



HØGSKOLEN
I BERGEN

BERGEN UNIVERSITY COLLEGE

Effekt av simuleringstrening på sykepleieres ferdigheter og kunnskaper

Effect of simulation-based training on nurses' knowledge and skills

Pål André Hegland

Master i kunnskapsbasert praksis i helsefag

Senter for kunnskapsbasert praksis

Avdeling for helse- og sosialfag

15.05.15

Effekt av simuleringstrening på sykepleieres ferdigheter og kunnskaper

Effect of simulation-based training on nurses' knowledge and skills

Kandidat:

Pål André Hegland

Veileder:

Gro Jamtvedt

15.05.15

Antall ord:

13314

Forord

Simuleringstrening som læringsform har fascinert meg fra første gang jeg fikk være med som deltaker. Måten det hele er organisert på kombinert med muligheten for å lære i trygge omgivelser, gjorde meg interessert i temaet. Jeg ble også opptatt av rollen som fasilitator, der en fikk være med på å utvikle ulike scenarier og se hvordan deltakerne løste utfordringene. Når jeg nå fikk muligheten til å utforske emnet gjennom en masteroppgave var det et ypperlig valg for meg. Gjennom studiet har jeg likt systematiske oversikter fordi de har gitt verdifull informasjon. I tillegg har det virket spennende å få være med i den prosessen det er å utarbeide en slik. Da det viste seg at det var grunnlag for å lage en systematisk oversikt om simuleringstrening var jeg bestemt på hva jeg ville fordype meg i gjennom mitt masterprosjekt. Tidvis har tanken på et slikt prosjekt vært veldig skremmende og virket uoverkommelig, men det har vært utrolig lærerikt og spennende.

I løpet av studiet har jeg fått lov til å bli kjent med en utrolig variert og kunnskapsrik gjeng av medstudenter. Kullet MAKP 2011 har vært en sammensveiset gjeng som har vært opptatt av å hjelpe hverandre fremover. Spesielt har jeg samarbeidet med min medstudent Hege Aarlie, som i de fleste skriftlige arbeid har kommet med konstruktive og kunnskapsrike tilbakemeldinger. Videre har jeg fått viktig hjelp til å få innsikt i systematiske søk av Hilde Strømme og Malene Gundersen. Hilde har gitt tips og råd som har vært uvurderlige i prosessen med oversikten. Jeg er også veldig glad for samarbeidet med Kristine Tilestad, som har vært utviklende for meg. I prosessen med masteroppgaven har jeg vært så heldig å ha Gro Jamtvedt som veileder. Med sin oppmuntrende veilederstil har hun guidet meg med oppgaven, og jeg har kunnet dra nytte av hennes erfaring og faglighet.

Men som familiefar med tre små barn hadde jeg ikke kunnet gjennomføre dette uten en kone som har pushet meg. Marianne har sørget for at jeg har fått fred til å skrive i en travel hverdag. Jeg hadde ikke sittet der jeg er i dag uten hennes innsats på hjemmefronten.

Førde 15.05.15

Pål André Hegland

Effekt av simuleringstrening på sykepleieres ferdigheter og kunnskaper

Bakgrunn: Simuleringstrening er en utbredt strategi for å bedre kvalitet i helsevesenet. Imidlertid har ikke effekten av simuleringstrening for sykepleiere blitt dokumentert i en systematisk oversikt.

Hensikt: Formålet med dette prosjektet er å evaluere effekten av simuleringstrening på sykepleieres ferdigheter og kunnskaper.

Metode: Vi utførte søk i CDSR, DARE, HTA, CENTRAL, CINAHL, MEDLINE, Embase, ERIC og SveMed+ etter randomiserte kontrollerte studier (RCT) som evaluerte effekten av simuleringstrening for sykepleiere. Søkene var gjennomført i september 2014. Screening av abstracts og fulltekst, dataekstraksjon og vurdering av risiko for systematiske skjevheter ble utført av to personer individuelt. Vi utførte sammenligningene simuleringstrening versus ingen simulering, høyteknologisk simulering versus andre simuleringsformer og ulik organisering av simuleringstreningen. Data ble analysert i en meta-analyse og narrative synteser. GRADE ble benyttet til å vurdere kvaliteten på dokumentasjonen til hver sammenligning og utfall.

Resultat: Syv RCTer møtte inklusjonskriteriene. For sammenligningen simuleringstrening versus ingen simulering på sykepleieres ferdigheter sammenstilte vi effektestimater i en meta-analyse som viste SMD -1, 23 (KI -2, 69 til 0, 23). Det var imidlertid stor heterogenitet (I^2 92 %). Det var høy grad av variasjon i resultatene for de andre sammenligningene. Den samlede dokumentasjonen bak alle sammenligningene ble gradert til lav kvalitet.

Konklusjon: Effekten av simuleringstrening varierer betydelig internt i og mellom de inkluderte studiene. Meta-analysen viste usikker effekt av simuleringstrening sammenlignet med ingen simulering. De andre sammenligningene viste ikke-signifikante forskjeller og inkonsistens i effektestimater. Basert på våre funn kan vi ikke trekke noen sikker konklusjon på om simuleringstrening er en effektiv strategi for å øke sykepleieres ferdigheter og kunnskaper.

Sammendrag: Denne masteroppgaven består av en systematisk oversikt i artikkelform og en kappe knyttet til denne. Kappen er utdypning av metode, resultat og en utvidet diskusjonsdel i forhold til artikkelen.

Nøkkelord: Simulering, Sykepleier, RCT, Ferdighet, Kunnskap, Systematisk oversikt.

Effect of simulation-based training on nurses' knowledge and skills

Background: Simulation-based training is a widespread strategy to improve health-care quality. However the effectiveness on registered nurses has previously not been established in systematic reviews.

Aim: The aim of this systematic review is to evaluate effectiveness of simulation-based training on nurse's skills and knowledge

Method: We searched CDSR, DARE, HTA, CENTRAL, CINAHL, MEDLINE, Embase, ERIC and SveMed+ for randomised controlled trials (RCTs) evaluating effect of simulation-based training among nurses. Searches were completed in September 2014. Screening of abstracts and full-text, data extraction and assessment of risk of bias were independently conducted by two reviewers. We made the comparisons simulation-based training versus no simulation, high-fidelity simulation versus other simulation strategy and different organisation of simulation training. Data were analysed through meta-analysis and narrative syntheses. GRADE was used to assess quality of evidence.

Results: Seven RCT's met the inclusion criteria. For the comparison simulation-based training to no simulation on nurses' skills we pooled effect sizes using a random effects model. Our meta-analysis showed no significant differences between the groups (SMD -1.23, CI -2.69 to 0.23). There was large heterogeneity (I^2 92 %). For the other comparisons there was large between-study variation in results. The quality of evidence for all comparisons was graded as low.

Conclusion: The effect of simulation-based training varies substantially between the included studies. Our meta-analysis showed no significant difference of simulation training compared to no simulation, but the quality of evidence was low indicating uncertainty. Other comparisons showed inconsistency in results. Based on our findings we cannot conclude if simulations-based training increases nurses' skills and knowledge.

Summary: This master's theses consist of a systematic review and an introduction section. The introduction is an elaboration of the method used, results and discussion of the systematic review.

Keywords: Simulation, Technology-enhanced, Nurses, RCT, Knowledge, Skill, Systematic Review.

Innholdsfortegnelse

Bakgrunn	8
Metode.....	13
Inklusjons og eksklusjonskriterier	13
Sammenligninger.....	15
Litteratursøk	16
Utvelgelse av studier	18
Dataekstraksjon	19
Kritisk vurdering	19
Gradering av dokumentasjonen.....	22
Analyse og sammenstilling av data	23
Resultater.....	25
Litteratursøk og utvelgelse	25
Beskrivelse av inkluderte studier.....	26
Beskrivelse av ekskluderte studier	28
Risiko for systematiske skjevheter	28
Analyse, sammenstilling og gradering	31
Effekt av simuleringstrening versus ingen simulering	31
Effekt av høyteknologisk simulering versus andre simuleringstrategier	34
Effekt av ulik organisering av simuleringssesjonen	37
Diskusjon.....	39
Drøfting av oversiktens resultater	39
Resultatenes sammenheng med tidligere publiserte oversikter	40
Betydning for forskning	41
Betydning for praksis	41
Styrker og svakheter ved den systematiske oversikten	42
Konklusjon	48
Referanseliste	50
Effect of simulation-based training on nurses' knowledge and skills – A systematic review and meta-analysis	
Appendix 1 – 7 Search strategies per database	
Appendix 8 Excluded studies	
Appendix 9a GRADE Evidence profile	
Appendix 9b GRADE Summary of findings table	
Appendix 9c GRADE Summary of findings table	
Appendix 10 Forest Plot comparisons	
Vedlegg 1 Søkeord PIO	
Vedlegg 2a-g RoB-skjema	

Oversikt over figurer og tabeller i oppgaven

Kappe:

Tabell 1 PICOS

Tabell 2 Karakteristika av inkluderte studier

Tabell 3 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av simuleringstrening versus ingen simulering

Tabell 4 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av høyteknologisk simuleringstrening versus andre simuleringstrategier

Tabell 5 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av ulike organisering av simuleringstreningen

Figur 1 Flytdiagram utvelgelse av studier

Figur 2 Risiko for systematiske skjevheter

Figur 3 Risk of bias graf

Figur 4 Meta-analyse simuleringstrening versus ingen simulering, utfall ferdigheter

Figur 5 Sensitivitetsanalyse simuleringstrening versus ingen simulering, utfall ferdigheter

Figur 6 Simuleringstrening versus ingen simulering, utfall kunnskaper

Figur 7 Høyteknologisk simulering versus andre former for simulering, utfall ferdigheter

Figur 8 Høyteknologisk simulering versus andre simuleringstrategier, utfall kunnskaper

Figur 9 Ulik organisering av simuleringstreningen, utfall ferdigheter

Artikkel:

Table 1 Characteristics of included studies

Table 2 Summary of findings effectiveness of simulations-based training compared to no simulation

Table 3 Summary of findings effectiveness of high-fidelity simulation compared to other simulation strategy

Table 4 Summary of findings effectiveness of different organisation of simulation training

Figure 1 Flow-chart screening and inclusion of studies

Figure 2 Risk of Bias in included studies

Bakgrunn

Helsevesenet tilbyr stadig mer kompleks utredning og behandling til pasienter. Pasientgrupper som tidligere ikke hadde behandlingstilbud, kan i dag motta kurerende eller livsforlengende tiltak. Dette skyldes i stor grad utvikling av avansert medisinsk-teknisk utstyr slik som avanserte respiratorer og dialysemaskiner. Sammen med nye behandlingsformer stiller dette store krav til kunnskap og ferdigheter hos helsepersonell (Stubberud, 2010, s. 32).

Kvalitet på helsetjenesten og sikkerhet for pasienten har sterkt fokus i dagens helsevesen, både nasjonalt og internasjonalt. Feil som begås av helsepersonell vies stor oppmerksomhet blant myndigheter og samfunn for øvrig (Helse- og Omsorgsdepartementet, 2012 ; Aspden et al., 2004, s. 17-20). Slike feil kan være knyttet til administrering av medisiner, mangelfull koordinering slik at pasienten ikke får nødvendig undersøkelse, eller at helsepersonellet mangler kunnskap om den behandlingen som gis (Forskrift om håndtering av medisinsk utstyr, 2014 ; Institute for Healthcare Improvement, 2015). Studier viser at behandling pasienter mottar kan variere fra sykehus til sykehus uten noen åpenbar forklaring, og at den beste tilgjengelige forskningen i en del tilfeller ikke ligger til grunn for de tiltakene pasientene mottar (Flottorp & Aakhus, 2013). Videre viser undersøkelser at en stor del av uønskede hendelser i helsevesenet kan forebygges dersom det drives systematisk kvalitetsforbedringsarbeid (Helsedirektoratet, 2011 ; Sosial- og Helsedirektoratet, 2005). Det er dermed en sterk sammenheng mellom kvalitet på tjenestene som tilbys, og sikkerheten til pasientene som mottar behandlingen (Grol & Wensing, 2013b, s. 6-7). Kvalitet i arbeidet kjennetegnes i kvalitetsstrategien til Sosial- og Helsedirektoratet (2005) av at faren for å begå feil er redusert til et minimum. Videre fremheves at kvalitet i helsetjenesten innebærer at tjenestene er virkningsfulle og trygge, involverer brukerne, er preget av kontinuitet ved å være samordnet, utnytter ressurser på en god måte og er rettferdig fordelt. I tillegg skal pasientsikkerhet være ivaretatt gjennom systematisk kvalitetsforbedringsarbeid (Sosial- og Helsedirektoratet, 2005).

Kvalitetsforbedring er systematisk arbeid for å avdekke svakheter ved tjenesten samt tiltak for å forbedre kvaliteten på tjenestene (Nasjonalt kunnskapsenter for helsetjenesten, 2014 ; Grol, Ouwens & Wollersheim, 2013, s. 65 ; Batalden & Davidoff, 2007). Batalden og Davidoff (2007) har definert kvalitetsforbedring som en kontinuerlig prosess der pasienter, helsepersonell, forskere, utdanningsinstitusjoner og ledelse jobber målrettet for å bedre utfall for pasienter, gi mer effektive tjenester og bedre læringsprosessene. Kvalitetsforbedring

krever dermed at alle nivå i helsetjenesten jobber målrettet for å tilby sikre tjenester. Dette innebærer at gode rammer må være til stede fra myndigheter og ledelse for at den enkelte helsearbeider kan utføre god praksis. I tillegg er det avgjørende at den enkelte helsearbeider er opptatt av kvalitet i sitt eget arbeid (Grol, Ouwens & Wollersheim, 2013, s. 65-74). Flottorp og Aakhus (2013) beskriver en modell kalt «Fra kunnskap til handling» først beskrevet av Graham med kolleger (Graham et al., 2006). Modellen beskriver prosessen som starter med å identifisere problemet og videre kunnskap. Kunnskapen tilpasses, før en vurderer hva som kan hemme eller fremme kunnskapen. Tiltak skal så skreddersys og iverksettes, før en videre monitorerer og vurderer resultatene. Modellen minner i stor grad om Kunnskapssenterets modell for kvalitetsforbedring som viser den kontinuerlige prosessen forbedringsarbeid er (Nasjonalt kunnskapssenter for helsetjenesten, 2014). Kvalitetsforbedring er ofte referert til som endringsarbeid eller kvalitetsprogrammer og har som mål om å bedre tjenestene som ytes. Kvalitetsprogram kan være lokale og knyttet til ett sykehus eller ett helseforetak, eller det kan være landsomfattende i form av nasjonale direktiv eller retningslinjer (Sosial- og Helsedirektoratet, 2005).

Systematisk kvalitetsforbedringsarbeid i helsetjenesten er krevende, og betinger ulike tiltak tilpasset settingen der endringene skal finne sted. Det er derfor sentralt at implementeringstiltakene som iverksettes er kunnskapsbaserte, som betyr at en benytter tiltak som gjennom forskning har vist seg å være effektive (Grol & Wensing, 2013b, s. 14-15 ; DiCenso, Guyatt & Ciliska, 2005, s. 14 ; Glasziou, Ogrinc & Goodman, 2011 ; Flottorp & Aakhus, 2013). Det finnes ulike teorier om hvordan endring bør initieres i helsevesenet, slik som kognitiv teori, teori om læring og motivasjonsteori. I tillegg finnes teorier som beskriver ledelsesfaktorer, sosiale faktorer, kontekst samt transteoretisk endringsmodell (Grol et al., 2013, s. 21-36 ; Nasjonalt kunnskapssenter for helsetjenesten, 2014 ; Flottorp & Aakhus, 2013). Disse ulike teoriene gir innsikt i hvordan endring kan skje i praksis. Teoriene skal således skape forståelse for prosessene, som igjen skaper grunnlag for forskning omkring hvordan teoriene virker i praksis. Studier utført på implementering viser at endring i komplekse systemer, slik som i helsevesenet, krever en tilnærming basert på flere ulike teorier. Det er også viktig at ansvarlige for endringsarbeid kjenner til de ulike teoriene slik at tiltakene er tilpasset konteksten (Grol et al., 2013, s. 19-21). Videre skal tiltakene være kunnskapsbaserte, slik at tilnærmingen en velger har vist seg å være effektiv i sammenlignbare settinger (Flottorp & Aakhus, 2013).

Kunnskapsbasert praksis er videre sammensatt av flere dimensjoner, definert som forskningsbasert kunnskap, erfaringsbasert kunnskap og brukermedvirkning. For å jobbe kunnskapsbasert er det i tillegg viktig at praksisen er tilpasset den gjeldende konteksten (Nortvedt et al., 2012, s. 16-20). For å tilby pasientene trygge og effektive tjenester er det dermed essensielt at helsearbeiderne integrerer alle elementene i sitt arbeid, samtidig som de reflekterer over egen praksis og aktivt søker etter forbedringsområder. Kunnskapen som ligger til grunn for helsearbeiderens handlinger kan dermed være hentet fra flere kilder, og i dagens helsevesen er det en dreining til oftere å søke etter svar fra forskning. Kunnskapsbasert praksis innebærer dermed at helsepersonellet utfører behandlingen ut i fra den beste tilgjengelige forskningen (Sackett et al., 1996 ; Nortvedt et al., 2012, s. 44-45).

Kunnskapsbasert praksis er allerede implementert i norsk helsetjeneste, blant annet gjennom offentlige føringer for helsearbeidere (Helse- og Omsorgsdepartementet, 2013). Samtidig er det variasjon i kvaliteten på tjenestene som tilbys, noe som tyder på at ikke alle helsearbeidere jobber ut i fra den beste tilgjengelige forskningen. Det er dermed behov for systematisk arbeid for å bedre kvaliteten på helsetjenestene (Nasjonalt kunnskapssenter for helsetjenesten, 2014). Et tiltak for kvalitetsforbedring kan være å gjennomføre en klinisk audit på et område for å avdekke om det er behov for forbedring, for så å bruke kunnskapsbaserte endringstiltak (Ivers et al., 2012). Det publiseres store mengder forskning som gjør det utfordrende for den enkelte helsearbeider å holde seg oppdatert (Flottorp & Aakhus, 2013). Forskingen en baserer seg på i praksis og i endringsarbeid bør derfor være hentet fra kunnskapsbaserte retningslinjer eller systematiske oversikter (DiCenso, Bayley & Haynes, 2009 ; Flottorp & Aakhus, 2013 ; Nortvedt et al., 2012, s. 44).

En systematisk oversikt blir av The Cochrane Collaboration definert som “*A systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made*” (Green et al., 2011, Kap. 1.2.2). Dette innebærer at metoden er utført og dokumentert på en eksplisitt og systematisk måte, slik at den lar seg reprodusere av andre (Bjørndal, Flottorp & Klovning, 2013, s. 42 ; Jamtvedt, 2013). Forfatterne har dermed fulgt systematiske og transparente prosedyrer for å identifisere, vurdere og sammenstille forskningen omkring et tema. Videre er styrken ved en systematisk oversikt at forskningen omkring temaet er kvalitetsvurdert, samlet og syntetisert. Analysen av resultatene kan være narrativ eller kvantitativ. En narrativ syntese innebærer en strukturert

beskrivelse av de inkluderte studiene, mens en kvantitativ syntese presenteres som statistikk. Vanligst er da å presentere kvantitative resultat som en meta-analyse (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 44-47 ; Smedslund, 2013). Presentasjon som meta-analyse har sine fordeler i at effektestimaterne blir mer presise, og at ekstern validitet blir større fordi effektestimaterne er et resultat av flere studier (Smedslund, 2013). En meta-analyse kan også være mer visuelt tilgjengelig for leseren, og dermed sikre at resultatene lettere blir tolket. En forutsetning for å gjennomføre en meta-analyse er at studiene er like nok til å sammenstilles med tanke på populasjon, intervensjon, sammenligning og utfall (Deeks, Higgins & Altman, 2011 ; Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 46-47 ; Moher et al., 2009). Kunnskap fra en systematisk oversikt gir mer presise svar og mer styrke enn informasjon fra en primærstudie. Dermed vil innhenting og bruk av kunnskap fra systematiske oversikter med større sannsynlighet bidra til kvalitetsforbedring av helsetjenestene enn om kunnskapen hentes fra primærstudier (DiCenso, Guyatt & Ciliska, 2005, s. 138-139 ; Nortvedt et al., 2012, s. 49-51).

I dagens norske helsevesen er flere tiltak igangsatt for å bedre kvaliteten av tjenestene, samt forebygge uønskede hendelser. Nasjonal pasientsikkerhetskampanje har for eksempel program som «Trygg kirurgi», «Samstemming av legemiddellister» og «Forebygging av trykksår» (Helsedirektoratet, 2011). I tillegg er det opprettet en nasjonal meldeordning for uønskede hendelser i helsevesenet. Dette for å registrere uønskede hendelser og bidra til at hele Helse-Norge skal lære av feil begått i pasientbehandlingen (Nasjonalt kunnskapssenter for helsetjenesten, 2012). Simuleringstrening er en kvalitetsforbedringsmetode som benyttes stadig mer i helsevesenet, blant annet som konseptet ALERT (Einarson et al., 2011). ALERT er en opplæringspakke med hensikt at helsearbeideren skal jobbe på en standardisert og strukturert måte (Einarson et al., 2011). Simuleringstrening er en læringsmetode der helsepersonell skal tilegne seg, samt vedlikeholde kompleks kompetanse (Gordon & Vozenilek, 2008). Simulering er trening på realistiske situasjoner der helsepersonellet får presentert en case. De skal løse utfordringer som oppstår med tiltak de ville ha brukt i en reell situasjon. Valgene helsearbeideren tar og hvordan disse utføres vil påvirke hvordan scenariet forløper (SAFER, 2015 ; Hviding et al., 2009). Simulering har vært mest brukt i forhold til trening på hjerte – lungeredning (HLR) ved hjertestans, men også på mer komplekse pasientsituasjoner og trening i tverrfaglige team (The International Nursing Association for Clinical Simulation & Learning, 2013 ; Ilgen, Sherbino & Cook, 2013). Simuleringstreningen kan være «high-fidelity», som innebærer bruk av meget avanserte simuleringsskinner som

etterligner menneskets fysiologi. Den kan også være «low-fidelity», som er trening der en benytter enkle simuleringsskinner eller at en frisk person spiller pasient. Til sist kan simuleringen være «computer-based», der det benyttes elektroniske simuleringssystemer (SAFER, 2015 ; Salas, Paige & Rosen, 2013 ; Society for Simulation in Healthcare, 2014). Internasjonalt brukes begrepene «high-fidelity» om avansert simulering og «low-fidelity» om enklere former av simulering (The International Nursing Association for Clinical Simulation & Learning, 2013). På norsk brukes ofte full-skala simulering for det som internasjonalt kalles «high-fidelity», mens vi mangler gode begrep for «low-fidelity» som ofte kalles ferdighetstrening eller lignende (SAFER, 2015 ; Høgskolen i Gjøvik, 2011). Jeg har derfor i denne oppgaven valgt å bruke begrepene høyteknologisk simulering som synonym til «high-fidelity» og lavteknologisk simulering for «low-fidelity». Computer-based har jeg valgt å kalle e-læringsbasert simulering. E-læring er et vidt begrep som kan innebære alt fra enkle instruksjonskurs til mer avanserte interaktive program. E-læring er et begrep også brukt om simulering (SAFER, 2015).

Systematiske oversikter viser at simuleringstrening versus ingen simulering er effektivt som pedagogisk metode for å bedre ferdigheter og kunnskaper blant studenter innen helsefag (Cook et al., 2011 ; Cant & Cooper, 2010). Videre viser en systematisk oversikt at virtuell trening er effektivt for tekniske ferdigheter hos leger, men usikker effekt i forhold til bedring av kvalitet på tjenestene (Hviding et al., 2009). En nylig publisert systematisk oversikt med meta-analyse ser på simuleringstrening for sykepleiere, leger og ambulanspersonell innen akuttmedisin (Ilgen, Sherbino & Cook, 2013). Denne konkluderer med at simuleringstrening øker kunnskap og tekniske ferdigheter, men har moderat effekt på pasientutfall (Ilgen, Sherbino & Cook, 2013). Forfatterne bak denne systematiske oversikten har søkt bredt etter forskning i relevante databaser og tidsskrift. De har også søkt etter forskning som evaluerer effekten av simuleringstrening for sykepleiere (Ilgen, Sherbino & Cook, 2013). Imidlertid har Ilgen og kolleger (2013) ikke gradert dokumentasjonen, og det foreligger dermed usikkerhet knyttet til disse resultatene. I tillegg presenterer Ilgen og kolleger (2013) i sin meta-analyse resultatene av sin systematiske oversikt for sykepleiere og sykepleierstudenter samlet. Dette er to populasjoner som sannsynligvis ikke er sammenlignbare. Selv om det er dokumentert effekt for sykepleierstudenter (Cook et al., 2011) er det ikke gitt at ferdig utdannede sykepleiere har samme utfall, blant annet fordi sykepleiere har et annet erfaringsgrunnlag enn studentene. Jeg har også gjennom søk identifisert nyere forskningsartikler som er publisert etter at Ilgen, Sherbino og Cook (2013) har avsluttet sine søk. Med bakgrunn i dette

mener jeg det er grunnlag for å gjøre en systematisk oversikt over effekt av simulering for sykepleiere. Målet med dette prosjektet er dermed å gjøre en systematisk oversikt over effekt av simuleringstrening for sykepleiere på utfallene ferdigheter og kunnskaper.

Denne oppgaven består av en kappe og en systematisk oversikt i artikkelform. Først presenteres kappen der jeg beskriver metoden jeg har brukt i utarbeidelsen av den systematiske oversikten. Deretter beskrives resultatene av litteratursøket, utvelgelse og analyse og sammenstilling av data. Videre diskuteres resultat, betydning for forskning samt styrker og svakheter ved denne systematiske oversikten, før jeg til slutt trekker sammen trådene i en konklusjon. Artikkelen er bygget opp etter standardene fra The PRISMA statement, som er en sjekklister utarbeidet for rapportering av systematisk oversikt og meta-analyse (Moher et al., 2009), samt retningslinjer fra tidsskriftet *Clinical Simulation in Nursing* sin forfatterveiledning. Artikkelen planlegges sendt til tidsskriftet for vurdering når masteroppgaven er godkjent. Det er noe overlapp mellom informasjon i kappen og artikkelen der jeg mener dette er viktig for sammenhengen.

