

Revisors bruk av profesjonelt skjønn i forbindelse med stikkprøvebasert testing av frekvente interne kontroller

Kulset, Ellen Hiorth Marthinsen^{1,2}

¹Institutt for regnskap, revisjon og rettsvitenskap - Norges Handelshøyskole

²Institutt for økonomi, historie og samfunnsvitenskap - Universitetet i Sørøst-Norge

Dette er den aksepterte versjonen av et bokkapittel publisert av Fagbokforlaget, og kan avvike fra den publiserte versjonen. Kapittelet er utgitt i *Trender og utfordringer i regnskap og revisjon* (2019), lenke til publisert versjon:

<https://www.fagbokforlaget.no/Trender-og-utfordringer-i-regnskap-og-revisjon/19788245025309>

Kulset, E. M. (2019). Revisors bruk av profesjonelt skjønn i forbindelse med stikkprøvebasert testing av frekvente interne kontroller. I L. Gårseth-Nesbakk, K. M. Baksaas, & T. Gustavsen (Red.), *Trender og utfordringer i regnskap og revisjon* (s. 151-164). Fagbokforlaget.

10. Revisors bruk av profesjonelt skjønn i forbindelse med stikkprøvebasert testing av frekvente interne kontroller

Sammendrag

Når revisor velger å bygge på en kundes interne kontroller, må disse kontrollene testes, for eksempel ved hjelp av stikkprøver. ISA 530 åpner for at valgene revisor tar ved slik testing baseres på skjønn. I dette kapittelet analyseres norske revisors valg av stikkprøvestørrelse ved testing av frekvente interne kontroller. Resultatene viser at skjønnsutøvelsen i stor grad samsvarer med internasjonal praksis og at små revisjonsselskap synes å ha en mer variert praksis enn store revisjonsselskap. Noen av stikkprøvestørrelsene som er i bruk er så lave at de tilsier bruk av et profesjonelt skjønn som ikke nødvendigvis er i samsvar med hva som kan forventes med utgangspunkt i statistisk teori.

10.1. Innledning

Revisors profesjonelle skjønn defineres i ISA 200, pkt. 13k (International Standard on Auditing 200), som «Anvendelsen av relevant opplæring, kunnskap og erfaring innenfor rammen av standarder for revisjon, regnskap og etikk når det fattes informerte beslutninger om hvilke handlingsplaner som er hensiktsmessige etter omstendighetene ved revisjonsoppdraget». Profesjonelt skjønn er en av hjørnesteinene i revisjonsfaget og spiller blant annet en viktig rolle når revisor henter inn bevis ved hjelp av stikkprøver. Revisors bruk av stikkprøver er regulert i ISA 530, og det følger av pkt. A9 at revisor selv avgjør hvorvidt han ønsker å gjennomføre en statistisk eller en ikke-statistisk stikkprøve, og at revisor tar denne beslutningen ved å bruke profesjonelt skjønn. I en statistisk basert stikkprøve, reguleres blant annet revisors valg av

stikkprøvestørrelse av statistisk teori. Dersom revisor velger å ikke gjennomføre en statistisk stikkprøve, kan han bruke profesjonelt skjønn når han bestemmer seg for passende stikkprøvestørrelse, men stikkprøvestørrelsen skal rent prinsipielt ikke avhenge av om revisor bruker skjønn eller statistisk teori, jf. ISA 530, pkt. A9 og A11. Hensikten med dette kapittelet er å bidra til en bedre forståelse av norske revisors bruk av skjønn i forbindelse med sin bruk av stikkprøver. Fokus vil særlig være på valg av stikkprøvestørrelse når revisor tester frekvente interne kontroller hos revisjonskundene.

Revisors bruk av stikkprøver i forbindelse med test av intern kontroll er tidligere særlig belyst på amerikanske data (se for eksempel Christensen, Elder og Glover 2015) og har i stor grad fokusert på praksis i de store revisjonsselskapene. Elder, Akresh, Glover, Higgs og Liljegren (2013) etterlyser derfor forskning på bruk av stikkprøver også i mindre selskaper. Det er ikke tidligere gjennomført norsk forskning på dette området. Finanstilsynet gjennomførte imidlertid et tematisyn om stikkprøver hos norske revisjonsselskaper i 2016, og i dette kapittelet analyseres data fra dette tematisynet for å svare på følgende forskningsspørsmål: *Hva kan norske revisors stikkprøvepraksis i forbindelse med valg av stikkprøvestørrelse ved test av frekvente interne kontroller fortelle oss om revisors utøvelse av skjønn, og hvilke konsekvenser kan denne skjønnsutøvelsen ha?*

I resten av dette kapittelet presenteres først bakgrunn for temaet som analyseres inklusiv resultater fra tidligere forskning. Deretter presenteres undersøkelsens design og funn før funnene fra analysedelen oppsummeres og betydningen av disse diskuteres. Revisjonsbransjen står i likhet med de fleste andre bransjer overfor store endringer i tiden fremover på grunn av pågående digitalisering av både kundenes regnskapsprosesser og revisors eget arbeid. Avslutningsvis diskuteres derfor revisors behov for og nytte av stikkprøver i tiden fremover.

