



HØGSKOLEN
I BERGEN

BERGEN UNIVERSITY COLLEGE

Prediksjon av pasienters voldsfare i psykiatriske akuttavdelinger: En systematisk oversikt og meta-analyse av tilgjengelige instrumenter.

Prediction of inpatient violence in psychiatric acute wards; A Systematic Review and Meta Analysis of Available Instruments.

Petter Jakobsen

Mastergrad i kunnskapsbasert praksis i helsefag

Senter for kunnskapsbasert praksis

Avdeling for helse- og sosialfag

14.12.2015

**Prediksjon av pasienters voldsfare i psykiatriske
akuttavdelinger: En systematisk oversikt og me-
taanalyse av tilgjengelige instrumenter.**

**Prediction of inpatient violence in psychiatric
acute wards; A Systematic Review and
Meta Analysis of Available Instruments.**

Petter Jakobsen

Veileder

Geir Smedslund

Innleveringsdato: 14.12.15

Antall ord: *13635*

Forord:

Jeg takker sjefen min, Ketil Ødegaard, ved Forskningsavdelingen i Helse Bergen HF Divisjon psykisk helsevern for å ha lagt forholdene til rette for å gjennomføre et mastergradsstudium på toppen av en full stilling, samt for finansiering av all pensum litteraturen. Jeg vil også takke Anne Kristin Snibsøer ved Senter for kunnskapsbasert praksis for å se masteroppgavepotensiale i en middelmådig eksamensoppgave fra andre semesteret på studiet. Jeg vil videre rette en stor takk til personene som har bidratt i prosessen med å skrive denne oppgaven. Gunhild Austrheim for råd og veiledning i utviklingen av en søkestrategi. Jill Bjarke for å ha vært den uavhengige andrepersonen i inklusjons- og kvalitetsvurderingsprosessen. Rolf Gjestad for hans bidrag med statistiske ekspertise. Tilslutt vil jeg rette en stor takk til min veileder Geir Smedslund ved Kunnskapscenteret, og takke for gode råd og et fruktbart samarbeid i skriveprosessen.

Prediksjon av pasienters voldsfare i psykiatriske akuttavdelinger: En systematisk oversikt og metaanalyse av tilgjengelige instrumenter.

Sammendrag:

Bakgrunn: Forekomsten av vold og aggresjon vesentlig hyppigere innen psykisk helsevern enn generelt i samfunnet, og inneliggende pasienters aggresjon blir håndtert med ulike preventive tiltak. Samtidig er det et krav fra politikere og pasientorganisasjoner at psykisk helsevern reduserer bruken av tvang. Da er det interessant at i en studie er det vist at systematisk bruk av voldsrisikoprediksjonsverktøy reduserer både forekomsten av voldsepisoder og bruk av preventive tiltak i akuttpsykiatrien.

Hensikt: Hensikten med denne systematiske oversiktsstudien er å identifisere validerte prediksjonsverktøy, samt kartlegge hvor nøyaktige verktøyene er til å predikere vold i en akuttpsykiatrisk kontekst.

Metode: Et systematisk søk etter litteratur ble gjennomført i databasene MEDLINE, EMBASE, PsycINFO og the NCJRS Abstracts Database. Prospektive studier som validere relevante verktøy ble identifisert, og kvalitetsvurdert med sjekklisten QUADAS-2. Til slutt ble det gjennomført en metaanalyse for å sammenligne hvor nøyaktig verktøyene predikerte voldsfare. Innledningsdelen inneholder en narrativ vurdering av samtlige artikler som beskriver de ulike verktøyene som ble inkludert, samt beskrivelser av studiene om de verktøyene som nesten ble inkludert.

Resultat: The Brøset Violence Checklist og The Extended Brøset Violence Checklist ble inkludert i metaanalysen. Resultatene er tilnærmet like for begge verktøyene. Prediksjonene er treffsikre ($AUC \geq 0,87$), men andelen falske positive prediksjoner er likevel vesentlig ($PPV < 15\%$). Innledningsdelen inneholder en narrativ vurdering av samtlige artikler som beskriver de ulike verktøyene som ble inkludert, samt beskrivelser av studiene om de verktøyene som nesten ble inkludert.

Konklusjon: Begge verktøyene er godt egnede kliniske verktøy til fortløpende vurdering av akutte risikofaktorer for fysisk aggresjon ved innleggelse på en akuttpsykiatrisk sengepost.

Nøkkelord: Violence, Aggression, Psychiatry, Risk Assessment, Review

**Prediction of inpatient violence in psychiatric acute wards;
A Systematic Review and Meta Analysis of Available Instruments.**

Abstract:

Background: The occurrence of violence and aggression is significantly more frequent in mental health care than in society in general. Inpatient aggression is managed by various preventive measures, but there is a demand from politicians and patient organizations that mental health services reduce the use of coercive measures. In this context, it is interesting that one study confirms that the systematic use of tools that predict the risk of violence can reduce the frequency of violence as well as the use of preventive measures.

Aims: The purpose of this systematic study is to identify valid predictive tools and assess how accurate they predict violence in an acute psychiatric context.

Method: A systematic literature search was conducted in the databases MEDLINE, EMBASE, PsycINFO and the NCJRS Abstracts Database. Prospective studies that validate relevant tools were identified, and the checklist QUADAS-2 was used to assess their quality. Finally a meta-analysis was conducted to compare how accurately the tools predict the risk for violence.

Results: The Brøset Violence Checklist and The Extended Brøset Violence Checklist were included in the meta-analysis. The results are similar for both tools. They accurately predict the risk for violence ($AUC \geq 0.87$), but the percentage of false positive predictions is high ($PPV < 15\%$). The Introductory section contains a narrative assessment of all the articles that describe the various included tools, as well as descriptions of studies of tools that were almost included.

Conclusion: Both checklists are well suited clinical tools for the continuous assessment of acute risk for physical aggression in patients when admitted to an acute psychiatric ward.

Keywords: Violence; Aggression; Psychiatry; Risk Assessment; Review

INNHold		Side
1. Innledning		7
2. Bakgrunn		7
3. Formål og problemstilling		9
4. Metode		
4.1 Inklusjons kriterier		9
4.2 Litteratursøk		
4.2.1 Generelt om litteratursøk		10
4.2.2 Hvor søke etter litteratur		11
4.2.3 Søkestrategi		12
4.3 Datasamling og analyse		
4.3.1 Utvalgelse av inkluderbare studier		13
4.3.2 Datauttrekk		15
4.3.3 Kvalitetsvurdering		17
4.3.4 Analyse		19
5. Funn		
5.1 Litteratursøket		
5.1.1 Presentasjon av ulike verktøy, samt endelig utvalgelse av inkluderte studier		23
5.1.2 Presentasjon av inkluderte verktøy og studier		24
5.1.3 Presentasjon av verktøyene som nesten ble inkludert, med begrunnelse for eksklusjonen av studiene.		26
5.2 Resultatet av kvalitetsvurderingen av studiene, samt endelig utvalgelse av artikler som kan inkluderes i metaanalysen.		30
5.3 Metaanalyse og statistiske resultater		33
6. Diskusjon		37
7. Konklusjon		41
8. Metodekritikk		41
9. Referanser		42

Artikkel med vedlegg

Vedlegg I - VII

1. Innledning

I denne innledningsdelen vil jeg presentere et prosjekt med tittel: Prediksjon av pasienters voldsfare i psykiatriske akuttavdelinger, en systematisk oversikt og metaanalyse av tilgjengelige instrumenter. I tillegg er det, med utgangspunkt i innledningsdelen, produsert en artikkel med tittel: Voldsprediksjon i psykiatriske akuttavdelinger, en metaanalyse av tilgjengelige verktøy. I artikkelen er fokuset utelukkende på de studiene og verktøyene som er inkludert i innledningsdelens metaanalyse. Artikkelen er ment for publisering i tidsskriftet Sykepleien Forskning. The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (DTA håndboken) (Deeks, Wisniewski & Davenport, 2013; Bossuyt & Leeflang, 2008; de Vet et al., 2008; Reitsma et al., 2009; Macaskill, et al., 2010) har vært veiledende for denne oppgaven. Dette har medført enkelte utfordringer. Håndboken er ennå under utarbeidelse, hvor kun utvalgte kapitler er gjort tilgjengelig. Dermed blir informasjonen om hvordan skrive en protokoll noe mangelfull. Grunnet at metoden er avvikende fra en mer tradisjonell tilnærming (Deeks, Wisniewski & Davenport, 2013, s. 4), blir den uferdige boken likevel hovedredskapet for denne prosjektbeskrivelsen. I tillegg har også Cochrane Handbook for Systematic Reviews of Interventions Version (Higgins & Green, 2011) vært nyttig støttelitteratur, og har gitt gode anbefalinger, på enkelte områder i løpet av prosessen.

2. Bakgrunn

I utgangspunktet er forekomsten av vold og aggresjon vesentlig hyppigere i psykiatrien enn generelt i samfunnet. Samtidig er helsetilstanden til sykepleiere som regelmessig eksponeres for aggresjon signifikant forringet i forhold til andre helsearbeidere (Cornaggia et al., 2011). En vesentlig del av en helsearbeiders hverdag i akuttpsykiatrien handler om å håndtere aggresjon blant inneliggende pasienter med forskjellige tiltak, alt fra personlig oppfølging av pasienter til iverksettelse av preventive tiltak (Björkdahl, Olsson & Palmstierna, 2006). Preventive tiltak er gjerne bruk av tvangsmidler som belter, remmer, fastholding, isolasjon og korttidsvirkende legemidler. Men behandlingstiltaket skjerming er også mye brukt. Skjerming handler om å holde pasienten adskilt fra avdelingen, og begrense kontakten med andre til kun et fåtalls pleiepersonell (Knutzen et al., 2014, s. 9). Bruk av tvang er nøye regulert av lovgivningen, men samtidig er det et krav fra politikere, helsebyråkrater og pasientorganisasjoner at psykisk helsevern generelt reduserer bruken av tvang (Knutzen et al., 2014, s. 8). I denne sammenhengen er det både relevant og interessant at en studie (Abderhalden et al., 2008) finner at systematisk bruk av et voldsrisikoprediksjonsverktøy i

akuttpsykiatrien reduserte både bruken av preventive tiltak og forekomsten av voldsepisoder. Effektene kom som et resultat av økt bevissthet omkring potensielle farer, og dermed mer fokusert bruk av preventive tiltak, men samtidig ble personalet også mer fokusert på konfliktdempende samhandling med pasientene. Samtidig anbefaler internasjonale retningslinjer (National Institute for Health and Clinical Excellence (NICE), 2005) at psykiatriske pasienter farlighetsvurderes, og gir konsensusbaserte anbefalinger om bruk av klinisk skjønn ved akutte voldsrisikovurderinger (s. 18). Men det foreligger god dokumentasjon på at klinisk skjønn er svært lite egnet til å forutse voldelig adferd (Singh, Grann & Fazel, 2011). Retningslinjens anbefaling (NICE, 2005, s. 9) gjelder voldsrisikovurderinger i et korttidsperspektiv, inntil 72 timer, for inneliggende psykiatriske pasienter. Anbefalingene må sees i sammenheng med den tids konsensus om at videre forskning er påkrevd, for å utvikle adekvate prediksjonsverktøy for risikovurdering av inneliggende psykiatriske pasienter (Woods & Ashley, 2007; Steinert, 2002; NICE, 2005, s. 66). Norske retningslinjer som behandler temaet gir enten juridisk fundamenterte anbefalinger (Sosial- og helsedirektoratet, 2007; Helsedirektoratet, 2012, s. 53), eller anbefalinger uten et oppdatert kunnskapsgrunnlag og med et uklart tidsperspektiv (Helsedirektoratet, 2013, s. 51). En mulig forklaring på tilstanden er at vurderinger om personers potensielle farlighet normalt har vært sikkerhets- og fengselspsykiatriens domene (Woods & Ashley, 2007; Hartvig et al., 2011). Fokuset har derfor hovedsakelig vært på langtidsprediksjon av risikoen når potensielt farlige personer returnerer til samfunnet. Men i løpet av de siste årene er det publisert en rekke studier som har fokus på prediksjon av voldsfare innen ordinær psykiatri, også med en akutt tidshorisont (Hartvig et al., 2011). Prediksjonsverktøyene som studeres kan deles inn i to grupper. En gruppe er verktøy som fortløpende kan brukes til å vurdere de akutte risikofaktorene ved en pasient. Den andre gruppen er verktøy som identifiserer potensielt farlige pasienter, for så å selekere dem til en grundigere utredning av farlighet. Det er i tillegg to forskjellige tilnærminger til hvordan strukturerte risikovurderinger er designet (Woods & Ashley, 2007; Singh et al., 2011; Hartvig et al., 2011). Sjekklistene, der sumskåren klassifiserer graden av voldsfare. Men også verktøy for å gjøre en strukturert klinisk vurdering, slik at systematisk innsamlet materiale kan brukes for å støtte behandlerens konklusjoner, basert på informert klinisk skjønn.

På tidspunktet prosjektbeskrivelsen for denne systematiske oversikten ble til, ble det ikke identifisert noen systematiske oversiktsartikler, publisert i løpet av de siste 10 årene, som omhandler temaet prediksjonsverktøy for voldsrisikovurdering av inneliggende psykiatriske

pasienter i et korttidsperspektiv. Det ble søkt i databasene The Cochrane Library, Medline, Embase og PsycINFO i juni 2013. Søkeordene var *violen**, *aggress**, *inpatient** og *psychiatric*, Søket ble avgrenset til "review articles".