Metode

Metoden for dette masterprosjektet er systematisk oversikt, på engelsk a systematic review. The Cochrane Collaboration (Higgins & Green, 2011) sin håndbok vært brukt som basis for metoden i prosjektet. Jeg har i tillegg hentet støtte fra håndbøkene til Centre for Reviews and Dissemination (2008) og Nasjonalt kunnskapssenter for helsetjenesten (2013).

Inklusjons og eksklusjonskriterier

Det ble utarbeidet eksplisitte inklusjons- og eksklusjonskriterier for å sikre at alle relevante studier ble inkludert i denne systematiske oversikten. Se tabell 1 for PICOS.

Tabell 1 PICOS

Populasjon	Intervensjon	Sammenligning	Utfall	Studiedesign
Sykepleiere	Simuleringstrening	Ingen simulering; Annen simuleringsform; Ulik organisering av simuleringstreningen	Ferdigheter Kunnskaper	Randomisert kontrollert studie

Følgende inklusjonskriterier ble benyttet:

Populasjon (P)

Populasjonen i denne systematiske oversikten er sykepleiere med bachelorgrad og sykepleiere under videreutdanning utover bachelorgrad. Studier som omhandlet trening i tverrfaglige team ble inkludert, forutsatt at dataene var presentert separat for sykepleiere.

Studier der populasjonen var bachelorstudenter i sykepleie ble ekskludert. Dersom studiens resultater ikke var spesifisert i forhold til yrkesgrupper ble disse ekskludert, fordi data for sykepleiere ikke kunne ekstraheres separat.

Intervensjon (I) og sammenligning (C)

Intervensjonen er simuleringstrening, ofte kalt medisinsk simulering (SAFER, 2015).

Simuleringen kan være med bruk av teknisk avanserte dukker (høyteknologisk), ved bruk av enkle dukker eller mennesker som pasient (lavteknologisk), eller e-læringsbasert simulering (Society for Simulation in Helathcare, 2014). Med simuleringstrening forstås settinger der deltakerne som trener får oppgitt en case som skal være så realistisk som mulig. Videre skal utøverne utføre handlinger og tiltak som ville ha vært naturlig i en reel situasjon. Hvordan scenariet forløper avhenger ofte av hvilke tiltak deltakerne setter i verk (SAFER, 2015 ; Society for Simulation in Helathcare, 2014). Simulering kan også være trening på praktiske prosedyrer, slik som hjerte-lungeredning og intubering¹.

Studier som evaluerte simuleringstrening versus ingen simulering ble inkludert. Disse studiene vil ofte sammenligne simulering mot tradisjonell undervisning eller kurs (The International Nursing Association for Clinical Simulation & Learning, 2013). Videre ble studier som sammenlignet ulike typer av simulering inkludert. Til sist ble sammenligninger av effekten av ulik organisering av simuleringstreningen inkludert.

Studier som tok for seg simulering der dyr var markøren ble ekskludert. Studier som brukte simulering som metode, men der formålet ikke var å evaluere effekt av selve simuleringen, ble også ekskludert. Dette var for eksempel studier som brukte simulering for å evaluere hvordan sykepleiere utførte en prosedyre (Society for Simulation in Helathcare, 2014).

¹ Intubering er en metode for å sikre frie luftveier hos pasienter, og vil si at et plastrør (en tube) føres ned til pasientens lunger (Bakkeland og Thorsen, 2010, s. 347).

Utfall (O)

Studiene måtte rapportere på sykepleieres kunnskaper og/eller ferdigheter for å inkluderes. Med ferdigheter menes sykepleiers evne til å sette i verk korrekte tiltak til pasientene i situasjonen, samt utføre tiltakene på korrekt måte (Benner et al., 2011, s. 2-4). Med kunnskaper menes hva sykepleier kan vise av fagforståelse i en test eller via samtale (Stubberud, 2010, s. 43-46). Ferdigheter kunne være målt ved observasjon av en utenforstående person, eller ved objektive data fra simuleringsdukken. Ved observasjon kunne det være benyttet forhåndsdefinerte sjekklister på ferdigheter. Kunnskaper kunne være målt med skriftlige eller muntlige tester. Både ferdigheter og kunnskaper er viktige for sykepleierens arbeid, og var derfor i denne systematiske oversikten primære utfallsmål. Studier som rapporterte ett eller begge disse utfallene enten som primære eller sekundære utfallsmål ble inkludert.

Studier som utelukkende rapporterte andre utfallsmål ble ekskludert. Eksempler på slike utfall kunne være økonomiske vurderinger, tilfredshet og så videre.

Studiedesign (S)

Randomiserte kontrollerte studier (RCT) ble inkludert. En randomisert kontrollert studie er et egnet design for å evaluere effekt av et tiltak. Randomiseringen er sentral, og skal sikre at fordelingen til de ulike gruppene skal være tilfeldig og dermed sikre grupper som kan sammenlignes (Bjørndal, Flottorp & Klovning, 2013, s. 65).

Andre studiedesign ble ekskludert.

Sammenligninger

For å analysere og presentere dataene på en hensiktsmessig måte, er det anbefalt å lage sammenligninger som det rapporteres på. Dette gjelder både der det er mulighet for å lage en meta-analyse, men også ved narrative synteser (Higgins & Deeks, 2011, Kap. 7). Jeg presenterer følgende sammenligninger:

- Effekt av simuleringstrening versus ingen simulering på kunnskap og ferdigheter
- Effekt av ulike typer simulering på kunnskap og ferdigheter. For eksempel høyteknologisk simulering versus andre former for simulering.
- Effekt av ulik organisering av simuleringstrening på kunnskap og ferdigheter.

Litteratursøk

Litteratursøket til en systematisk oversikt er en omfattende prosess. Et grundig litteratursøk er av stor viktighet for å identifisere alle relevante studier, og således viktig for å sikre god kvalitet på oversikten (DiCenso, Guyatt & Ciliska, 2005, s.139-140).

Valg av databaser

Valg av databaser ble gjort i samarbeid med en erfaren fagbibliotekar. Utvalgs Kriterium for valg av database var at de måtte inneholde forskning innen fagfeltet sykepleie eller medisin. For å øke sannsynligheten for å finne alle studier er det anbefalt å utføre søk i et bredt utvalg av relevante databaser (Lefebvre, Manheimer & Glanville, 2011, Kap. 6 ; DiCenso, Guyatt & Ciliska, 2005, s. 139).

Vi søkte først etter eksisterende oversikter i Cochrane Database of Systematic Reviews, og Database of Abstracts of Reviews of Effects for og indentifisere systematiske oversikter på emnet. I tillegg ble det utført søk i Health Technology Assessment (HTA) via Centre of Reviews and Dissemination, da denne databasen inneholder oppsummert forskning av helse-teknologiske tiltak. Oversiktsartikler som ble identifisert i disse databasene ble screenet etter aktuelle primærstudier, relevante søkeord, samt undersøke hvilke databaser disse hadde utført søk i. Databasen HTA kan også være en kilde til å identifisere grå litteratur (Centre for Reviews and Dissemination, 2015).

Søk etter primærstudier ble utført i Cochrane Central Register of Controlled Trials (CENTRAL), CINAHL, MEDLINE, Embase, ERIC og SveMed+. Databasene CENTRAL, MEDLINE og Embase er vurdert som et minimum av databaser som skal være med i et systematisk litteratursøk (Lefebvre, Manheimer & Glanville, 2011, Kap. 6). Databasen ERIC er primært en pedagogisk database som inneholder forskning om kvalitetsforbedring i læring (DiCenso, Guyatt & Ciliska, 2005, s. 39-42). Dette er relevant da populasjonen også er sykepleiere under videreutdanning. Søk i SveMed+ ble utført av to grunner. Denne databasen inneholder nordisk forskning, og fordi de nordiske landene er aktive med simuleringstrening og forskning på temaet valgte jeg å søke der. Databasen er videre relevant for å identifisere både tekstord og emneord (MeSH) til søkestrategiene.

Videre ble det gjennomført håndsøk i tidsskriftene Clinical Simulation in Nursing og Simulation in Healthcare for 2013 og 2014, fordi disse kun var indeksert i MEDLINE til januar 2013 på tidspunktet jeg gjennomførte det systematiske søket.

Ekspert² på feltet ble kontaktet per e-post med forespørsel om de kjente til publiserte eller upubliserte studier som var sentrale å inkludere i den systematiske oversikten.

Til sist ble det utført søk i Clinical Trials for å identifisere publiserte protokoller fra pågående studier, eller studier som ikke ble fanget opp i litteratursøket. Her ble søkeordet «simulation» benyttet.

Utarbeidelse av søkestrategier

Søkestrategiene ble tilpasset til de enkelte databasene. Databasenes veiledninger for trunkering, nærhetsoperatorer og boolske operatorer ble benyttet for å tilpasse søkene. Emneordenes funksjon «explode» ble benyttet for å utvide søket, der dette ble vurdert hensiktsmessig. Alle søkestrategiene ble kvalitetssikret av en fagbibliotekar med lang erfaring i søk etter forskningslitteratur, noe som anses som viktig for å sikre gode søkestrategier (Lefebvre, Manheimer & Glanville, 2011, Kap. 6). Det ble ikke satt noen begrensinger for søket i forhold til språk eller studiedesign. Databasenes filter for søk maksimalt tilbake i tid ble benyttet.

Søkeord

Det ble gjort et grundig forarbeid i å finne søkeord, da en systematisk søkestrategi skal bestå av alle aktuelle emne- og tekstord til temaet (Polit & Beck, 2014, s. 119 ; Lefebvre, Manheimer & Glanville, 2011, Kap. 6). Søkeord til elementene PIO ble identifisert, se Vedlegg 1.

Tekstord ble identifisert i flere ulike kilder. Alle kjente synonymer ble skrevet ned, i tillegg til ord fra artikler som omhandlet simulering. Videre ble førsteforfatter til to systematiske oversikter (Cook et al., 2011 ; Ilgen, Sherbino & Cook, 2013) kontaktet, med forespørsel om å få søkestrategien til deres prosjekt tilsendt. Emneord har også blitt brukt som tekstord i søkene.

Utarbeidelse av søkestrategi

Søkestrategiene til denne systematiske oversikten ble utarbeidet for å søke bredt for flest mulige treff, og dermed høy sensitivitet. En god søkestrategi er bygget opp slik at det søkes bredest mulig innen emnet for å øke sannsynligheten for at alle relevante artikler blir funnet. Dersom ikke alle relevante studier inkluderes vil den systematiske oversikten rammes av

² Høgskolen i Gjøvik, Simuleringssenteret ved Oslo Universitetssykehus Ullevål samt SAFER ble kontaktet, da de alle er anerkjente aktører innen simuleringstrening.

publikasjonsbias (Lefebvre, Manheimer & Glanville, 2011, Kap. 6). På den annen side er det nødvendig med visse begrensninger for at søket ikke skal generere for mange ikke-relevante treff. På et punkt vil et bredt søk ikke generere flere relevante treff, og tiden som brukes på screeningen står ikke i forhold til resultatet. I litteraturen er dette referert til som sensitivitet versus presisjon. Ved søk til en systematisk oversikt skal en søke å finne en balanse mellom disse (Lefebvre, Manheimer & Glanville, 2011, Kap. 6). Søkestrategiene til de respektive databasene ligger ved artikkelen som Appendix 1-7.

Søket ble så bygget opp steg for steg etter de forhåndsdefinerte elementene i PICOS. Det ble utarbeidet søkestrategier på Populasjon (P), Intervensjon (I) og Sammenlikning (C). Det ble ikke funnet hensiktsmessig å søke på Utfallet (O) og Studiedesignet (S), da det vil redusere sensitiviteten i søket, og medføre risiko for å snevre inn antallet treff i for stor grad (Lefebvre, Manheimer & Glanville, 2011, Kap. 6).

Da databasesøkene var gjennomført, ble referansene importert til referanseprogrammet EndNote X7. EndNote er et program designet for å håndtere en stor mengde referanser, og dermed hjelpe forfatteren til å ha bedre oversikt (Centre for Reviews and Dissemination, 2008, s. 21). Det ble opprettet et bibliotek for hver database, før referansene ble samlet i ett felles bibliotek der duplikater ble fjernet. Duplikatene ble fjernet både ved hjelp av programmets spesialfunksjon for å fjerne dubletter, samt manuelt.

Utvelgelse av studier

I prosessen med å velge hvilke studier som skal inkluderes eller ekskluderes har to personer uavhengig av hverandre gått igjennom titler, abstracts og fulltekst. Valgene er blitt drøftet, før en er kommet til enighet. Personen jeg samarbeidet med i denne fasen er medstudent på min masterutdanning. Det er anbefalt alltid å være to personer som uavhengig av hverandre går igjennom alle studier i alle faser, fra gjennomgang av titler i de endelige databasesøkene, til utvelgelse av hvilke studier som endelig skal inkluderes (Higgins & Deeks, 2011, Kap. 7).

Utvelgelse basert på tittel og abstract

Vi piloterte utvelgelsen av titler og abstracts ved å screene de 50 første referansene individuelt, før utvelgelsen ble drøftet. Denne piloteringen kalibrerte vår forståelse av inklusjons- og eksklusjonskriteriene (Higgins & Deeks, 2011, Kap. 7). De resterende titler og abstracts ble så gjennomgått av begge selvstendig for å beslutte hvilke studier som skulle leses i fulltekst. Som medstudent på min masterutdanning hadde personen jeg samarbeidet

med i denne fasen kunnskap om å lage en systematisk oversikt. Imidlertid er hun ikke sykepleier, og har ikke erfaring med simulering. Det er anbefalt å ha en tredje person tilgjengelig dersom en ikke kommer til enighet om studien skal med videre i prosessen. Det kan for eksempel være uenigheter som skyldes ulik forståelse av inklusjonskriterier (Higgins & Deeks, 2011, Kap. 7). Tredjepersonen i dette prosjektet var forhåndsbestemt til å være veilederen på masteroppgaven, noe som sikret at tredjepersonen var en fagperson med god erfaring i prosessen med å skrive systematiske oversikter.

I første utvelgelsesfase var det lav terskel for å ta studien med videre. Dersom det ikke åpenbart kom frem av tittel eller abstract at studien ikke dekket eksklusjonskriteriene ble den med videre for fulltekstlesing. Dette er anbefalt strategi i denne fasen, for å sikre at studier som skal inkluderes ikke blir feilaktig ekskludert (Higgins & Deeks, 2011, Kap. 7).

Fulltekstlesing av studier og endelig inklusjon

Alle artiklene ble lest i fulltekst av begge uavhengig av hverandre. Det er i denne fasen kritisk å inkludere de riktige studiene for å sikre at ingen studier blir feilaktig ekskludert, men også at studiene som er inkludert faktisk matcher inklusjonskriteriene (Higgins & Deeks, 2011, Kap 7 ; Liberati et al., 2009). De pre-definerte inklusjonskriteriene ble brukt aktivt for å velge hvilke studier som skulle videre i prosessen og hvilke som skulle ekskluderes. Det er her av stor betydning med eksplisitte inklusjonskriterier, og at begge personene har samme forståelse av dem (Higgins & Deeks, 2011, Kap. 7).

Dataekstraksjon

Dataekstraksjonsskjema ble utarbeidet for å hente ut samme informasjon fra alle studiene (Higgins & Deeks, 2011, Kap. 7 ; Centre for Reviews and Dissemination, 2008), og der følgende data ble innhentet: Forfatter og publikasjonsår, antall deltakere, formål, intervensjon og sammenligning, utfall, måleinstrument og resultater. I tillegg en rubrikk med link til studiens protokoll, dersom denne var publisert. Det ble søkt på nettstedet Clinical Trials etter om de inkluderte studiene hadde publisert sine forsknings protokoller (ClinicalTrials.gov, 2014).

Kritisk vurdering

Relevante studier ble inkludert i den systematiske oversikten uavhengig av metodisk kvalitet på studiene. Det er derfor viktig å vurdere kvaliteten på studiene for å gi leserne mulighet til å vurdere hvor stor tiltro de kan ha til resultatene (Higgins, Altman & Sterne, 2011, Kap. 8).

Dette gjøres ved å vurdere intern og ekstern validitet for de inkluderte studiene (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s.38)

Intern validitet; Risk of Bias

Cochrane Collaboration sin Risk of Bias Tool (RoB) ble valgt som verktøy for å identifisere mulig risiko for skjevheter i studiene (Higgins, Altman & Sterne, 2011, Kap. 8). RoB ble valgt da den dekker de viktigste metodiske element ved en randomisert kontrollert studie. I tillegg er det den sjekklisten som er anbefalt å bruke til en systematisk oversikt over kvantitative studier, da sjekklisten er utarbeidet for å oppdage systematiske feil ved forskningen. Slike systematiske feil, kalt bias, kan påvirke resultatene i studien slik at effekten som rapporteres ikke er et resultat av selve intervensjonen men snarere oppstått ved en tilfeldighet (Higgins, Altman & Sterne, 2011, Kap. 8). Sjekklisten er bygget opp slik at dersom det skåres high risk of bias på et spørsmål, er det avslørt en svakhet ved studien som kan ha påvirket vår tillit til resultatet. Motsatt vil low risk of bias bety at forskerne har benyttet gode metoder. Dersom en artikkel skårer high risk of bias på mange kriterier, vil dette minske resultatenes troverdighet og det er risiko for at studiens resultater skyldes tilfeldigheter. Motsatt blir det ved lav skår, der resultatene sannsynligvis skyldes selve intervensjonen, og ikke som et resultat av svakheter og tilfeldigheter i studien (Higgins, Altman & Sterne, 2011, Kap. 8).

Risk of bias tool inneholder 7 spørsmål, og i første punkt vurderer en randomiseringsprosedyren. Herunder vurderes om deltakerne er tilfeldig utvalgt, og om selve randomiseringsprosedyren er skjult for både forskere og deltakere. Dersom randomiseringen er utført tilfredsstillende øker dette sannsynligheten for at forskerne ikke påvirker hvilke deltakere som fordeles til de ulike gruppene. Hensikten er en tilfeldig fordeling, og at en dermed vil oppnå grupper som er sammenlignbare og utskiftbare. Det styrker troen på at en eventuelt observert forskjell mellom gruppene skyldes selve intervensjonen. Her stilles det krav om en tabell som viser relevante demografiske data fordelt på tiltak og kontrollgruppe (Higgins, Altman & Sterne, 2011, Kap. 8).

Neste punkt er allocation concealment. Her vurderes det om deltakere og forskere er blindet for gruppetilhørighet. Dersom en deltaker får visshet om hvilken gruppe han tilhører kan dette påvirke innstillingen hans til studien og dermed indirekte influere resultatet av studien. På samme vis kan opptreden til personellet i studien påvirkes av hvilken gruppe deltakeren tilhører (Higgins, Altman & Sterne, 2011, Kap. 8).

Videre vurderes om intervensjonen er blindet for deltakere og personell. Der det er mulig bør intervensjonen skjules for begge. Dette for at resultatene av intervensjonen ikke skal kunne påvirkes av deltakernes forforståelse av intervensjonen. Det er imidlertid ikke alle intervensjoner der det er praktisk mulig å blinde på grunn av intervensjonens art. Dette gir i seg selv en risiko for bias. På den annen side vil ikke intervensjonen kunne gjennomføres blindet. I slike tilfeller er det viktig at den som fordeler gruppene og den som analyserer dataene er blindet for hvilken gruppe deltakeren tilhører. Dette for å forebygge at forskerens ønske om resultat av studien påvirker analysen av resultatene (Higgins, Altman & Sterne, 2011, Kap. 8).

Neste punkt som skal evalueres er om ufullstendige data er tilstrekkelig ivaretatt. I en del studier vil det av ulike årsaker være deltakere som forlater studien, eller ulike data forskerne ikke har klart å fremskaffe hos enkelte deltakere. Dette punktet er for å sjekke om forskerne har ivaretatt frafall på en god måte. Dette kan blant annet gjøres ved å bruke en intention to treat analyse av deltakere som dropper ut, basert enten på deltakerens initiale data eller ved å bruke gjennomsnitt av de andre deltakerne. Intention to treat betyr at deltakeren analyseres i den gruppen han ble randomisert til, selv om han ikke mottar tiltaket (Hollis & Campbell, 1999 ; Higgins, Altman & Sterne, 2011, Kap. 8). Det er uansett viktig at forskerne har redegjort for hvordan de har håndtert manglende data, stort frafall og så videre (Higgins, Altman & Sterne, 2011, Kap. 8).

Videre gjøres en vurdering av om forskerne har rapportert de utfallsmålene som er relevante å rapportere, samt de utfallsmål de har beskrevet at de skal rapportere. Her vil det være en styrke om forskerne har publisert en forskningsprotokoll. En protokoll vil vise om forskerne har utført det de hadde til hensikt å utføre, og ikke endret på studien etter hvert som de så effekten av tiltaket (Higgins, Altman & Sterne, 2011, Kap. 8).

I neste punkt vurderes om studien kan ha andre systematiske skjevheter enn de som er vurdert i de foregående punkt.

Cochrane-håndbokens kriterier for kritisk vurdering (Risk of Bias) er brukt som støtte for scoren som er blitt gitt på hvert element, samt totalscore (Higgins, Altman & Sterne, 2011, Kap. 8).

Også i fasen med kritisk vurdering evaluerte to personer uavhengig av hverandre kvaliteten på studiene. Videre ble resultatene fra den kritiske vurderingen drøftet, og uenigheter ble løst ved diskusjon.

Ekstern validitet

De inkluderte studienes eksterne validitet ble vurdert ut i fra PICOS som var definert for den systematiske oversikten. Ekstern validitet betyr om studien er gyldig i de forhold der resultatene skal benyttes. Dette innebærer om intervensjonen er godt gjennomført, om den er overførbar, stemmer overens med de ressursene en har tilgjengelig og så videre (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 38). Ekstern validitet er også et element som inngår i graderingen av den samlede dokumentasjonen, beskrevet som direktet (Vist et al., 2013), se under.

Gradering av dokumentasjonen

Tilliten til, eller kvaliteten på dokumentasjonen bak de ulike sammenligningene simuleringstrening versus ingen simulering, høyteknologisk simulering versus andre simuleringstrategier, samt ulik organisering av simuleringstreningen, ble gradert ved hjelp av Grading of Recommendations Assessment, Development and Evaluation (GRADE) (Guyatt et al., 2008a). Jeg benyttet Guideline Development Tool (GDT) som er et verktøy utarbeidet av GRADE Working Group. Dette har til hensikt å gjøre vurderingen av kvaliteten på den samlede forskningen transparent ved blant annet å utarbeide evidens-profiler (GRADE Working Group, 2012 ; Mustafa et al., 2013). GRADE brukes for å vurdere hvor stor tillit en har til at effektestimater ligger nær den sanne effekt, og dermed robustheten av dokumentasjonen (Vist et al., 2013 ; Guyatt et al., 2008a ; Guyatt et al., 2008b ; Goldet & Howick, 2013).

Graderingen av dokumentasjonen evaluerer åtte element ved studiene, og gir en samlet vurdering på om dokumentasjonen har høy, moderat, lav eller veldig lav kvalitet (Guyatt et al., 2008a). Ved høy kvalitet er det stor tillit til at effektestimater ligger nær den samme effekt, ved middels kvalitet ligger effektestimater sannsynligvis nær det sanne resultat, men det er usikkerhet knyttet til det. Ved lav kvalitet er det begrenset tillit, mens det ved svært lav kvalitet er svært liten tillit til effektestimater (Nasjonalt kunnskapssenter for helsetjenesten, 2013, s. 49). Det første elementet er studiedesign der RCT starter på høy kvalitet, mens observasjonsstudier starter på lav kvalitet. RCTer kan så nedgraderes og observasjonsstudier kan oppgraderes alt etter kvaliteten på studien. Neste element er studiekvalitet, der faren for

systematiske feil vurderes. Videre vurderes konsistens, som er en vurdering av om resultatene i studiene viser samme retning. Deretter vurderes elementet direktehet der deltakere, intervensjon og utfall vurderes i forhold til relevans. Elementet presisjon ser på hvor presise studienes resultater er, og bredden på konfidensintervallet. Der hvor det kun er en studie, nedgraderes presisjonen. Rapporteringsskjevheter går på selektiv rapportering av utfall, eller mangelfull rapportering av negative resultat. Deretter vurderes om det er sterke sammenhenger mellom intervensjon og utfall, før en til sist vurderer dose-respons i studiene..

Det er utført gradering på alle tre sammenligninger for ferdigheter og kunnskaper. Det er i denne systematiske oversikten gjennomført en regular oppstilling i programmet Guideline Development Tool på sammenligningen høyteknologisk simulering versus ingen simulering på utfallet ferdigheter, mens det er gjort en narrativ oppstilling på de resterende sammenligninger og utfallsmål. Tabellene utarbeidet i Guideline Development Tool er vedlegg i artikkelen, Appendix 9a-9c.

Datatyper

For dette prosjektet var både kontinuerlige data og dikotome effektestimater relevante.

Kontinuerlige data kan innta en hvilken som helst verdi innen et visst intervall.

Effektestimater målt som kontinuerlige data er presentert som forskjeller i mean og 95 % konfidensintervall eller median og interkvartil bredde, og blir i den kvantitative oppsummeringen presentert som Standardized Mean Difference (SMD) (Deeks, Higgins & Altman, 2011, Kap. 9). Dikotome data er datatyper som har et enten eller utfall, slik som død – levende. Disse dataene blir presentert som andel og prosent, og rapportert i effektestimater som Risk Ratio (RR), Odds Ratio (OR), Risk Difference (RD) eller Numbers Needed to Treat (NNT) (Deeks, Higgins & Altman, 2011, Kap. 9).

Analyse og sammenstilling av data

Analyse av data i en systematisk oversikt har til hensikt å sammenfatte funnene fra primærstudier. Det skal gis en beskrivelse av de samlede resultatene, retningen på effektestimaterne, størrelsen på effektestimaterne, om resultatene er konsistente og kvaliteten på dokumentasjonen (Deeks, Higgins & Altman, 2011, Kap. 9).

Dataprogrammet Review Manager 5.3 (RevMan) (The Nordic Cochrane Centre, 2014) ble benyttet til å analysere og presentere samlede effektestimater, samt utarbeide figurer. RevMan er utarbeidet av Cochrane Collaboration og er godt egnet for å presentere kvantitative resultat

i systematiske oversikter, da det er mulighet for å lage ulike tabeller slik som Risk of Bias table, Forest Plot og så videre (Deeks, Higgins & Altman, 2011, Kap. 9).