10.2. Bakgrunn og tidligere forskning

10.2.1. Bakgrunn

Stikkprøver har tradisjonelt blitt regnet som et kraftfullt revisjonsverktøy ettersom man kan hente inn gode bevis om store populasjoner ved å kontrollere et relativt lite antall kontroller, transaksjoner, bilag, kontoer osv. Revisor anvender stikkprøver både når han skal kontrollere om en eller flere av et selskaps interne kontroller fungerer som ønsket (test av kontroller), og når han skal hente inn bevis om hvor store feil det er grunn til å anslå at befinner seg i et selskaps transaksjonsklasser eller balansekonti (detaljtesting).

For å bestemme hva som vil være en passende stikkprøvestørrelse (n), er det utviklet tabeller basert på tilnærming til binomiske sannsynligheter der n fastsettes basert på parametrene ønsket konfidensnivå, forventet avvikshyppighet og tolererbar avvikshyppighet, se for eksempel «Audit Sampling» (AICPA 2014). Av tabell 1 ser vi for eksempel at dersom vi ønsker et konfidensnivå på 90 %¹, en tolererbar avvikshyppighet på 5 % og ikke forventer noen avvik i populasjonen, vil det være passende med en stikkprøve på 45. Alternativt kan revisor benytte stikkprøvemodulen i en av revisjonsprogramvarepakkene som finnes på markedet, for eksempel ACL, IDEA osv. eller han kan benytte Excel.

[Sett inn tabell 1 her]

¹ Med et konfidensnivå på 90 % vil vi dersom vi anvender en test i vårt løpende revisjonsarbeid, konkludere korrekt på testen i gjennomsnitt 9 av 10 ganger som testen benyttes (Lillestøl 2000, side 150).

Når revisor henter inn bevis for ulike revisjonspåstander, kan han hente inn bevis gjennom tre ulike testtyper: risikovurderingshandlinger, test av kontroller og substanshandlinger.² Revisor velger hvilken kombinasjon av testing han finner hensiktsmessig ved hjelp av sitt profesjonelle skjønn og med utgangspunkt i de begrensninger som er nedfelt i ISA-ene (se blant annet ISA 330, pkt. 8, 18 og 21) for å oppnå en ønsket revisjonsrisiko.³ Revisor kan for eksempel ikke hente inn bevis for en påstand kun ved hjelp av en substanshandling dersom en slik substanshandling «alene ikke gir tilstrekkelig og hensiktsmessig revisjonsbevis», jf. ISA 330, pkt. 8b. For å styre revisjonsrisikoen benytter revisor «revisjonsrisikomodellen». I denne modellen er revisjonsrisiko en funksjon av risiko for vesentlig feilinformasjon og oppdagelsesrisiko⁴ slik at dersom risiko for vesentlig feilinformasjon er høy må revisor revidere med lavere oppdagelsesrisiko enn dersom risiko for vesentlig feilinformasjon er lav for å oppnå samme revisjonsrisiko. Videre er det slik at jo lavere oppdagelsesrisiko det er behov for, desto «mer» revisjon ved hjelp av substanshandlinger må revisor utføre. Rent konkret gir dette følgende sammenhenger:

- (1) Bruk av lavere stikkprøvestørrelse betyr at revisor må hente inn mer bevis ved hjelp av substanshandlinger enn dersom han hadde brukt en høyere stikkprøvestørrelse i forbindelse med kontrolltesten (gitt at revisor kan bygge på kontrollen) for å holde revisjonsrisikoen konstant. Dersom lavere stikkprøvestørrelse i forbindelse med kontrolltesting ikke kombineres med mer innhentet bevis fra substanshandlinger betyr dette at revisjonen utføres med en høyere revisjonsrisiko (underrevisjon).

² Substanshandlinger er definert i ISA 330 som «Revisjonshandling som er utformet for å avdekke vesentlig feilinformasjon på påstandsnivå» (ISA 330, pkt. 4a).

³ Revisjonsrisiko defineres i ISA 200, pkt. 13c som «risikoen for at revisor gir uttrykk for en uriktig mening i revisjonsberetningen når regnskapet inneholder vesentlig feilinformasjon».

⁴ Risiko for vesentlig feilinformasjon defineres i ISA 200, pkt. 13n som «risikoen for at regnskapet inneholder vesentlig feilinformasjon før det revideres». Oppdagelsesrisiko defineres i ISA 200, pkt. 13e som «risikoen for at revisjonshandlingene som utføres av revisor for å redusere revisjonsrisikoen til et akseptabelt lavt nivå, ikke vil avdekke eksisterende feilinformasjon som kan være vesentlig, enten alene eller sammen med annen feilinformasjon».