3. Formål og problemstilling

Formålet med dette prosjektet er å finne fram til et validert systematisk verktøy som kan benyttes til å fortløpende vurdere akutte risikofaktorer for pasienters potensiale for fysisk aggresjon ved innleggelse i akuttpsykiatrien. Et verktøy som vil gjøre det mulig å redusere stress i avdelingen, gjennom iverksettelse av hensiktsmessige preventive tiltak ovenfor individer som identifiseres. Problemstillingen blir dermed å identifisere hvilke ulike verktøy som finnes, samt forsøke å kartlegge hvor nøyaktige de er til å forutse voldsfare i et korttidsperspektiv. Dette gjøres ved å inkludere artikler som beskriver prediksjonsverktøy. En artikkel kan beskrive ett eller flere verktøy, og ett verktøy kan være beskrevet i flere artikler.

4. Metode

4.1 Inklusjons kriterier

Et diagnostisk verktøy sitt formål er å identifisere tilstedeværelsen av en tilstand. I en typisk studie av diagnostiske verktøy blir testen sammenlignet med en referansestandard, metoden som vanligvis brukes til å diagnostisere tilstanden. Denne typen studier tar som oftest utgangspunkt i et bestemt målepunkt, og kan beskrives som tverrsnittstudier (Bossuyt & Leeflang, 2008). For studier av voldsrisikoverktøy blir det annerledes, da slike verktøy ikke prøver å diagnostisere en tilstand, men å forutsi sannsynligheten for en fremtidig hendelse (Fazel et al., 2012). Skal man undersøke en gruppe mennesker over tid, er det en prospektiv kohortstudie som er den foretrukne design (Polit & Beck, 2012, s. 225). Siden voldsrisiko ikke kan påvises av for eksempel blodprøver eller hjerneskaning, finnes det ingen objektiv referansestandard på dette feltet. En eventuelt fremtidig hendelse må derfor betraktes som referansestandard. Ut fra en slik tanke bør også retrospektive studier egne seg til å belyse temaet. Men siden slike studier er betraktet som vesentlig metodologisk svakere enn prospektive (Polit & Beck, 2012, s. 224), vil de likevel ikke bli inkludert i dette prosjektet. En inkluderbar prospektiv kohortstudie bør foregå i et psykiatrisk miljø rimelig likt det man finner innen norsk akuttpsykiatrisk helsevern (Mellesdal et al., 2010). Dette innebærer en sammensatt pasientgruppe i forhold til diagnose, rusbruk, tilstand og sykdomshistorie. Pasientene kan være innlagt enten frivillig eller på tvang. Slike forhold vil trolig variere sterkt

over landegrensener, grunnet forhold som tradisjon og psykiatrilovgivning. Derfor vil den viktigste avgrensningen i inkluderbare studier være at studien ikke er gjennomført på en sikkerhetspsykiatrisk populasjon. Pasientgruppen er allerede definert som farlige, og dermed ikke representative for psykiatriske pasienter generelt (Kumar & Simpson, 2005). Deltakernes tilstand beskrives ikke mer utfyllende enn det som er nevnt i bakgrunns kapitlet og i forskningsspørsmålet. Det dreier seg om innlagte pasienters aggresjon, nærmere bestemt vold rettet mot både personale og medpasienter. I dette prosjektet har jeg inkludert studier om ulike prediksjonsverktøy for risikovurdering av inneliggende psykiatriske pasienter i et kortidsperspektiv. Verktøy måtte være egnet for bruk av sykepleiere, samt hensiktsmessig til sin intenderte bruk (Polit & Beck, 2012, s. 345). Et omstendelig, tidkrevende og omfattende verktøy vil ikke være særlig egnet for praktisk bruk, og dermed ikke for dette prosjektets formål.

4.2. Litteratursøk

4.2.1 Generelt om litteratursøk

For å skrive en systematisk oversiktsartikkel om et gitt tema er det nødvendig å få oversikten over samtlige relevante studier på feltet, slik at det grunnleggende forskningsspørsmålet best kan besvares. En søkestrategi må utvikles, og de Vet et al. (2008¹) gir omfattende anbefalinger om hvordan. Prosessen beskrives som en stor utfordring (kapitel 7.1), og det anbefales sterkt å konsultere en erfaren bibliotekar. Derfor fikk jeg bistand fra en av Høgskolen i Bergen sine bibliotekarer for å kvalitetssikre søkestrategien. Det kan ha vært problematisk at dette prosjektet kun forholdt seg til litteratur på engelsk og skandinaviske språk, fordi en slik språklig begrensning ofte er antatt å gi publikasjonsskjevheter i utvalget av studier. Det er slik at positive funn gjerne oftere publiseres internasjonalt, i motsetning til ikke signifikante funn (Polit & Beck, 2012, s. 657), men ikke nødvendigvis (de Vet et al., 2008). For dette prosjektet var begrensningen nødvendig grunnet økonomiske og lingvistiske ressurser. I praksis ble kun et fåtalls artikler ekskludert grunnet språk, nærmere beskrevet var det tre artikler som ikke ble lest i fulltekst grunnet utilgjengelig språk. Språkene var kinesisk, kroatisk og fransk.

¹ DTA Håndbokens 7. kapittel (de Vet et al., 2008) finnes kun tilgjengelig i versjon 0,4, og kapitlet inneholder ennå ikke sidetall. Siden referansestilen Harvard krever sidetall i siteringen, samt at teksten er omfangsrik, har jeg valgt å benytte underkapittelnummereringen som erstatning for sidetall.

4.2.2. Hvor søke etter litteratur

I følge anbefalingene til de Vet, et al. (2008, kapittel 7.2.1.3) er det å søke etter litteratur i databasene MEDLINE og EMBASE å betrakte som et minimumskrav når man prøver å identifisere studier om diagnostiske tester og lignende verktøys nøyaktighet. I tillegg anbefales det søk i mer fagspesifikke databaser. Dermed ble den psykologi- og psykiatrifaglige databasen PsycINFO selvskreven for dette prosjektet. I tillegg er det vurdert som hensiktsmessig å benytte databasen The US National Criminal Justice Reference Service Abstracts, siden akkurat denne fremstår som en god kilde til å identifisere studier innenfor dette prosjektets tematikk (Fazel et al., 2012; Singh et al., 2011). Videre sier teorien at i brukergrensesnittet Ovid, som benyttes for databasene MEDLINE, EMBASE og PsycINFO, foreslås det relaterte artikler for funnene, noe som er en god kilde til å identifisere flere relevante studier. En tilsvarende kilde er å sjekke referanser i artiklene som blir identifisert (de Vet et al., 2008, kapittel 7.2.2.1). I tillegg kan Ovid gjøre siteringssøk. Potensielt relevante artikler identifiseres ut i fra at de refererer til en allerede identifisert artikkel (de Vet et al., 2008, kapittel 7.2.1.6). Google Scholar kan benytte tilsvarende. Tilslutt må det gjøres et søk etter grå litteratur, dette er studier som ikke er formelt publisert, samt konferansemateriale. Dette er et viktig tiltak for å motvirke potensialet for publikasjonsskjevheter i materialet (de Vet et al., 2008, kapittel 7.2.1.8). OpenSIGLE, NTIS og BIOSIS er databaser som anbefales. I tillegg kan eksperter på fagfeltet også gi nyttig innsikt i upublisert materiale.

I praksis ble det gjort et systematisk litteratursøk i de fire databasene som omtales i starten av den teoretiske gjennomgangen, MEDLINE, EMBASE, PsycINFO og The US National Criminal Justice Reference Service Abstracts. Årsaken er at dette prosjektet kun har én medarbeider, så det ble vurdert å være for omfattende å gjøre siteringssøk og søk etter grå litteratur. Teorien beskriver strenge krav til prosessen med å skrive en systematisk oversiktsartikkel på Cochrane nivå, men samtidig gjøres normalt slikt arbeid av en større gruppe mennesker. Likevel ble Stål Bjørkly, professor og klinisk psykologspesialist, kontaktet underveis i prosessen. Mannen er av mange fagfolk vurdert som Norges store ekspert i psykologisk forståelse av aggresjon hos mennesker. I tillegg er Bjørkly medforfatter på flere av artiklene som vurderes i denne studien (Bjørkly & Moger, 2007; Hartvig et al., 2011; Knutzen et al., 2014). Han hadde lite å tilføye omkring upublisert materiale, men var fornøyd med utvalget av inkluderbare prediksjonsverktøy som denne studien hadde funnet fram til etter at litteratursøket var sortert.

4.2.3 Søkestrategi

Teorien hevder at å designe en søkestrategi vil by på store utfordringer, og det er flere essensielle premisser den bør utvikles i forhold til (de Vet et al., 2008, kapittel 7.4.1). Hovedårsaken til dette har sin forklaring i at det er få standardiserte emneord for tester i de ulike databasene, samt inkonsekvent bruk av disse. I tillegg er ofte innrapporteringen av relevante studier svært mangelfull. En hensiktsmessig søkestrategi består i utgangspunktet av to deler: identifisering av aktuelle tester, og beskrivelse av sykdomstilstanden testene er beregnet for (de Vet et al., 2008, kapittel 7.4.2). Det er viktig å være ordrik i beskrivelsen av testen, da fanges artikler opp, uavhengig av hvordan den er indeksert i databasen. Søket bør ha høy følsomhet og lav presisjon (de Vet et al., 2008, kapittel 7.4.4). Normalt vil lav presisjon gi en stor mengde irrelevante artikler, men dette er nødvendig for at relevante studier ikke skal unnsnippe. Både ulike metodologiske søkefiltre og andre inklusjonskriterier bør unngås, da slikt kan medføre at relevante studier ikke identifiseres (de Vet et al., 2008, kapittel 7.4.2). Å la seg inspirere er anbefalt (de Vet et al., 2008, kapittel 7.4.5). Samtidig anbefales det å benytte søkeord og relevante emneord i kombinasjon. Emneordene bør være utvidet (explode), og på minst mulig spesifikt nivå, dette for å identifisere flest mulig potensielt relevante artikler.

Med inspirasjon fra en systematisk oversiktsartikkel om prediksjon av psykiatriske pasienters voldsfare i etterkant av utskrivelse (Singh et al., 2011), ble oppgavens søkestrategi utviklet i flere etapper, i tett samarbeid med Senter for kunnskapsbasert praksis sin bibliotekar. Underveis i prosessen ble også søkestrategien presentert i undervisningen på høyskolen. Den faglige diskusjonen og innspillene omkring søkestrategien var nyttig. Den endelige søkestrategien ble utviklet og gjennomført lojalt etter kriteriene beskrevet ovenfor, dog ble det innført flere betingelser i søket. Dette ble gjort for å begrense antall treff til en håndterbar mengde, og det er god støtte i faglitteraturen for en slik tilnærming når antall treff blir for omfattende (de Vet et al., 2008, kapittel 7.4.7). De nye betingelsene er en beskrivelse av populasjonen for studiene, samt populasjonen plassert i en relevant kontekst. Det er under slike omstendigheter viktig å være svært åpen i beskrivelsen av de nye betingelsene, faren for at en for snever beskrivelse kan fjerne relevante studier fra resultatet av søket er stor.

Søkestrategien, utført i de tre databasene i brukergrensesnittet Ovid, beskrives her kun med søkeord, emneordene vil fremkomme av vedlagte søkehistorikk for de ulike databasene (se tabeller 1 - 3, vedlegg I, II & III). (violen* or aggress* or attack* or assault*) AND (risk* or

assess* or predict* or tool* or instrument* or measur* or scale* or test* or survey* or checklist* or questionnaire* or appraisal* or sensitivity or specificity) AND (mental* or bipolar* or schizo* or psychos* or personality* or manic* or psychiatric*) AND (psychiatric adj (inpatient* or hospital* or admission* or readmission*)) The US National Criminal Justice Reference Service Abstracts har vesentlig mer begrensede muligheter, og søkestrategien der er tilsvarende enkel: risk, violen* og inpatient*. (se tabell 4, vedlegg IV).

Søket ble gjennomført 11. mars 2014. Resultatene fra søkene i de ulike databasene ble importert inn i referanseverktøyet EndNote, og dubletter ble slettet. Resultatet ble at totalt 2183 ulike referanser ble identifisert gjennom dette første litteratursøket. Høsten 2015 ble det i tillegg gjort et oppdateringssøk i de samme databasene, et enkelt søk hvor det kun ble søkt på de inkluderte verktøyenes navn, både i fulltekst og forkortet form. Totalt ble 2462 referanser identifisert gjennom begge søkene. I boken til de Vet, et al (2008, kapittel 7.4.12) anbefales det gjerne at det gjøres et oppdateringssøk, for å finne studier publisert i tiden etter det opprinnelige søket. Men det anbefales ingen konkret tidshorisont for når et slikt søk er nødvendig. I utgangspunktet skal det gjøres et identisk søk med det opprinnelige, kun begrenset til tidsrommet etter det første søket. Men grunnet at behovet for å gjøre et oppdateringssøk er uklar, og at den valgte formen for et oppdateringssøk ble diskutert i undervisningen ved Høgskolen i Bergen, samt anbefalt av flere personer tilknyttet høgskolen faglig, ble det slik. Gjennom oppdateringssøket ble det identifisert to nyere artikler om relevante verktøy.

4.3 Datainnsamling og analyse

4.3.1 Utvelgelse av inkluderbare studier

Som tidligere omtalt er formålet med denne systematiske oversikten å finne fram til et verktøy som kan benyttes til fortløpende å vurdere akutte risikofaktorer for pasienters potensielle farlighet i akuttpsykiatrien. Problemstillingen er todelt. Først skal verktøy som er egnet for dette prosjektets formål identifiseres, videre skal verktøyene kvalitetsvurderes både i forhold til formålet, men også i forhold til verktøyets nøyaktighet i voldsfareprediksjonen. Nedenfor følger en beskrivelse av prosessen med å identifisere ulike prediksjonsverktøy som potensielt er egnet for dette prosjektets formål ut av de totalt 2462 referansene som ble identifisert gjennom litteratursøkene, se også vedlagte flytdiagram (figur 1, vedlegg V) som illustrerer prosessen.