Det ble laget meta-analyse der dette var hensiktsmessig og sensitivitetsanalyse.

Forskningsspørsmålet i denne systematiske oversikten er kvantitativt, og studiedesignet er randomisert kontrollert studie, noe som kvalifiserer til å gjennomføre en meta-analyse (Deeks, Higgins & Altman, 2011, Kap. 9). For noen av sammenligningene og utfallsmålene er det utført en narrativ syntese, og presentasjoner via forest plot.

Formålet med en meta-analyse er å styrke syntesen, og øke presisjonen ved å slå sammen resultatene fra flere sammenlignbare studier. Den presenteres så som et forest plot og der totalen av effektestimater presenteres. Denne totalen presenteres i form av en diamant som viser det samlede effektestimater til studiene. Det er dermed mulig og visuelt vurdere det samlede resultatet (Borenstein et al., 2009, s. 366-368). Før det utarbeides en meta-analyse må det vurderes hvorvidt populasjonen og intervensjonen lar seg sammenligne og sammenstille. Ulikheter i populasjon og intervensjon betegnes som kliniske forskjeller eller klinisk heterogenitet (Deeks, Higgins & Altman, 2011, Kap. 9). Videre må utfallsmålene være de samme, og målt på samme måte. I en meta-analyse vurderes det også statistisk heterogenitet, som innebærer forskjeller i studienes resultat som ikke kan forklares med tilfeldigheter (Deeks, Higgins & Altman, 2011, Kap. 9). Heterogeniteten i meta-analysen er målt med Tau^2 , Chi-kvadrat (Chi^2 / X^2), frihetsgrader (df), p-verdi og I^2 . Tau^2 er et estimat av variasjon mellom studier i en random-effects meta-analyse. CHI-kvadrat test gjør oss i stand til å vurdere om forskjellen i frekvens i de to gruppene er en reell effekt av intervensjonen, eller om forskjellen skyldes tilfeldigheter ut i fra gruppesammensetningen. En høy verdi angir heterogenitet. Frihetsgradene er knyttet til Chi^2 . Videre er P-verdi et mål på signifikans, og ved en lav p-verdi angir det heterogenitet. Til sist vil I^2 beskrive hvor mange prosent av variabiliteten av effektestimater som skyldes heterogenitet og ikke tilfeldigheter (Polit & Beck, 2012, s. 421 ; Deeks, Higgins & Altman, 2011, Kap. 9).

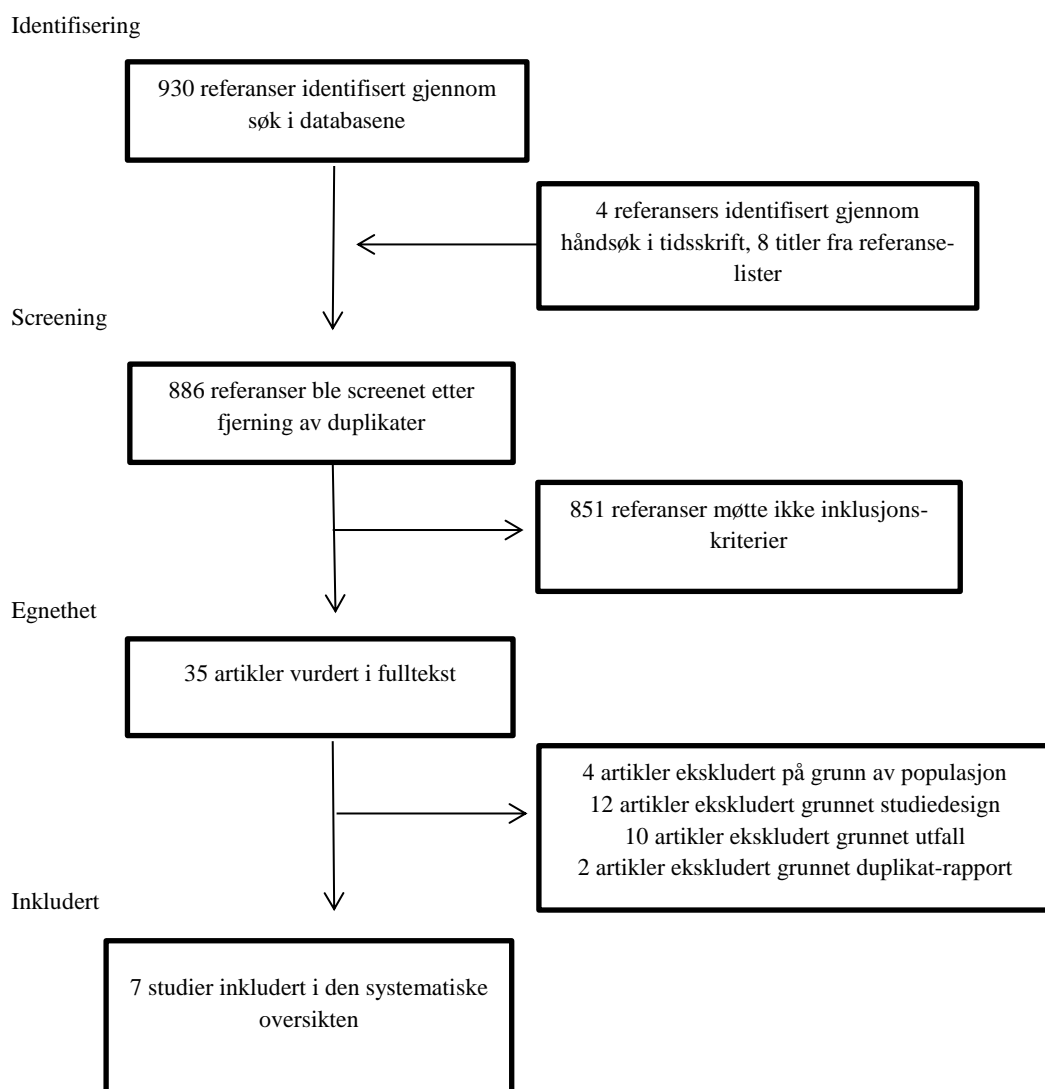
Formålet med en narrativ syntese er å presentere funnene som en samlet beskrivelse av resultatene, samt retningen på effektestimater (Nasjonalt kunnskapscenter for helsetjenesten, 2013, s.45). Resultatene blir presentert for hver enkelt sammenligning på de ulike utfallsmålene, henholdsvis ferdigheter og kunnskaper. Den narrative syntesen beskrives som tekst, og presenteres i tabellform som forest plot utarbeidet for hver enkelt sammenligning. Alle sammenligningene og utfallene presenteres i en oppsummeringstabell med gradering av

dokumentasjonen, da slike tabeller forenkler leserens tolkning av oversiktens resultater (Carrasco-Labra et al., 2015).

Resultater

Litteratursøk og utvelgelse

Vi identifiserte 942 publikasjoner. Figur 1 nedenfor viser et flytskjema over søk og håndtering av referansene.



Figur 1 Flytdiagram utvelgelse av studier

De systematiske søkene identifiserte 930 treff i referanser. I tillegg ble fire referanser identifisert gjennom håndøk i tidsskrifter og åtte studier fra referanselister. Håndøket i tidsskriftene *Clinical Simulation in Nursing* og *Simulation in Healthcare* identifiserte ingen nye randomiserte kontrollerte studier. Forespørselen til ekspertene på simulering resulterte heller ikke i noen nye studier. Etter at duplikater var fjernet gjenstod 886 referanser som ble screenet på tittel og abstract av to uavhengige personer. Totalt 851 referanser ble vurdert som ikke aktuelle og 35 referanser ble inkludert til vurdering i full tekst. Begge hadde valgt ut studier til lesing i fulltekst som den andre ikke hadde med. Dette viser viktigheten av å være to personer i utvelgelsesprosessen (Higgins & Deeks, 2011, Kap. 7).

I fase en av fulltekstlesingen ble 20 artikler ekskludert, og etter grundig gjennomgang av de resterende 15 artiklene ble ytterligere åtte ekskludert. Bakgrunnen for å ta bort to av artiklene var at de viste seg og være duplikatrapporter til studier som allerede var inkludert. Dette gjelder De Regge et.al. (2012) som refererer samme studien som Monsieus et.al (2012), samt Johnson og Johnson (2014) som refererer samme studie som Johnson et.al. (2012).

Syv studier (Arnold et al., 2013 ; Cioffi, Purcal & Arundell, 2005 ; De Regge et al., 2008 ; Johnson et al., 2012 ; Maneval et al., 2012 ; Monsieus et al., 2012 ; Weiner et al., 2011) ble inkludert i den systematiske oversikten da disse fullt ut møtte de pre-definerte inklusjonskriteriene.

Beskrivelse av inkluderte studier

De syv inkluderte studiene var gjennomført i henholdsvis USA, Australia og Belgia. Seks av studiene (Arnold et al., 2013 ; Cioffi, Purcal & Arundell, 2005 ; De Regge et al., 2008 ; Maneval et al., 2012 ; Monsieus et al., 2012 ; Weiner et al., 2011) er utført i en sykehussetting, mens en av de inkluderte studiene var i en militær setting (Johnson et al., 2012). To av de inkluderte studiene har utført studier på sykepleiere på spesialavdelinger (Arnold et al., 2013 ; Weiner et al., 2011), mens to studier inkluderte sykepleiere under videreutdanning i anestesisykepleie eller jordmorstudiet (Cioffi, Purcal & Arundell, 2005 ; Johnson et al., 2012). De resterende har studert ordinære sykepleiere på sykehus med erfaring fra nyutdannet til flere års arbeidserfaring (De Regge et al., 2008 ; Maneval et al., 2012 ; Monsieus et al., 2012). Totalt har studiene inkludert 378 sykepleiere.

I fire av de inkluderte studiene (Arnold et al., 2013 ; Johnson et al., 2012 ; Maneval et al., 2012 ; Weiner et al., 2011) var intervensjonen høyteknologisk simulering, mens tre studier (Cioffi, Purcal & Arundell, 2005 ; De Regge et al., 2008 ; Monsieus et al., 2012) brukte en

lavteknologisk strategi. Fire studier (Cioffi, Purcal & Arundell, 2005 ; Johnson et al., 2012 ; Maneval et al., 2012 ; Weiner et al., 2011) sammenlignet simuleringstrening med ingen simulering, to studier (Arnold et al., 2013 ; Johnson et al., 2012) sammenlignet høyteknologisk simulering mot lavteknologisk eller e-lærings simulering, og to studier (De Regge et al., 2008 ; Monsieurs et al., 2012) evaluerte ulike organisering av simuleringstreningen. Seks studier (Arnold et al., 2013 ; Cioffi, Purcal & Arundell, 2005 ; De Regge et al., 2008 ; Johnson et al., 2012 ; Monsieurs et al., 2012 ; Weiner et al., 2011) hadde utfallsmålet ferdigheter, mens fire studier (Arnold et al., 2013 ; Cioffi, Purcal & Arundell, 2005 ; Maneval et al., 2012 ; Weiner et al., 2011) målte kunnskaper. Utfallene ble målt med observasjonsskjemaer eller data fra simuleringsskjermer på ferdigheter, mens kunnskaper ble målt ved skriftlige tester. Tabell 2 viser de inkluderte studienes karakteristika.

Tabell 2 Karakteristika av inkluderte studier

Forfatter, År, Land	Deltakere, Metode, Setting	Intervensjon	Sammenligning	Utfall
Arnold et.al 2013 USA	28 sykepleiere. Randomisert kontrollert studie, i opplæringsprogram på et stort sykehus.	Full-skala (høyteknologisk) simulering	Lavteknologisk og E-læringsbasert simulering.	Ferdigheter målt med ERPT, kunnskaper målt med skriftlig test.
Cioffi et.al. 2005 Australia	36 jordmorstudenter. Randomisert kontrollert studie, jordmor-studie på universitetsnivå.	To simuleringssesjoner	Ordinære forelesninger.	Ferdigheter og kunnskaper målt gjennom egenutviklede skårings-skjemaer.
De Regge et.al. 2008 Belgia	103 sykepleiere fra ikke-akuttavdelinger. Randomisert kontrollert studie, i en sykehussetting.	En instruktør til en deltaker (IT)	En instruktør til seks deltakere (GT)	Ferdigheter målt med data fra simuleringsskjermer
Johnson et.al. 2012 USA	49 anestesisykepleierestudenter. Randomisert kontrollert studie, i et militært simuleringssenter	Høyteknologisk simulering CD-ROM	Ordinær opplæring	Ferdigheter målt gjennom et skjema utarbeidet for studien.
Maneval et.al. 2012 USA	26 nyutdannede sykepleiere. Randomisert kontrollert studie, på generelle sykehusavdelinger.	Standard opplæringsprogram pluss seks høyteknologiske simuleringssesjoner med Laerdal Sim-Man®.	Standard opplæringsprogram	Kunnskap målt med The health sciences reasoning test
Monsieurs et.al. 2012 Belgia	90 sykepleiere Randomisert kontrollert studie, i en sykehussetting.	Egentrening med dukke og instruksjonsbok (SL)	Instruktørledet trening (IL)	Ferdigheter målt med data fra simuleringsskjermer.

Forfatter, År, Land	Deltakere, Metode, Setting	Intervensjon	Sammenligning	Utfall
Weiner et.al. 2011 USA	46 sykepleiere. Randomisert kontrollert studie. Sykepleiere ansatt på en føde-avdeling.	Egentrening med dukke og instruksjonsbok før 90 min. simuleringssesjon med instruktør.	Tradisjonell klasseroms-undervisning.	Ferdigheter målt med data fra simuleringsdukken, kunnskap målt med skriftlig test.

Beskrivelse av ekskluderte studier

I den initiale utvelgelsen av artikler til fulltekstlesing var åtte studier inkludert, som så i fase to ble ekskludert. Fire av studiene ble ekskludert da det viste seg å være kontrollerte ikke-randomiserte studier, to studier ble ekskludert da de viste seg å være duplikatrapporter til to av studiene som allerede var inkludert. De to siste studiene ble ekskludert på bakgrunn av at utfallsmålene ikke møtte inklusjonskriteriene. De ekskluderte studiene er presentert i artikkelen som Appendix 8 med begrunnelse for ekskluderingen.

Risiko for systematiske skjevheter

Samlet vurdering av risiko for systematiske skjevheter i de inkluderte studiene er oppsummert i figur 2. Vurderingen for hver enkelt studie presenteres i vedlegg 3a-3g RoB-Skjema per studie.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Arnold et al, 2013	?	+	-	-	-	?	+
Cioffi et al, 2005	+	+	?	?	-	?	+
De Regge et al, 2008	+	+	?	?	?	?	+
Johnson et al, 2012	?	-	?	+	?	?	+
Maneval et al, 2012	+	+	+	+	+	?	+
Monsieurs et al, 2012	+	+	-	-	+	?	+
Weiner et al, 2011	+	?	+	+	+	?	+

Figur 2 Risiko for systematiske skjevheter

Tilfeldig fordeling

Fem av de inkluderte studiene (Cioffi et.al., 2005; De Regge et.al., 2008; Maneval et.al., 2012; Monsieurs et.al., 2012; Weiner et.al., 2011) har en tydelig beskrivelse av randomiseringsprosedyren. Cioffi et.al. (2005) og De Regge et.al.(2008) har brukt lukkede konvolutter, mens Maneval et.al. (2012), Monsieurs et.al (2012) og Weiner et.al. (2011) har fått tilfeldig fordeling fra et dataprogram. Disse er dermed gradert til lav risiko. Arnold et.al. (2013) og Johnson et.al. (2012) sier at deltakerne er randomisert, men mangler beskrivelse av selve randomiseringen. Dermed er disse gradet til uklar risiko for bias.

Skjult fordeling

Arnold et.al. (2013), Cioffi et.al. (2005), De Regge et.al. (2008), Maneval et.al. (2012) og Monsieurs et.al. (2012) har alle fått lav risiko da fordelingen har vært skjult på en tilfredsstillende måte. Weiner et.al. (2011) har fått uklar risiko fordi det mangler beskrivelse om når fordelingen ble gjort, og hvordan den ble holdt skjult. Johnson et.al. (2012) har fått høy risiko for bias da det fullstendig mangler informasjon om fordelingen, og forskerne ikke har vært skjult for fordelingen.

Blinding av deltakere og personell

To av studiene (Maneval et.al., 2012 og Weiner et.al., 2011) har lav risiko i forhold til blinding, da begge studiene hadde ivaretatt blinding på en god måte. Tre av studiene (Cioffi et.al., 2005; De Regge et.al., 2008; Johnson et.al., 2012) har uklar risiko da blindingen av deltakere og personell ikke er godt nok beskrevet. Arnold et.al. (2013) viser både intervensjons- og kontrollgruppen selve simuleringen, mens Monsieurs et.al. (2012) ikke har nevnt blinding. Begge får derfor høy risiko for bias.

Blinding av utfallsmålere

Både Johnson et.al.(2012), Maneval et.al. (2012) og Weiner et.al. (2011) har beskrevet at utfallsmåler er blindet, og får dermed lav skår. Cioffi et.al. (2005) og De Regge et.al. (2008) får begge uklar risiko da blindingen er uklart beskrevet, mens Arnold et.al. (2013) og Monsieurs et.al. (2012) også her får høy skår da utfallsmåler ikke er tilstrekkelig blindet.

Ufullstendige data og frafall fra studien

Tre av studiene (Maneval et.al., 2012; Monsieurs et.al., 2012; Weiner et.al. 2011) har alle fått lav skår da alle har med tabeller med demografiske data, samt at studiene har hatt et meget lavt frafall. To studier (De Regge et.al., 2008; Johnson et.al., 2012) har fått uklar risiko da de ikke beskriver om frafallet er tatt hensyn til i analysen. Johnson et.al. (2012) mangler også en tabell med demografiske data per gruppe slik at disse ikke er mulig å sammenligne. Til sist har Arnold et.al. (2013) og Cioffi et.al. (2005) høy risiko for bias da begge mangler tabell for demografiske data per gruppe. Arnold et.al. (2013) har et forholdsvis stort frafall uten å ha tatt hensyn til det i analysen, mens Cioffi et.al. (2005) ikke har redegjort for frafall.

Selektiv rapportering

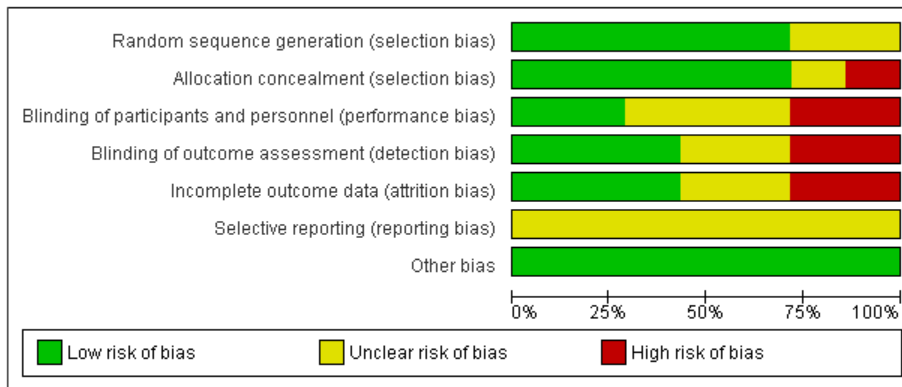
Alle syv studiene (Arnold et.al., 2013; Cioffi et.al., 2005; De Regge et.al., 2008; Johnson et.al., 2012; Maneval et.al., 2012; Monsieurs et.al., 2012; Weiner et.al., 2011) har uklar risiko fordi ingen av studiene har publisert en forskningsprotokoll. Det blir dermed vanskelig å vurdere om rapporteringen er selektiv eller ikke. Det er ikke grunnlag for å tro at studiene har unnlatt å presentere negative resultat.

Andre systematiske skjevheter

Jeg har vurdert alle de syv inkluderte studiene (Arnold et.al., 2013; Cioffi et.al., 2005; De Regge et.al., 2008; Johnson et.al., 2012; Maneval et.al., 2012; Monsieurs et.al., 2012; Weiner et.al., 2011) til å ha lav risiko for andre systematiske feil, utover hva som er redegjort for over.

Samlet vurdering

Den samlede vurdering av total risiko for systematiske skjevheter i studiene viser at to studier har lav risiko (Maneval et al., 2012 ; Weiner et al., 2011), tre studier har uklar risiko (Cioffi, Purcal & Arundell, 2005 ; De Regge et al., 2008 ; Johnson et al., 2012) og to studier (Arnold et al., 2013 ; Monsieurs et al., 2012) har høy risiko for systematiske feil. Den samlede vurderingen viser at seks av syv kriterier i RoB har studier med uklar eller høy risiko for skjevheter. Prosentvis risiko for skjevheter for de ulike kriteriene er presentert i figur 3.



Figur 3 Risk of bias graf

Analyse, sammenstilling og gradering

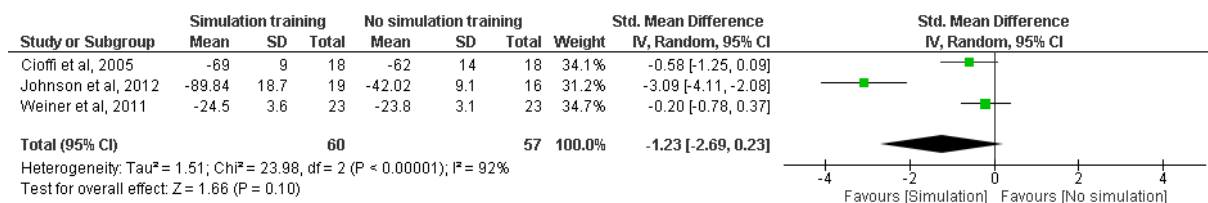
Vi utførte meta-analyse med sensitivitetsanalyse for sammenligningen simuleringstrening versus ingen simulering for utfallet ferdigheter. For de resterende sammenligningene og utfallene ble det utført en narrativ syntese. Gullstandarden for analyse og presentasjon av data i en systematisk oversikt er å utføre en meta-analyse når studiene er like nok til å sammenstilles. Dersom studiene har store kliniske forskjeller bør en derimot utføre en narrativ syntese (Deeks, Higgins & Altman, 2011, Kap 9). Jeg presenterer først simuleringstrening versus ingen simulering knyttet til henholdsvis ferdigheter og kunnskap, før jeg presenterer høyteknologisk simulering versus andre simuleringsformer. Til sist presenteres ulik organisering av simuleringstreningen knyttet til de samme utfallsmålene. Jeg beskriver min tiltro til de samlede resultatene av sammenligningene ved å knytte det til graderingen av dokumentasjonen.

Effekt av simuleringstrening versus ingen simulering

Fire studier hadde sammenlignet simuleringstrening versus ingen simulering. Tre av studiene (Cioffi, Purcal & Arundell, 2005 ; Johnson et al., 2012 ; Weiner et al., 2011) hadde målt ferdigheter ved observasjon og sjekklister, mens tre (Cioffi, Purcal & Arundell, 2005 ; Maneval et al., 2012 ; Weiner et al., 2011) hadde målt kunnskaper via skriftlige tester.

Ferdigheter

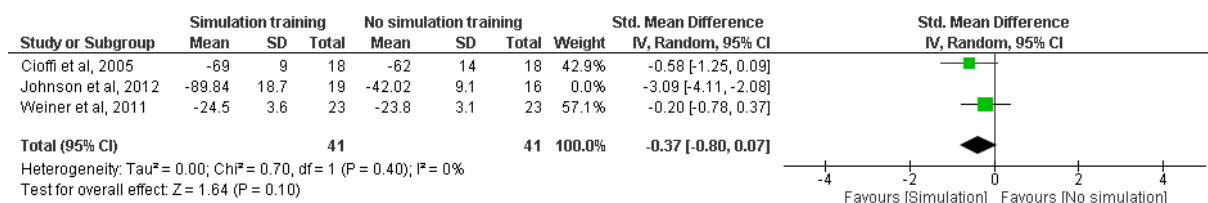
Studiene til Cioffi et.al (2005), Johnson et.al. (2012) og Weiner et.al. (2013) er vurdert som like nok til å sammenstilles i en meta-analyse. Meta-analysen viser ulikheter i resultatene, men samme retning av effektestimatene. Totalen av effektestimatene viser en ikke-signifikant p-verdi, og studiene har dermed ikke klart å vise effekt av intervensjonen. Meta-analysen er presentert som figur 4.



Figur 4 Meta-analyse simuleringstrening versus ingen simulering, utfall ferdigheter.

Cioffi et.al.(2005) og Weiner et.al (2013) viser begge sammenlignbare resultat der effektestimater krysser nulleffektslinjen, men er i favør av intervensjonsgruppen. Konfidensintervallene er omtrent like brede. For studien til Johnson et.al.(2012) skiller resultatene seg ut ved å være klart signifikante i favør av intervensjonsgruppen.

Beregningen av heterogenitet i meta-analysen viser høy grad av statistisk heterogenitet mellom studiene, med en p-verdi < 0,00001 og I² på 92 %. Jeg mener populasjonen og intervensjonen i de tre studiene er like nok til å sammenstilles i en meta-analyse. Deltakerne i studien til Johnson et.al.(2012) kan muligens tenkes å være mer innstilt for simuleringstrening da dette er sykepleiere som skal i en ny setting. I tillegg har denne studien noe høyere risiko for bias enn de to andre studiene. Jeg har på bakgrunn av disse momentene valgt å utføre en sensitivitetsanalyse der jeg tok ut studien til Johnson et.al. (2012). Denne analysen av Cioffi et.al.(2005) og Weiner et.al.(2013) er presentert som figur 5.



Figur 5 Sensitivitetsanalyse simuleringstrening versus ingen simulering, utfall ferdigheter.

Denne analysen viser ingen heterogenitet mellom studiene med p = 0,93 og I² på 0 %. Denne sammenligningen viser ikke signifikante effektestimater da konfidensintervallene krysser nulleffektslinjen.