(2) Bruk av høyere stikkprøvestørrelse betyr at revisor kan hente inn mindre bevis ved hjelp av substanshandlinger enn dersom han hadde brukt en lavere stikkprøvestørrelse i forbindelse med kontrolltesten (gitt at revisor kan bygge på kontrollen) for å holde revisjonsrisikoen konstant. Dersom høyere stikkprøvestørrelse i forbindelse med kontrolltesting ikke kombineres med mindre innhentet bevis fra substanshandlinger betyr dette at revisjonen utføres med en lavere revisjonsrisiko (overrevisjon).

10.2.2. Tidligere forskning

I 1970- og 1980-årene var amerikanske revisjonsoppdrag preget av en utstrakt bruk av statistiske tester (Elder, Akresh, Glover, Higgs og Liljegren 2013) både i forbindelse med kontroll- og detaljtesting (se blant annet Akresh 1980). Etter hvert endret dette seg, og ifølge Elder, Akresh, Glover, Higgs, og Liljegren (2013) finner for eksempel Sullivan (1992) at revisorene i «big 6⁵» nærmest utelukkende bruker ikke-statistiske tester. I den senere tid har de store revisjonsselskapene igjen begynt å bruke stikkprøvestørrelser som har grunnlag i statistisk teori i forbindelse med test av kontroller (Elder, Akresh, Glover, Higgs og Liljegren 2013, side 112–113), men der

«the firms' sampling policies and practice aids simplify many of the judgments necessary to determine sample size. The input choices, in terms of levels of assurance, importance of the control, and expected deviation rate, may contain only a few choices (e.g., moderate or high assurance, moderate or high importance) to simplify the judgments and to improve consistency in the application of the sampling approach».

⁵ I 90-årene fantes det seks store internasjonale revisjonsselskaper, gjerne kalt for «big 6». Etter fusjoner i bransjen brukes i dag ofte begrepet «big 4» som omfatter selskapene PwC, KPMG, Deloitte og EY. Finanstilsynet benytter i sin rapport begrepet «5 store». I tillegg til selskapene som omfattes av benevnelsen «big 4» inkluderes da selskapet BDO.

Elder, Akresh, Glover, Higgs og Liljegren (2013) oppsummerer tidligere forskning innenfor temaet revisors bruk av stikkprøver («audit sampling»), og vi viser til denne artikkelen for en grundig gjennomgang av emnet.

Christensen, Elder og Glover (2015) studerer stikkprøvepolicyen til «big 4»-revisjonsselskapene samt to andre store internasjonale revisjonsselskaper ved hjelp av en surveybasert studie og undersøker blant annet hvilke parameterverdier som selskapene typisk benytter for å fastslå passende stikkprøvestørrelse. Parameterverdiene som oppgis, er angitt i tabell 2. Kun et av selskapene oppgir at de benytter konfidensnivåer lavere enn 90 %. Basert på disse parameterverdiene kan man beregne at revisorene benytter en stikkprøvestørrelse i intervallet 22 til 59 ved konfidensnivå lik 90 eller 95 %. Forskjellene i parameterverdier kan altså gi systematisk ulike stikkprøvestørrelser mellom revisjonsselskapene i undersøkelsen. Respondentene ble videre bedt om å oppgi gjennomsnittlig stikkprøvestørrelse for frekvente kontroller (definert i denne studien som kontroller som utføres daglig eller oftere), og man finner at svarene varierer relativt mye: et selskap oppgir intervaller fra 12-23 når man skal stole mye på kontrollen, ellers 9-16, et selskap oppgir stikkprøvestørrelse lik 25 uansett, og et selskap oppgir stikkprøvestørrelse i intervallet 25-60, avhengig først og fremst av kontrollens viktighet.

[Sett inn tabell 2 her]

10.3. Data og metode

Finanstilsynet gjennomfører ulike typer tilsyn hos norske revisjonsselskaper. En type tilsyn er «tematilsyn» der

Formålet er å kartlegge hvordan en definert problemstilling faktisk blir håndtert av revisorbransjen, [...] Etter tematilsynet er gjennomført, blir det utarbeidet en fellesrapport som offentliggjøres. Der oppsummerer Finanstilsynet de funn som er gjort, og gir uttrykk for om situasjonen anses tilfredsstillende på det aktuelle området (Finanstilsynet 2017).

