 1 til 50, avhengig av revisjonsselskap og ønsket sikkerhetsnivå, se tabell 8.5.

Tabell 8.5 Norske revisorers stikkprøvestørrelser ved kontrolltesting*

Sikkerhet	5 store		Øvrige	
	Moderat ⁵	Høy	Moderat	Høy
Kontrollfrekvens				
Ca. 2000 ganger i året	14-25	25-60	8-50	8-80
Ca. 250 ganger i året	10-25	15-25	5-42	10-60
Ca. 50 ganger i året	2-5	5-8	1-30	1-50

* Innholdet er hentet fra tabell 3 i Finanstilsynets rapport *Stikkprøver i revisjon. Tematisyn 2016* (2016). Finanstilsynet ba i utgangspunktet selskapene legge til grunn at det var ønskelig med moderat sikkerhet.

- 5 De fem store revisjonsselskapene oppga i sine svar hvilke stikkprøvestørrelser de benytter ved ønske om både høyt og moderat sikkerhetsnivå. For de mindre selskapene har Finanstilsynet tolket de innsamlede dataene slik at dersom det er oppgitt et stikkprøveintervall, benyttes den laveste stikkprøvestørrelsen ved ønske om moderat sikkerhet og den høyeste oppgitte størrelsen ved ønske om høy sikkerhet. Dette er ikke nødvendigvis en korrekt fortolkning ettersom det kan tenkes at alle eller noen av disse selskapene bruker en stikkprøvestørrelse som er høyere enn det oppgitte intervallet når de ønsker høy sikkerhet.

De fem største revisjonsselskapene bruker en stikkprøvestørrelse i intervallet fem til åtte hvis de ønsker høy grad av sikkerhet, og en stikkprøvestørrelse fra to til fem hvis de ønsker moderat grad av sikkerhet. Alle disse stikkprøvestørrelsene er lavere enn størrelsen som foreslås av Jacoby og Hitzig (2011). Stikkprøvestørrelsen på to er også lavere enn den som foreslås av både AICPA og DnR.

Avhengig av hvilke forutsetninger vi legger til grunn knyttet til forventet og tolererbar avviksrate, vil stikkprøvestørrelser fra fem til åtte kunne føre til implisitte konfidensnivåer fra 27 til 84 prosent, se tabell 8.4. Som vi har vist tidligere, er disse oppnådde sikkerhetsnivåene ikke i samsvar med høy grad av sikkerhet slik dette defineres av AICPA, som legger til grunn at høy grad av sikkerhet først oppnås ved konfidensnivåer på 90 prosent eller høyere.

Hvis det spørres etter bare moderat grad av sikkerhet, benytter et av de store norske revisjonsselskapene en stikkprøvestørrelse på to. Hvis det i denne situasjonen tolereres tre avvik i populasjonen (det vil si at det opereres med en tolererbar avviksrate på 6 prosent), ser vi fra tabell 8.4 at disse forutsetningene fører til et implisitt konfidensnivå på 11 prosent. Hvis vi tolererer ti avvik (det vil si en tolererbar avviksrate på 19 prosent), gir forutsetningene et implisitt konfidensnivå på 35 prosent. Disse konfidensnivåene synes ikke kompatible med et moderat sikkerhetsnivå.

De mindre revisjonsselskapene oppgir at de bruker både mindre og større stikkprøvestørrelser enn de større revisjonsselskapene, se tabell 8.5. Mindre revisjonsselskaper synes å ha en svært variabel praksis: Der ett av selskapene oppgir at det benyttes en stikkprøvestørrelse på én ved test av ukentlig forekommende kontroller, benyttes det for eksempel også stikkprøvestørrelser på 30 og 50 (det vil si at svært store deler av populasjonen på 52 testes, og det er grunn til å tro at spørsmålet enten er misforstått, eller at selskapet over-reviderer i slike situasjoner).