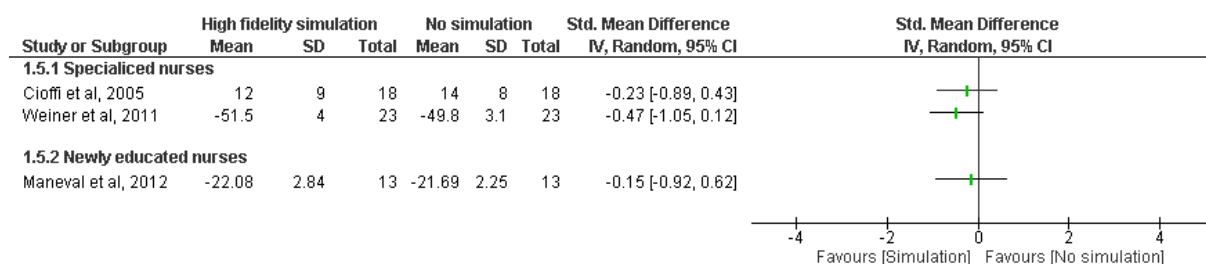
Kvaliteten på dokumentasjonen ble gradert til lav kvalitet. Bakgrunnen for dette var uklart til høy risiko for bias i forhold til randomisering og blinding i flere av de inkluderte studiene, samt at resultatene fra studiene var inkonsistente. Derfor er tilliten lav til at effektestimater fra denne meta-analysen viser den sanne effekt. Oppsummert kan en dermed ikke konkludere

om simuleringstrening har effekt på utfallet ferdigheter når det sammenlignes med ingen simulering.

Resultatene for effekt av simuleringstrening versus ingen simulering, og graderingen av dokumentasjonen er presentert i tabell 3.

Kunnskaper

Studiene til Cioffi et.al.(2005), Maneval et.al. (2012) og Weiner et.al. (2011) ble vurdert til å være for ulike klinisk til å sammenstilles i en meta-analyse, og blir derfor presentert som narrativ syntese. Dette skyldes at populasjonen i studien til Maneval et.al. (2012) skiller seg klart fra populasjonen i de to andre studiene, da deltakerne her er nyutdannede sykepleiere. Cioffi et.al.(2005) presenterer kontinuerlige data i forhold til hvor mange ganger deltakeren oppsøker fakta for å komme til en konklusjon (Data Review Segments) og finner mean 12 (SD 8) for intervensjonsgruppen og mean 14 (SD 8) for kontrollgruppen. Resultatene er dermed i favør av intervensjonsgruppen, men ikke signifikant med p-verdi 0,20. Maneval et.al.(2012) presenterer resultat fra The Health Sciences Reasoning Test som mean score og SD på testen. Intervensjonsgruppen har mean 22, 08 (SD 2, 84) mot mean 21, 69 (SD 2, 25) i kontrollgruppen. Resultatene viser dermed et bedre resultat for intervensjonsgruppen, men igjen ikke signifikant med p-verdi på 0,71. Til slutt har Weiner et.al. (2011) oppgitt effektestimat som mean score med SD for testresultat. For intervensjonsgruppen er mean 51, 5 (SD 4, 0) og kontrollgruppen mean 49, 8 (SD 3, 1). Også denne studien viser resultat svakt i favør av intervensjonsgruppen, men ikke signifikant forskjell mellom gruppene, se figur 6.



Figur 6 Simuleringstrening versus ingen simulering, utfall kunnskaper.

Oppsummert viser alle tre studiene resultater i favør av simuleringstrening på deltakernes kunnskaper. Imidlertid har ingen av studiene lyktes å vise signifikant forskjell, og effekten av simuleringstrening på kunnskaper er derfor usikker.

Graderingen av dokumentasjonen viser lav kvalitet på grunn av usikkerhet rundt fordeling og blinding. En av studiene oppgir heller ikke baseline karakteristika til deltakerne, og det nedgraderes av den grunn. Tiltroen til disse effektestimaterne er dermed lav. Se tabell 3.

Tabell 3 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av simuleringstrening versus ingen simulering.

Utfall	Resultat		Effektestimater (95 % KI)	Antall studier (Deltakere)	Kvalitet på dokumentasjon (GRADE)
	Intervensjon (I)	Kontroll (C)			
Ferdigheter	Mean 69 (SD 9)	Mean 62 (SD 14)	<i>Meta-analyse</i>	3 (117)	⊕⊕○○ Lav ^{1,2,3,4}
	Mean 89, 84 (SD 18, 7)	Mean 42, 02 (SD 9, 1)	SMD -1, 23 (-2, 69 to 0, 23)		
	Mean 24, 5 (SD 3, 6)	Mean 23, 8 (SD 3, 1)	<i>Sensitivitets analyse</i> SMD -0, 37 (-0, 80 to 0, 07)	2 (71)	
Kunnskap	Mean 12 (SD 9)	Mean 14 (SD 12)	SMD -0, 23 (-0, 89 to 0, 43)	3 (108)	⊕⊕○○ Lav ^{5,6,7}
	Mean 51, 5 (SD 4)	Mean 49, 8 (SD 3, 1)	SMD -0, 47 (1, 05 to 0, 12)		
	Mean 22, 08 (SD 2, 84)	Mean 21, 69 (2, 25)	SMD -0, 15 (-0, 92 to 0, 12)		

SD = Standard Deviation, KI = Konfidensintervall, SMD = Standardised Mean Difference

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Fotnoter:

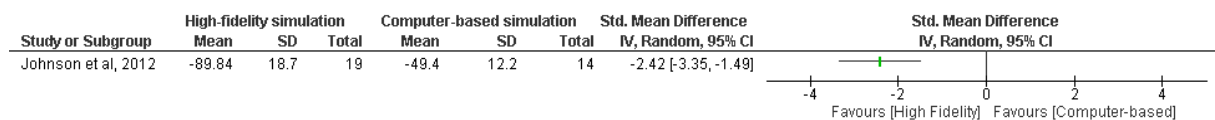
1. One of the included studies had unclear risk of bias on randomization
2. One of the included studies had High risk of bias, and one study had unclear risk regarding blinding
3. One of the included studies had high risk of bias, and one had unclear risk regarding incomplete outcome data
4. The results differ from significant to not significant
5. One of the studies had unclear risk of bias on allocation concealment
6. One of the studies had unclear risk of bias on blinding
7. One of the studies had high risk of bias due to no reported baseline characteristics

Effekt av høyteknologisk simulering versus andre simuleringstrategier

To studier sammenlignet høyteknologisk simulering versus andre simuleringstrategier for utfallsmålet ferdigheter (Arnold et al., 2013 ; Johnson et al., 2012), mens en studie hadde utført sammenligningen på utfallsmålet kunnskaper (Arnold et al., 2013).

Ferdigheter

Arnold et.al. (2013) og Johnson et.al (2012) har begge målt effekten av ulike simuleringstrategier på sykepleieres ferdigheter. Imidlertid har ikke Arnold et.al. (2012) presentert resultat for hver enkelt gruppe, grunnet manglende data fra den ene gruppen. Den deskriptive analysen blir dermed på bakgrunn av resultatene fra Johnson et.al. (2012) som har presentert resultater fra sjekklister som mean med standard avvik (SD). Johnson et.al (2012) sammenligner høyteknologisk simulering med e-lærings simulering, og måler mean 89, 84 (SD 18, 7) i intervensjonsgruppen og mean 49, 40 (SD 12, 2) i kontrollgruppen på en 107 poengs skala. Resultatene er klart signifikante i favør av intervensjonsgruppen med p-verdi <0,000, men i og med kun denne ene studien har vist resultater kan vi ikke konkludere effekten. Resultatene er presentert i figur 7.



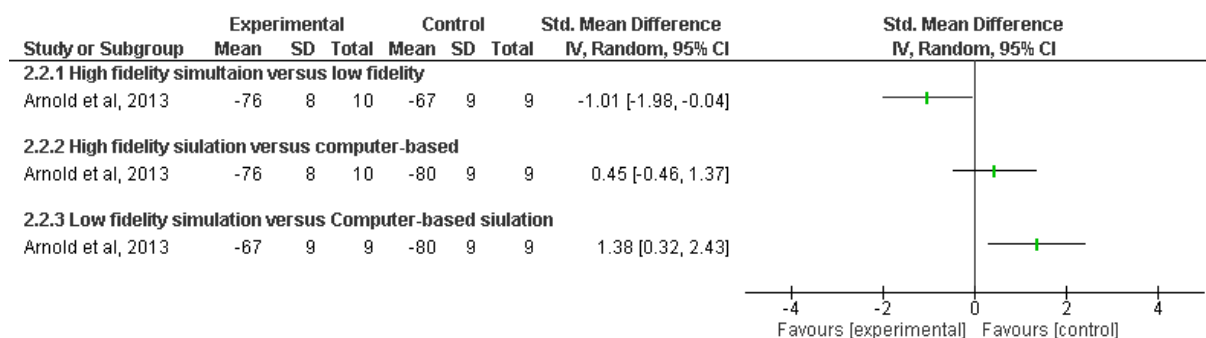
Figur 7 Høyteknologisk simulering versus andre former for simulering, utfall ferdigheter.

Også graderingen av høyteknologisk simulering versus andre simuleringstrategier viser lav kvalitet av dokumentasjonen. Begrunnelsen for nedgraderingen er også her risiko for bias i studien, samt resultat fra kun en studie. Se tabell 4.

Kunnskaper

For kunnskaper har en av de inkluderte studiene utført målinger (Arnold et al., 2013). Arnold et.al.(2013) har utført tre sammenligninger der de har sett på høyteknologisk simulering, lavteknologisk simulering og e-læringssimulering. For sammenligningen høyteknologisk simulering versus lavteknologisk simulering er mean 76 (SD 8) i intervensjonsgruppen og mean 67 (SD 9) i kontrollgruppen. Resultatene er dermed i signifikant favør av intervensjonsgruppen. Standardavviket er imidlertid nær opp ved nulleffektslinjen (SD -1, 98 - -0, 04), og den sanne effekten av intervensjonen er dermed usikker. For sammenligningen høyteknologisk simulering versus e-læringssimulering viser resultatene mean 76 (SD 8) i intervensjonsgruppen og mean 80 (SD 9) i kontrollgruppen. Resultatene her er i favør kontrollgruppen, men ikke signifikante. Til sist viser sammenligningen lavteknologisk simulering versus e-læringssimulering mean 67 (SD 9) i intervensjonsgruppen og mean 80 (SD 9) i kontrollgruppen. Igjen resultater i favør av e-læringssimulering, og resultatene er

signifikante med p-verdi = 0,01. Studien til Arnold et.al. (2013) kan dermed gi retning av at e-lærings simulering er den mest effektive læringsformen i forhold til kunnskaper. Se figur 8.



Figur 8 Høyteknologisk simulering versus andre simuleringerformer, utfall kunnskaper.

Kvaliteten på dokumentasjonen er gradert til lav grunnet resultat fra kun en studie med få deltakere. I tillegg er det risiko for bias knyttet til blinding og frafall av data. Se tabell 4.

Tabell 4 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av høyteknologisk simuleringstrening versus andre simuleringstrategier.

Utfall	Resultat		Effektestimert (95 % KI)	Antall studier (Deltakere)	Kvalitet på dokumentasjon (GRADE)
	Intervensjon (I)	Kontroll (C)			
Ferdigheter	Mean 89, 84 (SD 18, 7)	Mean 49, 4 (SD 12, 2)	SMD -2, 42 (-3, 35 to -1, 49)	1 (46)	⊕⊕○○ Lav ^{1,2,3,5}
Kunnskap	<i>Høyteknologisk versus lavteknologisk</i>		SMD -1, 01 (-1, 98 to -0, 04)	1 (26)	⊕⊕○○ Lav ^{4,5}
	<i>Høyteknologisk versus e-lærings simulering</i>		SMD 0, 45 (-0, 46 to 1, 37)		
	<i>Lavteknologisk versus e-lærings simulering</i>		SMD 1, 38 (0, 32 to 2, 43)		

SD = Standard Deviation, KI = Konfidensintervall, SMD = Standardised Mean Difference

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Fotnoter:

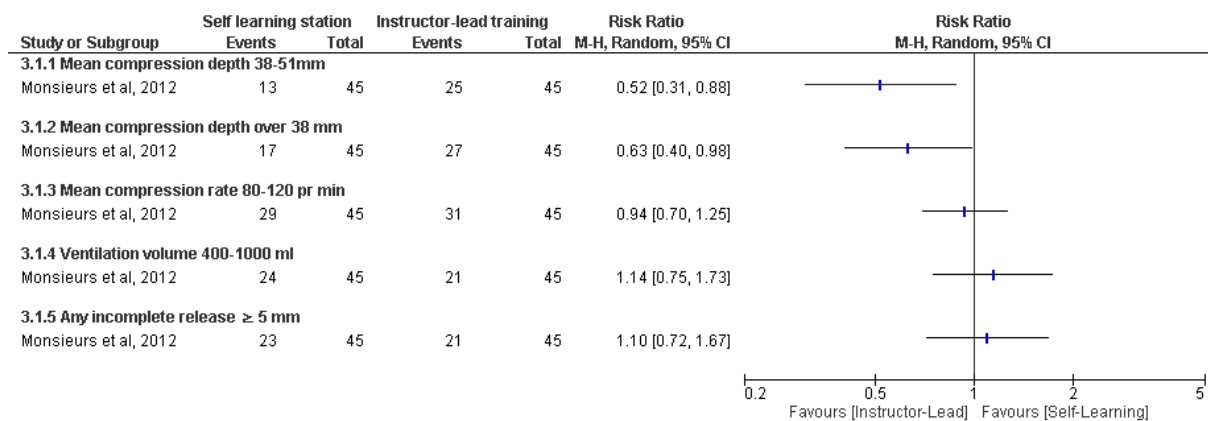
1. The study has high risk of bias on allocation concealment
2. The study has unclear risk of bias on blinding
3. The study has unclear risk of bias on incomplete outcome data
4. The study has high risk of bias on blinding and incomplete outcome data, and unclear risk of bias on allocation concealment
5. The study has unclear risk of bias on random sequence generation

Effekt av ulik organisering av simuleringssesjonen

To studier har utført denne sammenligningen og målt ferdigheter via data fra simuleringssesjonen samt observasjon. Studiene presenteres hver for seg grunnet ulike effektmål.

De Regge et.al. (2008) har sammenlignet ulik organisering av simuleringstreningen. Dataene er oppgitt som kontinuerlige utfall som median og interkvartil bredde, uten å oppgi grunndataene. Jeg har på bakgrunn av usikkerheten rundt manglende bakgrunnstall valgt å ikke lage forest plot for denne studien. De Regge et.al. (2008) har presentert ferdigheter som «antall ventilasjoner», «volum på ventilasjoner», «antall ventilasjoner med korrekt volum», «antall kompresjoner», «dybde på kompresjoner» og «antall kompresjoner med korrekt dybde», «kompresjons rate» og «hands off time». Jeg velger å beskrive «antall ventilasjoner med korrekt volum» og «antall kompresjoner med korrekt dybde», da disse to målingene etter mitt syn best dekker utfallet ferdigheter. For «antall ventilasjoner med korrekt volum» har intervensjonsgruppen median 4 med interkvartil bredde 0-7, mens kontrollgruppen har median 5 med interkvartil bredde 1-7. For «antall kompresjoner med korrekt dybde» har intervensjonsgruppen median 59 med interkvartil bredde 9-86 og kontrollgruppen median 33 med interkvartil bredde 8-82. Resultatene gir ingen retning til om intervensjonen har effekt eller ei, da resultatene peker i favør av kontrollgruppen på antall ventilasjoner med korrekt volum, mens resultatene er i favør av intervensjonsgruppen på antallet kompresjoner med korrekt dybde.

Monsieurs et.al (2011) har sammenlignet egentrening på HLR med dukke og «voice feedback» mot trening i små tradisjonelle HLR-grupper med instruktør. De har målt kompresjons dybde, kompresjons rate, ventilasjonsvolum og «incomplete release», og presentert resultatene som antall og prosent. For kompresjons dybde 38-51 mm er relativ risiko 0, 52, for kompresjons dybde ≥ 38 mm er relative risiko 0, 63, for kompresjons frekvens er relativ risiko 0, 94, for ventilasjonsvolum 400-1000 ml er relativ risiko 1, 14 og for incomplete release ≥ 5 mm er relative risiko 1, 10. Resultatene er dermed inkonklusive med resultat i favør av kontrollgruppen på de tre første målingene, og i favør av intervensjonsgruppen på de to siste. Resultatene er presentert som figur 9.



Figur 9 Ulik organisering av simuleringstreningen, utfall ferdigheter.

For denne sammenligningen totalt sett er resultatene høyst varierende med resultat i favør av både intervensjons- og kontrollgruppen. Kvaliteten på dokumentasjonen er nedgradert til lav grunnet inkonsistens i resultater, samt høy risiko for systematiske skjevheter. Min tiltro til effektestimaterne i studiene er dermed lav.

Tabell 5 Oppsummeringstabell og dokumentasjonsgrunnlag for effekten av ulike organisering av simuleringstreningen.

Utfall	Resultat		Effektestimater (95 % KI)	Antall studier (Deltakere)	Kvalitet på dokumentasjon (GRADE)
	Intervensjon (I)	Kontroll (C)			
Ferdigheter	<i>Ventilations with correct volume</i> Median 4 (0-7) Median 5 (1-7)			2 (193)	⊕⊕○○ Lav ^{1,2,3,4}
	<i>Compressions with correct depth</i> Median 59 (9-86) Median 33 (8-82)				
	<i>Mean compression depth 38-51 mm</i> Events 13/45 Events 25/45		RR 0, 52 (0, 31 to 0, 88)		
	<i>Mean compression depth over 38 mm</i> Events 17/45 Events 27/45		RR 0, 63 (0, 40 to 0, 98)		
	<i>Mean compression rate 80 – 120 pr. Min</i> Events 29/45 Events 31/45		RR 0, 94 (0, 70 to 1, 25)		
	<i>Ventilation volume 400-1000 ml</i> Events 24/45 Events 21/45		RR 1, 14 (0, 75 to 1, 73)		
	<i>Any incomplete release ≥ 5mm</i> Events 23/45 Events 21/45		RR 1, 10 (0, 72 to 1, 67)		

RR = Risk Ratio, KI = Konfidensintervall

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Fotnoter:

1. The study has unclear risk of bias on blinding
2. The study has unclear risk of bias on incomplete outcome data
3. The study has unclear risk of bias on selective reporting
4. The results of the study are not conclusive because they are in favor of both the experimental and control group. The results are not significant.

Diskusjon

Drøfting av oversiktens resultater

Det kan ikke trekkes noen klare konklusjoner om effekten av simuleringsbasert trening for sykepleiere basert på denne systematiske oversikten. Sammenligningene viser ingen signifikant forskjell mellom gruppene på noen av utfallene, og vi har lav tillit til estimatene. Resultatene i flere av studiene går i favør av intervensjonen men har effektestimater som krysser nulleffektlinjen, og forskjellene er dermed ikke signifikante. For enkelte av studiene er resultatene inkonsistente, med resultat i favør av både intervensjons- og kontrollgruppen for

samme tiltaket. Dette gjelder blant annet studien til Monsieurs et.al. (2012) som sammenligner ulik organisering av simuleringstrening på utfallet ferdigheter. Den eneste studien som viser klart signifikante resultat av høyteknologisk simulering er studien til Johnson et.al. (2012). Imidlertid har denne studien uklar risiko for systematiske feil, og det er dermed usikkerhet knyttet til om disse effektestimaterne ligger nær den sanne effekten. Videre er tilliten lav til at effektestimaterne i de ulike analysene viser den sanne effekt, fordi graderingen av den samlede dokumentasjonen bak alle sammenligningene er vurdert til lav kvalitet.

Resultatenes sammenheng med tidligere publiserte oversikter

Tidligere publiserte systematiske oversikter konkluderer med at simuleringstrening har effekt på ferdigheter og kunnskaper. Oversiktene til Cook et.al. (2011) samt Cant og Cooper (2010) har sammenfattet funn fra studenter innen helsefag, mens Ilgen, Sherbino og Cook (2013) har sykepleiere innen akuttmedisin og studenter innen akuttsykepleie som populasjon. Disse oversiktene har resultat som skiller seg fra resultatene jeg har funnet i min systematiske oppsummering. Dette kan ha flere årsaker. Jeg har identifisert studier som ikke er inkludert i disse systematiske oversiktene, publisert både før og etter disse utførte sine søk. Et annet element kan være at studenter har et annet erfaringsgrunnlag enn sykepleiere, og dermed har en større effekt av praktisk øvelse en hva sykepleiere har. Det kan også være at sykepleiere med erfaring vil ha større barriere til simuleringstrening, da de føler at de utleverer seg mer enn hva studenter gjør. Det kan skyldes at de selv tenker de må prestere mer grunnet sitt erfaringsgrunnlag. Dette er i alle fall min erfaring gjennom mitt arbeid med simuleringstrening. Det kan også være en sammenheng mellom at studenter er søkende etter ny kunnskap, og dermed har større glede av simuleringstrening enn hva sykepleiere i ordinært arbeid har. Dette kan blant annet understøttes av resultatene i studien til Johnson et.al. (2012), der populasjonen er sykepleiere under videreutdanning. Dette blir imidlertid kun spekulasjoner. Så langt jeg har funnet har de andre systematiske oversiktene vurdert faren for systematiske skjevheter i de inkluderte studiene, men har brukt andre sjekklister enn hva jeg har. Jeg har ikke funnet at de har gradert kvaliteten på dokumentasjonen ved hjelp av GRADE, slik jeg har gjort. Dette kan bety at effekten av simuleringstrening vist i tidligere oversikter har lav eller middels tillit på bakgrunn av kvaliteten av dokumentasjonen. Der har ikke forfatterne klart å være transparente.

Betydning for forskning

Kvalitetsforbedring i helsevesenet er viktig for å øke sikkerheten for pasientene og yte bedre tjenester. Ofte kan endring i helsevesenet være styrt av impulsive tiltak og tilfeldig kunnskap. Slike tiltak kan også føre til bedre tjenester for pasientene, men forutsetningen er da at tiltaket reelt fører til forbedring og at sterke fagpersoner leder endringen (Grol & Wensing, 2013b, s. 8 ; Flottorp & Aakhus, 2013). For endringstiltak generelt er det viktig at kunnskapsbaserte tiltak ligger til grunn for endringsarbeidet, og at forskning støtter helsearbeiderne til å ta gode beslutninger til beste for pasientene (Grol & Wensing, 2013a, s. 81-82). Ulike tiltak kan være effektive i endringsarbeid. Kurs og arbeidsgrupper er et hyppig brukt tiltak i helsevesenet, og har vist noe effekt på å bedre helsearbeideres praksis (Forsetlund et al., 2009). Videre har trykket materiale vist seg å være lite effektivt som forbedringstiltak (Giguère et al., 2012), mens klinisk audit har vist noe effekt for å oppnå endring (Ivers et al., 2012). Felles for implementeringstiltakene er at de har liten effekt alene, mens dersom flere tiltak tilpasses den settingen de skal brukes i har de oftest økt effekt (Baker et al., 2015). Det er derfor viktig at det utføres studier på kvalitetsforbedringsprosjekt, slik som simuleringstrening, for å evaluere effekten og organiseringen. Forskerne bør da hente støtte fra forskning rundt kvalitetsforbedring i planleggingen av studien.

De studiene jeg har inkludert i denne systematiske oversikten er av varierende kvalitet. Mange av studiene har svakheter i design og studienes resultater er varierende og til dels inkonsistente. Jeg mener dermed det er behov for ytterligere forskning på temaet. Først burde forskningen søke å dokumentere om simuleringstrening er effektivt for sykepleieres kunnskaper og ferdigheter. Dersom simulering dokumenteres effektivt, vil det være grunnlag for å sammenligne ulike strategier og organiseringer. Studiene bør i tillegg evaluere kostnader som et av utfallene. Fremtidige studier bør være av designet randomisert kontrollert studie og forskerne må tilstrebe og følge reglene som følger dette designet, slik at risiko for systematiske feil reduseres til et minimum.

Betydning for praksis

Simuleringstrening er som vist en metode for kvalitetsforbedring brukt ved de fleste norske sykehus (Einarson et al., 2011 ; Høgskolen i Gjøvik, 2011). Imidlertid er det behov for flere tiltak for å bedre kvaliteten på heletjenestene. For å implementere nye kliniske retningslinjer, eller få sykepleiere til å utføre god praksis slik som for eksempel forebygging av trykksår, vil det være aktuelt med tilnærminger som etter mitt syn ikke er godt egnet for simulering. I slike

tilfeller kan andre kvalitetsforbedringstiltak være aktuelle. Dette være seg kurs for å øke kunnskap, eller workshops for å innarbeide og gjøre kjent de nye retningslinjene (Forsetlund et al., 2009). Imidlertid kan simuleringstrening være en del av en skreddersydd implementeringsstrategi for en ny retningslinje (Baker et al., 2015).

Funnene i denne systematiske oversikten gir ikke grunnlag for å endre praksis med simuleringstrening, da resultatene av sammenligningene er forbundet med stor usikkerhet. Dersom helseinstitusjoner nytter simuleringstrening til kvalitetsforbedring er investeringer i utstyr gjort, og rutinen allerede etablert. Flere av studiene viser resultat i favør av simuleringstrening, og jeg mener derfor det vil være tilrådelig og fortsette denne læringsformen. Dette baserer jeg også på funn fra systematiske oversikter som har oppsummert en positiv effekt for studenter innen helseprofesjoner, selv om jeg har vist at studenter og sykepleiere ikke nødvendigvis har samme effekt av simulering. På en annen side gir ikke funnene i denne systematiske oversikten grunnlag for å starte opp med simulering, eller investere i utstyr dersom denne læringsformen ikke allerede er etablert. Jeg mener denne systematiske oversikten allikevel er et viktig bidrag til praksis, da det er den første som oppsummerer forskning av designet RCT begrenset til populasjonen sykepleiere.

Styrker og svakheter ved den systematiske oversikten

Å utarbeide en systematisk oversikt er i alle håndbøker beskrevet til å være et teamarbeid. Det pekes på viktigheten av å være flere som sammen jobber med prosjektet (Green & Higgins, 2011, Kap. 2). Dette er den største svakheten ved prosjektet, i og med den systematiske oversikten er et ledd i min masteroppgave. Jeg har igjennom hele prosessen samarbeidet med en medstudent, og har dermed ivaretatt dobbeltkontrollen på de kritiske punkt slik som utvelgelse, kritisk vurdering og så videre. Svakheten er at vi begge er uerfarne med prosessen, da det er første gang vi gjennomfører en systematisk oversikt. Allikevel mener jeg vi har ivaretatt kvaliteten ved å følge håndbøkene til Cochrane Collaboration og CRD tett i de ulike fasene av studien. Jeg har også hatt støtte av min veileder, som er erfaren med å skrive systematiske oversikter.