Finanstilsynet gjennomførte i 2016 et tematilsyn hos 65 norske revisjonsselskaper inkludert de fem store, for å «få et bredere bilde av hvordan ISA 530 er innarbeidet i revisjonsselskapenes metodikk og hvordan den etterlevs. Formålet var også å øke revisorenes oppmerksomhet knyttet til bruk av stikkprøver i revisjonen» (Finanstilsynet 2016, side 2). Tematilsynet besto blant annet av en spørreundersøkelse der respondentene ble bedt om å svare på spørsmål knyttet til «hvordan de planlegger, gjennomfører og dokumenterer stikkprøvekontroller» (Finanstilsynet 2016, side 3). Finanstilsynet ba i sin undersøkelse revisjonsselskapene om å oppgi hvilken stikkprøvestørrelse de bruker dersom de ønsker et henholdsvis moderat eller høyt sikkerhetsnivå ved test av frekvente kontroller. Finanstilsynet definerer begrepet frekvente kontroller som kontroller som utføres mer enn 2 000 ganger per år eller kontroller som utføres daglig (250 ganger per år). Begrepene moderat eller høyt sikkerhetsnivå tallfestes ikke, og revisjonsselskapene blir ikke bedt om å oppgi hvilke forutsetninger de legger til grunn med tanke på forventet og tolererbar avvikshyppighet. Dataene knyttet til norske revisorerers stikkprøvepraksis som analyseres i dette kapittelet er hentet fra Finanstilsynets tematilsyn (2016).

Beslutningsanalyse (the study of judgment and decision making, JDM) handler ifølge Baron (2004) om «comparison of judgments to standards, standards that allow evaluation of the judgments as better or worse». For å undersøke hvordan revisor utøver skjønn ved testing av frekvente kontroller, gjennomføres i delkapittel 10.4. ulike sammenligninger. Først sammenlignes norsk stikkprøvepraksis med internasjonal praksis, deretter brukes statistisk teori

som norm for å evaluere noen av de stikkprøvestørrelsene norske revisorer benytter. I denne siste analysen har vi særlig fokus på noen av de minste utvalgene som norske revisjonsselskaper oppgir at de benytter, se tabell 3, ettersom disse representerer de beslutningene revisjonsselskapene oppgir at de tar i praksis som det er minst sannsynlig at er i samsvar med statistisk teori.⁶ Selve analysen gjennomføres ved at statistisk teori benyttes for å beregne «hvor mye» bevis revisor henter inn (målt som implisitt oppnådd konfidensnivå) når han benytter ulike stikkprøvestørrelser for ulike forutsetninger om tolererbar avvikshyppighet og en forventet avvikshyppighet lik null. Dette implisitte konfidensnivået sammenlignes så med revisjonsselskapenes informasjon om ønsket sikkerhetsnivå (konfidensnivå).

[Sett inn tabell 3 her]

10.4. Funn og analyser

Ved test av kontroller som utføres 2000 ganger per år bruker de 5 store revisjonsselskapene i Norge en stikkprøvestørrelse fra 14 til 60 avhengig av ønsket sikkerhet. Mindre revisjonsselskaper rapporterer mer variasjon med stikkprøvestørrelser i intervallet 8 til 80. Ved test av kontroller som utføres daglig (250 ganger per år), bruker de store revisjonsselskapene en stikkprøvestørrelse fra 10 til 25 mens de små benytter en stikkprøvestørrelse fra 5 til 60 avhengig av ønsket sikkerhet, se tabell 3.

10.4.1. Sammenligning med utenlandsk praksis

⁶ Anekdotiske bevis tilsier videre at noen av de største stikkprøvestørrelsene i tabell 3 benyttes i situasjoner der revisor legger til grunn en forventet avvikshyppighet større enn null.

De store norske revisjonsselskapene rapporterer om bruk av tilnærmet samme stikkprøvestørrelser som de store revisjonsselskapene i Christensen, Elder og Glovers (2015) studie for frekvente kontroller: for høy sikkerhet rapporterer de store norske revisjonsselskapene om stikkprøvestørrelser i intervallet 25 til 60, mens Christensen, Elder og Glover (2015) rapporterer om intervallet 22 til 59 ved forutsetningene konfidensnivå lik 90 %, tolererbar avvikshyppighet lik 10 % og forventet avvikshyppighet lik null, se tabell 2 og 3. Dette er ikke overraskende ettersom de globale revisjonsselskapene i utgangspunktet benytter en felles internasjonal metodikk i forbindelse med sine revisjonsoppdrag. Stikkprøvestørrelser fra 25 til 60 tilsvarer implisitt et konfidensnivå fra 90 % eller høyere hvis man ønsker en tolererbar avvikshyppighet i størrelsesorden fra 5 % til 10 %, se tabell 4.