To av de små selskapene oppgir at det kan benyttes en stikkprøvestørrelse på så mye som 20, dvs. en stikkprøvestørrelse som er i overkant av intervallet som foreslås av Jacoby og Hitzig (2011). Denne stikkprøvestørrelsen impliserer at det planlegges med konfidensnivåer i intervallet 78 til tilnærmet 100 prosent, og den er i samsvar med AICPA's krav til høy

sikkerhet dersom det forutsettes en tolererbar avviksrate på 10 prosent eller høyere, se tabell 8.4. 23 av de små revisjonsselskapene oppgir at de tidvis benytter en stikkprøvestørrelse på 10 (implisitt konfidensnivå fra 48 til 90 prosent). Bruk av denne stikkprøvestørrelsen er ikke overraskende tatt i betraktning DnRs forslag fra 2010. De resterende små revisjonsselskapene i undersøkelsen har en stikkprøvepraksis som enten er i samsvar med praksis i de store revisjonsselskapene, eller som innebærer lavere stikkprøvestørrelser enn i de store selskapene.

Tabell 8.6 Maksimal stikkprøvestørrelse - ukentlig forekommende kontroller*

	Maksimal stikkprøvestørrelse	Antall
Store revisjonsselskap	8	2
	5	3
Små revisjonsselskap	20	2
	10	23
	8	4
	5	22
	Lavere	3

*Totalt antall revisjonsselskap i tabellen er 59. Totalt antall revisjonsselskap i Finanstilsynets undersøkelse er 65. De 6 små revisjonsselskapene som ikke er inkludert i tabellen svarte «test hele populasjonen» (1), «bruker profesjonelt skjønn» (2), foreslo bruk av større stikkprøvestørrelse når populasjonen ble mindre (det vil hadde trolig misforstått spørsmålet) (2) eller svarte ikke (1).

8.4.2 Konsistente stikkprøvestørrelser

I denne delen av kapittelet, har vi først fokusert på de store revisjonsselskapenes stikkprøvepraksis når det er ønskelig med høy grad av sikkerhet. I slike situasjoner benytter de store norske revisjonsselskapene en stikkprøvestørrelse fra fem til åtte. Det følger fra tidligere utregninger at dette, avhengig av forutsetninger knyttet til forventede og tolererbare avvik, gir implisitte konfidensnivåer på fra 27 til 84 prosent, se tabell 8.4.

Ifølge Finanstilsynet (2016) benytter norske revisorer i de samme revisjonsselskapene stikkprøvestørrelser fra 15 til 60 når de reviderer frekvente kontroller (det vil si når den interne kontrollen utføres én eller flere ganger per dag) dersom de ønsker en høy grad av sikkerhet, det vil

si at de planlegger med et konfidensnivå fra omtrent 60 (hvis vi antar ingen forventede avvik, tolererbare avvik på 6 prosent og n lik 15) til nær 100 prosent (hvis vi antar ingen forventede avvik, tolererbare feil på 19 prosent og n lik 60). Vi ser altså at med de samme forutsetningene knyttet til forventede og tolererbare feil, vil en stikkprøvestørrelse fra fem til åtte for ukentlig forekommende kontroller og en stikkprøvestørrelse på 15 til 60 for frekvente kontroller tilsi at det planlegges med et lavere konfidensnivå når man reviderer ukentlig forekommende kontroller enn når man reviderer frekvente kontroller. Dette impliserer en inkonsistent stikkprøvepraksis ettersom man systematisk henter inn mer bevis ved testing av frekvente kontroller enn ved tester av ukentlig forekommende kontroller.