Inklusjons og eksklusjonskriterier PICOS

En annen svakhet ved denne systematiske oversikten er de forholdsvis snevre inklusjonskriteriene jeg har satt. En systematisk oversikt etter standarden fra Cochrane Collaboration søker å svare på alle relevante utfallsmål, samt fokusere på et bredt spørsmål, for at den systematiske oversikten skal være anvendbar i flere settinger (O'Connor, Green &

Higgins, 2011, Kap. 5). Dette prosjektet har hatt en begrenset populasjon, samt kun to typer av utfallsmål. Begrensingene har jeg gjort av kapasitetshensyn siden denne systematiske oversikten er en del av min masteroppgave. På den annen side mener jeg at ved kun å inkludere sykepleiere i denne systematiske oversikten, er resultatene presise nettopp på denne populasjonen. Det kunne videre vært relevant å evaluere utfallsmål som team-samarbeid og økonomi, dersom vi hadde vært flere som utarbeidet denne systematiske oversikten sammen. Jeg mener allikevel jeg har valgt viktige utfallsmål, da ferdigheter og kunnskap er primære utfallsmål i andre systematiske oversikter over dette temaet (Cant & Cooper, 2010 ; Ilgen, Sherbino & Cook, 2013 ; Cook et al., 2011).

Videre har jeg valgt å kun inkludere randomiserte kontrollerte studier. Fordelen med dette er at resultatene fra en RCT er mer pålitelige, på bakgrunn av oppbyggingen av designet (Bjørndal, Flottorp & Klovning, 2013, s. 65). Svakheten ved kun å inkludere RCT studier er at antallet studier er begrenset, og jeg kan ha oversett nyttig kunnskap evaluert i andre design, slikt som kontrollerte før-etter studier eller tidsserier. Inklusjon av ikke-randomiserte kontrollerte studier kunne ha bidratt til at de ulike sammenligningene hadde et større evidensgrunnlag. Jeg valgte allikevel bort disse studiedesignene, da de innebærer en risiko for systematiske feil grunnet svakheter i design. Dette gjenspeiles også i GRADE da andre design enn RCT starter på en lavere skår. Til sist har jeg valgt å kun inkludere studier på språk jeg selv kan forstå. Dette har innebåret at jeg har ekskludert studier på spansk, kinesisk og fransk som kunne ha vært aktuelle å ha med i denne systematiske oversikten. Dette ser jeg på som den største faren for inklusjons-bias i denne systematiske oversikten.

Litteratursøk

Jeg mener at litteratursøket til denne systematiske oversikten har vært grundig og av god kvalitet. Relevante søkeord er identifisert i flere ulike kilder og jeg har hatt en god prosess i utvelgelse av databaser. Søkene er også tilpasset til de ulike databasenes funksjonaliteter med tanke på emneord, operatører og syntakser. Alle litteratursøkene har også vært kvalitetssikret av en fagbibliotekar. Allikevel er det noen svakheter ved søkene.

Det er i denne systematiske oversikten utført søk i et begrenset antall databaser, da det i samråd med fagbibliotekar ble vurdert som liten sannsynlighet for å få relevante treff i andre databaser, eller fordi databasene ikke lå tilgjengelig til å gjennomføre søk i. Et eksempel på database som ikke er tilgjengelig via norske søkekanaler er Joanna Briggs Institute, som er en database som inneholder forskning av sykepleiere (Nortvedt et.al., 2012, s. 50-51). Det er

dermed en risiko for at relevante studier ikke er identifisert gjennom søkene, selv om jeg anser denne risikoen for å være liten. Søkene i denne systematiske oversikten er også utført i tilsvarende databaser som i de andre systematiske oversiktene jeg har funnet.

Videre er det ikke utført søk etter grå litteratur i dette prosjektet. Jeg kan dermed ha gått glipp av resultat som kun er presentert på konferanser, eller som ikke er publisert i vitenskapelige tidsskrift. Ved ikke å ha søkt på grå litteratur kan denne systematiske oversikten være rammet av bias ved at studier som ikke viser signifikante funn ikke er tatt med. Dette har spesielt betydning dersom en intervensjon viser effekt, da effekten vil overestimeres fordi de ikke-signifikante funnene ikke er tatt hensyn til i analysen (Polit & Beck, 2014, s. 358-359).

Allikevel mener jeg at jeg har belegg for ikke å ha søkt etter grå litteratur. De fleste studiene jeg har inkludert i oversikten viser ikke-signifikante resultat, noe som tyder på at også disse studiene publiseres på dette temaet. I følge Cochrane Handbook, og håndboken til Nasjonalt kunnskapssenter for heletjenesten er det heller ikke påkrevet å søke etter grå litteratur. På den annen side har jeg gjennomført søk på intervensjon i Clinical Trials for eventuelt å finne studier som ikke ble identifisert gjennom de systematiske søkene. Selv om søket i Clinical Trials ikke identifiserte relevante studier, styrker det dette prosjektet fordi jeg ikke fant protokoller på ikke-publiserte eller pågående studier. I tillegg mener jeg det styrker tiltroen til at litteratursøkene har vært grundige. Til sist kontaktet jeg eksperter på feltet med forespørsel om deres kjennskap til publisert og upublisert materiale. Jeg anerkjenner risikoen for inklusjons-bias i min systematiske oversiktsartikkel, men mener risikoen for slik bias er liten.

I litteratursøkene i de ulike databasene hadde jeg med tekstordet «nurs*». Jeg valgte å fravike dette i databasen CINAHL, da dette tekstordet ga en stor andel treff, og de fleste av disse var ikke relevante for emnet. Jeg valgte å beholde emneordet «Nurses». Denne avgjørelsen ble tatt i samråd med fagbibliotekar. Det er en risiko for at jeg har gått glipp av relevante studier, men jeg mener at jeg allikevel etter all sannsynlighet har fått med alle relevante treff.

Det ble gjennomført systematiske søk i alle de tidligere nevnte databasene utenom SveMed+, der det kun ble søkt på emne- og tekstord som omhandlet intervensjonen. Begrunnelsen for dette var at søket ga få treff i utgangspunktet, og det var dermed håndterbart og screene alle treffene. Dersom det hadde vært søkt på flere PICO element, hadde det vært en risiko for å ekskludere bort relevante studier.

Det er anbefalt et oppdateringssøk i de ulike databasene dersom det har gått lang tid fra det endelige søket er utført og til den systematiske oversikten publiseres (Nasjonalt

kunnskapssenter for helsetjenesten, 2013, s. 32-33). Det er ikke utført noe oppdateringssøk i dette prosjektet, men det planlegges å utføre et slikt søk før artikkelen skal publiseres.

Utvelgelse

Det er en styrke ved prosjektet at vi var to personer som gikk igjennom EndNote biblioteket med alle treffene fra litteratursøkene. Utvelgelsesfasen er et av de kritiske element når en systematisk oversikt utarbeides. Det er i denne fasen stor fare for å overse studier som skal inkluderes i oversikten. Derfor er det anbefalt alltid å være to personer som gjør denne utvelgelsen uavhengig av hverandre, før en i fellesskap går igjennom sine vurderinger (Higgins & Deeks, 2011, Kap. 7).

I søket identifiserte vi to studier som var duplikater av allerede inkluderte studier. Det er viktig at duplikatrapporter blir fanget opp, og ikke rapportert dobbelt. Dette gjelder spesielt når en lager meta-analyse, men er også viktig ved en narrativ syntese (Higgins & Deeks, 2011, Kap. 7). Om dubletter ikke blir fanget opp vil resultatene fra studien kunne få stor påvirkning på det samlede resultatet.

Det lyktes å skaffe til veies alle 35 artiklene som skulle leses i fulltekst, og dermed ble ingen studier ekskludert med bakgrunn i at vi ikke fikk tak på dem. Dersom noen artikler ikke hadde latt seg skaffe, ville den systematiske oversikten rammes av inklusjonsbias, som betyr at resultatene fra en studie ikke blir rapportert fordi studien ikke er med i analysen (Centre for Reviews and Dissemination, 2008, s. 23-24).

Vi sikret at vi hadde samme forståelse av inklusjons- og eksklusjonskriteriene ved å pilotere utvelgelsen med de 50 første referansene. Jeg kunne valgt å regne ut kappa-verdi for ulikhetene, men fordi vi begge skulle gå igjennom alle titlene videre i prosessen valgte jeg og ikke gjøre disse beregningene. Kappa er en enkel statistisk beregning der en score mellom 0,6-0,7 reflekterer god enighet, mens score over 0,75 viser stor grad av enighet (Higgins & Deeks, 2011, Kap. 7). Det er ikke standard prosedyre å regne ut kappa i systematiske oversikter, og jeg mener at det tette samarbeidet jeg og min medstudent hadde var tilstrekkelig for å sikre at vi fanget opp de relevante studiene.

Min medstudent og jeg har ulik yrkesbakgrunn, og ulik bakgrunn med simuleringstrening. Dette medfører at vi er en fagekspert, og en ikke-ekspert som velger ut relevante studier. Det er anbefalt en slik strategi, da det øker sannsynligheten for at alle studier som skal inkluderes

blir inkludert i den systematiske oversikten. Dette blant annet fordi den som ikke er ekspert på faget leser titler og abstract på en mindre forutinntatt måte (Higgins & Deeks, 2011, Kap. 7).

Jeg mener at vår grundighet i utvelgelsesprosessen, samt at vi var to personer i alle ledd gjør det overveiende sannsynlig at alle studiene som skulle inkluderes faktisk er blitt inkludert. Dermed er dette en styrke for min systematiske oversikt. Allikevel er det alltid en mulighet for at vi har oversett studier som burde vært med, noe som gjør at en aldri kan si med sikkerhet at alle relevante studier er inkludert i en systematisk oversikt.

Jeg hadde en forventning om at jeg ville finne flere studier som hadde sett på effekten av simuleringstrening for sykepleiere. Jeg hadde også antatt at det ville være studier fra de nordiske landene, samt flere studier fra akuttmedisinske settinger. En grunn til at jeg ikke fant dette, kan være at jeg begrenset inklusjonen til designet RCT. Jeg antar jeg ville funnet flere studier ved å ha en bredere inklusjon på studiedesign., for eksempel studier uten kontrollgrupper.

Kritisk vurdering

Kritisk vurdering av de inkluderte studiene er gjennomført ved bruk av Risk of Bias Tool utarbeidet av Cochrane Collaboration. Dette er et verktøy som er godt egnet til å oppdage systematiske svakheter ved studiene, og det er derfor en styrke ved prosjektet at dette er benyttet (Higgins, Altman & Sterne, 2011, Kap. 8). Jeg kunne valgt å tilpasse dette verktøyet til mitt prosjekt, men dette ble ikke gjort da jeg mente verktøyet dekket de viktige element ved vurderingen. Tidligere systematiske oversikter har benyttet Medical Education Research Study Quality Instrument (MERSQI) (Cook et al., 2011 ; Ilgen, Sherbino & Cook, 2013) og Critical Appraisal Skills Programme (CASP) (Cant & Cooper, 2010) til kritisk vurdering. MERSQI (Reed et al., 2007) og CASP (Critical Appraisal Skills Programme, 2014) er sjekklister som er forholdsvis likt oppbygget som RoB, og jeg mener verktøyet jeg har brukt er godt tilpasset dette prosjektet.

Det er også en styrke ved prosjektet at vi er to personer som uavhengig av hverandre har kritisk vurdert artiklene, før vi diskuterte resultatene. I forkant av den kritiske vurderingen fikk vi også veiledning av en person med erfaring i bruk av RoB, slik at vi var bedre rustet til denne delen av prosessen.

Jeg gjennomførte søk etter de inkluderte studienes protokoller i Clinical Trials, noe som også har vært en styrke for den kritiske vurderingen med tanke på selektiv rapportering i studiene.

Gradering av dokumentasjon

GRADE er benyttet for å vurdere tiltroen til effektestimaterne fra de ulike sammenligningene. Dermed har dette vært en styrke i vurderingen av resultatene. I tillegg vil dette være en fordel når leserne skal vurdere resultatene i den systematiske oversikten (Guyatt et al., 2008a). Svakheten ved dette prosjektet er igjen uerfarenhet med bruk av GRADE, og dermed en risiko for å være for mild i vurderingen av kvaliteten. De andre systematiske oversiktene jeg har identifisert har ikke gradert dokumentasjonen. Jeg mener det er en styrke ved denne oversikten at GRADE er benyttet.

Analyse

Dataekstraksjon fra de ulike studiene er utført av en person, og det er foretatt en kvalitetskontroll av uttrekket av en annen. Det ville ha styrket studien om to personer uavhengig av hverandre hadde trukket ut data uavhengig av hverandre, for så å diskutere resultatet (Higgins & Deeks, 2011, Kap. 7).

Det er gjennomført en meta-analyse på sammenligningen simuleringstrening versus ingen simulering. Selv om det kun inngår tre studier i meta-analysen og to studier i sensitivitetsanalysen blir effektestimaterne mer presise enn om disse var blitt presentert alene (Borenstein et al., 2009, s. 9). Det har blitt vurdert mulighet for meta-analyse for de andre sammenligningene knyttet til ferdigheter eller kunnskaper, men jeg har funnet at studiene i disse sammenligningene er for ulike til å sammenstilles. Dette blant annet fordi populasjonen skiller seg vesentlig. Eksempelvis for sammenligningen simuleringstrening versus ingen simulering på utfallet kunnskaper, er populasjonen i studiene til Cioffi et.al. (2005) og Weiner et.al. (2013) spesialsykepleiere, mens Maneval et.al. (2012) har studert nyutdannede sykepleiere. Jeg har derfor laget et forest plot med subgruppe uten å sammenstille dataene. Jeg mener det er en styrke for den systematiske oversikten at jeg har laget en meta-analyse inkludert en sensitivitetsanalyse. Samtidig mener jeg det er en styrke at det ikke er utarbeidet en meta-analyse på de andre sammenligningene. Dette fordi en da ville ha utført en sammenligning på studier som ikke er like nok klinisk til å sammenstilles. Jeg har allikevel valgt å lage forest plot på disse sammenligningene. Dermed vil leseren få et visuelt inntrykk av studienes resultater, uten av effektestimaterne er sammenstilt (pooled). Sammenligningene som er utført i denne systematiske oversikten er tilsvarende sammenligninger gjort i andre systematiske oversikter. Dette støtter at riktige sammenligninger er gjort.

Meta-analysen og forest plotene er utarbeidet av meg, som er uerfaren med slike statistiske sammenstillinger. Dette gir en mulighet for at feil data er trukket ut, eller at feil analysemetode er valgt. For å sikre kvaliteten på disse analysene har veileder med kompetanse i meta-analyse blitt konsultert for vurdering av analysene. Jeg føler meg dermed sikker på at riktige data er tatt inn i beregningene. Jeg har videre utført en random effects analyse selv om det er få studier i meta-analysen. Jeg mener allikevel det er belegg for dette, i og med studiene har målt utfallet med ulike verktøy og i litt ulike settinger, og jeg dermed kan forvente en viss variasjon i estimatene. Jeg forsøkte også både fixed effect og random effect, og fikk ingen forskjell i resultater (Borenstein et al., 2009, s. 83-86).

Forfatterne av primær-artiklene er ikke blitt kontaktet for manglende data. Det kan være aktuelt å kontakte forfatterne dersom data mangler i artiklene for å gjennomføre analysen (Higgins & Deeks, 2011, Kap. 7). Dette kunne f.eks. vært aktuelt for studien til De Regge et.al. (2008), da disse har presentert resultatene i median med interkvartil bredde, uten å oppgi alle dataene. På bakgrunn av denne usikkerheten ønsket jeg ikke å gjøre disse dataene om til mean, og jeg fikk dermed ikke utarbeidet forest plot for denne studien. Kontakt med forfatterne kunne vært utført for å øke kvaliteten på denne systematiske oversikten, og er dermed en svakhet ved prosjektet.

Konklusjon

Funnene i denne systematiske oversikten gir ikke konklusive svar på om simuleringstrening er effektivt for å øke sykepleieres kunnskaper og ferdigheter. Meta-analysen som er utført viser resultat i favør av simuleringstrening på sykepleieres ferdigheter, når simulering er sammenlignet med standard opplæringsprogram. Imidlertid er kvaliteten på dokumentasjonen lav. For utfallet kunnskap viser effektestimaterne ingen forskjell mellom gruppene som får simuleringstrening eller ikke. De andre sammenligningene vi har gjort viser heller ingen konklusive resultater på de utfallene vi har målt. Kvaliteten på dokumentasjonen bak sammenligningene blitt gradert til lav. Denne nedgraderingen skyldes moderat til høy risiko for systematiske feil i flere av studiene, samt inkonsistens i studienes resultater. I tillegg har flere av studiene få deltakere, noe som kan ha medført at de ikke har klart å vise signifikante resultat. Den lave kvaliteten på dokumentasjonen gjør at vi ikke har tiltro til at effektestimaterne på sammenligningene vi har gjort viser den sanne effekten. I og med simuleringstrening viser usikker effekt, kan det også være aktuelt å vurdere andre kvalitetsforbedringsstrategier, slik som arbeidsgrupper eller audit.

Simulering som kvalitetsforbedringsstrategi er en av flere ulike metoder etablert i norsk helsevesen. I slike komplekse system, er det behov for flere ulike tiltak tilpasset konteksten der endringen som skal skje. Sammen med for eksempel klinisk audit og workshops har simuleringstrening etter mitt syn en viktig plass.

Denne systematiske oversikten er utført som et ledd i min masteroppgave, noe om medfører en risiko for flere svakheter i gjennomføringen. Jeg mener allikevel at kvaliteten er ivaretatt på en god måte ved å følge anbefalingene fra håndbøkene jeg har brukt som veiledere. Jeg har hatt samarbeidspartner i de delene av prosessene en skal være to, og jeg har innhentet veiledning fra eksperter der dette har vært nødvendig. Jeg mener dermed at jeg leverer en systematisk oversikt av god kvalitet. Basert på dette mener jeg at funnene i den systematiske oversikten ikke har redusert min tiltro til at simuleringstrening er effektivt som læringsform. For ferdigheter synes det som om høyteknologisk simulering gir en bedre effekt, mens for kunnskaper synes effekten å være best med e-lærings simulering. Imidlertid er dette usikkert grunnet lav kvalitet på dokumentasjonen, og det er behov for nye randomiserte kontrollerte studier av god kvalitet for å dokumentere om simuleringstrening er effektiv læringsform for sykepleiere.

Referanseliste

Arnold, J. J., Johnson, L. M., Tucker, S. J., Chesak, S. S. & Dierkhising, R. A. (2013) Comparison of Three Simulation-Based Teaching Methodologies for Emergency Response. *Clinical Simulation In Nursing*, 9 (3), s. e85-e93.

Aspden, P., Corrigan, J. M., Wolcott, J. & Erickson, S. M. red. (2004) *Patient Safety: Achieving a New Standard for Care*. Institute of Medicine, Washington, DC, The National Academies Press.

Baker, R., Camosso-Stefinovic, J., Gillies, C., Shaw, E. J., Cheater, F., Flottorp, S., Robertson, N., Wensing, M., Fiander, M., Eccles, M. P., Godycki-Cwirko, M. & van Lieshout J, J. C. (2015) Tailored interventions to address determinants of practice. *Cochrane Database of Systematic Reviews 2012*, (4),

Bakkeland, J. & Thorsen, B. H. (2010) Lungesvikt. I: Gulbrandsen, T. & Stubberud, D.-G. red. *Intensivsykepleie*. Oslo, Akribe, s. 311-442.

Batalden, P. B. & Davidoff, F. (2007) What is “quality improvement” and how can it transform healthcare? *Quality and Safety in Health Care*, 16 (1), s. 2-3.

Benner, P. E., Hooper-Kyriakidis, P. L., Stannard, D. & Benner, P. E. (2011) *Clinical wisdom and interventions in acute and critical care : a thinking-in-action approach*. New York, Springer Pub.

Bjørndal, A., Flottorp, S. & Klovning, A. (2013) *Kunnskapshåndtering i medisin og helsefag*. Oslo, Gyldendal akademisk.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009) *Introduction to meta-analysis*. Chichester, Wiley.

Cant, R. P. & Cooper, S. J. (2010) Simulation-based learning in nurse education: systematic review. *Journal of Advanced Nursing*, 66 (1), s. 3-15.

Carrasco-Labra, A., Brignardello-Petersen, R., Santesso, N., Neumann, I., Mustafa, R., Mbuagbaw, L., Ikobaltzeta, I., De Stio, C., McCullagh, L., Alonso-Coello, P., Meerpohl, J., Vandvik, P., Brozek, J., Akl, E., Bossuyt, P., Churchill, R., Glenton, C., Rosenbaum, S., Tugwell, P., Welch, V., Guyatt, G. & Schunemann, H. (2015) Comparison between the standard and a new alternative format of the Summary-of-Findings tables in Cochrane review users: study protocol for a randomized controlled trial. *Trials*, 16 (1), s. 164.

Centre for Reviews and Dissemination (2008) *Systematic Reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination, University of York, 2008.

Centre for Reviews and Dissemination (2015) *About the databases; About HTA* [Internett], National Institute for Health Research, University of York. Tilgjengelig fra: <<http://www.crd.york.ac.uk/CRDWeb/AboutPage.asp>> [Nedlastet 03.03.15].

Cioffi, J., Purcal, N. & Arundell, F. (2005) A pilot study to investigate the effect of a simulation strategy on the clinical decision making of midwifery students. *Journal of Nursing Education*, 44 (3), s. 131-4.

ClinicalTrials.gov (2014) [Internett], The National Library of Medicine; The National Institutes of Health. Tilgjengelig fra: <<https://clinicaltrials.gov/>> [Nedlastet 03.03.15].

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J. & Hamstra, S. J. (2011) Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA*, 306 (9), s. 978-988.

Critical Appraisal Skills Programme (2014) *CASP Checklists* [Internett], Oxford. CASP. Tilgjengelig fra: <<http://www.casp-uk.net/>> [Nedlastet 03.04.15].

De Regge, M., Calle, P. A., De Paepe, P. & Monsieurs, K. G. (2008) Basic life support refresher training of nurses: individual training and group training are equally effective. *Resuscitation*, 79 (2), s. 283-7.

De Regge, M., Monsieurs, K. G., Valcke, M. & Calle, P. A. (2012) Training Nurses in a Self-Learning Station for Resuscitation: Factors Contributing to Success or Failure. *JEN: Journal of Emergency Nursing*, 38 (4), s. 386-391.

Deeks, J. J., Higgins, J. P. T. & Altman, D. G. (2011) Analysing data and undertaking meta-analyses. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration 2011.

DiCenso, A., Bayley, L. & Haynes, R. B. (2009) Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. *Evidence Based Nursing*, 12 (4), s. 99-101.

DiCenso, A., Guyatt, G. & Ciliska, D. (2005) *Evidence-based nursing: a guide to clinical practice*. St. Louis, Elsevier Mosby.

Einarson, E., Jaeger, M., Melby, A. K. & Brudvik, M. (2011) *ALERTTM* [Internett], Kvalitetsforbedring - Pasientsikkerhet- Helsebiblioteket.no. Tilgjengelig fra: <file:///C:/Users/Acer/Downloads/www_helsebiblioteket_no_Kvalitetsforbedring_Pasientsikkerhet.pdf> [Nedlastet 16.03.15].

Flottorp, S. & Aakhus, E. (2013) Implementeringsforskning: Vitenskap for forbedring av praksis. *Norsk Epidemiologi*, 23 (2), s. 187-196.

Forsetlund, L., Bjørndal, A., Rashidian, A., Jamtvedt, G., O'Brien, M. A., Wolf, F. M., Davis, D., Odgaard-Jense, J. & Oxman, A. D. (2009) Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews 2009*, (2),

Forskrift om håndtering av medisinsk utstyr. (2014) *Forskrift om håndtering av medisinsk utstyr*. Justis- og beredskapsdepartementet, Helse- og omsorgsdepartementet. Fastsatt ved kgl.res. 29. november 2013. Nedlastet 12.04.15 fra: <https://lovdata.no/dokument/SF/forskrift/2013-11-29-1373>

Giguère, A., Légaré, F., Grimshaw, J., Turcotte, S., Fiander, M., Grudniewicz, A., Makosso-Kallyth, S., Wolf, F. M., Farmer, A. P. & Gagnon, M.-P. (2012) Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* (10),

Glasziou, P., Ogrinc, G. & Goodman, S. (2011) Can evidence-based medicine and clinical quality improvement learn from each other? *BMJ Quality & Safety*, 20 Suppl 1, s. i13-17.

Goldet, G. & Howick, J. (2013) Understanding GRADE: an introduction. *Journal of Evidence-Based Medicine*, 6, s. 50-54.

Gordon, J. A. & Vozenilek, J. A. (2008) 2008 Academic Emergency Medicine Consensus Conference. *On Behalf of the Saem Simulation Task Force Interest Group the Technology in Medical Education, Committee, Academic Emergency Medicine*, 15 (11), s. 971-977.

GRADE Working Group (2012) *Guideline Development Tool (GDT)* [Internett]. Tilgjengelig fra: <<http://www.guidelinedevelopment.org/>> [Nedlastet 16.02.2015].

Graham, I. D., Logan, J., Harrison, M. B., Straus, S. E., Tetroe, J., Caswell, W. & Robinson, N. (2006) Lost in knowledge translation: time for a map? *Journal of Continuing Education in the Health Professions*, 26 (1), s. 13-24.

Green, S. & Higgins, J. P. T. (2011) Preparing a Cochrane Review. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration 2011.

Green, S., Higgins, J. P. T., Alderson, P., Clarke, M., Mulrow, C. D. & Oxman, A. D. (2011) Introduction. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0* The Cochrane Collaboration 2011.

Grol, R., Ouwens, M. & Wollersheim, H. (2013) Planning and organizing the change process. I: Grol, R., Wensing, M., Eccles, M. & Davis, D. red. *Improving patient care: the implementation of change in health care*. 2. utg. Chichester,, Wiley-Blackwell.

Grol, R. & Wensing, M. (2013a) Characteristics of successful innovations. I: Grol, R. & Wensing, M. red. *Improving patient care: the implementation of change in health care*. 2. utg. Chichester, Wiley-Blackwell.

Grol, R. & Wensing, M. (2013b) Implementation of change in healthcare: a complex problem. I: Grol, R., Wensing, M., Eccles, M. & Davis, D. red. *Improving patient care: the implementation of change in health care*. 2. utg. Chichester, Wiley-Blackwell.