Higgins & Green (2011, kapittel 7.2) beskriver en prosess for utvelgelse av relevante studier fra et omfattende litteratursøk. Dette er en stegvis prosess som starter med å importere samtlige funn fra søket inn i referanseverktøyet EndNote. Deretter gjennomgås titler og abstrakter, og alt materialet som fremstår som irrelevant fjernes. I denne sammenhengen ble dette steget gjort noe utradisjonelt i to faser. Først ble titler og abstrakter gjennomgått av en person, og de definitivt irrelevante studiene ble fjernet. Ved den minste usikkerhet ble artiklene ikke fjernet. Det er akseptabelt at kun en person står for denne grovsorteringen, men ideelt sett bør denne fasen gjøres av to uavhengige personer (Higgins & Green, 2011, kapittel 7.2). Problemet med en slik tilnærming er at en person kan være for streng i utfasingen av artikler, men i arbeidet med denne masteroppgaven var en annen løsning ikke praktisk mulig å få til. Andre fase var et forsøk på å kompensere for denne svakheten ved at to personer uavhengig av hverandre gikk gjennom titler og abstrakter på de siste 394 referansene i dette første steget av utvelgelsesprosessen. Denne andre personen er en medstudent fra masterstudiet og kollega på arbeidsplassen. Hun er ikke å betrakte som en faglig ekspert på temaet oppgaven omhandler, noe som er å betrakte som gunstig når man deltar i en slik utvelgelsesprosess (Higgins & Green, 2011, kapitel 7.2). Tilslutt ble 57 referanser lest i fulltekst av de samme to personene uavhengig av hverandre, og i første omgang ble de kun vurdert i forhold til enkelte utvalgte inklusjonskriterier. Kriteriene var at studien måtte handle om en form for risikovurdering av inneliggende voksne akuttpsykiatriske pasienter, og med et tidsperspektiv for vurderingen som var rimelig kort. Studier gjort utelukkende på en sikkerhetspsykiatrisk populasjon, samt studier om barn og ungdom ble ekskludert. Fjorten ulike verktøy, som i utgangspunktet tilsynelatende fremsto som utviklet for å vurdere akutte risikofaktorer for voksne pasienters potensielle farlighet i akuttpsykiatrien, ble identifisert. Samtlige studier og verktøy presenteres i kapitlet om resultater. I den narrative presentasjonen av de ulike verktøyene gjøres samtidig en grundig vurdering av artiklene i forhold til samtlige inklusjonskriterier. Inklusjonskriteriene er, i tillegg til de tidligere nevnte, om studiedesignet er relevant, om verktøyet blir validert, om verktøyet er egnet for bruk av sykepleiere, samt om det er hensiktsmessig til sin intenderte bruk. Utover at dette blir en mer omstendelig utvelgelsesprosess enn teorien krever (Higgins & Green, 2011, kapitel 7.2), så er det intet som tilsier at valg av en slik fremgangsmåte svekker kvaliteten på identifiseringsarbeidet. Begrunnelsen for valg av en slik fremgangsmåte er dels for å imøtekomme oppgavens problemstilling, som ønsker å identifisere samtlige verktøy som er utviklet for å predikere akutte risikofaktorer. Samtidig er inklusjonskriteriene rimelig komplekse, og en grundig beskrivelse av de ulike verktøyene sett opp mot inklusjonskriteriene gir dermed også en

dypere innsikt i, og god forståelse av, den komplekse utvelgelsesprosessen. Valg av metode er også inspirert av et foredrag av professor, psykiater og lærebokforfatter Ulrik Malt, der han er svært kritisk til hvordan arbeidet med en bestemt retningslinje forholdt seg ukritisk til det faglige innholdet i mye av den forskningen retningslinjen bygget på. Det tas på ingen måte stilling til Malts påstander her, men ønsket hans om at man må grundigere inn i materien før man konkluderer har vært inspirerende i dette arbeidet.

Det oppstod ingen uenighet mellom de to personene, som vurderte materialet uavhengig av hverandre, omkring utvelgelsen av relevante artikler som tilfredsstilte inklusjonskriteriene. Men om uenighet hadde oppstått, så var planen i første omgang å diskutere seg fram til en akseptabel løsning for begge parter. Om uenigheten for eksempel skyldtes manglende rapportering i en artikkel, så kunne en eventuell avgjørelse utsettes til utfyllende informasjon var innhentet fra artikkelforfattere. Uansett, om partene ikke skulle lykkes i å oppnå enighet seg i mellom, så kunne i tilfelle en tredjepart blitt konsultert.

4.3.2 Datauttrekk

Etter at inklusjonen av artikler er gjennomført så skal data trekkes ut av materialet. Et datauttrekk bør gjøres av to personer, tilsvarende som beskrevet ovenfor (Higgins & Green, 2011, kapittel 7.6). Hensikten med en slik fremgangsmåte er å eliminere feilkilder som personlige oppfatninger og ignoranse. En utfordring er at en enkelt studie kan generere utallige artikler. Dette løses praktisk ved at data fra hver artikkel trekkes ut separat, for deretter å bli sammenfattet. En forutsetning for at denne prosessen kan gjennomføres er at det utvikles et verktøy hvor man registrerer ønskete data. Premissene er inklusjonskriteriene og forskningsspørsmålet, samt hva som er nødvendig til metaanalysen (Higgins & Green, 2011, kapittel 7.5.1). Det fremstår som hensiktsmessig å utvikle et elektronisk skjema for dette formål. Higgins & Green (2011, kapittel 7.5.3) gir generelle gode tips om hvordan utvikle et datainnsamlings-skjema, men fordi metoden for en systematisk oversikt om diagnostiske tester er avvikende fra tilsvarende metode for systematiske oversikter over intervensjonsstudier (Deeks, Wisniewski & Davenport, 2013, s. 4), valgte jeg kun å forholde meg til enkelte overordnede tips. Dette blir å versjonshandtere skjemaene, ha muligheten til å nummerere unike studier, samt gi et unikt nummer for de forskjellige artiklene. Ellers var STARD sjekklisten utgangspunktet for å designe et datainnsamlingsverktøy. STARD (Bossuyt et al., 2003) er utviklet for å forbedre kvaliteten og innholdet i artikler om diagnostiske tester. Verktøyet er en 25 punkters sjekklister med fokuset på at all nødvendig informasjon, til bruk i

systematiske oversiktsartikler, er til stede i publikasjoner. At en artikkel inneholder mangelfull informasjon har kun to mulige forklaringer: Enten er artikkelen av dårlig kvalitet, eller så er studien som beskrives av mangelfull kvalitet?. Uten kunnskap om hvordan alt henger sammen, kan man ikke annet enn å anta at studien er av dårlig metodologisk kvalitet (Reitsma et al., 2009, s. 4). Som nevnt tidligere er det visse forskjeller mellom studier om diagnostiske verktøy og prognostiske verktøy (Bossuyt & Leeflang, 2008), men likevel er STARD egnet til dette prosjektets formål. Visse tilpasninger er trolig nødvendig, men slikt er å betrakte som en naturlig del av prosessen. Det anbefales å pilotere et datauttrekksskjema, både for å samkjøre personene som skal gjennomføre datainnsamlingen, og for at selve skjemat skal få en hensiktsmessig form (Higgins & Green, 2011, kapittel 7.6.3).

Det ble utviklet et datauttrekksskjema, etter metoden beskrevet ovenfor, i Excel . Verktøyet er vedlagt i tabell 5 (vedlegg VI). I tillegg ble dette skjemaet brukt som utgangspunkt for databasen som ble utarbeidet for de statistiske analysene. Relevante data ble overført fra datauttrekksskjema, og en syntaks ble utviklet for å generere de ulike statistiske variablene, som omtales under 4.3.4 Analyse. Datauttrekket ble gjort i fellesskap med statistikeren som bidro til dette prosjektet, for deretter å bli kontrollert av begge parter uavhengig av hverandre. Denne metoden er ikke helt i tråd med hvordan Higgins & Green (2011, kapittel 7.6) krever den skal være, men det ble vurdert slik at den uavhengige etterkontrollen sikret kvaliteten. Fordelen med den valgte tilnærmingen er at datauttrekksskjemaet ble dynamisk i datauttrekksprosessen, og en pilotering av skjemaet ble dermed ikke nødvendig i forkant.

Datauttrekksskjemaet tok som planlagt utgangspunkt i sjekklisten STARD (Bossuyt & Leeflang, 2008), men ble tilpasset slik at skjemaet fremsto som relevant for dette prosjektets formål. Hovedfokuset for designet av skjemaet var at vi både skulle trekke ut relevante data til analysen, samt til den påfølgende kvalitetsvurderingen av de inkluderte studiene. I etterkant av dette arbeidet ble vi presentert for rapporteringsretningslinjen RAGEE (Singh et al., 2015). RAGEE er absolutt mer egnet for oppgavens formål enn STARD, siden RAGEE spesielt er utviklet for å forbedre kvaliteten og innholdet i artikler om prognostiske voldsrisiko prediksjonsverktøy. Hensikten med RAGEE er å innføre kunnskapsbaserte metoder og dermed heve kvaliteten på publiserte studier, innen et fagfelt som generelt har vært svak på dette feltet (Singh et al., 2015). Datauttrekksskjemaet (se tabell 5, vedlegg VI) ble derfor til slutt vurdert opp mot sjekklisten RAGEE. Datauttrekksskjemaet ble vurdert til å kvalitativt være i samsvar med prinsippene i RAGEE, dog noe forenklet, men hensiktsmessig til sitt intenderte formål.

Det ble gjort en avgrensning i datauttrekket for en studie (Abderhalden et al., 2006). Her rapporteres data med utgangspunkt i to utfallsmål, det første er voldsepisode, mens det andre utfallsmålet er voldsepisode kombinert med iverksetting av preventive tiltak, som skjerming og medisinerings. For det sistnevnte utfallsmålet er det ikke trukket ut data. Årsaken er at selv om hensikten med en slik tilnærming er å håndtere problemet med falske positive identifikasjoner som skyldes iverksettelse av preventive tiltak. Så kan ikke et slikt utfallsmål betraktes som uavhengig av prediksjonen, siden det er prediksjonen som selekterer til tiltakene. Dermed får denne tilnærmingen et preg av selvoppfyllende profeti, noe Abderhalden, et al. (2006) også til en vis grad problematiserer i artikkelen. Til slutt så må det påpekes at det innimellom er små avvikende verdier mellom datauttrekket og i de statistiske analysene. Dette forklarer statistikerne med at det praktiseres ulike avrundingsrutiner i forskjellige statistiske programmer.

4.3.3 Kvalitetsvurdering

Formålet med kvalitetsvurderingen av de inkluderte studiene er å vurdere om de er valide (Reitsma et al., 2009, s. 2). Valideringen av enkeltstudiene er todelt, både intern og ekstern validitet må vurderes. Intern validitet handler om faren for systematiske feil i en studie, og at feilkilden preger resultatene (Polit & Beck, 2012, s. 176). Feilkilden kan ligge i alt fra studiedesignet, sammensetningen av deltakergruppen til dårlig datasamlingsmetode. Tilfeldige feilmålinger er noe annet enn systematiske feil. Slike antas ikke å påvirke en studie i en bestemt retning, slik en systematisk feil vil gjøre (Reitsma et al., 2009, s. 3). Ekstern validitet kan beskrives som generaliserbarheten av en studie, om funnene har overføringsverdi utover selve studiesituasjonen (Polit & Beck, 2012, s. 250). Med andre ord om studiedeltakerne, studiesettingen og valg av måleverktøy, samt bruken av verktøyet, har relevans for denne prosjektbeskrivelsens inklusjonskriterier.

Reitsma et al. (2009, s. 5) anbefaler at sjekklisten QUADAS benyttes til å kvalitetsvurdere de inkluderte studiene. QUADAS er oppdatert til QUADAS-2 (Whiting et al., 2011), og jeg velger derfor å benytte sistnevnte til å kvalitetsvurdere de inkluderte studiene. QUADAS-2 tar utgangspunkt i en definisjon av kvalitet som et forhold mellom fraværet av potensielle systematiske feil i studien, og anvendbarheten av resultatene i forhold til det relevante forskningsspørsmålet. Kvalitetsvurderingen er inndelt i 4 kjerneområder, inklusjonen av pasienter, indekstest, referansestandard, og det siste området som er forløpet i studien, samt timingen mellom indekstest og referansestandard. Anvendbarheten vurderes kun på de tre

første kjerneområdene (Whiting et al., 2011). Konkret gjennomføres QUADAS-2 ved at for hvert kjerneområde presenterer man først informasjonen som vurderingen skal bygge på, så besvares noen konkrete spørsmål omkring informasjonen med enten ja, nei eller uklart. Dette oppsummeres så i en vurdering av om faren for systematiske feil i studien enten er lav, høy eller uklar for kjerneområdet. Er alle spørsmålene besvart med ja for et kjerneområde, så er det lav fare for systematisk feil i artikkelen/studien på akkurat dette området. Deretter vurderes anvendbarheten omkring hvordan studiepopulasjon, verktøy og referansestandarder i studien er relevant i forhold til forskningsspørsmålet som skal besvares gjennom den systematiske oversikten (QUADAS, 2011; Whiting et al., 2011). Det er ikke nødvendig å følge sjekklisten slavisk, en viss tilpasning er akseptert, spesielt vedrørende spørsmål som er irrelevante i forhold til testen som utprøves sine egenskaper. Men ved eventuelle justeringer bør minst to personer diskutere og evaluere disse før de gjennomføres. Alle vurderingene som gjøres i prosessen bør dokumenteres. Selve kvalitetsvurderingen skal også utføres av to personer uavhengig av hverandre, og uenigheter skal diskuteres. Eventuelt skal en tredje person konsulteres. All uenighet skal dokumenteres og rapporteres (Reitsma et al., 2009, s. 8). I artikkelen til Whiting, et al. (2011) stilles et minimumskrav om at en systematisk oversikt bør presentere en oppsummert oversikt over alle vurderte studier. Reitsma et al., (2009, s. 21) krever i tillegg en narrativ oppsummering av kvalitetsvurderingen i den systematiske oversikten. En slik oppsummering presenteres i resultatdelen.