[Sett inn tabell 4 her]

10.4.2. Statistisk teori som benchmark

Tabell 4 viser eksempler på hvilke stikkprøvestørrelser statistisk teori tilsier at revisor bør benytte dersom en intern kontroll utføres 2000 ganger per år og det forventes en feilfri populasjon, for ulike forutsetninger om ønsket konfidensnivå og tolererbar avvikshyppighet. Vi ser av denne tabellen at stikkprøvestørrelsene som norske revisorer benytter for test av frekvente kontroller i utgangspunktet er å anse som passende ved ulike forutsetninger om ønsket konfidensnivå og tolererbar avvikshyppighet. Dersom vi for eksempel legger til grunn at revisor tolererer en avvikshyppighet på 10 % (det vil si han vil konkludere med at kontrollen fungerer godt så lenge kontrollen utføres korrekt mer enn 90 % av alle de gangene den utføres) og forventer ingen avvik, oppnår revisor med en stikkprøvestørrelse på 10 et implisitt konfidensnivå på 60 % mens han oppnår et implisitt konfidensnivå på 95 % ved en

stikkprøvestørrelse på 29, jf. tabell 4. Ifølge AICPA er et konfidensnivå lik eller høyere enn 90 % kompatibelt med et *høyt* sikkerhetsnivå (AICPA 2014, pkt. 3.42). Et konfidensnivå på 60 % er altså langt unna AICPAs krav til konfidensnivå dersom man ønsker en høy grad av sikkerhet. Videre merker vi oss at i samsvar med revisjonsrisikomodellen, må revisor dersom han velger en stikkprøvestørrelse som gir lite sikkerhet, for eksempel ved at han benytter et konfidensnivå på 60 %, hente inn mye bevis fra andre tester (må kombinere detaljtester og analytiske substanshandlinger) dersom han skal ende opp med samme revisjonsrisiko som dersom han hadde valgt en stikkprøvestørrelse som gir større sikkerhet. Dersom han velger en stikkprøvestørrelse som gir en høy grad av sikkerhet, kan han for eksempel kanskje klare seg med å kun supplere kontrolltesten med en analytisk substanshandling.

Norske revisorer rapporterer at de benytter lavere stikkprøvestørrelser for test av kontroller som utføres 250 ganger per år enn for test av kontroller som utføres 2000 ganger per år, jf. tabell 3. Dette er i samsvar med statistisk teori ettersom tabellverket som tradisjonelt benyttes for å bestemme passende stikkprøvestørrelser i forbindelse med kontrolltesting gir i overkant store stikkprøver for mindre populasjoner (det vil si populasjonsstørrelse lavere enn ca. 500), se for eksempel Eilifsen, Messier, Glover og Prawitt (2014) side 285. Eilifsen, Messier, Glover og Prawitt (2014) foreslår derfor bruk av en korreksjonsformel⁷ for å oppnå en mer passende stikkprøvestørrelse når populasjonen stikkprøven trekkes fra er liten. Forutsetningene konfidensnivå lik 90 %, tolererbar avvikshyppighet lik 10 % og forventet avvikshyppighet lik null gir en stikkprøvestørrelse på 22 ved tabelloppslag, se tabell 1. Denne stikkprøvestørrelsen kan reduseres til 21 ved bruk av korreksjonsformelen ($22 \sqrt{1 - (22/250)} = 21$). Alternativt kan man benytte den hypergeometriske fordelingen istedenfor binomisk tilnærming. Dette vil

⁷ Vi kan benytte følgende korreksjonsformel: Stikkprøvestørrelse etter korreksjon = $n * \sqrt{1 - \left(\frac{n}{N}\right)}$ der n er stikkprøvestørrelsen før korreksjon og N er populasjonsstørrelse.

gi de statistisk korrekte stikkprøvestørrelsene, og vi finner ved bruk av den hypergeometriske fordelingen også en passende stikkprøvestørrelse lik 21.⁸

Når det gjelder en stikkprøvestørrelse på 15, som minst et av de store selskapene i Finanstilsynets rapport oppgir at de benytter for daglige kontroller, ser vi av diskusjonen over at høy sikkerhet (med forutsetning om 10 % tolererbar avvikshyppighet) krever en stikkprøvestørrelse på minimum 21. Implisitt betyr dette at selskapet opererer med et lavere konfidensnivå enn 90 % selv om det ønskes et høyt sikkerhetsnivå, alternativt at det tolereres en høyere avvikshyppighet enn 10 %, se tabell 4, og det synes dermed ikke å være en unison forståelse blant de store norske revisjonsselskapene om hvilke forutsetninger som skal legges til grunn når man ønsker et høyt sikkerhetsnivå. Christensen, Elder og Glover (2015) rapporterer at et av selskapene i deres undersøkelse oppgir en stikkprøvestørrelse på 12-23 for frekvente kontroller når det skal stoles mye på testen. Dette indikerer at det ikke bare er i norske revisjonsselskaper at det synes å herske en viss usikkerhet rundt hvilke stikkprøvestørrelser som er compatible med et høyt sikkerhetsnivå.