Grol, R., Wensing, M., Bosch, M., Hulscher, M. & Eccles, M. (2013) Theories on implementation of change in healthcare. I: Grol, R., Wensing, M., Eccles, M. & Davis, D. red. *Improving patient care: the implementation of change in health care*. 2. utg. Chichester, Wiley-Blackwell.

Guyatt, G. H., Oxman, A. D., Kunz, R., Vist, G. E., Falck-Ytter, Y. & Schünemann, H. J. (2008a) What is “quality of evidence” and why is it important to clinicians? 336 (7651), s. 995-998.

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P. & Schünemann, H. J. (2008b) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. 336 (7650), s. 924-926.

Helse- og Omsorgsdepartementet. (2012) *God kvalitet – trygge tjenester. Kvalitet og pasientsikkerhet i helse- og omsorgstjenesten. St.meld. 10 (2012-2013)*. Oslo, Helse- og Omsorgsdepartementet. Nedlastet 02.04.15 fra:
<https://www.regjeringen.no/nb/dokumenter/meld-st-10-20122013/id709025/>

Helse- og Omsorgsdepartementet. (2013) *Kvalitet og pasientsikkerhet 2013; St. meld. 11 (2014-2015)*. Oslo, Helse- og Omsorgsdepartementet. Nedlastet 03.05.15 fra:
<https://www.regjeringen.no/contentassets/4db4ebe7d94a4687946baa063cd683d5/no/pdfs/stm201420150011000dddpdfs.pdf>

Helsedirektoratet (2011) *Den nasjonale pasientsikkerhetskampanjen I trygge hender 24-7* [Internett]. Tilgjengelig fra: <<http://www.pasientsikkerhetsprogrammet.no/>> [Nedlastet 16.03.15].

Higgins, J. P. T., Altman, D. G. & Sterne, J. A. C. (2011) The Cochrane Collaboration's tool for assessing risk of bias. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration 2011.

Higgins, J. P. T. & Deeks, J. J. (2011) Selecting studies and collecting data. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0*. The Cochrane Collaboration 2011.

Higgins, J. P. T. & Green, S. (2011) *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. The Cochrane Collaboration 2011.

Hollis, S. & Campbell, F. (1999) *What is meant by intention to treat analysis? Survey of published randomised controlled trials*.

Hviding, K., A, A.-D., Holte, H. H. & Flottorp, S. (2009) Simulering i spesialistutdanning av leger - evaluering av effekt og kostnadseffektivitet. Kunnskapsoppsummering. Rapport Nr 4-2009. Oslo. Nasjonalt kunnskapssenter for helsetjenesten

Høgskolen i Gjøvik (2011) *Simulering* [Internett]. Tilgjengelig fra: <http://www.hig.no/simsenter/simulering> [Nedlastet 26.04.15].

Ilgen, J. S., Sherbino, J. & Cook, D. A. (2013) Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. *Acad Emerg Med*, 20 (2), s. 117-27.

Institute for Healthcare Improvement (2015) *Patient Safety* [Internett], Improving Health and Health Care Worldwide. Tilgjengelig fra: <http://www.ihl.org/Topics/PatientSafety/Pages/default.aspx> [Nedlastet 24.04.15].

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M. A., Johansen, M., Grimshaw, J. & Oxman, A. D. (2012) Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012, (6),

Jamtvedt, G. (2013) Systematiske oversikter om effekt av tiltak. *Norsk Epidemiologi*, 23 (2), s. 119-124.

Johnson, D., Corrigan, T., Gulickson, G., Holshouser, E. & Johnson, S. (2012) The effects of a human patient simulator vs. a CD-ROM on performance. *Military medicine*, (10), s. 1131-5.

Johnson, D. & Johnson, S. (2014) The Effects of Using a Human Patient Simulator Compared to a CD-ROM in Teaching Critical Thinking and Performance. *US Army Medical Department Journal*, s. 59-64.

Lefebvre, C., Manheimer, E. & Glanville, J. (2011) Searching for studies. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration 2011.*

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J. & Moher, D. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 339:b2700,

Maneval, R., Fowler, K. A., Kays, J. A., Boyd, T. M., Shuey, J., Harne-Britner, S. & Mastrine, C. (2012) The effect of high-fidelity patient simulation on the critical thinking and clinical decision-making skills of new graduate nurses. *Journal of Continuing Education in Nursing*, 43 (3), s. 125-134.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6 (7), s. 1-6.

Monsieurs, K. G. a., De Regge, M. b., Schelfout, S. a., D'Hondt, F. a., Mpotos, N. a., Valcke, M. c. & Calle, P. A. a. (2012) Efficacy of a self-learning station for basic life support refresher training in a hospital: a randomized controlled trial. *European Journal of Emergency Medicine*, 19 (4), s. 214-219.

Mustafa, R. A., Santesso, N., Brozek, J., Akl, E. A., Walter, S. D., Norman, G., Kulasegaram, M., Christensen, R., Guyatt, G. H., Falck-Ytter, Y., Chang, S., Murad, M. H., Vist, G. E., Lasserson, T., Gartlehner, G., Shukla, V., Sun, X., Whittington, C., Post, P. N., Lang, E., Thaler, K., Kunnamo, I., Alenius, H., Meerpohl, J. J., Alba, A. C., Nevis, I. F., Gentles, S., Ethier, M. C., Carrasco-Labra, A., Khatib, R., Nesrallah, G., Kroft, J., Selk, A., Brignardello-Petersen, R. & Schunemann, H. J. (2013) The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *Journal of Clinical Epidemiology*, 66 (7), s. 736-42.

Nasjonalt kunnskapssenter for helsetjenesten (2012) *Meldeordningen* [Internett]. Tilgjengelig fra: <<http://www.melde.no/>> [Nedlastet 24.04.15].

Nasjonalt kunnskapssenter for helsetjenesten (2013) *Slik oppsummerer vi forskning. Håndbok for Nasjonalt kunnskapssenter for helsetjenesten*. 3.2 revidert. utg. Oslo, Nasjonalt kunnskapssenter for helsetjenesten.

Nasjonalt kunnskapssenter for helsetjenesten (2014) *Modell for kvalitetsforbedring* [Internett]. Tilgjengelig fra: <<http://www.helsebiblioteket.no/Kvalitetsforbedring/Slik-kommer-du-i-gang/modell-for-kvalitetsforbedring>> [Nedlastet 16.03.15].

Nortvedt, M. W., Jamtvedt, G., Graverholt, B., Nordheim, L. V. & Reinar, L. M. (2012) *Jobb kunnskapsbasert!: en arbeidsbok*. [Oslo], Akribe.

O'Connor, D., Green, S. & Higgins, J. P. T. (2011) Defining the review question and developing criteria for including studies. I: Higgins, J. P. T. & Green, S. red. *Cochrane Handbook for systematic reviews of interventions version 5.1.0* The Cochrane Collaboration 2011.

Polit, D. F. & Beck, C. T. (2012) *Nursing research: generating and assessing evidence for nursing practice*. 9. utg. Philadelphia, Pa., Wolters Kluwer Health.

Polit, D. F. & Beck, C. T. (2014) *Essentials of nursing research: appraising evidence for nursing practice*. 8. utg. Philadelphia, Wolters Kluwer/Lippincott Williams & Wilkins.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E. & Wright, S. M. (2007) Association between funding and quality of published medical education research. *Jama*, 298 (9), s. 1002-9.

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B. & Richardson, W. S. (1996) Evidence-based medicine: What it is and what it isn't. *British Medical Journal*, 321, s. 71-72.

SAFER (2015) *Hva er medisinsk simulering?* [Internett], Stavanger Acute medicine Foundation for Education and Research (SAFER). Tilgjengelig fra: <<http://www.safer.net/simulering/>> [Nedlastet 11.03.15].

Salas, E., Paige, J. T. & Rosen, M. A. (2013) Creating new realities in healthcare: the status of simulation-based training as a patient safety improvement strategy. *BMJ Quality and Safety*, 22, s. 449-452.

Smedslund, G. (2013) Metaanalyse. *Norsk Epidemiologi*, 23 (2), s. 147-149.

Society for Simulation in Healthcare (2014) [Internett]. Tilgjengelig fra: <<http://www.ssih.org/>> [Nedlastet 03.04.15].

Sosial- og Helsedirektoratet. (2005) ... *OG BEDRE SKAL DET BLI! Nasjonal strategi for kvalitetsforbedring i Sosial- og helsetjenesten*.
http://www.ogbedreskaldetbli.no/Om_kvalitetsstrategien/IS-1162_4390a.pdf

Stubberud, D.-G. O. (2010) Intensivsykepleierens funksjons- og ansvarsområder. I: Gulbrandsen, T. & Stubberud, D.-G. red. *Intensivsykepleie*. Oslo, Akribe, s. 32-52.

The International Nursing Association for Clinical Simulation & Learning (2013) [Internett]. Tilgjengelig fra: <<http://www.inacsl.org/i4a/pages/index.cfm?pageid=1>> [Nedlastet 03.04.15].

The Nordic Cochrane Centre (2014) Review Manager 5.3 (RevMan). The Cochrane Collaboration. I.

Vist, G. E., Sæterdal, I., Vandvik, P. O. & Flottorp, S. A. (2013) Gradering av kvaliteten på dokumentasjonen. *Norsk Epidemiologi*, 23 (2), s. 151-156.

Weiner, G. M., Menghini, K., Zaichkin, J., Caid, A. E., Jacoby, C. J. & Simon, W. M. (2011) Self-directed Versus Traditional Classroom Training for Neonatal Resuscitation. *Pediatrics*, 127 (4), s. 713-719.

Effect of simulation-based training on nurses' knowledge and skills;
A systematic review and meta-analysis

Kandidat: Pål André Hegland

Masterprogram: Kunnskapsbasert praksis i helsefag

Avdeling: Avdeling for helse- og sosialfag, Senter for kunnskapsbasert praksis

Utdanningsinstitusjon: Høgskolen i Bergen

Planlagt publisert i: Clinical Simulation in Nursing

Forfatterveiledningen til tidsskriftet Clinical Simulation in Nursing har vært brukt som basis i utarbeidelse av artikkelen. Forfatterveiledningen er tilgjengelig fra:

<http://www.nursingsimulation.org/content/authorinfo#idp1378096>

Effect of simulation-based training on nurses' knowledge and skills; A systematic review and meta-analysis

Hegland, Pål A.³, Aarlie, Hege⁴, Strømme, Hilde⁵, Jamtvedt, Gro⁶

Centre of Evidence-Based Practice, Faculty of Health and Social Sciences Bergen University College, Bergen, Norway

Background: Simulation-based training is a widespread strategy to improve health-care quality. However the effectiveness on registered nurses has previously not been established in systematic reviews. The aim of this systematic review is to evaluate effectiveness of simulation-based training on nurse's skills and knowledge

Method: We searched CDSR, DARE, HTA, CENTRAL, CINAHL, MEDLINE, Embase, ERIC and SveMed+ for randomised controlled trials (RCTs) evaluating effect of simulation-based training among nurses. Searches were completed in September 2014. Screening of abstracts and full-text, data extraction and assessment of risk of bias were independently conducted by two reviewers. We made the comparisons simulation-based training versus no simulation, high-fidelity simulation versus other simulation strategy and different organisation of simulation training. Data were analysed through meta-analysis and narrative syntheses. GRADE was used to assess quality of evidence.

Results: Seven RCT's met the inclusion criteria. For the comparison simulation-based training to no simulation on nurses' skills we pooled effect sizes using a random effects model. Our meta-analysis showed no significant differences between the groups (SMD -1.23, CI -2.69 to 0.23). There was large heterogeneity (I^2 92 %). For the other comparisons there was large between-study variation in results. The quality of evidence for all comparisons was graded as low.

Conclusion: The effect of simulation-based training varies substantially between the included studies. Our meta-analysis showed no significant difference of simulation training compared to no simulation, but the quality of evidence was low indicating uncertainty. Other comparisons showed inconsistency in results. Based on our findings we cannot conclude if simulations-based training increases nurses' skills and knowledge.

Keywords: Simulation, Technology-enhanced, Nurses, RCT, Knowledge, Skill, Systematic Review.

³ Corresponding author – pal.a.helgand@hotmail.com

⁴ Second reviewer. Bergen University College, Bergen, Norway

⁵ Librarian, Norwegian Knowledge Centre for the Health Services, Oslo, Norway

⁶ Norwegian Knowledge Centre for the Health Services, Oslo, Norway

Introduction

Healthcare services offer complex and advanced treatment for patients. Therefore highly competent and skilled healthcare providers are needed to secure patient safety (Grol, Berwick, & Wensing, 2008). Studies show that errors in healthcare are a risk for patient safety, and can in many cases be prevented (Patel, Kannampallil, & Shortliffe, 2015). Patient safety and quality improvement therefore have a large focus in today's society (Institute for Healthcare Improvement, 2015). There are several tools for quality improvement such as evidence based guidelines or clinical audits (Ivers et al., 2012; NICE, 2014). Another strategy used to improve performance among health care workers and students is simulation-based training.

Simulation-based training is practicing realistic scenarios using a specialized manikin, computer software or humans playing the role as patient (Society for Simulation in Helathcare, 2014; The International Nursing Association for Clinical Simulation & Learning, 2013). The setting can be high-fidelity where manikins and equipment are advanced, also called technology-enhanced simulation. It can also be low-fidelity where the equipment is less advanced (Healthy Simulation, 2014; Salas, Paige, & Rosen, 2013). The most advanced manikins today simulate the physiology of humans with pulse, blood pressure and secretion of sweat and tears. The facilitator has the ability to regulate the parameters according to the actions initiated by the healthworkers, using specialized computer software (Healthy Simulation, 2014).

Previously published systematic reviews on simulation-based training for students in health-profession education, showed large effects on students' knowledge and skills and moderate effects on patient-related outcomes (Cant & Cooper, 2010; Cook et al., 2011). Simulation-based training for critical care nurses in continuing education programmes, can improve adherence to recommendations about safe medication (Jansson, Kääriäinen, & Kyngäs, 2012). Further technology-enhanced simulation in emergency medicine seems to have effect on several outcomes (Ilgen, Sherbino, & Cook, 2013).

One systematic review has summarized evidence for graduated nurses separately, and found only one cohort study. Since this learning strategy is widely used for nurses, it is relevant to evaluate its effectiveness. The aim of this systematic review is therefore to summarize the effect of simulation-based training on nurses' knowledge and skills.

Methods

Inclusion Criteria

To be relevant for inclusion in the systematic review, the studies had to be randomised controlled trials (RCTs) evaluating effectiveness of simulation-based training for graduated nurses or graduated nurses in continuing education. Skills and/or knowledge had to be the primary outcomes in the trials. The studies were eligible for inclusion if they were written in English, German, Norwegian, Swedish or Danish.

Comparisons

Relevant comparisons for the systematic review are simulation-based training to no simulation, different simulation strategies compared to each other, or different organisation of the simulation training.

Identification of studies

Systematic searches was performed in The Cochrane Database of Systematic Reviews (Wiley), Database of Abstracts of Reviews of Effects (CRD), Health Technology Assessment Database (CRD), Cochrane Central Register of Controlled Trials (Wiley), CINAHL (EBSCO), MEDLINE (OVID), Embase (OVID), ERIC (EBSCO) and SveMed+. Final searches were performed 5 September 2014. Search strategies were reviewed by an experienced research librarian (Sampson et al., 2009). Search terms included “simulation”, “technology-enhanced simulation”, “computer-based simulation”, “nurs*”, “skills”, “knowledge” among others. No limitations pertaining to language, publication year, or study design were applied. Complete search strategies are presented in Appendices 1-7. Hand searches were performed in the journals *Clinical Simulation in Nursing* and *Simulation in Healthcare* for the years 2013 and 2014 as they were indexed in MEDLINE until January 2013 at the time of search. Previously identified systematic reviews and primary studies were screened for relevant references. Experts in the field were contacted for additional published or unpublished research. Finally Clinical Trials (ClinicalTrials.gov, 2014) were searched with the text word “simulation”, to identify unpublished or ongoing studies.

Screening and selection of studies

All titles and abstracts were screened by two review authors independently. The selection process was piloted by reading the first 50 titles and abstracts to calibrate understanding of

inclusion- and exclusion criteria (Higgins & Deeks, 2011, Ch. 7). We obtained full-text articles of all studies that did not clearly meet the exclusion criteria. The same two reviewers read all full-text articles for final inclusion. Disagreements in all stages were solved by discussion until consensus was reached (Higgins & Deeks, 2011, Ch. 7).

Data extraction

A pre-defined data-extraction form was developed. Data from the included studies were extracted by one person and quality checked by a second person. We extracted the following data: author name, publication year, number of participants, interventions and comparisons, outcomes, country and effect measures.

Assessing risk of bias and grading the evidence

The Cochrane Collaborations Risk of Bias Tool (Higgins, Altman, & Sterne, 2011, Ch. 8.5), was used to evaluate risk of bias in included studies. We used Guideline Development Tool (GRADE Working Group, 2012; Guyatt et al., 2011) to assess quality of the evidence for the following comparisons: Simulation-based training versus no simulation strategy, high-fidelity simulation versus other simulation strategy and different organisation of simulation training. The grading was made per comparison for each outcome, and was assessed as high, moderate, low or very low quality (Guyatt et al., 2008).

Data synthesis

We planned to do a quantitative synthesis, by conducting meta-analyses when there was low clinical diversity in the studies (Deeks, Higgins, & Altman, 2011, Ch. 9). We also planned narrative syntheses if meta-analyses were not possible to conduct. The statistical data were entered to Review Manager 5.3 (The Nordic Cochrane Centre, 2014).

For continuous measures we calculated Standardized Mean Difference (Inverse Variance, random effects model) and 95% confidence interval (CI), and for dichotomous measures we calculated Risk Ratio with 95% CI (Mantel-Haenszel, random effects model) (Deeks et al., 2011, Ch. 9).

In the meta-analysis between-study consistency (heterogeneity) was calculated with I^2 statistics, which estimates the percentage of the variability not due to chance. An I^2 value $>50\%$ indicates substantial heterogeneity. A p-value was also calculated, and $p < 0.05$ indicates significant between-study heterogeneity (Deeks et al., 2011, Ch. 9).

Results

Identification of studies and study selection

942 potentially relevant studies were identified through the database searches, hand searching and screening of reference lists. 35 articles were selected for full-text reading. Seven studies met the inclusion criteria and were included in this systematic review, see Figure 1. Excluded studies are presented in Appendix 8.

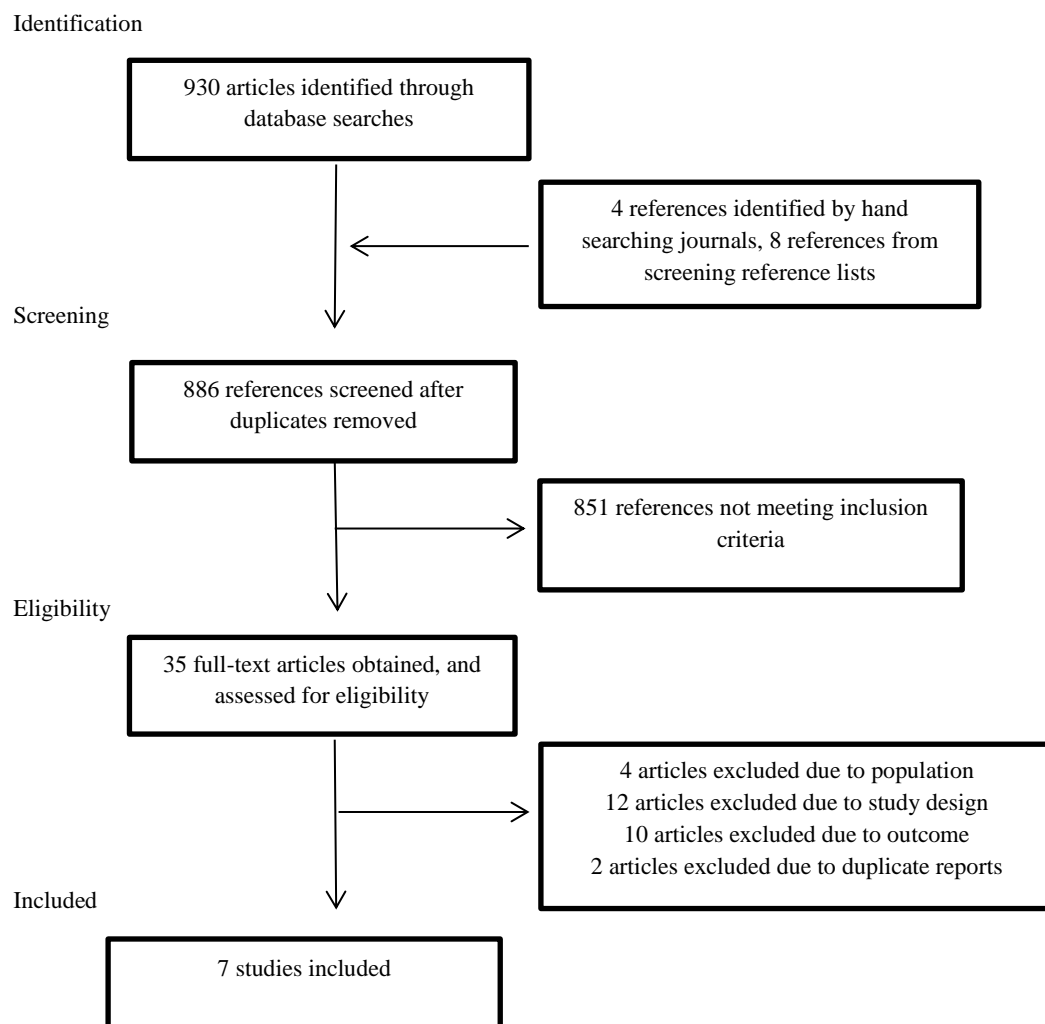


Figure 1 Flow-chart screening and inclusion of studies

Study characteristics

Included studies were published between 2005 and 2013, two of which before 2010. Four studies were conducted in the USA (Arnold, Johnson, Tucker, Chesak, & Dierkhising, 2013; Johnson, Corrigan, Gulickson, Holshouser, & Johnson, 2012; Maneval et al., 2012; Weiner et al., 2011), one in Australia (Cioffi, Purcal, & Arundell, 2005) and two in Belgium (De Regge,

Calle, De Paepe, & Monsieurs, 2008; Monsieurs et al., 2012). The studies had enrolled a total of 479 registered nurses. Six studies were conducted in hospitals (Arnold et al., 2013; Cioffi et al., 2005; De Regge et al., 2008; Maneval et al., 2012; Monsieurs et al., 2012; Weiner et al., 2011), whereas one study was in a military training programme (Johnson et al., 2012). Two studies had enrolled nurses in specialised wards (Arnold et al., 2013; Weiner et al., 2011) and two had enrolled anaesthesia- or midwifery students (Cioffi et al., 2005; Johnson et al., 2012). The final three had enrolled regular nurses (De Regge et al., 2008; Maneval et al., 2012; Monsieurs et al., 2012). Four trials compared simulation-based training to no simulation strategy (Cioffi et al., 2005; Johnson et al., 2012; Maneval et al., 2012; Weiner et al., 2011), and two studies compared high-fidelity simulation to other simulation strategies like computer-based or low-fidelity (Arnold et al., 2013; Johnson et al., 2012). Finally, two studies compared alternative organisation of basic-life support training, compared to standard group training (De Regge et al., 2008; Monsieurs et al., 2012). Manikins were used for training in six studies (Arnold et al., 2013; De Regge et al., 2008; Johnson et al., 2012; Maneval et al., 2012; Monsieurs et al., 2012; Weiner et al., 2011), and one study used a fellow participant playing the role as patient (Cioffi et al., 2005). Skills were evaluated in six trials (Arnold et al., 2013; Cioffi et al., 2005; De Regge et al., 2008; Johnson et al., 2012; Monsieurs et al., 2012; Weiner et al., 2011), whereas four studies evaluated knowledge (Arnold et al., 2013; Cioffi et al., 2005; Maneval et al., 2012; Weiner et al., 2011). The studies are presented in Table 1.

Table 1 Characteristics of included studies

Author; Year, Country	Participants , Method, Setting	Intervention	Comparison	Outcome measures
Arnold et.al 2013 USA	28 Graduated nurses. A randomised controlled trial, in critical care orientation programme in a large hospital.	Full-scale (high-fidelity) simulation	Low fidelity and computer-based simulation	Skills assessed with ERPT, knowledge measured with written exam
Cioffi et.al. 2005 Australia	36 midwifery students. A randomised controlled trial, in midwifery education at university level	Two simulation sessions	Lectures as usual	Skills and knowledge measured by a self-developed master sheet
De Regge et.al. 2008 Belgium	103 registered nurses from non-critical care wards. A randomised controlled trial, in a hospital setting	One instructor to one trainee (IT)	One instructor to six trainees (GT)	Skills measured by data from the manikin ten months after
Johnson et.al. 2012 USA	49 anaesthesia students. A randomised controlled trial, in a military nursing simulation centre	High-fidelity simulation CD-ROM	No simulation strategy	Skills measured by a self-developed performance instrument

Author; Year, Country	Participants , Method, Setting	Intervention	Comparison	Outcome measures
Maneval et.al. 2012 USA	26 new graduate nurses. A randomised controlled trial, in a hospital setting in non-specialised wards	Standard orientation programme plus six high-fidelity patient simulation experiences with the Laerdal Sim-Man®.	Standard orientation programme	Knowledge measured with The health sciences reasoning test
Monsieurs et.al. 2012 Belgium	90 registered nurses. A randomised controlled trial, in a hospital setting	Self-learning training (SL)	Instructor lead training (IL)	Skills measured with data from the manikin
Weiner et.al. 2011 USA	46 registered nurses. A randomised controlled trial, in a hospital setting, nurses working in a postpartum ward	Self-directed with textbook and access to simulation room before a 90 min. simulation session with instructor.	Traditional classroom training	Skills measured by data from manikin and knowledge measured by written tests

Risk of Bias (RoB) in included studies

Risk of bias assessment is presented in Figure 2. The risk of bias in the included studies was overall unclear to high due to issues with allocation concealment, blinding and incomplete outcome data. Two studies (Arnold et al., 2013; Monsieurs et al., 2012) have overall high risk of bias, three studies (Cioffi et al., 2005; De Regge et al., 2008; Johnson et al., 2012) have unclear risk, and only two (Maneval et al., 2012; Weiner et al., 2011) studies have overall low risk of bias.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Arnold et al, 2013	?	+	+	+	+	+	+
Cioffi et al, 2005	+	+	?	?	+	?	+
De Regge et al, 2008	+	+	?	?	?	?	+
Johnson et al, 2012	?	+	?	+	?	?	+
Maneval et al, 2012	+	+	+	+	+	?	+
Monsieurs et al, 2012	+	+	+	+	+	?	+
Weiner et al, 2011	+	?	+	+	+	?	+

Figure 2 Risk of Bias in included studies

The quality of evidence for simulation-based training versus no simulation strategy, high-fidelity simulation versus other simulation strategy and different organisation of simulation-based training was assessed low quality for both outcomes. The grading of documentation is presented in Appendices 9a-9c GRADE Evidence profiles or Summary of findings tables.