Før arbeidet med kvalitetsvurderingen startet ble det nødvendig å presisere visse aspekter ved forskningsspørsmålet. Dette for at QUADAS-2 skal kunne benyttes hensiktsmessig, spesielt i forhold til anvendbarheten av resultatene fra de vurderte studiene. Sammenhengen verktøyene skal benyttes i er i en akuttpsykiatrisk sengepost, med en sammensatt pasientgruppe i forhold til diagnose, rusbruk, tilstand, sykdomshistorie og alder. Referansestandarder er fysisk aggresjon, altså voldshandlinger rettet mot andre mennesker, enten helsepersonell eller medpasienter. Verbal aggresjon inkluderes ikke i denne presiseringen. Det ble ikke vurdert som nødvendig å gjøre endringer eller tilpasninger med QUADAS-2, utover enkelte som berørte referansestandarder. Om dette utdypes det nedenfor i teksten. Kvalitetsvurderingen ble utført av to personer uavhengig av hverandre, de samme som er beskrevet i utvelgelsesprosessen etter litteratursøket. Den eneste uenigheten som oppstod i prosessen var om faren for systematiske feil ved vurdering av referansestandarder skulle settes som høy eller uklar. Siden akkurat denne vurderingen berører den største utfordringen ved å benytte QUADAS-2 som sjekkliste for kvalitetsvurdering av studier om et prediktivt verktøy, så ble

uenigheten løst ved å erkjenne denne utfordringen. Siden QUADAS-2 er primært utviklet for å kunne kvalitetsvurdere studier om diagnostiske verktøys nøyaktighet (Whiting et al., 2011), blir referansestandarder betraktet som en objektiv standard, med 100 % sensitivitet og spesifisitet, som det studerte verktøyet vurderes i forhold til. Dette er problematisk i forhold til studier om prediksjonsverktøy, hvor fremtidig vold benyttes som referansestandard, av flere årsaker. Fordi en vesentlig del av en helsearbeiders hverdag i akuttpsykiatrien handler både om å forholde seg til og håndtere aggresjon blant inneliggende pasienter. For å unngå at potensielt voldelige pasienter faktisk blir voldelige, så finnes det en rekke tilnærminger, fra personlige samtaler til forskjellige preventive tiltak. Så det er ikke urimelig å anta at andelen falske positive delvis er et resultat av godt klinisk arbeid og preventive tiltak (Abderhalden et al., 2006; Rechenmacher et al., 2014; Yao et al., 2014). Etisk sett vil det være svært betenkelig å ikke iverksette tiltak ovenfor en potensielt farlig pasient, bare for å unngå en økning i falske positive prediksjoner. Tilsvarende ville det også være uetisk om personalet er blindt for resultatet av en indekstest, da slik kunnskap krever preventiv handling (Abderhalden et al., 2004; Rechenmacher et al., 2014). Konsekvensen av denne realiteten blir i praksis at det tredje kjerneområdet i QUADAS-2 ikke kan vektlegges i kvalitetsvurderingen av studiene.

4.3.4 Analyse

Utgangspunktet for å gjennomføre en metaanalyse med troverdige resultater, er at det kun analyseres studier av god metodologisk kvalitet, som konkret omhandler problemstillingen (Macaskill et al., 2010, s. 5). Formålet med analysen er å fremskaffe informasjon om hvor god en diagnostisk test er til å skille mellom pasienter med en sykdomstilstand og de uten. Dette gjøres ved å vurdere nøyaktigheten (accuracy), eller treffsikkerheten, til en test i forhold til to faktorer. Sensitiviteten; et uttrykk for andelen som korrekt positivt identifiseres av testen som syke, og spesifisiteten; andelen som korrekt ikke identifiseres av testen som syke (Macaskill, et al., 2010, s. 4). Innen studier om voldsprediksjon er en tilsvarende rapportering standard. Mossman (1994) etablerte at metoden som benyttes av studier om diagnostiske tester også er metoden for å måle treffsikkerheten til voldsprediksjoner.

En metaanalyse kan enten vurdere hvor nøyaktig et enkelt verktøy er, ved å sammenligne flere studier om det samme verktøyet, og rapportere gjennomsnittsverdiene for sensitivitet og spesifisitet på tvers av studiene. Eller metaanalysen kan også brukes til å sammenligne prediktive egenskaper ved flere ulike tester, for så å identifisere den testen som predikerer

med størst treffsikkerhet (Macaskill et al., 2010, s. 6). Formålet med denne oppgavens metaanalyse er å identifisere det verktøyet som predikerer med størst treffsikkerhet. Uansett, uavhengig av innfallsvinkel, så er det viktig å teste for heterogenitet, eller graden av ulikhet mellom studiene som skal inngå i analysen. Dette gjøres ved å vurdere hvordan treffsikkerheten varierer mellom studier av den samme testen, sett i forhold til forskjeller i faktorer som studiepopulasjon, studiedesign, studiekontekst og utførelse av testen. Er det heterogenitet, altså at forskjellen mellom studiene er for stor, anbefales ikke en metaanalyse for materialet (Polit & Beck, 2012, s. 657). Men selv om en vurdering av faktorene nevnt ovenfor både kan identifisere og forklare noe av heterogeniteten, så anbefaler Macaskill, et al., (2010, s. 29) at det gjøres en grundig statistisk analyse av heterogenitet. En analyse som benytter all data som er tilgjengelig på tvers av studiene. Det finnes flere ulike statistiske metoder og innfallsvinkler, og det anbefales sterkt å benytte statistisk ekspertise til å gjennomføre analysene (Macaskill et al., 2010, s. 4). I tillegg kan mulig heterogenitet observeres i et forest plot, i forhold til forskjellene i sensitivitet og spesifisitet mellom studiene. Dog er hovedpoenget med et forest plot i denne sammenhengen først og fremst å gi innsyn i to ganger to tabellene som sensitivitet og spesifisitet beregnes ut i fra (Macaskill et al., 2010, s. 16).

Prinsippet bak å analysere nøyaktigheten til en test tar utgangspunkt i at det kun skilles mellom om sykdomstilstanden er tilstede eller ikke (Macaskill, et al., 2010, s. 8). Å definere grenseterskelen for sykdom blir dermed av stor betydning for testens treffsikkerhet (Macaskill, et al., 2010, s. 11). Med utgangspunkt i sett med binære variabler og en definert grenseterskel, kan resultater settes opp i en to ganger to tabell, som vist i figur 2.

Fig. 2:

		<i>Faktisk tilstand (i virkeligheten):</i>	
		<i>Har tilstand (A+C)</i>	<i>Har ikke tilstand (B+D)</i>
<i>Identifisert av test som positiv (A+B):</i>	Sanne positive (A)	Falske positive (B)	
<i>Identifisert av test som negativ (C+D):</i>	Falske negative (C)	Sanne negative (D)	

(Macaskill, et al., 2010, s. 9)

Med utgangspunkt i to ganger to tabellen kan de ulike statistiske estimatene som vil ligge til grunn for en analyse presenteres. Sensitiviteten er et estimat som sier noe om sannsynligheten for at testen vil identifisere en person som faktisk har tilstanden som positiv. Den utregnes som følgende: $\text{Sanne positive} / (\text{Sanne positive} + \text{Falske negative}) (A/A+C)$. Spesifisiteten er et

tilsvarende estimat som sier noe om sannsynligheten for at testen vil identifisere en person som faktisk ikke har tilstanden som negativ. Den utregnes som følgende: $\frac{\text{Sanne negative}}{\text{Sanne negative} + \text{falske positive} + \text{sanne negative}} = \frac{D}{B+D}$. Youden's Index er sensitiviteten og spesifisiteten uttrykt som en enkelt verdi. Den gir et generelt uttrykk for en test sin treffsikkerhet ved at begge feilkildene til testen behandles likeverdig, altså de falske positive og de falske negative. Youden's Index utregnes som følgende: sensitiviteten + spesifisiteten - 1. Resultatet blir et tall mellom null og en. Et resultat på en tilsier en perfekt test, altså at det er verken falske negative eller falske positive prediksjoner. Et resultat på null indikerer en ubrukelig test, der prediksjonen er lik uansett om tilstanden er til stede eller ikke (Macaskill, et al., 2010, s. 10). Noe annet er prediksjonsverdier, som sier noe om sannsynligheten for at svar fra testen har sammenheng med hvordan tilstanden faktisk er. Det er positive og negative prediksjonsverdier, og de oppgis som enten en prosentverdi eller et desimaltall. Dermed blir disse verdiene meningsfulle for den generelle helsearbeider uten statistisk kompetanse, fordi prediksjonsverdien uttrykker prosentvis hvor stor sannsynlighet det er for at resultatet fra en prediksjon faktisk er korrekt. En positiv prediksjonsverdi (PPV) er sannsynligheten for at et positivt svar på testen tilsvarer at tilstanden er til stede, og verdien estimeres som $\frac{A}{A+B}$. Den negative prediksjonsverdien (NPV) viser i hvor stor grad et negativt testsvar har sammenheng med andelen av de som faktisk ikke har tilstanden, og verdien estimeres som $\frac{D}{C+D}$. (Macaskill, et al., 2010, s. 10; Polit & Beck, 2012, s. 342). Videre kan sannsynlighetsratio (LR) beskrives som svaret på et spørsmål om hvor mye mer sannsynlig det er for at konklusjonen på testen blir positiv for de som har tilstanden, sammenlignet med for dem som ikke har tilstanden. LR uttrykker forholdet mellom sensitiviteten og spesifisiteten som et enkelt tall (Polit & Beck, 2012, s. 343). Det skilles mellom positiv (LR+) og negativ (LR-) sannsynlighetsratio. LP+ er forholdet mellom sanne positive og falske positive utfall, og estimeres som $\frac{A/(A+C)}{B/(B+D)}$. Resultatet bør være større enn 1, fordi LP+ uttrykker hvor mange ganger mer sannsynlig det er at et positivt testsvar tilsvarer et sant positivt tilfelle, enn et falskt positivt tilfelle. LP- bør derimot gi et svar mindre enn 1, fordi estimatet uttrykker hvor mange færre ganger det er sannsynlig at et negativt testsvar tilsvarer et falskt negativt tilfelle i forhold til et sant negativt tilfelle. I praksis betyr dette at verdiene for begge sannsynlighetsratioene bør være avvikende fra 1, og samtidig at størrelsen på avviket fra 1 sier noe om testens brukbarhet. LP- estimeres som $\frac{C/(A+C)}{D/(B+D)}$ (Macaskill, et al., 2010, s. 11). Til slutt er det estimatet diagnostisk odds ratio (DOR) som uttrykker hvor mye større oddsene er for at et positivt resultat oppnås for en som har tilstanden, i forhold til for en som ikke har tilstanden. Estimatet kan uttrykkes som $\frac{LR+}{LR-}$. DOR er

først og fremst et estimat for statistisk bruk, og relevant når det er varierende grenseterskler i materialet (Macaskill, et al., 2010, s. 11; s. 15).

Å definere en bestemt grenseterskel for en test er ikke bare av stor betydning for testens treffsikkerhet, men også for å kunne evaluere kvaliteten til testens prediktive egenskaper. Ulike grenseterskler påvirker balansen mellom sensitivitet og spesifisitet, og dermed andelen falske positive og falske negative predikert av testen (Macaskill, et al., 2010, s. 11). Det har seg slik at prediksjoner alltid vil medføre en viss risiko for at det predikeres feil. Fastsetting av en grenseterskel innebærer å ta et valg. Et valg om hvilken type feil som er å foretrekke at potensielt gjøres (Mossman, 1994). Med andre ord, en bestemt grenseterskel velges utfra en vurdering av forholdet mellom sensitiviteten og spesifisiteten på de ulike grenseverdiene. Om studiene inkludert i en metaanalyse praktiserer ulike grenseterskler, så vil dette medføre statistiske utfordringer, siden estimatene på sensitivitet og spesifisitet ikke er direkte sammenlignbare for studiene av den samme testen (Macaskill, et al., 2010, s. 19). Det er vanlig å rapportere resultater fra studier som validerer en test i en Receiver Operating kurve (ROC). Dette er en grafisk modell som viser forholdet mellom sensitivitet og spesifisitet for alle grenseverdiene. ROC viser forholdet mellom sensitiviteten og én minus spesifisiteten ($1 - (D/B+D)$) (Macaskill, et al., 2010, s. 13). Nøyaktigheten til en test, på tvers av alle grenseterskler, presenteres som areal under kurven (AUC) (Polit & Beck, 2012, s. 345). En AUC verdi på 1 tilsvarer en 100 % treffsikker test, mens verdien 0,5 innebærer at testen ikke predikerer bedre enn tilfeldighet. Med andre ord er AUC et estimat for sannsynligheten for at en tilfeldig utvalgt voldelig pasient vil ha skåret høyere på prediksjonstesten, enn en tilfeldig utvalgt fredelig pasient (Mossman, 1994). I metaanalyser er det vanligst å fremstille resultatene i en oppsummerings ROC (SROC), der hver studie er plottet som et område, basert på et forhold mellom sensitivitet og spesifisitet. Med andre ord på bakgrunn av en bestemt grenseterskel (Macaskill, et al., 2010, s. 15). Ved varierende grenseterskler mellom studiene må SROC estimeres som en kurve tilsvarende ROC.

Som nevnt i avsnittet om datauttrekk, bidro en statistiker tilknyttet min arbeidsplass med sin kompetanse til dette prosjektet. Det har vært visse utfordringer i forhold til datamaterialet, spesielt vanskelig var det å gjennomføre en statistisk heterogenitetstest, slik beskrevet ovenfor. Årsaken er at fire studier er få, og når disse videre blir delt opp i to subgrupper, gitt ved to forskjellige instrumenter, så vil statistiske analyser i stor grad være påvirket av dette lave antallet studier. Deskriptiv statistikk som gjennomsnitt og standardavvik vil kunne være påvirket av avvikende studier. Statistisk styrke til å finne eventuelle statistisk signifikante

funn vil være svært begrenset, og estimater vil generelt være ustabile på grunn av dette. Et så lite utvalg vil videre vanskeliggjøre estimering av standard resultater fra metaanalyser, som mål på heterogenitet. Konsekvenser er at resultatene i denne oppgaven kun vil være deskriptiv, og beskrivende i forhold til de enkelte studier.