Revisorene i de mindre norske revisjonsselskapene benytter både mindre og større stikkprøver enn de som benyttes i de store revisjonsselskapene, jf. tabell 3. Den laveste stikkprøvestørrelsen som benyttes av små revisjonsselskaper når en intern kontroll utføres 2000 ganger per år for å oppnå høy sikkerhet, det vil si 8, tilsvarer et implisitt konfidensnivå på 60 % ved en tolererbar avvikshyppighet noe lavere enn 10 %, jf. tabell 4. Dette er ikke i samsvar med en naturlig oppfatning av et høyt sikkerhetsnivå og tilsier at det brukes en stikkprøvestørrelse som er altfor lav.

⁸ Vi finner n lik 21 ved å løse ut for minimum n for $CHyp(k_e, n, L_T, N) \leq \beta$ der k_e er forventet antall avvik i stikkprøven (0), n er stikkprøvestørrelsen, L_T er tolererbart populasjonsavvik (25), N er populasjonsstørrelse (250) og β er risiko (0,10), se Stewart 2012.

Når det gjelder daglige kontroller, er mønsteret knyttet til valg av stikkprøvestørrelser det samme som for mer frekvente kontroller: små selskaper rapporterer både om stikkprøvestørrelser som gir minst og størst grad av sikkerhet, jf. tabell 3. Den minste stikkprøvestørrelsen som rapporteres med et ønsket høyt sikkerhetsnivå er på 10, jf. tabell 3, og vi ser i tabell 4 at med forutsetninger om en forventet avvikshyppighet lik null og en tolererbar avvikshyppighet lik 10 % gir en stikkprøvestørrelse på 10 et tilhørende implisitt konfidensnivå på 60 %. Dette konfidensnivået er ikke i samsvar med AICPA's forståelse av hva som er et høyt sikkerhetsnivå (minimum 90 %). Det er overraskende at selskapet som rapporterer om mulig bruk av en stikkprøvestørrelse så lav som 8 for de mest frekvente kontrollene, synes å oppgi at de bruker en høyere stikkprøvestørrelse for mindre populasjoner. Dette er ikke i samsvar med statistisk teori.

10.5. Diskusjon og oppsummering

I denne undersøkelsen har vi analysert revisors bruk av profesjonelt skjønn når han tester frekvente interne kontroller. Når det gjelder de fem store norske revisjonsselskapene, har vi sett at skjønnsutøvelsen tilsvarende den som utøves internasjonalt ved at stikkprøvestørrelsene som benyttes i det store og hele er de samme.

Noen av stikkprøvestørrelsene som er i bruk, både i de små og de store revisjonsselskapene, se tabell 4, er imidlertid så lave at våre analyser tilsier bruk av et profesjonelt skjønn som ikke nødvendigvis er i samsvar med hva vi ville forvente med utgangspunkt i statistisk teori. For kontroller som utføres 250 ganger per år rapporteres det i gruppen «store revisjonsselskaper» for eksempel om stikkprøvestørrelser så lave som 15 ved et ønsket høyt sikkerhetsnivå.

Sammenlignet med i de store revisjonsselskapene åpnes det i de mindre revisjonsselskapene for både mindre og større stikkprøvestørrelser. Mindre stikkprøvestørrelser kan ha ulike årsaker og

ulike implikasjoner. Små revisjonsselskaper har for det første gjerne små kunder med mindre formalisert internkontroll enn store selskaper som det er mindre passende bygge på i revisjonsarbeidet. Dette kan tilsi at revisor tester et lite antall utførte kontroller og kombinerer denne testen med en mer omfattende detaljtest enn dersom kontrolltesten var mer omfattende. Dersom en slik mer omfattende detaljtest ikke utføres, vil imidlertid konsekvensen være at revisor reviderer med en høyere revisjonsrisiko. Dette kan også være en god løsning, men kun dersom revisor gjør dette bevisst i selskaper som har lav samfunnsmessig betydning. Det kan i denne sammenheng bemerkes at Finanstilsynet blant annet konkluderer med at de i sine kontroller har avdekket at «stikkprøvestørrelsen i flere tilfeller var betydelig lavere enn forventet uten at dette var begrunnet med andre revisjonshandlinger som i sum sammen med stikkprøvetesting skulle gi tilstrekkelig og hensiktsmessig revisjonsbevis. Dette funnet er klarest i tilfeller der metodikken gir liten støtte for revisors skjønnsutøvelse⁹», (Finanstilsynet 2016, side 14). Stikkprøvestørrelsene som minst et av de mindre selskapene benytter ved frekvente kontroller når det ønskes høy sikkerhet, er så lave at de implisitt innebærer et konfidensnivå på 60 % ved tolererbar avvikshyppighet lik 10 %. Dette gjelder for både høyfrekvente og noe mindre frekvente kontroller og er ikke i samsvar med en naturlig forståelse av hva et høyt sikkerhetsnivå er (jf. AICPA 2014).