Når det gjelder de mindre revisjonsselskapene, oppgir 23 at de kan benytte en stikkprøvestørrelse på 10 når de reviderer ukentlig forekommende kontroller. Dette impliserer et planlagt konfidensnivå på fra 48 til 90 prosent gitt forventede avvik lik null og tolererbare avvik i intervallet 6 til 19 prosent, jf. tabell 8.4. Den laveste stikkprøvestørrelsen som benyttes av disse selskapene når det testes frekvente interne kontroller, er 25, noe som impliserer et konfidensnivå i intervallet 78 til 99,5 prosent⁶ gitt samme forutsetninger om forventede og tolererbare feil som i tilfellet med ukentlig forekommende kontroller. Et selskap benytter en stikkprøvestørrelse på 20 for ukentlige kontroller (implisitt konfidensnivå på 78 til tilnærmet 100 prosent) og 50 for frekvente kontroller (implisitt konfidensnivå i intervallet 95 til tilnærmet 100 prosent). Vi ser at heller ikke disse revisjonsselskapene har en konsistent stikkprøvepraksis.

Resten av de små revisjonsselskapene i undersøkelsen benytter stikkprøvestørrelser i intervallet én til åtte. Vi ser fra tabell 8.4 at dette gir implisitte konfidensnivåer i intervallet 6 til 84 prosent. Den laveste stikkprøvestørrelsen som benyttes av disse selskapene ved test av frekvente kontroller, er 15, noe som gir implisitte konfidensnivåer i intervallet 60 til 96 prosent.

6 Ikke tabulert.

Tabell 8.7 Stikkprøvestørrelse som en funksjon av konfidensnivå og tolererbar avviksrate og en forventet avviksrate på null*

N = 52		Konfidensnivå		
Tolererbar avviksrate	50 %	70 %	90 %	
6 %	10	17	28	
10 %	5	9	16	
19 %	3	5	10	

N = 1000		Konfidensnivå		
Tolererbar avviksrate	50 %	70 %	90 %	
6 %	12	20	37	
10 %	7	12	22	
19 %	4	6	11	

*Beregningene er basert på den hypergeometriske fordelingen.

For å oppsummere ser vi at ingen av revisjonsselskapene har en konsistent revisjonspraksis slik vi definerer begrepet i dette kapitlet. Vi ser i tabell 8.7 at en slik konsistent praksis vil innebære at stikkprøvestørrelsen som benyttes i forbindelse med test av ukentlig forekommende kontroller, kan være noe lavere enn ved test av frekvente kontroller, men ikke så mye som den forskjellen som forekommer i praksis.

8.4.3 Når vil dagens stikkprøvestørrelser medføre/bidra til tilstrekkelig bevis?

Funnene i Finanstilsynets tematilsyn om norske revisorers stikkprøvepraksis viser at det er klare forskjeller mellom stikkprøvestørrelsene som benyttes i de mindre revisjonsselskapene og dem som benyttes i de store revisjonsselskapene. Men mens det ikke er noen av de store selskapene som benytter stikkprøvestørrelser med implisitte konfidensnivåer på 90 prosent eller høyere, er det 23 av de små selskapene som tidvis opererer med implisitte konfidensnivåer på 90 prosent eller høyere.

Lav stikkprøvestørrelse innebærer rent generelt at det isolert sett hentes inn lite bevis fra testen, og som det fremheves i *Audit Sampling Guide* er stikkprøvestørrelsene i tabell 8.3 ikke passende «if the internal control test is the sole source of evidence regarding this control» (AICPA, 2014, s. 43).

Stikkprøvestørrelsene som foreslås av AICPA, dvs. i intervallet fem til ni, kan imidlertid være passende i ulike situasjoner, og vi vil i det følgende presentere noen slike. (Husk at vi har sett at det øvre intervallet i DnRs tommelfingerregler, det vil si en stikkprøvestørrelse på 10, gir et konfidensnivå som tilsier at en slik test ikke nødvendigvis må kombineres med ytterligere bevis.)