Som en konsekvens av problemet med estimering av mål på heterogenitet måtte det gjøres en vurdering av om det skulle gjennomføres en metaanalyse i denne oppgaven. Det ble besluttet å gjennomføre en metaanalyse, og vurderingen bygger på følgende. Utgangspunktet er at samtlige inkluderte studier ble vurdert til å være av god metodologisk kvalitet i kvalitetsvurderingen. I tillegg ble det praktisert en høy terskel for inklusjon til metaanalysen. Samtidig er det få tegn på heterogenitet i materialet, samtlige inkluderte studier er svært like i design, og resultatene fra dem kan ikke beskrives som sterkt avvikende (Polit & Beck, 2012, s. 657). Dessuten er det liten sannsynlighet for klinisk heterogenitet i materialet. Samtlige studier er utført på akuttpsykiatriske avdelinger, og pasientene ble fortløpende inkludert ved innleggelse fra avdelingens opptaksområde. Tilslutt så var det ingen tegn på metodologisk heterogenitet i materialet, fordi prediksjonsverktøyene er benyttet med lik grenseterskel, samt med rimelig lik tidshorisont for prediksjonen i samtlige studier (Polit & Beck, 2012, s. 663). At en studie (Abderhalden et al., 2004) opererer med 12 timers tidshorisont for prediksjonen, i motsetning til de andres 8 timer, ble vurdert til å være innenfor begrepet rimelig likt.

5. Funn

5.1 Litteratursøket

5.1.1 Presentasjon av ulike verktøy, samt endelig utvelgelse av inkluderte studier

Det ble i første omgang identifisert totalt 24 artikler som beskriver 14 forskjellige potensielt egnede verktøy for voldsrisikoprediksjon av inneliggende ordinære psykiatriske pasienter i et korttidsperspektiv. I første omgang ble verktøyene kun vurdert etter kriteriene om at studien måtte handle om en form for risikovurdering av inneliggende voksne akuttpsykiatriske pasienter, og at tidsperspektivet for vurderingen måtte være rimelig kort. En eksakt fastsettelse av tidsperspektivet for prediksjonene fastsettes ikke på dette stadiet, men et utgangspunkt for forståelsen av begrepet korttidsperspektiv har vært NICE retningslinjens (2005, s. 9) tre døgns tidshorisont. Så konkret har kriteriet om vurdering av inneliggende pasienter har vært det mest fundamentale i denne omgang. Andre eventuelle studier om verktøy som utelukkende er utviklet for barn- og ungdomspsykiatri, sikkerhetspsykiatri eller

for bruk utenfor psykiatrien var allerede ekskludert tidligere i utvelgelsesprosessen (se vedlagte figur 1), og presenteres derfor ikke her. Av de totalt 24 artiklene om 14 ulike verktøy var det kun tre verktøy og syv artikler som passerte en grundigere vurdering i forhold til de resterende inklusjonskriteriene, som var om studiedesignet er relevant, om verktøyet valideres, om verktøyet er egnet til bruk av sykepleiere, og om verktøyet er hensiktsmessig til sin intenderte bruk. Først presenteres de tre inkluderte verktøyene, og samtlige studier av verktøyet vurderes individuelt opp mot samtlige inklusjonskrav. Deretter blir de elleve ikke inkluderbare verktøyene narrativt beskrevet, og artiklene vurdert opp mot samtlige inklusjonskriterier. Hensikten med denne presentasjonsformen er å gi leseren både god innsikt i og forståelse av den komplekse utvelgelsesprosessen.

5.1.2 Presentasjon av inkluderte verktøy og studier

The Brøset Violence Checklist (BVC) er en sjekklister utviklet ved den sikkerhetspsykiatriske avdelingen på Brøset ved St Olavs hospital i Trondheim, opprinnelig ment kun til bruk i sikkerhetspsykiatrien. Grunnet gode prediksjonsegenskaper ble det også antatt at sjekklisten var egnet innen generell psykiatri, og antakelsen er bekreftet gjennom flere studier (Woods et al., 2008) Prediksjonsverktøyet BVC har seks variabler, forvirret, irritabel, brautende atferd, verbale trusler, fysiske trusler og angrep på gjenstander (Almvik, Woods & Rasmussen, 2000). Basert på observasjon og samhandling med pasienten skåres variablene som enten fraværende (0) eller tilstede (1), deretter oppsummeres resultatet. Det er sumskåren av sjekklisten som klassifiserer graden av voldsfare. Sumskåren null innebærer liten voldsfare, mens en og to indikerer moderat voldsfare. Sumskårer over tre klassifiseres som høy voldsfare, og samtidig anbefales det iverksettelse av preventive tiltak. I praktisk bruk er det erfart at verktøyet kun tar få minutter å administrere (Yao et al., 2014; Clarke, Brown & Griffith, 2010). Dermed fremstår BVC å være i tråd med det jeg etterspør i dette prosjektet, et prediksjonsverktøy egnet for bruk av sykepleiere. BVC har i tillegg et kortidsperspektiv på prediksjonene, i de ulike studiene varierer tidshorisont for prediksjonen på mellom 8 og 24 timer.

Søket identifiserte syv artikler om BVC (Almvik, Woods & Rasmussen, 2000; Abderhalden et al., 2004; Björkdahl, Olsson & Palmstierna, 2006; Abderhalden et al., 2006; Woods et al., 2008; Yao et al., 2014; Clarke, Brown & Griffith, 2010). Fire artikler beskriver fire ulike prospektive kohortstudier gjennomført på psykiatriske akuttavdelinger hvor BVC valideres. Studiene til Almvik, Woods & Rasmussen (2000), Abderhalden, et al. (2004; 2006) og Yao,

et al., (2014) inkluderes dermed i denne systematiske oversikten. De resterende tre artiklene vurderes til ikke å være inkluderbare. Björkdahl, Olsson & Palmstierna (2006) beskriver en retrospektiv journalstudie omkring BVCs prediktive egenskaper. Retrospektive studier inkluderes ikke i denne systematiske oversikten. Woods et al. (2008) og Clarke, Brown & Griffith (2010) er begge studier omkring sjekklistens praktiske egenskaper i klinisk bruk, og har fokus på helsepersonells erfaringer med sjekklisten, ikke dens prediktive egenskaper.

The Extended Brøset Violence Checklist (BVC-CN) er som navnet antyder sjekklisten BVC utvidet med en VAS-skala. BVC-CN gjennomføres med at personalet, etter å ha skåret BVC på en pasient, gir uttrykk for sitt subjektive syn på sannsynligheten for at pasienten fysisk utagerer i perioden prediksjonen er gyldig for (Rechenmacher et al., 2014). VAS-skalaen går fra ingen risiko (0) til veldig høy risiko (6), resultatet summeres med sumskåren fra BVC, og totalsummen kan dermed maksimalt bli tolv. En totalsum på under fire vurderes som svært liten sannsynlighet for fysisk aggresjon, mellom fire og seks vurderes som liten sannsynlighet, mellom syv og ni som moderat risiko, og en totalsum over ni vurderes som at det er høy sannsynlighet for fysisk utagering. I likhet med BVC fremstår BVC-CN å være et prediksjonsverktøy egnet for bruk av sykepleiere, og tidshorisont for prediksjonen er på 8 timer.

Det ble identifisert fire artikler om BVC-CN gjennom litteratursøket (Abderhalden et al., 2006; Needham et al., 2004; Abderhalden et al., 2008; Rechenmacher et al., 2014). To av artiklene beskriver valideringsstudier av BVC-CN med en prospektiv kohort design, og anses dermed som inkluderbare i denne systematiske oversikten. Artiklene er Abderhalden, et al. (2006) og Rechenmacher, et al. (2014). Abderhalden, et al. (2006) gjør en studie som både sammenligner og validerer begge utgavene av Brøset Violence Checklist, dermed er artikkelen inkludert for begge verktøyene. Det rapporteres også selvstendige utfallsmål for VAS i artikkelen, men denne inkluderes ikke. Riktignok må VAS i kombinasjon med BVC kunne betraktes som et uttrykk for en informert subjektiv vurdering. Men som selvstendig verktøy vurderes VAS som lite egnet til denne oppgavens formål, fordi da vil skåringen bare være et uttrykk for subjektivt klinisk skjønn. Hvilket det er god dokumentasjon på at er svært lite egnet til å predikere voldelig adferd (Singh, Grann & Fazel, 2011). To artikler vurderes som ikke inkluderbare i den systematiske oversikten (Needham et al., 2004; Abderhalden et al., 2008). Begge artiklene beskriver studier som undersøker om innføringen av et allerede validert voldsrisikoprediksjonsverktøy i en klinisk kontekst, reduserer forekomsten av vold og tvangsbruk på psykiatriske sengeposter.

The Dynamic Appraisal of Situational Aggression (DASA-IV) er en syv punkters sjekkliste som det tar maksimalt fem minutter å administrere. Skåringen av DASA-IV gjøres som en oppsummering av sykepleierens observasjoner omkring en pasient gjennom et arbeidsskift (Ogloff & Daffern, 2006). DASA-IV er sammensatt av fire punkter hentet fra andre prediksjonsverktøy, samt tre nye punkter identifisert av Ogloff & Daffern (2006) i tidligere studier. To punkter er fra BVC, irritabilitet og verbale trusler. To punkter som har vist seg å være assosiert med inneliggende pasienters voldsfare, er hentet fra langtidsprediksjonsverktøyet HCR-20 (Historical Clinical Risk Factors). Punktene er negative holdninger og impulsivitet. De tre unike punktene for DASA-IV er om pasienten har lav terskel for å bli provosert, om pasienten blir sint når ønsker ikke oppfylles, og om pasienten er uvillig til å følge instruksjoner. Hvert punkt skåres tilsvarende BVC, altså som enten fraværende (0) eller tilstede (1), deretter oppsummeres resultatet. En sumskår på null antyder at voldsfaren er liten, en skår mellom en og tre antyder en moderat voldsfare og at preventive tiltak bør vurderes, og en sumskår på fire eller høyere indikerer stor voldsfare, samt behov for preventive tiltak. Hensikten med DASA-IV er foruten å predikere voldsfare, å være et hjelpemiddel til klinisk håndtering av voldsrisikoen slik at voldsepisoder ikke oppstår (Ogloff & Daffern, 2006). Oppsummert så fremstår DASA-IV som å være et verktøy egnet til bruk av sykepleiere, og et verktøy hensiktsmessig til bruk i akuttpsykiatrien. Tidshorizonten for prediksjonen er i utgangspunktet et døgn.

Litteratursøket identifiserte to artikler om studier som validerer DASA-IV for bruk i psykisk helsevern utenfor sikkerhetspsykiatrien. Dumais, et al. (2012) rapporterer om en prospektiv kohortstudie, og er dermed inkluderbar. I studien er det to tidshorisonter for prediksjonen, 8 og 24 timer. Tilsvarende to tidshorisonter er det i studien til Griffith, Daffern & Godber (2013). Denne studien har et eksperimentelt design der deltakerne ble randomisert til to grupper for å sammenligne bruken av ustrukturerte kliniske vurderinger av voldsfare opp mot bruk av DASA-IV som prediksjonsverktøy. Siden studien er prospektiv, og resultatene fra intervensjonsgruppen rapporteres som en valideringsstudie av DASA-IV, inkluderes også denne artikkelen i den systematiske oversikten.

5.1.3 Presentasjon av verktøyene som nesten ble inkludert, med begrunnelse for eksklusjonen av studiene.

Violence Screening Checklist (VSC) er opprinnelig utviklet for å være en enkel 5 punkters sjekkliste som skal selekttere potensielt farlige psykiatriske pasienter til en grundige

voldsrisikokartlegging (McNiel & Binder, 1994; McNiel et al., 2003). I den opprinnelige valideringsstudien (McNiel & Binder, 1994) viste VSC seg også å ha gode prediksjonsegenskaper for å forutse voldsutøvelse i den nære fremtiden for inneliggende psykiatriske pasienter. Funnene ble bekreftet i den andre valideringsstudien (McNiel et al., 2003). Men grunnet studienes design, retrospektiv journalstudie og retrospektiv kasus-kontrollstudie, vurderes derfor ikke studiene å være inkluderbare i denne systematiske oversikten. Samtidig er det flere andre svakheter både ved selve sjekklisten og studiene. I andre sammenhenger stilles det spørsmål omkring overføringsverdien til funnene omkring VSC til andre mer generelle psykiatriske avdelinger, mye grunnet at studiene kun ble utført på små populasjoner av kun de alvorligste psykiske lidelsene, som mani og schizofreni (Hartvig et al., 2011; Daffern, 2007). McNiel et al. (2003) lanserer i sin studie en ny forkortet versjon av VSC-R, som hovedsakelig spør etter voldshistoriske og diagnostiske data, men samtidig også om pasienten er mann. Hvilken verdi har egentlig dette punktet? Spørsmålet stilles fordi i en prospektiv kohortstudie utført i Bergen (Mellesdal, 2003) om temaet aggresjon på psykiatriske akuttavdelinger, finner man en tendens til at kvinner innlagt i psykiatrien utøver vold mot personalet hyppigere enn sine mannlige medpasienter. Tilsvarende funn er også vist i annen internasjonal forskning, mens andre studier ikke finner kjønnsforskjell i voldeligheten (Grassi et al., 2001, Monahan et al., 2001, Barlow, Greyer & Ilkiw-Lavalle, 2000, Lam, McNiel & Binder, 2000, i: Mellesdal, 2003).