På den annen side ser vi at noen av de små revisjonsselskapene også åpner opp for å bruke større stikkprøvestørrelser enn de som benyttes av de store revisjonsselskapene. På samme måte er det to hovedforklaringer som kan ligge bak: (1) revisor ønsker å bygge mye på internkontrollen og kombinerer en grundig test av kontroller med en mindre omfattende substanshandling eller (2) revisor reviderer den aktuelle påstanden med en lavere revisjonsrisiko (grundig test av kontroller fører ikke til at man reduserer type, tidspunkt og

⁹ Dette utsagnet retter seg spesielt mot de mindre revisjonsselskapene som ifølge Finanstilsynet har en metodikk som «gir liten støtte for revisors skjønnsutøvelse», Finanstilsynet 2016, side 14.

omfang av substanshandlingene). Med tanke på hvilken kundeportefølje de små revisjonsselskapene er i besittelse av, synes det imidlertid ikke naturlig at disse selskapene skal revidere med en lavere revisjonsrisiko enn de store revisjonsselskapene. Fremtidig forskning bør derfor undersøke nærmere hvorfor norske revisorers stikkprøvepraksis synes å variere mellom små og store revisjonsselskaper.

Det er kjent fra tidligere forskning (se for eksempel Christensen, Elder og Glover 2015) at valgene den enkelte revisor i de store revisjonsselskapene har mulighet til å ta ved valg av stikkprøvestørrelse, gjerne er svært begrenset (for eksempel valg mellom moderat eller høy sikkerhet) slik at praksis blir ensrettet på tvers av revisjonsoppdrag og -team. Som en konsekvens kan vi si at deler av det profesjonelle skjønn som det i ISA 530 er åpnet opp for at revisor kan benytte, er begrenset av revisjonsselskapene revisorene er ansatt i. Konsekvensen av denne ensrettingen kan være at de enkelte revisjonsoppdragene i realiteten ikke baseres på foretatte risikovurderinger ettersom man stort sett benytter samme stikkprøvestørrelse uansett. I hvilken grad revisorene i store revisjonsselskaper har mulighet til å velge andre stikkprøvestørrelser enn de som fremgår av selskapenes manualer, er et spørsmål som er interessant å belyse i videre forskning. Videre er det interessant å undersøke hvorvidt man i de mindre revisjonsselskapene i større grad åpner opp for at den enkelte revisor kan benytte sitt eget profesjonelle skjønn når han beslutter hvilke stikkprøvestørrelser han ønsker å benytte. Finanstilsynet påpeker jo også i sin rapport at noen av revisjonsselskapene som undersøkes synes å ha en metodikk som «gir liten støtte for revisors skjønnsutøvelse», (Finanstilsynet 2016, side 14).

Basert på analysene som er utført i dette kapitlet synes det å eksistere et behov for å undersøke hvorvidt norske revisjonsselskaper, og da særlig de mindre, i tilstrekkelig grad ser bevis innhentet fra ulike kilder i sammenheng slik at revisjonsoppdrag blir gjennomført med en passende revisjonsrisiko. Videre bør revisor i større grad enn i dag bør gjøres kjent med hvilke

stikkprøvestørrelser som kan benyttes for å oppnå *høy* sikkerhet ved testing av frekvente interne kontroller.

Den pågående digitaliseringsprosessen i revisjonsbransjen åpner opp for at revisor fremover i stadig mindre grad vil ha behov for å gjennomføre stikkprøver, da han i større grad enn tidligere vil kunne (1) teste enten et lite antall enheter i en populasjon ettersom datamaskiner utfører like handlinger likt hver gang eller (2) vil kunne teste alle enheter i en populasjon. Alternativ (1) er spesielt aktuell i forbindelse med test av kontroller, men revisor må tidvis bruke stikkprøver isteden da metoden bygger på at revisor kan stole på kundens generelle IT-kontroller (for eksempel fungerende tilgangskontroller). Anekdotiske bevis indikerer videre at man ved bruk av alternativ (2) i praksis gjerne sitter igjen med et relativt stort antall «mistenkelige» enheter som må kontrolleres, kanskje ved at man må gjennomføre en stikkprøve. Også i fremtiden vil det være naturlig å nyttiggjøre seg stikkprøver av andre årsaker, se for eksempel Eilifsen, Messier, Glover og Prawitt (2014); for det første vil revisor ha et behov for å kontrollere ulike manuelle kontroller som en kunde benytter. Videre kan det være behov for at revisor fysisk undersøker ulike eiendeler, blant annet i forbindelse med varetelling. Når det gjelder innhentede bevis fra tredjeparter, for eksempel saldoforespørsler, kan det også være naturlig å nyttiggjøre seg stikkprøvemethodikk. Avslutningsvis er det dermed viktig å poengtere at selv om revisor fremover trolig vil benytte færre stikkprøver enn i dag, vil stikkprøvemethodikk også fremover være metoder revisor bør beherske.

Referanser

AICPA. 2019. <https://www.aicpa.org/>

Akresh, A. 1980. Statistical sampling in public accounting. *The CPA Journal* 50 (7): 20-26.