For det første, og som nevnt i kapitlets innledning, kan revisor benytte profesjonelt skjønn når det foretas avgjørelser knyttet til stikkprøvestørrelser, jf. ISA 530, pkt. A4, men stikkprøvestørrelsen skal ikke avhenge av om det benyttes skjønn eller statistikkbaserte metoder, jf. ISA 530, pkt. A11. Det følger av dette at en stikkprøvestørrelse på åtte kan være passende hvis revisor ikke forventer noen avvik, tolererer omtrent 10 prosent avvik i populasjonen og ikke ønsker et høyere konfidensnivå enn ca. 60 prosent. For å hente inn tilstrekkelig bevis knyttet til den konkrete påstanden som revideres, må nå revisor bruke revisjonsrisikomodellen og *kombinere* de innhentede bevisene fra internkontrolltesten med bevis hentet inn ved hjelp av andre revisjonshandlinger. Vi fokuserer i denne studien ikke på i hvilken grad norske revisorer faktisk henter inn slikt tilstrekkelig revisjonsbevis ved å se bevis innhentet fra test av kontroller og andre typer tester i sammenheng. Dette spørsmålet overlater vi til senere forskning.

For det andre kan det som diskutert i Jacoby og Hitzig (2011), være slik at ikke-frekvente kontroller har *andre egenskaper* enn mer frekvente kontroller, og dette kan ha betydning for hvilke stikkprøvestørrelser det er passende å benytte ved test av disse kontrollene. Dersom internkontrollene er akkumulerende av natur, kan det for eksempel være tilstrekkelig å teste bare den siste gangen kontrollen utføres i løpet av året. Dette vil blant annet kunne gjelde kvartalsvise kontroller. Stikkprøvestørrelsene som foreslås i tabell 8.3, kan da faktisk være større enn det egentlig er behov for, og eventuelt større stikkprøvestørrelse vil kunne ha mer preg av å være kundeservice enn av å være nødvendig for å hente inn tilstrekkelige bevis. Hvorvidt den interne kontrollen er automatisert eller manuell, vil naturligvis også ha betydning for valg av stikkprøvestørrelse. Ved automatiserte kontroller vil det normalt være tilstrekkelig å teste kontrollene én eller noen få ganger ettersom slike kontroller vil fungere på samme måte hver gang dersom de er korrekt satt opp. Slik redusert testing krever

imidlertid at det er foretatt tilstrekkelig testing ved første gangs implementering, jf. ISA 330, pkt. A129. Dersom de generelle IT-kontrollene i et selskap ikke vurderes som tilfredsstillende, må imidlertid automatiserte kontroller testes som om de skulle være manuelle.

For det tredje vil en del interne kontroller i et selskap gjerne være ganske like eller helt like over en lengre tidsperiode. Dette tilsier at vi fra et statistisk ståsted kan argumentere for at det, med en ukentlig forekommende kontroll, ikke er nødvendig å ta utgangspunkt i en populasjon på 52, men i en populasjon som består av utførte kontroller over en lengre tidsperiode, for eksempel tre år, noe som vil gi en populasjonsstørrelse på ca. 150. Med en populasjon bestående av 156 enheter (tre år kombinert), et konfidensnivå på 90 prosent, ingen forventede avvik og en tolererbar avviksrate på 10 prosent er 22 en passende stikkprøvestørrelse, se tabell 8.1. Dette gir en gjennomsnittlig årlig stikkprøvestørrelse på 7, dvs. i midten av intervallet som foreslås av AICPA, jf. tabell 8.3. Følgende må være på plass dersom man skal gjennomføre en stikkprøve på denne måten med et konfidensnivå på 90 prosent: (1) Kontrollen i stikkprøven må være den samme, og den må utføres på samme måte i hele tidsperioden som populasjonen det trekkes fra, omfatter. (2) For at man skal bygge på egenskapene til den hypergeometriske fordelingen, må stikkprøven som trekkes være tilfeldig. Årsaken til at vi fremhever at stikkprøven vi benytter skal være trukket tilfeldig, er at de matematiske teoremene som ligger til grunn for stikkprøvestørrelsene i tabell 8.1 og 8.2, bare vil gjelde i tilfeldige utvalg. (3) Ettersom revisor utsteder sin revisjonsberetning på årlig basis, må revisor foreta nødvendige tilpasninger i revisjonsplanen slik at han/hun henter inn tilstrekkelig revisjonsbevis i år 1 og 2 ettersom han/hun i disse årene ikke kan bygge på revisjonsbevis innhentet i senere perioder (en stikkprøvestørrelse på 7⁷ er ikke kompatibelt med et konfidensnivå på 90 prosent). (4) Revisjonsplanen må være i samsvar med ISA 330, pkt. 13 og 14 om bruk av revisjonsbevis innhentet i tidligere perioder.