Short-Term Assessment of Risk and Treatability (START) er et vurderingsverktøy utviklet for at behandlere i psykiatrien skal kunne vurdere om pasienter er til fare for seg selv eller andre i et korttidsperspektiv (Gray et al., 2011) START har form som en tjue punkters liste, men det presiseres samtidig i artikkelen at START på ingen måte er en sjekkliste hvor meningen er at sumskåren skal gi en antakelse om pasientens farlighet. START er utviklet for å være et hjelpemiddel for behandlere, som skal sikre systematisk innhentet informasjon til å foreta en strukturert klinisk vurdering av pasientenes tilstand. Så til tross for at START viser oppløftende prediksjonsevner i en prospektiv kohort valideringsstudie, dog på en blandet populasjon av sikkerhets- og ordinære psykiatriske pasienter, så er ikke dette et hensiktsmessig verktøy egnet for praktisk klinisk bruk av sykepleiere. Dessuten er START heller ikke validert for bruk av annet enn behandlere, psykologer og psykiatere, og tilfredsstillende dermed ikke inklusjonskriteriene.

The Alert Assessment Form (M55) er utviklet på et psykiatrisk sykehus i British Columbia (Kling et al., 2006). Verktøyet inngår som en del av deres omfattende varslingsystem (the

Alert system), og den opprinnelige funksjonen er å identifisere pasienter med behov for en mer omfattende voldsrisikokartlegging. Men i denne studien er fokuset kun på M55s prediktive egenskaper til å forutse voldsutøvelse fra inneliggende pasienter. M55 er en enkel sjekkliste, sterkt inspirert av BVC (Almvik, Woods & Rasmussen, 2000; Kling et al., 2006). Til tross for oppløftende funn om god prediksjonsevne, så vurderes likevel ikke M55 som inkludert for denne studien. Årsaken er at valideringsstudien som beskrives i artikkelen er en retrospektivt kasus-kontroll studie.

Risk Screening Instrument (RSI) (Holdsworth, Collis & Allott, 1999) ble utviklet i England ved flere allmennpsykiatriske sengeposter på siste halvdel av 1990-tallet. Verktøyet er en todelt sjekkliste som administreres av sykepleiere ved innleggelsen av nye akutt pasienter. Den første delen av sjekklisten har fokus på her-og-nå opplevelsen av pasienten, mens andre del har et mer historisk fokus på pasientens farlighet. RSI har et noe bredere fokus enn utelukkende voldsprediksjon, da formålet er å identifisere om pasienten enten er til fare for seg selv eller andre. Men formålet til Holdsworth, Collis & Allott (1999) var ikke å validere et allerede utviklet verktøy. Formålet med studien var å videreutvikle, samt undersøke verktøyet tekniske kvaliteter i klinisk bruk. Artikkelen konkluderer i så måte med at til tross for potensielt positive funn i studien av RSI, så må verktøyet videreutvikles og testes for det er relevant for klinisk bruk. Studien tilfredsstiller derfor ikke inklusjonskriteriene.

The Preliminary Scheme 33 (PS33) (Bjørkly & Moger, 2007) er et verktøy utviklet og utprøvd i Norge på akutt psykiatriske inneliggende pasienter. PS33 har både et kort- og langtidsperspektiv. Verktøyet låner elementer fra både et verktøy for strukturert klinisk risikovurdering i langtidsperspektiv HCR-20, og et korttids prediksjonsverktøy, The Brøset Violence Checklist. Verktøyet er utviklet for bruk av behandlere i psykiatrien, og dermed faller det trolig utenfor denne oppgavens formål. Uansett så er V-RISK-10 å betrakte som en videreutvikling av PS33 (Hartvig et al., 2011), så en videre omtale og vurdering av verktøyet ansees derfor ikke som påkrevd.

Violence Risk Screening-10 (V-RISK-10) (Hartvig et al., 2011) er en videreutvikling av PS33, verktøyet er satt sammen av de 10 punktene som viste best prediksjonsverdi i studien av det mer omfattende verktøyet PS33 (Bjørkly & Moger, 2007). Tilsvarende PS33 har V-RISK-10 både et kort og langtidsperspektiv. Til tross for at dette er en studie med en prospektiv kohortdesign, så vurderes dette verktøyet som lite egnet for denne oppgavens formål. Årsaken er at verktøyet er utviklet og ment for bruk av psykiatere, eller behandlere på

tilsvarende nivå, samt at skåringen av V-RISK-10 skal gjøres på bakgrunn av et omfattende klinisk intervju med pasienten (Hartvig et al., 2011). Verktøyet er heller ikke ment brukt som utelukkende en sjekkliste, men mer som en klinisk retningslinje for en strukturert klinisk vurdering av pasientens potensielle voldsfare.

The Phipps Aggression Screening Tool (PAST) er et verktøy utviklet i konsensus mellom kliniske og vitenskapelige miljøer ved Johns Hopkins Hospital i Baltimore (Jayaram, Samuels & Konrad, 2012). Formålet med verktøyet er å identifisere potensielt aggressive personer ved innleggelse i psykisk helsevern, fokuset er på pasienter som det er behov for å skjerme (seclusion). Målet er å bedre miljøet i posten, samt redusere belastningen for og skader på personalet. PAST er en 11 punkter sjekkliste til bruk av sykepleiere og annet helsepersonell som har tett kontakt med pasienten. Men studien til Jayaram, Samuels & Konrad (2012) har hovedsakelig fokus på inter-rater reliabilitet for verktøyet, samt identifisering av pasientprofiler i forhold til voldsfare. PAST valideres ikke for klinisk bruk i artikkelen. Verktøyet kan derfor ikke inkluderes i denne systematiske oversikten.

Classification of Violence Risk (COVR) (McCusker, 2007) er en validert sjekkliste for vurdering av pasienters potensielle voldsfare ved utskrivelse fra psykiatrisk sykehus. En videre presentasjon av verktøyet ansees ikke som påkrevd, da COVR definitivt faller utenfor denne oppgavens formål.

Verktøyet the Clinical assessment of risk decision support (CARDS) er et to stegs verktøy utviklet i England for å gjøre voldsrisikovurderinger i en klinisk kontekst (Watts et al, 2004). Det første steget er en fem punkts sjekkliste, det andre steget er et grundigere verktøy for videre utredning av individer som det første steget har identifisert som potensielle risikoer. Dette verktøyet valideres ikke, og vil derfor ikke bli videre inkludert i denne systematiske oversikten. Artikkelen beskriver kun en kompleks konsensus utviklingsprosess av verktøyet, samt en klinisk brukerevaluering i form av en kvalitativ studie. Studien tilfredsstillter derfor ikke inklusjonskriteriene.

The Brief Risk Profile (BRP) har form som et strukturert intervju for vurdering av psykiatriske pasienters potensielle farlighet (Presecki et al., 2010). Form og innhold i intervjuet er noe diffust beskrevet, og samtidig har artikkelen mangelfull informasjon om ulike utfallsmål, samt verktøyets prediksjonsevne. Hovedårsakene til at studien ekskluderes, er likevel at BRP fremstår som utviklet til bruk av psykiatere, samt at tidshorizonten er svært

diffus. Det er ingen bekreftelse i artikkelen på at dette verktøyet har et spesifikt kortidsperspektiv, og studien tilfredsstillende derfor ikke inklusjonskriteriene.

Bruk av en Visual Analogue Scale (VAS) som prediksjonsverktøy argumenteres det sterkt for i en Nederlandsk studie (Nijman et al., 2002). VAS brukes her utelukkende som en rapportering av helsepersonellens subjektive vurdering av om en pasient kan være farlig. Metoden er utviklet på tross av at det eksisterer god dokumentasjon på at bruk av ustrukturert klinisk skjønn er svært lite egnet til å forutse voldelig adferd (Singh, Grann & Fazel, 2011). Nijman et al. (2002) bestrider dog ikke den etablerte kunnskapen på et generelt grunnlag, men argumenter godt for at erfarne spesialiserte fagarbeidere er klinisk kompetente til å gjøre slike vurderinger. Spesielt i en liten lukket sengepost tilsvarende den studien er utført i, en spesialisert enhet for å håndtere særskilt farlige og problematiske pasienter. Det kan stilles spørsmål med hvor stor overføringsverdi funnene i denne studien har til andre mer generelle psykiatriske sengeposter. Det er også svært tvilsomt om denne metoden er forenelig med hva denne oppgaven etterspør, et validert systematisk verktøy. Uansett, så ekskluderes denne studien grunnet tidshorizonten for gyldigheten av vurderingen, som er i løpet av oppholdet. Studiepopulasjonen har en gjennomsnittlig liggetid på 28 døgn. Samlet sett vurderes denne studien til å ikke tilfredsstillende denne oppgavens formål, som er å finne fram til et konkret verktøy som kan benyttes til og fortløpende vurdere akutte risikofaktorer for pasienters potensielle farlighet i akuttpsykiatrien generelt.

5.2 Resultatet av kvalitetsvurderingen av studiene, samt endelig utvelgelse av artikler som kan inkluderes i metaanalysen.

Som en konsekvens av kvalitetsvurderingen av artiklene, som ble beskrevet i metod delen, blir tre av de syv inkluderte studiene ekskludert fra metaanalysen. En fjerde studie blir også vurdert til at den skal ekskluderes, men da er også andre momenter i artikkelen medvirkende til valget. Det er dette prosjektets mål å gjennomføre en metaanalyse med troverdige resultater, og i følge Macaskill, et al. (2010, s. 5) anbefales det da kun å analysere studier av god metodologisk kvalitet, som også konkret omhandler problemstillingen. Whiting, et al. (2011) gir mer ambivalente anbefalinger omkring dette i sin artikkel om QUADAS-2, men er ikke avvisende til å bare inkludere studier med liten sannsynlighet for systematiske feil i analysene. Studiene som ekskluderes er Yao, et al. (2013), Almvik, Woods & Rasmussen (2000), Griffith, Daffern, & Godber. (2013), samt Dumais, et al. (2012). Eksklusjonen vil bli grundig begrunnet i den påfølgende narrative oppsummeringen av kvalitetsvurderingen. Der

finner jeg det hensiktsmessig å presentere resultatene for hver artikkel samlet, for hvert av QUADAS-2 sine kjerneområder, i stedet for artikkelvis. Kjerneområdene er inklusjonen av pasienter, indekstesten, referansestandard og det siste området som er forløp og timing, vurdert i forhold til fare for systematiske feil og anvendbarheten av funnene. Den oppsummerte oversikten over alle de vurderte studiene (figur 3) vil være utgangspunktet for oppsummeringen.

Fig. 3:

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Abderhalden 2004	+	+	?	+	+	+	+
Abderhalden 2006	+	+	?	+	+	+	+
Almvik 2000	+	+	?	+	+	+	-
Dumais 2012	+	+	?	+	-	+	+
Griffith 2013	+	+	?	+	-	+	-
Rechenmacher 2014	+	+	?	+	+	+	+
Yao 2013	+	+	?	+	+	+	-

- High ? Unclear + Low

(Review Manager 5.3)

Når det gjelder inklusjonen av pasienter i de forskjellige studiene er det ingen bekymring for at inklusjonen av pasienter kan representere en fare for utvalgsskjevhet i studiene. Samtlige studier oppgir å ha konsekvent inkludert samtlige innlagte pasienter i perioden studiene pågikk. Det er heller ikke rapportert om at pasienter, tilstander eller diagnosegrupper er blitt ekskludert fra analysene. Når det gjelder spørsmålet om hvor relevant studielokaliseringen og pasientgruppen er for denne systematiske oversiktens forskningsspørsmål, er det grunn til å stille spørsmål omkring to studier (Dumais et al., 2012; Griffith, Daffern, & Godber 2013). De resterende studiene er utført på et variabelt antall akuttpsykiatriske poster som betjener tilhørende opptaksområder, noe som vurderes å være helt i tråd med forskningsspørsmålet. Griffith, Daffern, & Godber (2013) er utført på to små poster beskrevet som ikke sikkerhetspsykiatriske, hvor den ene fremstår å være en form for spesialisert langtids psykosepost. Den andre posten omtales som akutt i teksten, men fremstår som å ha en selektert diagnostisk pasientgruppe. Men dette er uklart beskrevet i artikkelen. Antall

pasienter inkludert og studiens tidsrammer er heller ikke oppgitt. Oppsummert så er det god grunn til å stille spørsmål omkring denne studiens relevans for mitt forskningsspørsmål. Dumais, et al. (2012) har utført sin studie i en 12 sengs psykiatrisk intensivavdeling. Hva som karakteriserer en intensivavdeling i Canada er vanskelig å vite, men en pasientstrøm på 77 innleggelser i løpet av 7 måneder gjør det problematisk å betrakte dette som en akuttavdeling tilsvarende det forskningsspørsmålet etterspør. For å sette det hele i perspektiv, ved det psykiatriske akuttmottaket på Haukeland Universitetssykehus er det nærmere to tusen innleggelser i løpet av året fordelt på 19 sengeplasser (Mellesdal et al., 2010). Studien til Dumais, et al. (2012) vurderes derfor til å ha liten overføringsverdi.

Omkring studienes bruk av indekstesten, i dette tilfelle BVC eller BVC-CN, så eksisterer det ingen bekymringer verken i forhold til gjennomføring eller overføringsverdi av resultatene. Samtlige studier er prospektive, ergo er indekstesten gjennomført i forkant av hendelsen den eventuelt predikerer. Det eksisterer heller ingen bekymring for at bruken av indekstesten avviker fra det denne oppgavens forskningsspørsmål etterspør.