Simulation-based training versus no simulation

Three studies were eligible for meta-analysis on this comparison for nurses' skills (Cioffi et al., 2005; Johnson et al., 2012; Weiner et al., 2011). Appendix 10 Forest Plot 1 presents the meta-analysis of this comparison on nurses' skills. Summary of findings are presented in table 2. All three studies measured effect with predefined scoring-sheets, all unique to the individual study. Cioffi et.al. (2005) found a 7 point difference in an ungraded scale on skills for data collection segments. Johnson et.al. (2012) had a 47.77 point difference on a 107 point scale, and Weiner et al. (2011) found no difference between the groups with only one point between the groups. We pooled the results in a random effects meta-analysis and found a standardized mean difference (SMD) -1.23 (CI -2.69 to 0.23). The pooled effect size for these studies shows a non-significant difference between the groups.

However the heterogeneity between the studies is high with I^2 92 % and $p < 0.00001$, and the results from the study of Johnson et.al. (2012) differs from the other studies being clearly significant. We performed a sensitivity analysis without Johnson et.al. (2012), presented as Appendix 10 Forest Plot 2. These results were more concise with SMD -0.37 (CI -0.80 to 0.07), however not statistically significant. The effects of simulation-based training compared to no simulation strategy on nurses' skills therefore still remains uncertain due to $p = 0.10$. Further the grading of the evidence showed low quality on this comparison, and our confidence in these finding is therefore limited.

Three studies evaluated effect of simulation on nurses' knowledge, all measured by tests, see table 2. The population varied from newly educated nurses, to specialised nurses. Because of that we assessed clinical diversity between the studies to be too significant to pool the total effect sizes. Cioffi et.al. (2005) found a 2 point difference on an ungraded scale in favour of simulation training. Weiner et.al. (2011) found a 1.8 point difference on a 55 point scale in favour of intervention and finally Maneval et.al. (2012) found a 0.39 point difference on a 33 point scale. None of the studies shows significant difference between the groups since the CI crosses the zero-effect line, and are also quite broad. The grading of evidence also showed low quality. This makes our confidence low in the results of this analysis.

Table 2 Summary of findings effectiveness of simulations-based training compared to no simulation

Outcome	Results		Effect sizes (95 % CI)	Number of studies (Participants)	Quality of evidence (GRADE)
	Intervention (I)	Control (C)			
Skills	Mean 69 (SD 9)	Mean 62 (SD 14)	<i>Meta-analysis</i> SMD -1.23 (-2.69 to 0.23)	3 (117)	⊕⊕○○ LOW ^{1,2,3,4}
	Mean 89.84 (SD 18.7)	Mean 42.02 (SD 9.1)			
	Mean 24.5 (SD 3.6)	Mean 23.8 (SD 3.1)	<i>Sensitivity analysis</i> SMD -0.37 (-0.80 to 0.07)	2 (71)	
Knowledge	Mean 12 (SD 9)	Mean 14 (SD 12)	SMD -0.23 (-0.89 to 0.43)	3 (108)	⊕⊕○○ LOW ^{5,6,7}
	Mean 51.5 (SD 4)	Mean 49.8 (SD 3.1)	SMD -0.47 (1.05 to 0.12)		
	Mean 22.08 (SD 2.84)	Mean 21.69 (2.25)	SMD -0.15 (-0.92 to 0.12)		

SD = Standard deviation, CI = Confidence Interval, SMD = Standardised Mean Difference

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Footnotes:

1. One of the included studies had unclear risk of bias on randomization
2. One of the included studies had high risk of bias, and one study had unclear risk regarding blinding
3. One of the included studies had high risk of bias, and one had unclear risk regarding incomplete outcome data
4. The results differ from significant to not significant
5. One of the studies had unclear risk of bias on allocation concealment
6. One of the studies had unclear risk of bias on blinding
7. One of the studies had high risk of bias due to no reported baseline characteristics

High-fidelity simulation versus other simulation strategies

Two studies compared high-fidelity simulation to other simulation strategies nurses' skills. However Arnold et.al. (2013) did not report results due to missing data from one group. The analysis is therefore based on one study (Johnson et.al. 2012), comparing high-fidelity simulation to CD-ROM. Johnson et.al. (2012) found a 40.44 point difference on a 107 point scale in favour of high-fidelity group. These results are statistically significant with $p < 0.0001$ and SMD -2.42 (CI -3.35 to -1.49). However the evidence was graded low quality on this comparison, and we cannot draw any conclusions from this one study with relatively few participants. Summary of findings are presented in table 3.

One study had made this comparison on nurses' knowledge. Arnold et.al. (2013) made three comparisons. For the comparison high-fidelity simulation versus low-fidelity simulation they found a SMD -1.01 (CI – 1.98 to -0.04). The results are statistically significant in favour of high-fidelity simulation. However the results are uncertain due to a broad CI close up to the zero-effect line. The comparison high-fidelity simulation versus computer-based simulation shows a SMD 0.45 (-0.46 to 1.37). The results are in favour of the computer-based group, however not statistically significant. Finally low-fidelity simulation versus computer-based simulation found a SMD 1.38 (CI 0.32 to 2.43), and results statistically in favour of computer-based simulation. See Appendix 10 Forest Plot 3.

The results indicate that computer-based simulation might be more effective for improving nurses' knowledge when compared to both high- and low-fidelity simulation. Further high-fidelity simulation might be superior to low-fidelity simulation. However the quality of evidence is low and our confidence in these results is therefore low, see table 3.

Table 3 Summary of findings effectiveness of high-fidelity simulation compared to other simulation strategy.

Outcome	Results		Effect measure (95 % CI)	Number of studies (Participants)	Quality of evidence (GRADE)
	Intervention (I)	Control (C)			
Skills	Mean 89.84 (SD 18.7)	Mean 49.4 (SD 12.2)	SMD -2.42 (-3.35 to -1.49)	1 (46)	⊕⊕○○ LOW ^{1,2,3,5}
Knowledge	<i>High-fidelity versus low fidelity</i>		SMD -1.01 (-1.98 to -0.04)	1 (26)	⊕⊕○○ LOW ^{4,5}
	Mean 76 (SD 8)	Mean 67 (SD 9)			
	<i>High-fidelity versus computer-based</i>		SMD 0.45 (-0.46 to 1.37)		
	Mean 76 (SD 8)	Mean 80 (SD 9)			
	<i>Low-fidelity versus computer-based</i>		SMD 1.38 (0.32 to 2.43)		
	Mean 67 (SD 9)	Mean 80 (SD 9)			

SD = Standard deviation, CI = Confidence Interval, SMD = Standardised Mean Difference

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Footnotes:

1. The study has high risk of bias on allocation concealment
2. The study has unclear risk of bias on blinding
3. One study has unclear risk of bias on incomplete outcome data
4. The study has high risk of bias on blinding and incomplete outcome data, and unclear risk of bias on allocation concealment
5. The study has unclear risk of bias on random sequence generation

Comparing different organisation of simulation

Two studies compared different organisation of simulation on nurses' skills. Summary of findings for this comparison are presented in table 4. De Regge et.al. (2008) compared basic life-support training with one participant to one instructor in the intervention group, and six participants to one instructor in the control. For number of ventilations with correct volume De Regge et.al. (2008) found Median 4 with an inter-quartile range 0-7 in the intervention group, and Median 5 and inter-quartile range 1-7 in the control group. For number of compressions with correct depth they found Median 59 with inter-quartile range 9-89 for the intervention group and Median 33 and inter-quartile range 8-82 in the control group. The effect estimates in this study are highly imprecise with results in favour of the intervention in one skill measure and in favour of control in the other skill measure.

Monsieurs et.al. (2012) have compared a self-learning station versus traditional basic life-support training. They have presented results for five skills, and reported the results as dichotomous outcomes. We present these results in Appendix 10 Forest Plot 4. The Risk Ratio varies from 0.52 in one skills-measure to 1.14 in another. Also for this study the results are highly inconsistent with results in favour of both intervention and control group.

Because of a high degree of inconsistency in the results, we cannot draw any overall clear conclusions based on the current studies. This is also due to low quality of the evidence.

Table 4 Summary of findings effectiveness of different organisation of simulation training.

Outcome	Results		Effect measure (95 % CI)	Number of studies (Participants)	Quality of evidence (GRADE)
	Intervention (I)	Control (C)			
Skills	<i>Ventilations with correct volume</i>			2 (193)	⊕⊕○○ LOW ^{1,2,3,4}
	Median 4 (0-7)	Median 5 (1-7)			
	<i>Compressions with correct depth</i>				
	Median 59 (9-86)	Median 33 (8-82)			
	<i>Mean compression depth 38-51 mm</i>		RR 0.52		
	Events 13/45	Events 25/45	(0.31 to 0.88)		
	<i>Mean compression depth over 38 mm</i>		RR 0.63		
Events 17/45	Events 27/45	(0.40 to 0.98)			
<i>Mean compression rate 80 – 120 pr. Min</i>		RR 0.94			
Events 29/45	Events 31/45	(0.70 to 1.25)			
<i>Ventilation volume 400-1000 ml</i>		RR 1.14			
Events 24/45	Events 21/45	(0.75 to 1.73)			
<i>Any incomplete release ≥ 5mm</i>		RR 1.10			
Events 23/45	Events 21/45	(0.72 to 1.67)			

SD = Standard deviation, CI = Confidence Interval, RR = Risk Ratio

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Footnotes:

1. The study has unclear risk of bias on blinding
2. The study has unclear risk of bias on incomplete outcome data
3. The study has unclear risk of bias on selective reporting
4. The results of the study are not conclusive because they are in favor of both the experimental and control group. The results are not significant.

Discussion

Based on our findings in this systematic review of seven included studies, we cannot draw any conclusions regarding the effect of simulation-based training on skills and knowledge for registered nurses. The meta-analysis did not show significant effect of simulation-based training compared to no simulation, even though the results indicated a favour of simulation. This could be due to only three eligible studies with small sample-sizes in the analysis. Further the grading of evidence showed low quality, which makes limited confidence in these findings. For the other comparisons several of the included studies presents results in favour of simulation. However there is also inconsistency in results between studies. Only one study

(Johnson et al., 2012) showed a significant effect of high-fidelity simulation compared to no simulation training on nurses' skills. There also seem to be a trend towards computer-based simulation as the most effective strategy on nurses' knowledge, but the results are uncertain and not definitive. The quality of evidence is graded as low on these comparisons as well. The lack of conclusive results in and between the studies can in our opinion be a result of small sample-sizes in most of the studies, as well as weakness in design for several of the trials.

In our meta-analysis there was large statistical heterogeneity with I^2 92 %. We could not find any clinical diversity between the studies, nor any major differences in the risk of bias of the studies that could explain the differences in results. The total number of participants is also approximately the same. We therefore performed a sensitivity analysis without the study of Johnson et.al. (2012). The results were still not statistically significant, but they had more concise CI with no heterogeneity between the estimates.

One reason the results of Johnson et.al. (2012) differs from the other studies, could be the motivation of the participants for this learning strategy. The nurses who are about to be health workers in a war setting might acknowledge the need of training on realistic situations. The situations they are about to face is substantially different from what they have met in their practice, and of a kind they will not meet in a secure hospital setting in the USA. This could make them more motivated to train with simulation, to be prepared for the situations they are faced in a war. However this is only our assumption, and there could be other reasons for the results.

Simulation-based training is one of the strategies used to improve quality in healthcare. However simulation-based training may not be the best strategy for implementing a new guideline, or government directives. Then workshops or clinical audits are frequently used strategies to increase health workers adherence to the guidelines. These strategies have shown small to moderate effect when used alone (Forsetlund et al., 2009; Ivers et al., 2012). However evidence show that the strategies used in quality improvement should be tailored to the context where they are implemented (Baker et al., 2015). We believe that quality improvement in healthcare need several approaches to be successful, depending on what the improvement is about. Simulation-based training is in our view most feasible for practical issues, whereas implementation of e.g. guidelines should use other approaches. Anyway

simulation-based training could be a part of a tailored implementation strategy, for example in combination with printed educational material (Giguère et al., 2012).

The inclusion criteria in this systematic review were relatively narrow, which may have contributed to our findings of non-conclusive results. If a broader range of study-designs and a wider population had been included, we would have had more studies to synthesize. On the other hand we chose the narrow criteria to make a highly focused review.

This systematic review has a thorough literature search, reviewed by an experienced librarian. In phases of the process which are critical, we have been two review authors screening and selecting independently. We are therefore confident that all relevant studies have been identified and included in the review. We have not contacted any of the authors of the primary studies for additional information. Our analyses have been based solely on information found in the articles. This might have influenced the risk of bias assessment and grading of evidence, since the authors could have contributed with information. However this was a choice we made, because we then had the same procedure collecting data from all studies. Further we only have two outcomes in this review. It could be relevant to evaluate patient outcomes and economic implications of simulation-based training as well.

The other systematic reviews we identified had the same comparisons and outcomes as this review. The sources searched in the other reviews were to a great extent the same as the ones we searched. The other reviews identified more references than we did from their searches, but this could be due to our relatively restricted population of registered nurses. Some of the reviews had searched for unpublished literature in databases designed for such studies. We only searched for protocols of unpublished studies in Clinical Trials. Further we have assessed risk of bias using the Cochrane' RoB-tool, whereas the other reviews have used different tools. However we find that the tools are comparable. We have not found that the other reviews have graded the quality of evidence using GRADE. Further the results from the other systematic reviews we identified differ from ours, in the sense they have found significant effect of simulation training (Cook et al., 2011; Ilgen et al., 2013). This difference could be due to the fact that we only included RCTs. Finally we have identified and included RCT-studies not included in the other reviews.

Simulation-based training is rather resource-demanding, and the equipment needed for high-fidelity simulation is costly. Based on the trends in this review, and the positive effect of simulation-based training for students shown in other systematic reviews, we assume that

further RCTs of good quality might conclude that simulation-based training is an effective learning-strategy for nurses. Based on our experience with simulation this is the case. However our findings in this review are not conclusive and we cannot claim that simulation-based training is effective on nurses' skills and knowledge.

Further studies should use RCT-designs, and they should aim to include a larger sample size of registered nurses. The purpose of these studies should be to establish whether or not simulation-based training for nurses is an effective learning strategy on nurses' skills and knowledge. When this is established, different simulation-strategies should be compared to each other to establish which strategy or organisation is the most effective.

Conclusion

Based on the findings in this systematic review no clear conclusions can be drawn. The methodical quality of the included studies varies too much, and the inconsistencies in results are significant. Even though the results in several studies are in favour of simulation-based training, the results on each comparison are not statistically significant. The quality of the evidence was also graded low for all comparisons. In our opinion the effectiveness of simulation-based training for nurses has not been established. There is a need of new, good quality RCT's of reasonable size to establish a confidence in the effect-measures. However we believe this systematic review is an important contribution to the knowledge in this topic.

Acknowledgement

There are no conflicts of interest in this systematic review except from a genuine interest in the topic. The experts in the field have only contributed with references, and there have been no economic funding. Kjetil Brurberg and Birgitte Espehaug at Bergen University College have contributed with advice in the statistical analysis in this review.

Key points:

- Quality improvement in healthcare is important to increase patient safety
- Simulation-based training is one of several quality-improvement strategies used in healthcare
- Effect of simulation-based training on nurses' skills and knowledge is still uncertain due to lack of robust evidence

References

- Arnold, J. J., Johnson, L. M., Tucker, S. J., Chesak, S. S., & Dierkhising, R. A. (2013). Comparison of Three Simulation-Based Teaching Methodologies for Emergency Response. *Clinical Simulation In Nursing*, 9(3), e85-e93. doi: 10.1016/j.ecns.2011.09.004
- Baker, R., Camosso-Stefinovic, J., Gillies, C., Shaw, E. J., Cheater, F., Flottorp, S., . . . van Lieshout J, J. C. (2015). Tailored interventions to address determinants of practice. *Cochrane Database of Systematic Reviews* 2012(4).
- Cant, R. P., & Cooper, S. J. (2010). Simulation-based learning in nurse education: systematic review. *Journal of Advanced Nursing*, 66(1), 3-15.
- Cioffi, J., Purcal, N., & Arundell, F. (2005). A pilot study to investigate the effect of a simulation strategy on the clinical decision making of midwifery students. *Journal of Nursing Education*, 44(3), 131-134.
- ClinicalTrials.gov. (2014). Retrieved 03.03.15, from <https://clinicaltrials.gov/>
- Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA*, 306(9), 978-988.
- De Regge, M., Calle, P. A., De Paepe, P., & Monsieurs, K. G. (2008). Basic life support refresher training of nurses: individual training and group training are equally effective. *Resuscitation*, 79(2), 283-287. doi: <http://dx.doi.org/10.1016/j.resuscitation.2008.06.015>
- Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2011). Analysing data and undertaking meta-analyses. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for systematic reviews of interventions version 5.1.0: The Cochrane Collaboration* 2011.
- Forsetlund, L., Bjørndal, A., Rashidian, A., Jamtvedt, G., O'Brien, M. A., Wolf, F. M., . . . Oxman, A. D. (2009). Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database of Systematic Reviews* 2009(2).
- Giguère, A., Légaré, F., Grimshaw, J., Turcotte, S., Fiander, M., Grudniewicz, A., . . . Gagnon, M.-P. (2012). Printed educational materials: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*(10).
- GRADE Working Group. (2012). Guideline Development Tool (GDT). Retrieved 16.02.2015, 2015, from <http://www.guidelinedevelopment.org/>
- Grol, R., Berwick, D. M., & Wensing, M. (2008). On the trail of quality and safety in health care. *BMJ*, 336(7635), 74-76. doi: 10.1136/bmj.39413.486944.AD
- Guyatt, G. H., Oxman, A. D., Akl, E., Kunz, R., Vist, G., & Brozek, J. (2011). GRADE guidelines 1. Introduction – GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64, 383-394.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *336(7650)*, 924-926. doi: 10.1136/bmj.39489.470347.AD
- Healthy Simulation. (2014). Medical Simulation. Retrieved 06.04.15, from <http://healthysimulation.com/medical-simulation/>
- Higgins, J. P. T., Altman, D. G., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for systematic reviews of interventions version 5.1.0: The Cochrane Collaboration* 2011.
- Higgins, J. P. T., & Deeks, J. J. (2011). Selecting studies and collecting data. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for systematic reviews of interventions version 5.1.0 The Cochrane Collaboration* 2011.
- Ilgen, J. S., Sherbino, J., & Cook, D. A. (2013). Technology-enhanced simulation in emergency medicine: a systematic review and meta-analysis. *Acad Emerg Med*, 20(2), 117-127. doi: 10.1111/acem.12076
- Institute for Healthcare Improvement. (2015). Patient Safety. Retrieved 24.04.15, from <http://www.ihl.org/Topics/PatientSafety/Pages/default.aspx>

- Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., . . . Oxman, A. D. (2012). Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012(6). doi: 10.1002/14651858.CD000259.pub3.
- Jansson, M., Kääriäinen, M., & Kyngäs, H. (2012). Effectiveness of Simulation-Based Education in Critical Care Nurses' Continuing Education: A Systematic Review. *Clinical Simulation In Nursing*, 9(9), e355-e360. doi: 10.1016/j.ecns.2012.07.003
- Johnson, D., Corrigan, T., Gulickson, G., Holshouser, E., & Johnson, S. (2012). The effects of a human patient simulator vs. a CD-ROM on performance. *Military medicine*, (10), 1131-1135. Retrieved from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-3113.2012.04840.x>
- Maneval, R., Fowler, K. A., Kays, J. A., Boyd, T. M., Shuey, J., Harne-Britner, S., & Mastrine, C. (2012). The effect of high-fidelity patient simulation on the critical thinking and clinical decision-making skills of new graduate nurses. *Journal of Continuing Education in Nursing*, 43(3), 125-134. doi: 10.3928/00220124-20111101-02
- Monsieurs, K. G. a., De Regge, M. b., Schelfout, S. a., D'Hondt, F. a., Mpotos, N. a., Valecke, M. c., & Calle, P. A. a. (2012). Efficacy of a self-learning station for basic life support refresher training in a hospital: a randomized controlled trial. *European Journal of Emergency Medicine*, 19(4), 214-219.
- NICE. (2014). What we do. Retrieved 03.05.15, from <https://www.nice.org.uk/about/what-we-do>
- Patel, V. L., Kannampallil, T. G., & Shortliffe, E. H. (2015). Role of cognition in generating and mitigating clinical errors. Retrieved 03.05.15, from <http://qualitysafety.bmj.com/content/early/2015/05/02/bmjqs-2014-003482.abstract>
- Salas, E., Paige, J. T., & Rosen, M. A. (2013). Creating new realities in healthcare: the status of simulation-based training as a patient safety improvement strategy. *BMJ Quality and Safety*, 22, 449-452.
- Sampson, M., McGowan, J., Cogo, E., Grimshaw, J., Moher, D., & Lefebvre, C. (2009). An evidence-based practice guideline for the peer review of electronic search strategies. *Journal of Clinical Epidemiology*, 62(9), 944-952. doi: 10.1016/j.jclinepi.2008.10.012
- Society for Simulation in Healthcare. (2014). Retrieved 03.04.15, from <http://www.ssih.org/>
- The International Nursing Association for Clinical Simulation & Learning. (2013). Retrieved 03.04.15, from <http://www.inacsl.org/i4a/pages/index.cfm?pageid=1>
- The Nordic Cochrane Centre, T. C. C. (2014). Review Manager (RevMan) (Version 5.3). Retrieved from <http://tech.cochrane.org/revman>
- Weiner, G. M., Menghini, K., Zaichkin, J., Caid, A. E., Jacoby, C. J., & Simon, W. M. (2011). Self-directed Versus Traditional Classroom Training for Neonatal Resuscitation. *Pediatrics*, 127(4), 713-719. doi: 10.1542/peds.2010-2829

Appendix 1 Search strategy CDSR and CENTRAL (Wiley)

Search Name: CDSR 05.09.14

Date Run: 05/09/14 09:52:55.977

Description:

ID	Search	Hits
#1	MeSH descriptor: [Nurses] this term only	280
#2	MeSH descriptor: [Nurse Anesthetists] this term only	45
#3	MeSH descriptor: [Nurse Clinicians] this term only	175
#4	MeSH descriptor: [Nurse Practitioners] this term only	291
#5	(acute or intensive or registered) next nurs*	460
#6	nurse next (anesthetist* or clinician* or practitioner*)	1091
#7	#1 or #2 or #3 or #4 or #5 or #6	1741
#8	MeSH descriptor: [Computer Simulation] this term only	1347
#9	MeSH descriptor: [Patient Simulation] this term only	253
#10	MeSH descriptor: [Manikins] this term only	346
#11	(manikin* or mannequin*)	526
#12	simulation*	5291
#13	(virtual or manikin or mannequin or technology or computer or patient) near (simulation)	2069
#14	#8 or #9 or #10 or #11 or #12 or #13	5681
#15	MeSH descriptor: [Professional Competence] explode all trees	1865
#16	skill*	11800
#17	(clinical or professional or technical or nurs*) near (skill* or competence or expertise)	3170
#18	#15 or #16 or #17	13092
#19	#7 and #14 and #18	41

Appendix 2 Search strategy HTA and DARE (Centre of Reviews and Dissemination)

Search date 05.09.14

	Line	Search	Hits	
<input type="checkbox"/>	1	MeSH DESCRIPTOR nurses EXPLODE ALL TREES IN DARE,HTA	65	Delete
<input type="checkbox"/>	2	((acute or intensive or critical or registered) NEAR3 Nurs*) IN DARE, HTA	55	Delete
<input type="checkbox"/>	3	(Nurse NEXT (anesthetist* or clinician* og practitioner*)) IN DARE, HTA	0	Delete
<input type="checkbox"/>	4	#1 OR #2 OR #3	115	Delete
<input type="checkbox"/>	5	MeSH DESCRIPTOR patient simulation IN DARE,HTA	9	Delete
<input type="checkbox"/>	6	MeSH DESCRIPTOR computer simulation IN DARE,HTA	29	Delete
<input type="checkbox"/>	7	MeSH DESCRIPTOR manikins IN DARE,HTA	4	Delete
<input type="checkbox"/>	8	((manikin* or mannequin*)) IN DARE, HTA	8	Delete
<input type="checkbox"/>	9	(simulation*) IN DARE, HTA	138	Delete
<input type="checkbox"/>	10	((virtual or manikin or mannequin or technology or computer or patient) NEAR3 (simulation)) IN DARE, HTA	45	Delete
<input type="checkbox"/>	11	#5 OR #6 OR #7 OR #8 OR #9 OR #10	142	Delete
<input type="checkbox"/>	12	MeSH DESCRIPTOR professional competence EXPLODE ALL TREES IN DARE,HTA	110	Delete
<input type="checkbox"/>	13	(skill*) IN DARE, HTA	962	Delete
<input type="checkbox"/>	14	((clinical or professional or technical or nurs*) NEAR3 (skill* or competence or expertice)) IN DARE, HTA	158	Delete
<input type="checkbox"/>	15	#12 OR #13 OR #14	1049	Delete
<input type="checkbox"/>	16	#4 AND #11 AND #15	2	Delete

Appendix 3 Search strategy CINAHL (EBSCO)

#	Query	Results
S21	S7 AND S16 AND S20	722
S20	S17 OR S18 OR S19	93,544
S19	TX skill*	68,374
S18	TX (technical or professional or nursing or clinical) N2 (competence* or expertise)	26,373
S17	(MH "Professional Competence+")	32,084
S16	S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15	24,428
S15	TX patient N1 simulation	1,534
S14	TX Computer N1 simulation	3,832
S13	TX (Technology-Enhanced) N1 simulation	4
S12	TX (virtual reality) N1 simulation	41
S11	TX mannequin* or manikin*	732
S10	TX simulat*	18,462
S9	(MH "Models, Anatomic+")	3,238
S8	(MH "Simulations+")	13,959
S7	S1 OR S2 OR S3 OR S4 OR S5 OR S6	253,524
S6	TX nurs* N1 (anaesthe* or clinician* or practitioner*)	1,036
S5	TX registered N1 nurse*	34,060
S4	TX (acute or intensive or critical) N3 nurs*	34,410
S3	(MH "Medical-Surgical Nursing+")	71,215
S2	(MH "Nursing Practice+")	40,122
S1	(MH "Nurses+")	144,323