For å trekke et slikt tilfeldig utvalg på 21 i en populasjon som består av alle ukentlig forekommende kontroller av en bestemt type over en

7 En stikkprøvestørrelse på 7 gir et implisitt konfidensnivå på 50 prosent hvis forventet avviksrate er lik null og tolererbar avviksrate er lik 10 prosent.

periode på tre år, kan vi benytte en tilfeldig-nummer generator til å generere 21 tall fra 1 til 156, der hvert tall representerer en gang kontrollen skulle ha vært utført. For å oppnå ønsket konfidensnivå også i år 2 og år 3, kan vi gjøre som følger: Etter å ha generert 21 tall genererer vi tall fra 1 til 104 inntil vi har en stikkprøvestørrelse på 21 også i år 2, og til slutt genererer vi tall fra 1 til 52 inntil vi har oppnådd en stikkprøvestørrelse på 21 også i dette intervallet. I praksis har vi følgende problem med en slik stikkprøvepraksis: ISA 330 synes å begrense antall perioder som kan sees i sammenheng til tre, og hvordan skal vi da trekke en stikkprøve i periode fire som ikke ødelegger de tilfeldige egenskapene til stikkprøven?

8.5 Konklusjon, implikasjoner, bidrag og begrensninger

I dette kapittelet finner vi at små norske revisjonsselskaper synes å ha en mer uensartet stikkprøvepraksis enn større revisjonsselskaper når de reviderer ukentlig forekommende kontroller. Mens de store revisjonsselskapene oppgir at de anvender stikkprøvestørrelser fra to til åtte, oppgir et av de små revisjonsselskapene at de benytter en stikkprøvestørrelse på én, mens et annet av de små selskapene oppgir at det benytter en stikkprøvestørrelse på 50, noe som resulterer i at det hentes inn helt ulik mengde bevis fra disse testene i de ulike revisjonsselskapene. Implisitte konfidensnivåer vil som en konsekvens variere mye. Noen av stikkprøvestørrelsene som benyttes, synes å være i minste laget, for eksempel et av de store selskapenes bruk av en stikkprøvestørrelse på to når det ønskes moderat sikkerhet. Mens ingen av stikkprøvestørrelsene som benyttes av de store revisjonsselskapene, isolert sett synes å være tilstrekkelige, benytter en stor andel av de små revisjonsselskapene tidvis stikkprøvestørrelser som isolert sett kan sies å være tilstrekkelige (stikkprøvestørrelser på ti eller høyere). Videre finner vi at stikkprøvestørrelsene som benyttes av både små og store revisjonsselskaper, er inkonsistente ettersom det hentes inn mer bevis fra stikkprøver når disse benyttes i forbindelse med testing av frekvente interne kontroller, enn når de brukes ved testing av ukentlig forekommende kontroller. Hvorvidt stikkprøvestørrelsene bør være konsistente, er et spørsmål vi ikke konkluderer på.