Referansestandarder er definert som uklare for samtlige studier, og årsaken er utfyllende beskrevet i metoddelen. For resultatet av kvalitetsvurderingen kan ikke dette betraktes som problematisk. I forhold til anvendbarheten er tre studier vurdert å ikke være i samsvar med hva forskningsspørsmålet etterspør, og dette er av avgjørende betydning for kvalitetsvurderingen. Yao, et al. (2013) og Griffith, Daffern, & Godber (2013) benytter både verbal og fysisk aggresjon som referansestandard. Resultatene har derfor liten overføringsverdi, siden denne oppgavens forskningsspørsmål utelukkende etterspør studier med fysisk aggresjon som referansestandard. Almvik, Woods & Rasmussen (2000) benytter en referansestandard som både er fysisk aggresjon og kasting av gjenstander, der det kun er verbale trusler som ikke klassifiseres som voldsepisoder. Dette er trolig årsaken til at en annen studie av samme verktøy (Abderhalden et al., 2004) problematiserer forekomsten av hyppige voldsepisoder i denne studien. Samlet sett er det derfor problematisk å vurdere referansestandarder benyttet av Almvik, Woods & Rasmussen (2000) som relevant for denne systematiske oversiktens formål.

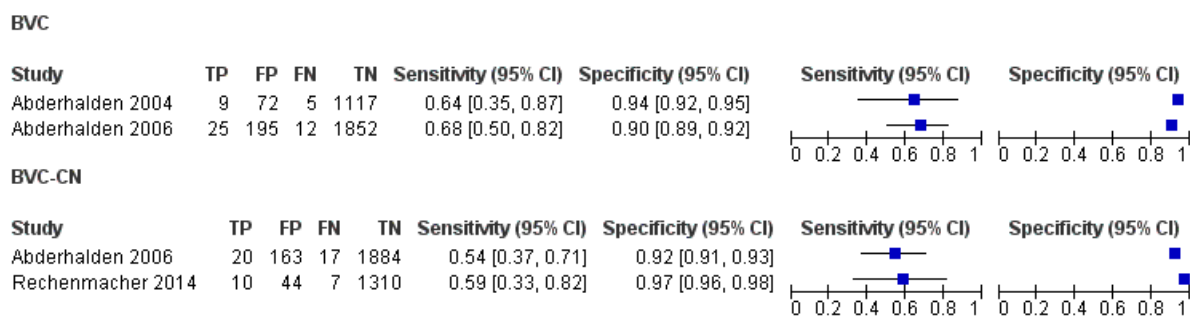
Om det siste kjerneområdet i QUADAS-2, forløp og timing, vurderes samtlige studier å være av god kvalitet. Samtlige inkluderte pasienter og verktøyprediksjoner i samtlige studier ble både redegjort for og inkludert i analysene.

Resultatet av kvalitetsvurderingen er altså at tre studier (Yao et al., 2013; Almvik, Woods & Rasmussen, 2000; Griffith, Daffern, & Godber, 2013) ekskluderes fra metaanalysen grunnet bruk av en referansestandard som ikke er egnet til å besvare denne oversiktens formål. I tillegg er populasjonen i Griffith, Daffern, & Godber (2013) sin studie vurdert som ikke å være tilsvarende en akuttpsykiatrisk pasientgruppe. Tilsvarende gjelder også studiepopulasjonen i Dumais, et al. (2012) sin studie, men den artikkelen oppgir samtidig svært få relevante utfallsmål. For uten informasjon om resultater som sensitivitet og spesifisitet, er det ikke mulig å gjøre analyser omkring verktøyets prediktive egenskaper i denne studien (Macaskill, et al., 2010, s. 6). Derfor, grunnet disse to årsakene, ekskluderes artikkelen til Dumais, et al. (2012) fra metaanalysen.

5.3 Metaanalyse og statistiske resultater

I dette avsnittet presenteres resultatene av metaanalysen, og samtidig vil resultatene bli benyttet til å si noe om mulig heterogenitet i materialet. Konklusjonen av denne vurderingen vil være av avgjørende betydning for om resultatene fra metaanalyse kan betraktes som troverdige, og dermed ha overføringsverdi til å besvare oppgavens formål. I presentasjonen av resultatene har jeg hovedsakelig valgt ikke å gi en nærmere tolkning og beskrivelse av estimater som først og fremst er til statistisk bruk. Variabiliteten og resultatene i materialet illustreres godt uten å bruke dem, samt at estimatene inneholder lite informasjon som ikke er uttrykt gjennom estimatene sensitivitet, spesifisitet og prediksjonsverdiene.

Fig. 4: Forest plot



(Review Manager 5.3)

Figur 4 viser et kombinert Forest Plot, som foruten å gi innsyn i to ganger to tabellene som er grunnlaget for beregning av sensitivitet og spesifisitet, også kan si noe om graden av heterogenitet mellom studiene. At spesifisitet verdiene ligger samlet nært verdien 1 tyder på liten forskjell mellom studiene. Samtidig viser dette at testene generelt er gode til å identifisere personene som ikke utøver vold, altså de sanne negative. Det er større grad av variasjoner mellom studiene i forhold til testenes sensitivitet, og i tillegg fremstår estimatet for sensitivitet som mer usikkert grunnet bredt konfidensintervall. Dette innebærer at testene generelt sett er mindre gode til å identifisere individene som faktisk vil utøve vold, altså at andelen falske positive er betydelig. Variasjonen mellom studiene for sensitiviteten kan være et tegn på at det er en viss grad av heterogenitet i materialet, noe som vil bli nærmere belyst i beskrivelsen av de deskriptive resultatene nedenfor. Den vedlagte SROC kurven (se figur 5, vedlegg VII) er en alternativ visualisering av resultatene som illustreres i Forest Plotet.

Tabell 6 viser gjennomsnitt og variasjon samlet over alle studiene, noe som kan sees på som et mål på heterogenitet. Her bekreftes funnet fra Forrest Plotet, at det er større grad av variasjoner mellom studiene i forhold til sensitiviteten (MIN = .54 og MAX = .68, SD = .06) enn for spesifisitet. Dette uttrykkes også med at vi ser en tilsvarende sterkere grad av variasjon for PPV (MIN = .11 og MAX = .19, SD = .04) enn for NPV. Videre er det noe variasjon rundt en gjennomsnittlig høy AUC verdi (MIN = .83 og MAX = .93, SD = .04).

Tabell 6:

Descriptive Statistics						
	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Sens	4	.54	.68	.6118	.05966	.004
Spes	4	.90	.97	.9330	.02702	.001
Youden	4	.46	.58	.5448	.05725	.003
PPV	4	.11	.19	.1298	.03696	.001
NPV	4	.99	1.00	.9937	.00195	.000
LRplus	4	6.79	18.10	10.6497	5.26288	27.698
LRneg	4	.36	.50	.4159	.06221	.004
AUC	4	.83	.93	.8753	.04247	.002
Valid N (listwise)	4					

(IBM SPSS Statistics 22.0.0.2)

Grunnet at det samlet sett er en viss antydning til heterogenitet i materialet, er det nødvendig å gjøre en grundigere vurdering av heterogeniteten. Dette gjøres ved å se nærmere på metaanalysens to undergrupper. Men først bør det kort beskrives hvordan verktøyene gjennomsnittlig presterer. At spesifisiteten ligger nærmere 1 (snitt= .93) bekrefter som tidligere vist at testene generelt er gode til å identifisere de sanne negative. Hvilket også understrekes av gjennomsnittet for den negative prediksjonsverdien (NPV, snitt= .99), som viser at det er 99 % sannsynlighet for at et negativ testsvar samsvarer med at personen faktisk ikke blir voldelig. Gjennomsnittet for sensitiviteten (snitt= .61) bekrefter at testene er langt mindre treffsikre når det gjelder positive prediksjoner. Dette kommer også tydelig fram av den positive prediksjonsverdien (PPR, snitt= .13), som viser at bare 13 % av de positive prediksjonene samsvarer med at tilstanden faktisk er til stede, og dermed blir 87 % av prediksjonene falske positive. Gjennomsnittet for AUC (snitt= .87) må betraktes som høyt, altså at verktøyene predikerer rimelig godt på tvers av alle grenseverdiene. Det har vært vanskelig å finne noen konkret skala for tolkning av kvaliteten på AUC i litteraturen, men det antas at en AUC på over 0,75 er å betrakte som et godt resultat for diagnostiske verktøy generelt sett. Over 0,95 er å betrakte som et suverent resultat (Abderhalden et al., 2006).

Tabell 7 differensierer deskriptive resultater for de to testene. I begge gruppene ser vi den samme hovedtendensen som beskrevet tidligere, med lav sensitivitet og høy spesifisitet. Studiene om BVC (1) viser liten spredning i samtlige estimater basert på 2 ganger 2 tabellen, spesielt for PPV (snitt= .11, SD= .00) og NPV (snitt= .99, SD= .00). Resultatene viser videre høy AUC for begge studier (snitt= .87). Når det i tillegg er liten variasjon i variablene alder, spredning i alder og kjønn, så kan studiene antas å være homogene, i det minste i forhold til disse beskrevne karakteristika. For BVC-CN (2) er variansen i estimatene større, spesielt i forhold til PPV (snitt = .15, SD = 0.05) og AUC (snitt = .88, SD = 0.07). Det er også vesentlig mer heterogenitet i bakgrunnsopplysningene alder (snitt= 42, SD= 7,5), aldersspredning (snitt= 15, SD= 2,9), og kjønn, sammenlignet med for BVC. Det er derfor rimelig å anta at det eksisterer heterogenitet i studiene av BVC-CN, og dermed blir det også legitimt å stille spørsmål ved troverdigheten til resultatene i metaanalysen.

Tabell 7:

Descriptive Statistics						
Select		N	Minimum	Maximum	Mean	Std. Deviation
1	Sensitivitet	2	.64	.68	.6593	.02321
	Spesifisitet	2	.90	.94	.9221	.02454
	Youden	2	.58	.58	.5814	.00133
	PPV	2	.11	.11	.1124	.00179
	NPV	2	.99	1.00	.9946	.00140
	LRplus	2	7.09	10.62	8.8545	2.49129
	LRneg	2	.36	.38	.3693	.01534
	AUC	2	.86	.88	.8700	.01414
	Age	2	36.40	39.90	38.1500	2.47487
	Age_SD	2	13.00	13.30	13.1500	.21213
	Male	2	.61	.61	.6100	.00000
	Valid N (listwise)	2				
	2	Sensitivitet	2	.54	.59	.5644
Spesifisitet		2	.92	.97	.9439	.03333
Youden		2	.46	.56	.5083	.06705
PPV		2	.11	.19	.1472	.05367
NPV		2	.99	.99	.9929	.00257
LRplus		2	6.79	18.10	12.4449	7.99974
LRneg		2	.43	.50	.4624	.05205
AUC		2	.83	.93	.8805	.07142
Age		2	36.40	47.04	41.7200	7.52362
Age_SD		2	13.30	17.39	15.3450	2.89207
Male		2	.50	.61	.5530	.08061
Valid N (listwise)		2				

1 = BVC, 2 = BVC-CN

(IBM SPSS Statistics 22.0.0.2)

Til slutt vil det være relevant å vurdere gjennomsnittresultatene i tabell 7 i forhold til formålet med denne oppgavens metaanalyse, som er å identifisere det verktøyet som predikerer med størst treffsikkerhet. Bekymringene om heterogenitet i materialet sees helt bort fra i denne avsluttende vurderingen, men vil bli tema i den påfølgende diskusjonen. Sammenligningen viser at BVC og BVC-CN predikerer rimelig likt på tvers av alle grensetersklene (AUC= .87 vs. .88). Begge testene er også like gode til å identifisere de sanne negative tilfellene (NPV= .99), og spesifisiteten er tilsvarende høy. Forskjeller av betydning mellom testene, er den positive prediksjonsverdien, hvor BVC-CN predikerer med lavest andel falske positive (PPV= .15), i forhold til BVC (PPV= .11). Men samtidig er

sensitiviteten vesentlig høyere for BVC (snitt= .66) enn for BVC-CN (snitt= .56). Dermed oppstår det et paradoks, grunnet at resultatene er motstridene om hvor gode testene er til å skille ut individene som faktisk vil utøve vold. Forklaringen er trolig å finne i utregningen av estimatene, siden sensitiviteten estimeres som et forhold mellom sanne positive og falske negative. Mens PPV estimeres som et forhold mellom sanne positive og falske positive. Youden's Index kan benyttes til å oppklare paradokset, siden estimatet forholder seg likeverdig til begge feilkildene, altså de falske positive og de falske negative. Resultatene viser avvikende verdier for BVC (snitt= .58) og BVC-CN (snitt= .51). Siden Youden's Index estimerer et tall mellom null og en, og størst tallverdi tilsier færrest feilkilder, fremstår BVC å være verktøyet som predikerer med størst treffsikkerhet. Men om dette argumentet kan være avgjørende for en konklusjon er usikkert, og må absolutt vurderes i lys av antydningene om heterogenitet i materialet. Som en avrunding er det verdt å vurdere om de to sannsynlighets ratioene, som også er estimater som estimeres i forhold til alle verdiene i to ganger to tabellen, inneholder relevant oppklarende informasjon. Men her er resultatene tilsvarende avvikende. LR+ viser at BVC-CN er best til å predikere positivt testsvar som tilsvarer et sant positivt tilfelle, mens LR- viser at BVC er best på at sannsynligheten for at et negativt testsvar tilsvarer et sant negativt tilfelle.

6. Diskusjon

I presentasjonen av funnene i metaanalysen kom det fram en svak mistanke om heterogenitet i materialet. Men denne heterogeniteten må også betraktes i forhold til at metaanalysen er gjort på svært få studier, og at det i tillegg er to rimelig like instrumenter som sammenlignes. Dermed blir det et spørsmål om forskjellene i resultatene enten kan forklares med tilfeldigheter i utvalgene av personer som var med i studiene. Eller om forskjellene har sin årsak i potensielle systematiske feil i studiene, og at denne feilkilden dermed preger resultatene. I resultatdelen ble det funnet varians i bakgrunnsopplysningene, som er variablene som beskriver sammensetningen av deltakergruppene. Dermed er det av avgjørende betydning for validiteten til denne systematiske oversikten at forskjellene i studiepopulasjonene skyldes rene tilfeldigheter, og ikke at deltakergruppene i studiene har ulik sammensetning, og dermed ikke er sammenlignbare. Dette er også avgjørende for generaliserbarheten av funnene i metaanalysen, altså om funnene har overføringsverdi til å besvare denne oppgavens formål. Det var ulikhetene i variablene alder, spredning i alder og kjønn som dannet grunnlag for mistanken om heterogenitet i materialet. Men det er vanskelig å finne bekreftelse for denne mistanken, når variablene vurderes i forhold til kunnskapsgrunnet innen voldsprediksjon.