Appendix 4 Search strategy MEDLINE (OVID)

Search date: 05.09.14

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

#	Searches	Results	Search Type
1	exp Nurses/	69034	Advanced
2	(nurse adj1 (clinician* or anesthetist* or practitioner*)),tw.	1341	Advanced
3	((registered or acute or intensive or critical) adj3 nurse*),tw.	13116	Advanced
4	1 or 2 or 3	80132	Advanced
5	patient simulation/	2991	Advanced
6	Computer Simulation/	132519	Advanced
7	manikins/	2975	Advanced
8	(manikin* or mannequin*),tw.	2467	Advanced
9	((virtual or technology) adj3 simulation*),tw.	1134	Advanced
10	simulat*,tw.	285130	Advanced
11	5 or 6 or 7 or 8 or 9 or 10	359328	Advanced
12	exp Professional Competence/	81898	Advanced
13	skill*,tw.	113655	Advanced
14	expertise.tw.	23865	Advanced
15	((nursing or technical) adj2 competence*),tw.	752	Advanced
16	12 or 13 or 14 or 15	198580	Advanced
17	4 and 11 and 16	252	Advanced

Appendix 5 Search strategy Embase (OVID)

Embase 1974 to 2014 February 04

#	Searches	Results	Search Type
1	exp nurse/	108663	Advanced
2	((anesthetist or clinician) adj3 nurse*).tw.	728	Advanced
3	((intensive or critical or acute) adj3 nurse*).tw.	6679	Advanced
4	(registered adj1 nurse*).tw.	7890	Advanced
5	1 or 2 or 3 or 4	118783	Advanced
6	exp simulation/	167444	Advanced
7	computer simulation/	88873	Advanced
8	simulator/	5063	Advanced
9	simulat*.tw.	311638	Advanced
10	(mannequin* or manikin*).tw.	3505	Advanced
11	6 or 7 or 8 or 9 or 10	368244	Advanced
12	exp competence/	91505	Advanced
13	skill/	38647	Advanced
14	skill*.tw.	149516	Advanced
15	((technical or professional or nursing) adj1 competence*).tw.	1583	Advanced
16	12 or 13 or 14 or 15	233264	Advanced
17	5 and 11 and 16	521	Advanced
18	limit 17 to exclude medline journals	28	Advanced

Appendix 6 Search strategy ERIC (EBSCO)

#	Query	Limiters/Expanders	Results
S21	S9 AND S16 AND S20	Search modes - Boolean/Phrase	34
S20	S17 OR S18 OR S19	Search modes - Boolean/Phrase	117,419
S19	DE "Skills" OR DE "Agricultural Skills" OR DE "Basic Skills" OR DE "Business Skills" OR DE "Communication Skills" OR DE "Daily Living Skills" OR DE "Decision Making Skills" OR DE "Home Economics Skills" OR DE "Homemaking Skills" OR DE "Information Skills" OR DE "Interpretive Skills" OR DE "Job Skills" OR DE "Language Skills" OR DE "Locational Skills (Social Studies)" OR DE "Map Skills" OR DE "Mathematics Skills" OR DE "Mechanical Skills" OR DE "Minimum Competencies" OR DE "Parenting Skills" OR DE "Psychomotor Skills" OR DE "Research Skills" OR DE "Salesmanship" OR DE "Science Process Skills" OR DE "Study Skills" OR DE "Teaching Skills" OR DE "Thinking Skills" OR DE "Visual Literacy"	Search modes - Boolean/Phrase	87,227
S18	DE "Skill Development"	Search modes - Boolean/Phrase	18,700
S17	DE "Competence" OR DE "Interpersonal Competence" OR DE "Minimum Competencies" OR DE "Teacher Competencies"	Search modes - Boolean/Phrase	26,554
S16	S10 OR S11 OR S12 OR S13 OR S14 OR S15	Search modes - Boolean/Phrase	85,860
S15	TX (technology-enhanced simulat*)	Search modes - Boolean/Phrase	3
S14	TX (virtual reality simulat*)	Search modes - Boolean/Phrase	64
S13	TX manikin*	Search modes - Boolean/Phrase	18
S12	TX mannequin*	Search modes - Boolean/Phrase	38
S11	TX simulat*	Search modes - Boolean/Phrase	21,890
S10	DE "Simulation" OR DE "Computer Simulation" OR DE "Markov Processes" OR DE "Models" OR DE "Monte Carlo Methods" OR DE "Role Playing"	Search modes - Boolean/Phrase	79,284
S9	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8	Search modes - Boolean/Phrase	5,368
S8	TX nurs* N3 practitioner*	Search modes - Boolean/Phrase	414
S7	TX nurs* N3 clinician*	Search modes - Boolean/Phrase	49
S6	TX nurs* N3 anaesthesia*	Search modes - Boolean/Phrase	26
S5	TX (acute care nurs*)	Search modes - Boolean/Phrase	42
S4	TX (critical care nurs*)	Search modes - Boolean/Phrase	26
S3	TX (intensive care nurs*)	Search modes - Boolean/Phrase	49
S2	DE "Nursing"	Search modes - Boolean/Phrase	1,950
S1	DE "Nurses" OR DE "School Nurses"	Search modes - Boolean/Phrase	3,695

Appendix 7 Search strategy SveMed+

Sök artiklar i SveMed+

Enkel sökning Avancerad sökning **Kombinera sökningar**

#1 OR #2 OR #3 OR #4

Peer reviewed tidskrifter Läs online

Sökningar  

Kombinera sökningarna med boolesk logik med hjälp av kryssboxarna eller direkt i sökrutan, t ex. **#1 AND (#2 OR #3)**

Markera/avmarkera alla

Nr	Söksträng	Antal träffar	Tid
<input type="checkbox"/> 1	exp:"Patient Simulation"	45	2014-09-05 12:33:27
<input type="checkbox"/> 2	exp:"Computer Simulation"	90	2014-09-05 12:33:49
<input type="checkbox"/> 3	simulat*	167	2014-09-05 12:34:20
<input type="checkbox"/> 4	simul*	237	2014-09-05 12:49:37
<input type="checkbox"/> 5	#1 OR #2 OR #3 OR #4	237	2014-09-05 12:50:33

Appendix 8 Excluded studies

Study	Excluded due to
Andrighetti, T. P., Knestruck, J. M., Marowitz, A., Martin, C. & Engstrom, J. L. (2012) Shoulder dystocia and postpartum hemorrhage simulations: student confidence in managing these complications. <i>Journal of midwifery & women's health</i> (1), s. 55-60.	Excluded due to study design, not randomised.
De Regge, M., Monsieurs, K. G., Valcke, M. & Calle, P. A. (2012) Training nurses in a self-learning station for resuscitation: factors contributing to success or failure. <i>Journal of Emergency Nursing</i> , 38 (4), s. 386-91.	Duplicate report from the already included study of Monsieurs et.al. (2012).
Ford, D.G., Seybert, A. L., Smithburger, P. L., Kobulinsky, L. R., Samosky, J. T. & Kane-Gill, S. L. (2010) Impact of simulation-based learning on medication error rates in critically ill patients. <i>Intensive Care Medicine</i> , 36, s. 1526-1531.	Excluded due to study design, not randomised.
Harvey, E. M., Echols, S. R., Clark, R. & Lee, E. (2014) Comparison of Two TeamSTEPPS® Training Methods on Nurse Failure-to-Rescue Performance. <i>Clinical Simulation In Nursing</i> , 10 (2), s. e57-e64.	Excluded due to study design. A quasi-experimental design, not randomised.
Jansson, M. M., Ala-Kokko, T. I., Ohtonen, P. P., Merilainen, M. H., Syrjala, H. P. & Kyngas, H. A. (2014) Human patient simulation education in the nursing management of patients requiring mechanical ventilation: a randomized, controlled trial. <i>Am J Infect Control</i> , 42 (3), s. 271-6.	Excluded due to no real intervention group. Both groups get the same training, the only difference is the voice feedback
Johnson, D. & Johnson, S. (2014) The Effects of Using a Human Patient Simulator Compared to a CD-ROM in Teaching Critical Thinking and Performance. <i>US Army Medical Department Journal</i> , s. 59-64.	Duplicate Report from the already included study Johnson et.al. (2012).
Pozner, C. N., Almozlino, A., Elmer, J., Poole, S., McNamara, D. A. & Barash, D. (2011) Cardiopulmonary resuscitation feedback improves the quality of chest compression provided by hospital health care professionals. <i>American Journal of Emergency Medicine</i> , 29 (6), s. 618-625.	Excluded due to Outcome measure. Do not measure effect of simulation-training.
Szlachta, J. (2013) Peer Instruction of First-Year Nurse Anesthetist students: A Pilot Study of a Strategy to Use Limited Faculty Resources and Promote Learning. <i>Journal of Nurse Education</i> , 52 (6), s. 355-359.	Excluded due to outcome measure. Do measure feedback from a fellow student, rather than the effect of simulation training.

Appendix 9a GRADE Evidence profile

Question: Simulation-based training compared to No simulation training for Nurses

Bibliography (systematic reviews):

Quality assessment							№ of participants		Effect		Quality	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	High Fidelity Simulation	No simulation training	Relative (95% CI)	Absolute (95% CI)		
Skills (assessed with: Scoring instruments) ⁴												
3	randomised trials	serious ^{1,2,3}	serious ⁴	not serious	not serious	none	60	58	-	SMD 1.12 lower (2.63 lower to 0.39 higher)	⊕⊕○○ LOW	CRITICAL
Knowledge (assessed with: Questionnaire or written test)												
3	randomised trials	serious ^{5,6,7}	serious	not serious	not serious	none	-/54	-/54	None of the studies has significant results, but two studies have results in favour of the intervention group, whereas one study's results are in favour of the control group		⊕⊕○○ LOW	CRITICAL

MD – mean difference, RR – relative risk

1. One of the included studies had unclear risk of bias on randomization
2. One of the included studies had High risk of bias, and one study had unclear risk regarding blinding
3. One of the included studies had high risk of bias, and one had unclear risk regarding incomplete outcome data
4. The results differ from significant to not significant
5. One of the studies had unclear risk of bias on allocation concealment
6. One of the studies had unclear risk of bias on blinding
7. One of the studies had high risk of bias due to no reported baseline characteristics

Appendix 9b GRADE Summary of findings table

Summary of findings:

High Fidelity Simulation compared to Other forms of simulation strategy for nurses

Patient or population: nurses

Intervention: High Fidelity Simulation

Comparison: Other forms of simulation strategy

Outcomes	Impact	No of participants (Studies)	Quality of the evidence (GRADE)
Skills assessed with: Scoring forms	One study reports significant effect of High fidelity Simulation, whereas one study lacks data from one of the comparison groups and therefore cannot report on the outcome. ⁴	63 (2 RCTs)	⊕⊕○○ LOW ^{1 2 3 6}
Knowledge assessed with: Test	The study shows a better, but not significant result of High fidelity Simulation on knowledge.	28 (1 RCT)	⊕⊕○○ LOW ⁵

*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; OR: Odds ratio;

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

1. One study has high risk of bias on allocation concealment
2. One study has high risk of bias on blinding, whereas one study has unclear risk
3. One study has high risk of bias on incomplete outcome data, whereas one has unclear risk
4. Two studies show significant effect, one shows better but not significant difference, and one study shows no difference between the groups
5. The study has high risk of bias on blinding and incomplete outcome data, and unclear risk of bias on allocation concealment
6. Two studies has unclear risk of bias on random sequence generation

Appendix 9c GRADE Summary of findings table

Summary of findings:

Individual Simulation Training or Voice feedback compared to Group Training or No voice feedback for basic life-support training for nurses

Patient or population: basic life-support training for nurses

Setting:

Intervention: Individual Simulation Training or Voice feedback

Comparison: Group Training or No voice feedback

Outcomes	Impact	No of participants (Studies)	Quality of the evidence (GRADE)
Skills assessed with: Data collected from manikin, plus observation	The results are inconclusive. On some measurements the intervention is in favour of the experimental group, whereas on other measurements the results are in favour of the control group.	193 (2 RCTs)	⊕⊕○○ LOW ¹²³⁴

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

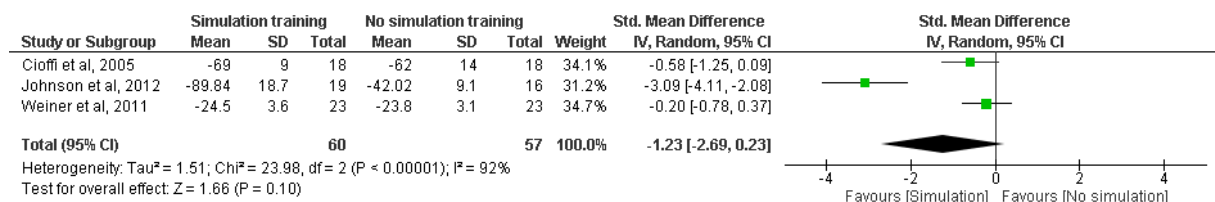
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

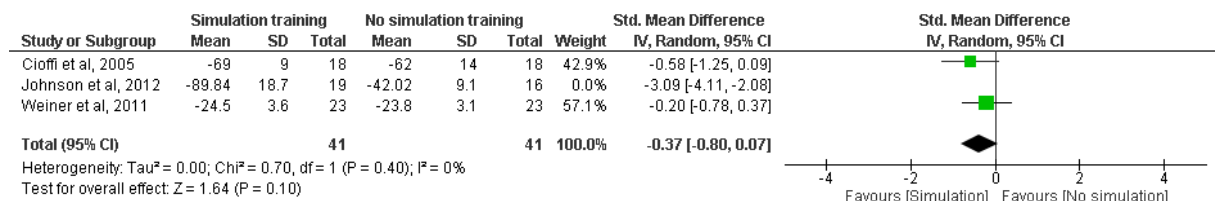
1. The study has unclear risk of bias on blinding
2. The study has unclear risk of bias on incomplete outcome data
3. The study has unclear risk of bias on selective reporting
4. The results of the study are not conclusive because they are in favor of both the experimental and control group. The results are not significant.

Appendix 10 Forest Plot

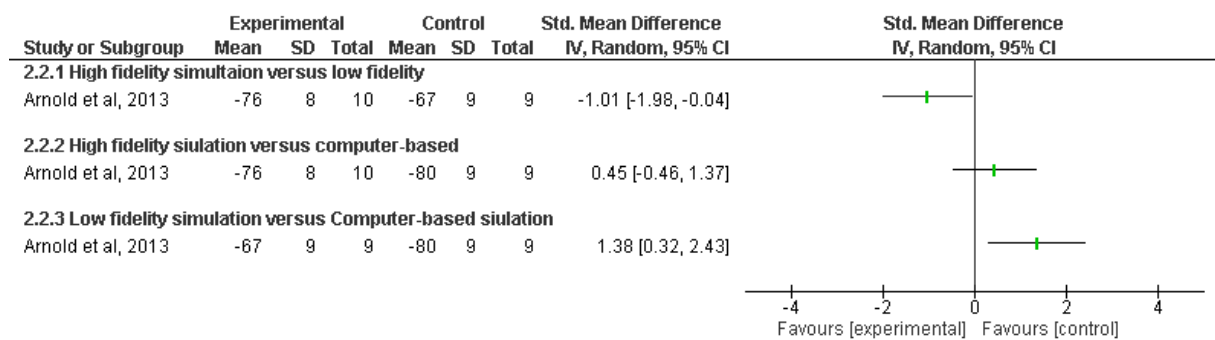
Forest Plot 1 Meta-analysis Simulation-based training versus no simulation, outcome skills



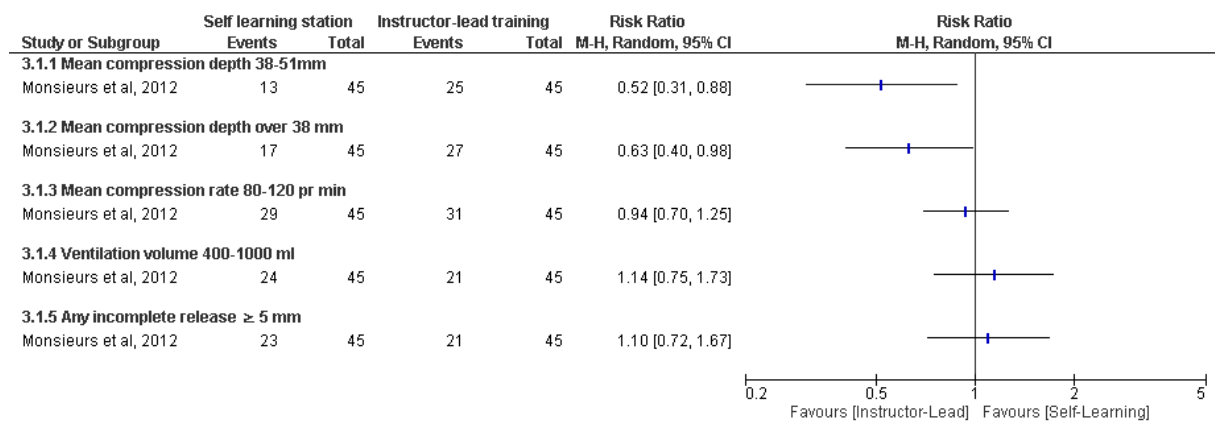
Forest Plot 2 Sensitivity analysis simulation-based training versus no simulation, outcome skills



Forest Plot 3 comparison different types of simulation-based strategies, outcome knowledge



Forest Plot 4 different organisation of simulation-training, outcome skills



Vedlegg 1 Søkeord PIO

Emneord	Populasjon	Intervensjon	Utfall
MeSH (MEDLINE, CDSR, DARE, HTA, CENTRAL, SveMed+)	Nurses Nurse Anesthetists Nurse Clinicians Nurse Practitioners	Computer Simulation Patient Simulation Problem-Based Learning Teaching Manikins Competency-based education.	Clinical Competence Professional competence
Emtree (Embase)	Nurse Registered nurse Nurse Anesthetists Nurse Clinicians Nurse Practitioner Clinical Nurse Specialist Acute nurse practitioner Emergency nurse practitioner	Simulation Simulator Computer simulation	Competence Clinical Competence Nursing competence Professional competence
CINAHL	Nurse Practitioners Nurses Nursing practice Clinical Nurse Specialists Registered nurses Critical care nursing Intensive care nursing	Simulations Computer simulation Patient simulation Models, anatomic Teaching methods, Clinical	Clinical competence Professional competence Nursing skills
ERIC	Nurses Nursing	Simulation Computer simulation	Competence Skill development
Tekstord – Alle databaser (Engelske)	<ul style="list-style-type: none"> • Nurse • Nursing • Acute care nurse • Intensive care nurse • Intensive care nursing • Registered nurse • Nurse Anaesthetists • Nurse Clinicians • Nurse Practitioners 	<ul style="list-style-type: none"> • Simulation. • Virtual reality simulation. • Mannequin – based simulation. • Technology-enhanced simulation training. • Computer simulation. • Training. • Mannequin. • Manikin. • Patient simulation • Problem-Based Learning • Teaching • Competency-based education. • Competency based education. 	<ul style="list-style-type: none"> • Skill. • Learning. • Clinical competence • Clinical skill • Professional competence • Technical expertise • Nursing skills
Tekstord SveMed+ (Norske)	<ul style="list-style-type: none"> • Sykepleier 	<ul style="list-style-type: none"> • Simulering • Medisinsk simulering 	<ul style="list-style-type: none"> • Kunnskap • Ferdighet

Vedlegg 2a RoB-skjema Arnold et.al. 2013

Methods	RCT	
Participants	28 registered nurses	
Interventions	Full-scale simulation, low-fidelity simulation and computer-based simulation	
Outcomes	Knowledge	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Randomisation procedure not described. All the participants enrolled to the Critical Care Orientation programme over a year were invited to participate.
Allocation Concealment (selection bias)	Low risk	Randomisation after informed consent was obtained. Then the names were entered into a randomised table, and assigned to one of the three groups.
Blinding of participants and personnel (performance bias)	High risk	The researchers who analysed the video were blinded to group assignment, but they could recognise some of the participants. The rest of the assessments were blinded to the investigators.
Blinding of outcome assessment (detection bias)	High risk	The participants were shown the simulation centre and manikin before the start of the interventions. This due to the final test. Group assignments were concealed until the demographic data were obtained.
Incomplete outcome data (attrition bias)	High risk	Demographic data is reported in a table, and said not to be statistically significant. The table only shows the total, and not data per group. 33 RN were enrolled, but only 28 completed. The five who dropped out is not described. Another one of the 28 did not complete the final assessment due to human error with the recording, this is explained. The score is high because 5 persons out of 33 is quite a lot. The same measuring instruments are used.
Selective reporting (reporting bias)	Unclear risk	Knowledge, Confidence, Satisfaction and Performance were reported both pre- and post-intervention. No published study-protocol.
Other bias	Low risk	Few participants makes the results weaker

Vedlegg 2b RoB-skjema Cioffi et.al. 2005

Methods	RCT	
Participants	26 midwife students	
Interventions	Two simulation sessions versus lectures as usual.	
Outcomes	Skills and Knowledge	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Students were randomised to intervention and control group by drawing names from a box. Students at the same university in their third semester.
Allocation Concealment (selection bias)	Low risk	The intervention is not possible to start in advance. No info on when the allocation was made due to the intervention
Blinding of participants and personnel (performance bias)	Unclear risk	No information on blinding in the study. Participants and personnel are not possible to blind due to the art of the intervention.
Blinding of outcome assessment (detection bias)	Unclear risk	No info if the ones who assessed data were blinded to group assignment, but they were independent
Incomplete outcome data (attrition bias)	High risk	Base-line demographics are not reported. It is neither reported if there is any loss of follow up in the study. It looks like the same measuring instrument is used in both groups.
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	No

Vedlegg 2c RoB-skjema De Regge et.al. 2008

Methods	RCT	
Participants	120 nurses in non-acute departments. 103 completed the study	
Interventions	IT – one participant to one instructor versus GT- six participants to one instructor	
Outcomes	Skills	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Assigned to groups by sealed envelopes
Allocation Concealment (selection bias)	Low risk	Randomisation kept secret until the start of intervention
Blinding of participants and personnel (performance bias)	Unclear risk	The intervention is not possible to blind. But the instructors were blinded to the results and the first test (T0)
Blinding of outcome assessment (detection bias)	Unclear risk	The outcome assessors were blinded to the results of test T0. But it is not described if they were blinded to group assignment
Incomplete outcome data (attrition bias)	Unclear risk	Reasons for dropout are described, but no information on how these are handled, etc. Intention to treat in the analysis. But a small dropout-rate
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	Not a big risk of contamination because the simulation is the same, it is the participants per instructor who varies.

Vedlegg 2d RoB-skjema Johnson et.al. 2012

Methods	RCT	
Participants	60 anaesthesia-nurse students in a military setting. 49 completed the study	
Interventions	High-fidelity simulation, Low-fidelity simulation or computer-based simulation	
Outcomes	Skills	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No information on the recruiting process. The authors describe they have randomised the participants, but the randomisation process is not described.
Allocation Concealment (selection bias)	High risk	No information
Blinding of participants and personnel (performance bias)	Unclear risk	Not possible to blind the intervention, and it cannot be started in advance due to the nature of the intervention
Blinding of outcome assessment (detection bias)	Low risk	The outcome assessors is blinded to group assignment
Incomplete outcome data (attrition bias)	Unclear risk	Baseline characteristic is reported, but not by group. No table on demographic data. The number in each group and dropout is described.
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	No

Vedlegg 2e RoB-skjema Maneval et.al. 2012

Methods	RCT	
Participants	26 new graduate nurses recruited to work in hospital	
Interventions	High-fidelity simulation in addition to standard orientation programme versus standard orientation programme	
Outcomes	Knowledge and skills	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomization via a computer
Allocation Concealment (selection bias)	Low risk	Not possible to start intervention in advance
Blinding of participants and personnel (performance bias)	Low risk	Not possible to blind due to the nature of the intervention
Blinding of outcome assessment (detection bias)	Low risk	The outcome assessors were blinded. The code that linked the questionnaires to participants was kept by an independent researcher.
Incomplete outcome data (attrition bias)	Low risk	Demographic data table is presented. The groups are not statistically different. In the control group the difference in age is a bit bigger, and more of the participants in the control group had experience as assistants in healthcare.
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	No

Vedlegg 2f RoB-skjema Monsieurs et.al. 2012

Methods	RCT	
Participants	235 nurses recruited, 191 completed	
Interventions	Individual simulation training with a manikin and text book, versus simulation with six participants to one instructor	
Outcomes	Skills	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A broad recruiting process. Not described in detail. The randomization is done by random numbers by a computer
Allocation Concealment (selection bias)	Low risk	Random generation
Blinding of participants and personnel (performance bias)	High risk	No information on blinding in the study
Blinding of outcome assessment (detection bias)	High risk	No information on blinding in the study
Incomplete outcome data (attrition bias)	Low risk	Baseline characteristics are presented. The groups are comparable. The reason for dropout is described, but not if the dropout is taken care of in the analysis. But a modest dropout-rate
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	No

Vedlegg 2g RoB-skjema Weiner et.al. 2011

Methods	RCT	
Participants	46 nurses in a postpartum ward	
Interventions	Self-directed with a textbook and simulation with an instructor versus traditional class	
Outcomes	Skills and knowledge	
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Random via computer
Allocation Concealment (selection bias)	Unclear risk	No information on when the group assignment and randomisation was revealed. Not possible to start the intervention in advance. Risk of contamination since it is colleagues in the same hospital.
Blinding of participants and personnel (performance bias)	Low risk	Blinding ok.
Blinding of outcome assessment (detection bias)	Low risk	The outcome assessors were blinded to group assignments
Incomplete outcome data (attrition bias)	Low risk	Baseline-characteristic is presented in a table, the groups is comparable. Does not seem to have been any dropout after the study started
Selective reporting (reporting bias)	Unclear risk	No published study-protocol
Other bias	Low risk	No