En stikkprøvestørrelse i intervallet fem til åtte når man tester ukentlig forekommende kontroller, bidrar altså med lite bevisverdi isolert sett knyttet til påstanden som revideres. Stikkprøvestørrelser i dette intervallet kan allikevel benyttes når revisor er klar over hvilken grad av sikkerhet testen bidrar til, kombinert med at én eller flere av følgende omstendigheter er på plass: (1) Bevisene som hentes inn fra testen, kombineres med bevis som innhentes ved hjelp av en eller flere andre tester. (2) Kontrollen som testes, har spesielle egenskaper som tilsier at et lavere sikkerhetsnivå er tilstrekkelig, den er f.eks. en automatisert kontroll, den er akkumulerende i sin natur, osv. (3) Stikkprøven er en del av en større utvalgsbasert plan hvor kontroller utført over en lengre periode enn ett år slås sammen slik at de til sammen utgjør en større populasjon. Det følger av diskusjonene i dette kapitlet at stikkprøvestørrelsene som anvendes av norske revisorer når de tester ukentlig forekommende kontroller, kan gi revisor tilstrekkelig bevis for de aktuelle påstandene som testes, men ettersom stikkprøvestørrelser som anvendes ofte er relativt små, vil mer beslutningsstøtte kunne gi en mer passende stikkprøvepraksis.

De viktigste bidragene fra dette kapitlet er som følger: Dataene som analyseres, kommer fra både små og store revisjonsselskaper og er hentet inn i et annet land enn USA. De kompletterer på denne måten funnene i Christensen mfl. (2015). Videre viser analysene som utføres, at norske revisorer tidvis benytter stikkprøvestørrelser som er i laveste laget når de tester ukentlig forekommende kontroller. Dette er funn som er av interesse for norske revisjons-selskap så vel som for reguleringsmyndighetene ettersom det indikeres at norske revisorer tidvis kan ha behov for tilgang til ytterligere beslutningsstøtte for å bedre sin bruk av stikkprøver. Mens Jacoby og Hitzig (2011) kritiserer nåværende revisjonspraksis (det vil si bruk av stikkprøvestørrelser i intervallet 5 til 9) for å være både ikke-tilstrekkelig og inkonsistent, har denne studien også sett på situasjoner der dette kan være en god praksis. Denne innfallsvinkelen er i liten grad benyttet i tidligere forskning.

Analysene som utføres i dette kapitlet, er basert på hva norske revisjonsselskap har oppgitt til Finanstilsynet om egen praksis, det vil si hva som er nedfelt i de ulike revisjonsselskapenes manualer når det gjelder stikkprøvebasert testing av ukentlig forekommende kontroller. I praksis

vil den enkelte revisor i stor grad måtte foreta egne valg knyttet til hvorvidt en identifisert kontroll skal testes, og, dersom kontrollen skal testes, hvilket sikkerhetsnivå som er ønskelig (for eksempel høyt eller moderat). I noen revisjonsselskaper kan revisor etter å ha valgt sikkerhetsnivå i tillegg velge stikkprøvestørrelse fra et spesifisert intervall. Den enkelte revisor må altså i stor grad benytte eget profesjonelt skjønn. Dette innebærer at en rekke spørsmål er av interesse for videre forskning: Hvilke valg tar revisor i praksis knyttet til revisjonskundernes ukentlige utførte kontroller, og hvilke faktorer har innvirkning på disse valgene? Velger revisor å teste ukentlige kontroller i det hele tatt? I den grad revisjonsselskapet revisor jobber i, skiller mellom moderat og høy sikkerhet: Hvilket sikkerhetsnivå velger revisor å legge til grunn? I den grad revisor kan velge mellom ulike stikkprøvestørrelser innenfor et intervall: legger han/hun seg høyt eller lavt i dette intervallet? Vi overlater disse spørsmålene til senere forskning. Videre vil det være interessant å undersøke nærmere revisors praksis når det gjelder mulighetene til å benytte bevis innhentet i forbindelse med kontrolltesting i tidligere år, sammen med bevis innhentet i inneværende år.