American Institute of Certified Public Accountants. 2014. *Audit Sampling*. New York: AICPA

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler

& N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making*: 19–36.

London: Blackwell.

Christensen, B. E., Elder, R. J. og S. M. Glover. 2015. Behind the numbers: Insights into Large Audit Firms Sampling Policies. *Accounting Horizons*. Vol. 29, No. 1: 61-81.

Eilifsen, Aa, Messier, W. F., Glover, S. M. og D. F. Prawitt. 2014. *Auditing and assurance services*. Berkshire: Mc Graw Hill Education.

Elder, R. J., Akresh, A. D., Glover, S. M., Higgs, J. L. og J. Liljegren. 2013. Audit Sampling Research: A Synthesis and Implications for Future Research. *Auditing: A Journal of Practice & Theory*. Vol. 32, Suppl. 1: 99-129.

Finanstilsynet. 2016. Stikkprøver i revisjon. Tematilsyn 2016. Hentet 02.05.2019 fra <https://www.finanstilsynet.no/contentassets/d3e9938d72314f4f84db450629766128/stikkprover-revisjon-tematilsyn-2016.pdf>

Finanstilsynet. 2017. Ulike tilsynsmetoder revisorer og revisjonsselskaper. Hentet 02.05.2019 fra <https://www.finanstilsynet.no/tilsyn/arkiv-fellessider-tilsyn/ulike-tilsynsmetoder-revisorer-og-revisjonsselskaper/>

ISA 200 Overordnede mål for den uavhengige revisor og gjennomføringen av en revisjon i samsvar med de internasjonale revisjonsstandardene

ISA 330 Revisors håndtering av anslåtte risikoer

ISA 530 Stikkprøver i revisjon

Kock Johansen, Ø. 2015. *Verdien av tillit. Revisjonens historie*. Oslo: Kagge Forlag AS.

Lillestøl, J. 2000. *Statistiske metoder i revisjon*. Oslo: Cappelen Akademisk Forlag.

Stewart, T. R. 2012. *Technical Notes on the AICPA Audit Guide Audit Sampling*. New York: AICPA.

Sullivan, J. 1992. Litigation risk broadly considered. In *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.

Ubøe, J. 2012. *Statistikk for økonomifag*. Oslo: Gyldendal Akademisk.

Tabeller

Tabell 1 – Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 90 % (hentet fra AICPAs 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 2 – Amerikanske revisorers bruk av parameterverdier i forbindelse med test av kontroller *)

Respondent	Konfidensnivå	Forventet avvikshyppighet	Tolererbar avvikshyppighet	Beregnet stikkprøvestørrelse**)
1	60-90 %	0	10 %	10-22
2	90 %	0	10 %	22
3	90 %	0	6-9,5 %	23-49
4	90-95 %	0	5-10 %	22-59
5	Manglende numeriske svar			
6	90-95 %	0-1	6-10 %	22-49

*) Innholdet i tabellen (med unntak av siste kolonne) er hentet fra Christensen Elder og Glover (2015).

***) Beregnet stikkprøvestørrelse fremkommer ved oppslag i tabell A-1 og A-2 i AICPAs «Audit Sampling Guide» (2014) med parameterverdier som oppgitt i tabellen. Anekdotiske bevis tilsier at tallene 22 og 23 i praksis rundes av nedover til 20 eventuelt oppover til 25 og at tallene 49 og 59 i tabellen over i praksis rundes opp til henholdsvis 50 og 60.

Tabell 3 – Norske revisorers stikkprøvestørrelser ved kontrolltesting *)

Risiko/sikkerhet	5 store		Øvrige	
	Moderat	Høy	Moderat	Høy
Kontrollen utføres:				
Ca. 2000 ganger/år	14-25	25-60	8-50	8-80
Ca. 250 ganger/år	10-25	15-25	5-42	10-60
Antall aksepterte avvik	0-2	0-2	0-3	0-3

*) Innholdet er hentet fra tabell 3 i Finanstilsynets rapport «Stikkprøver i revisjon. Tematilsyn 2016»

Tabell 4 – Stikkprøvestørrelser ved ulike kombinasjoner av konfidensnivå og tolererbar avvikshyppighet ved en forventet avvikshyppighet lik null ved kontroller utført 2000 ganger per år

Tolererbar avvikshyppighet	Konfidensnivå			
	60 %*)	80 %*)	90 %**)	95 %***)
4 %	23	41	57	74
5 %	19	33	45	59
10 %	10	17	22	29
15 %	7	11	15	19
20 %	5	9	11	14

*) Stikkprøvestørrelsene er beregnet ved hjelp av revisjonsprogramvaren ACL

***) Stikkprøvestørrelsene er hentet fra AICPAs tabell A-2 (AICPA 2014)

****) Stikkprøvestørrelsene er hentet fra AICPAS tabell A-1 (AICPA 2014)