Samtidig som den enkelte revisor tar egne valg, viser både funnene i Finanstilsynets rapport (2016) og tidligere forskning (Christensen et al., 2015) at stikkprøvestørrelsene som benyttes i den ulike revisjonsselskapene, varierer relativt mye når det gjelder revisjon av ukentlig forekommende kontroller. Dersom revisorer i revisjonsselskaper som systematisk benytter lavere stikkprøvestørrelser i forbindelse med kontrolltesting, ikke systematisk innhenter mer bevis fra andre typer tester enn revisorer som systematisk benytter høyere stikkprøvestørrelser, vil dette kunne innebære at revisjonsrisikoen på tvers av revisjonsselskap kan bli ulik. Hvordan revisor i praksis kombinerer bevis fra kontrolltesting og substanstesting både internt i det enkelte revisjonsselskap og på tvers av ulike revisjonsselskap, er dermed også et område som vil være av interesse for videre forskning.

Referanser

- AICPA (American Institute of Certified Public Accountants). (2014). *Audit sampling*. New York: AICPA.
- Christensen, B. E., Elder, R. J. & Glover, S. M. (2015). Behind the numbers: Insights into large audit firms sampling policies. *Accounting Horizons*, 29(1), 61–81.

- Eilifsen, A., Messier, W. F., Glover, S. M. & Prawitt, D. F. (2014). *Auditing and assurance services*. Berkshire: McGraw Hill Education.
- Elder, R. J., Akresh, A. D., Glover, S. M., Higgs, J. L. & Liljegren, J. (2013). Audit sampling research: A synthesis and implications for future research. *Auditing: A Journal of Practice & Theory*, 32, Suppl. 1, 99–129.
- Finanstilsynet. (2016). *Stikkprøver i revisjon. Tematilsyn 2016*. Hentet fra https://www.finanstilsynet.no/contentassets/d3e9938d72314f4f84db450629766128/stikkprover_-revisjon_tematilsyn_2016.pdf
- Finanstilsynet. (2017a). *Tilsyn revisor*. Hentet 28.04.2020 fra <https://www.finanstilsynet.no/tilsyn/revisor/>
- Finanstilsynet. (2017b). *Ulike tilsynsmetoder revisorer og revisjonsselskaper*. Hentet 28.4. 2020 fra <https://www.finanstilsynet.no/tilsyn/arkiv-fellessider-tilsyn/ulike-tilsynsmetoder-revisorer-og-revisjonsselskaper/>
- Jacoby, J. E. & N. B. Hitzig (2011). Auditing internal controls in small populations. *The CPA Journal*, December, 34–36.
- Johansen, Ø. K. (2015). *Verdien av tillit. Revisjonens historie*. Oslo: Kagge Forlag.
- Kulset, E. (2019). Revisors bruk av profesjonelt skjønn i forbindelse med stikkprøvebasert testing av frekvente interne kontroller. I L. Gårseth-Nesbakk, K. M. Baksaa & T. Gustavsen (Red.), *Trender og utfordringer i regnskap og revisjon*. Bergen: Fagbokforlaget.
- Kulset, E. & Lillestøl, J. (2020). Test av internkontroll ved hjelp av en stikkprøve: Kan revisor stole på kunders interne kontroller? *Regnskap og Revisjon*, 1, 42–49.
- Lillestøl, J. (1996). *Statistiske metoder i revisjon*. Oslo: Cappelen Akademisk Forlag.
- Den norske Revisorforening, (2010). *Veiledning: Dokumentasjon ved revisjon av små foretak*.
- Sullivan, J. (1992). Litigation risk broadly considered. I *Auditing symposium XI: Proceedings of the 1992 Deloitte and Touche/University of Kansas symposium on auditing problems*: 49–59. Lawrence, KS: The University of Kansas School of Business.
- The International Auditing and Assurance Standards Board / Den norske Revisorforening. (2009). *ISA 530 Stikkprøver i revisjon*. Norsk utgave av internasjonal revisjonsstandard 530. New York, NY: The International Federation of Accountants (IFAC).
- The International Auditing and Assurance Standards Board / Den norske Revisorforening. (2010). *ISA 330 Revisors håndtering av anslåtte risikoer*. Norsk utgave av internasjonal revisjonsstandard 330. New York, NY: The International Federation of Accountants (IFAC).