Singh, Grann & Fazel (2011) finner at til tross for at yngre mennesker oftere er voldelige enn eldre, så er pasientens alder ikke avgjørende for kvaliteten på voldsprediksjoner.

Kjønns sammensetningen kan også ha betydning for frekvensen av vold i en psykiatrisk sengepost, men det er ikke grunnlag for å anta at kjønn er en faktor av betydning når det gjelder kvaliteten til prediksjon av voldsfare. Det er derfor rimelig å anta at forskjellene i resultatene mellom studiene skyldes tilfeldigheter, og at variablene kjønn og alder ikke er egnet til å forklare variansen i estimatene basert på to ganger to tabellen.

I tillegg finnes det informasjon om den diagnostiske sammensetningen i studiepopulasjonene. Men dette gjelder bare for to av studiene, og derfor inngikk ikke denne informasjonen i heterogenitetsanalysen. Opplysninger om den diagnostiske sammensetningen i studiepopulasjonene fremkommer av tabell 5 (se vedlegg VI). I de to studiene der informasjon er oppgitt (Abderhalden et al., 2006; Rechenmacher et al., 2014), fremstår forholdet mellom psykose- og affektive diagnoser å være avvikende, mens for rusdiagnoser er forskjellen minimal. En sammenligning med populasjonen i Mellesdal, et al., (2000) sin studie på en norsk akuttpsykiatrisk populasjon, tidligere omtalt under inklusjonskriteriene, viser tilsvarende avvikende tall for forholdet mellom psykose- og affektive diagnoser. I denne studien er i tillegg andelen rusdiagnoser vesentlig avvikende fra de to studiene i metaanalysen. Til tross for avvik i den diagnostiske sammensetningen av studiepopulasjonene, er det ingen informasjon i materialet som tyder på at det gjennom inklusjonen til de ulike studiene ble gjort systematiske feil. Samtlige studier er utført på akuttpsykiatriske avdelinger, og pasientene ble fortløpende inkludert ved innleggelse fra avdelingens opptaksområde. Det er dermed grunnlag for å anta at forskjellene skyldes tilfeldigheter, og samtidig anta at manglende informasjon om diagnoser i Abderhalden, et al. (2004) ikke har betydning for kvaliteten på metaanalysen. Støttende for en slik konklusjon er også funn fra en randomisert kontrollert multisenterstudie om bruk av voldsrisikoprediksjonsverktøy. Abderhalden, et al., (2008) fant en tilsvarende diagnostisk varians mellom de ulike akuttenehetene i studien. Også her ble pasientene fortløpende inkludert ved innleggelse fra avdelingens opptaksområde. Likevel er det relevant å se nærmere på om det er en sammenheng mellom diagnoser og forekomsten av voldsepisoder innen psykisk helsevern. Finnes det en eventuell sammenheng, så kan det ikke utelukkes at denne har påvirket funnene i enkeltstudiene. Cornaggia, et al., (2011) finner i en systematisk oversikt om aggresjon i psykiatrien, at vold er særskilt assosiert med forekomsten av akutte psykoser og rusmisbruk. Forskjellene i rusdiagnoser er minimale mellom studiene. Men det

kliniske begrepet akutte psykoser er dessverre lite egnet til å forklare eventuelle forskjeller mellom studiene, grunnet at akutte psykoser kan forekomme både innen psykose- og affektive diagnoser. Og uten mer spesifikk informasjon omkring sammensetningen av undergrupper innenfor de overordnede diagnosenivåene, spesielt i forhold til de affektive diagnosene, blir kunnskapen meningsløs i denne sammenhengen. Det er dermed ikke grunnlag for å anta annet enn at den diagnostiske sammensetningen i pasientpopulasjonene ikke kan forklare forskjeller i resultatene mellom studiene. Man kan altså anta at funnene i metaanalysen har overføringsverdi til å besvare denne oversiktens forskningsspørsmål.

Samtlige inkluderte studier i metaanalysen viser den samme tendensen til at testene generelt er gode til å identifisere personene som ikke utøver vold, men at testene er mindre gode til å identifisere individene som faktisk vil utøve vold. Til tross for forskjeller mellom de ulike studiene og verktøyene er andelen falske positive prediksjoner generelt rimelig stor, knapt mer enn 1 av 10 prediksjoner er korrekte i forhold til å identifisere en fremtidig voldsutøver. Umiddelbart fremstår akkurat dette forholdet som en stor svakhet ved testene. At testene generelt er gode til å identifisere personene som ikke utøver vold, er nødvendigvis ikke formildende, for som det fremgår av figur 4 (Forest plot) er forekomsten av voldsepisoder svært få i forhold til det totale antallet prediksjoner. Samtidig skriver Singh, et al. (2011) at ved prediksjon av sjeldne forekomne tilfeller er det uunngåelig at andelen falske positive blir vesentlig. Andelen falske positive må også sees i lys av at det å predikere noe alltid vil medføre en viss risiko for at det predikeres feil, og at det derfor må tas en beslutning om hvilken type feil som er minst ønskelig å gjøre (Mossman, 1994). Konsekvensen av disse betingelsene blir naturligvis at testene er best egnet til å identifisere individene som ikke vil utøve vold. Dette fremstår som hensiktsmessig, siden en test med en høy andel falske negative neppe vil ha noen nytteverdi i praktisk klinisk arbeid. En slik test kunne resultere i at farlige situasjoner oppsto, grunnet redusert beredskap omkring de falske negative prediksjonene. Men samtidig er den høye andelen falske positive etisk betenkelig, i hvert fall om pasienter utsettes for unødvendig tvangsbruk som en konsekvens av prediksjonen. Men nå er det ikke nødvendigvis slik at samtlige potensielt farlige pasienter blir usatt for ulike tvangstiltak i preventivt hensyn. Mye handler om helsepersonellens bevissthet omkring risikoen, samt den personlige oppfølgingen og samhandlingen med pasientene. Likevel bør det være en viss bekymring omkring den høye andelen falske positive prediksjoner. For mange pasienter kan det oppleves som stigmatiserende å bli identifisert som potensiell voldsutøver. Men samtidig kan det ikke utelukkes at det er en sterk sammenheng mellom andelen falske positive, og

personalets kliniske arbeid. Dette bekreftes av Cornaggia, et al., (2011), som viser at kvaliteten og fagligheten i interaksjonen mellom pasient og sykepleier har stor betydning for aggresjonsnivået i en psykiatrisk sengepost. Tilsvarende er også omtalt tidligere, i forhold til hvordan bruk av fremtidig voldshandling som referansestandard kan være problematisk. For det er ikke urimelig å anta at klinisk arbeid, samt gjennomføring av relevante preventive tiltak øker andelen falske positive prediksjoner. Dette forholdet kan illustreres med et teoretisk eksempel, bare for å spissformulere poenget. Sett at man har det perfekte voldsrisikoprediksjonsverktøyet, og prøver det ut ved en perfekt avdeling med usedvanlige dyktige og erfarne sykepleiere. En avdeling hvor pasientene håndteres slik at fysisk vold aldri forekommer. Resultatet ville blitt et verktøy med spesifisitet i toppskiktet. Mens den positive prediksjonsverdien ville vist et verktøy som predikerte med 0 % sannsynligheten for at et positivt svar på testen tilsvarte at sann positiv prediksjon. Dermed er det vist at en stor andel falske positive prediksjoner nødvendigvis ikke er problematisk. Og samtidig at andelen falske positive er et resultat av tre forhold, at det predikeres sjeldne forekomne tilfeller, et bevisst valg om hvilken type feil som bør unngås, samt kvaliteten på det kliniske arbeidet. Estimaten for kvaliteten på prediksjonene må også sees i lys av at de er basert på en statistisk metode som forutsetter en objektiv referansestandard. Og fremtidig vold er ikke en objektiv referansestandard.

Jeg har så langt vist at selv om begge verktøyene er langt fra perfekte, så kan man til en viss grad stole på dem, og begge prediksjonsverktøyene, BVC og BVC-CN, fremstår som gode hjelpemidler i klinisk psykiatrisk arbeid. Dog ikke for å selektere pasientene direkte til preventive tiltak, men som et systematisk hjelpemiddel å støtte seg på i klinikken. Rechenmacher, et al. (2014) konkluderer også tilsvarende om verktøyet BVC-CN, at den største gevinsten med verktøyet er funksjonen som rådgivende hjelpemiddel i den kliniske beslutningsprosessen, samt som et nyttig redskap for sykepleierne i vurderinger omkring vold og aggresjon. Tilsvarende finner Abderhalden, et al. (2008) at bruk av et voldsrisikoprediksjonsverktøy i akuttpsykiatrien reduserte både forekomsten av preventive tiltak og voldsepisoder. Resultatet forklares med økt bevissthet omkring potensielle farer, og dermed mer fokusert bruk av preventive tiltak. I tillegg ble personalet mer fokusert på konfliktdependende samhandling med pasientene. Med et slikt utgangspunkt er det beklagelig at studiene om prediksjonsverktøyet DASA-IV ikke passerte kvalitetsvurderingen, og derfor ikke inngår i metaanalysen. For et verktøy som foruten å predikere voldsfare, også har i seg å

være et hjelpemiddel til klinisk håndtering av voldsrisiko, fremstår definitivt som et verktøy med potensiell stor klinisk relevans.

7. Konklusjon

Med utgangspunkt i resultatene fra metaanalysen kan det ikke trekkes noen konkret konklusjon om hvilket verktøy, BVC eller BVC-CN, som er best egnet for å predikere fremtidige voldsepisoder. Men gjennom diskusjonen, med støtte i resultatene, har jeg vist at begge verktøyene er godt egnede kliniske verktøy til fortløpende vurdering av akutte risikofaktorer for fysisk aggresjon ved innleggelse i akuttpsykiatrien. I tillegg fremstår verktøyene å ha en klinisk nyttefunksjon utover bare voldsprediksjonen, ved at personalet blir mer fokusert på konfliktdempende samhandling med pasientene. Samtidig er det vist at bruk av et systematisk prediksjonsverktøy fører til redusert bruk av preventive tiltak, samt redusert forekomst av voldsepisoder. I praksis innebærer dette at sykepleierne får et bedret arbeidsmiljø, og at pasientene blir utsatt for færre preventive tiltak. Det er dermed en rimelig antakelse at også tvangsbruken kan reduseres tilsvarende innen psykisk helsevern, som en konsekvens av at et voldsrisikoprediksjonsverktøy implementeres på en akuttpsykiatrisk sengepost.

8. Metodekritikk

Gjennom fremstillingen av denne innledningsdelen har jeg forsøkt å gi leseren begrunnet innsikt i hvordan arbeidsprosessen har vært, og om hvilke valg som er tatt underveis. Så svakheter og begrensningene i metoden for dette prosjektet er på mange måter beskrevet under de enkelte kapitlene i metodedelen. Likevel vil det her bli gjort en generell kritisk gjennomgang av metoden i denne systematiske oversikten. For metaanalysen sin del var den største utfordringen at det ble inkludert få studier, noe som resulterte i at en statistisk heterogenitetstest ikke var hensiktsmessig. Antallet inkluderte studier kom som et resultat av en streng kvalitetsvurdering, hvor det ble praktisert en høy terskel for inklusjon. Hensikten med dette var at metaanalysen skulle gi troverdige resultater. DTA håndboken gir også støtte for andre innfallsvinkler, men å redusere terskelen for inklusjon til metaanalysen ville samtidig påvirket den eksterne validiteten til metaanalysen. Grunnen til dette er at funnene da får liten overføringsverdi, ved at studiedeltakerne, studiesettingen og bruken av verktøyet ikke blir relevant i forhold til forskningsspørsmålets premisser. Kvalitetsvurderingen ble utført og er beskrevet i tråd med anbefalingene i DTA håndboken, samt at QUADAS-2 ble benyttet i vurderingen.

Inklusjonsprosessen som ble gjennomført i denne oppgaven representerer ubestridt en metodologisk svakhet. Første del av utvelgesprosessen ble gjennomført av kun et individ, og dermed er det en viss fare for at enkelte relevante referanser er blitt borte på veien. Men at den avsluttende utvelgesprosessen var omstendelig oppdelt i to faser, kan ikke betraktes som en metodologisk svakhet. Videre så kan også selve søkeprosessen betraktes som metodologisk svak ved at anbefalingene i DTA håndboken ikke ble fulgt fullstendig. Det ble riktignok søkt i flere databaser enn minimumskravet er, men at ikke siteringssøk og søk etter grå litteratur ble gjennomført er en stor metodologisk svakhet. Tilsvarende problematisk er det at litteratur på andre språk enn engelsk ikke ble lest. Samlet sett kan det derfor ikke utelukkes at materialet kan være preget av publikasjonsskjevheter. Tilslutt, så representerer også søkestrategien en metodologisk svakhet, ved at det ble innført flere betingelser enn anbefalt i søkestrategien. Tiltaket, som var for å begrense resultatet av søket til en håndterbar mengde treff, kan også ha fjernet relevante referanser fra resultatene.

9. Referanser

Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H. J., & Fischer, J. (2006). Predicting inpatient violence using an extended version of the Broset-Violence-Checklist: instrument development and clinical application. *BMC Psychiatry*, 6 [Internett] Tilgjengelig fra: <http://www.biomedcentral.com/1471-244X/6/17> [nedlastet 09. sept. 2013]

Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H. J., & Fischer, J. (2008) Structured risk assessment and violence in acute psychiatric wards: randomized controlled trial. *The British Journal of Psychiatry* 193, s. 44 - 50.

Abderhalden, C., Needham, I., Miserez, B., Almvik, R., Dassen, T., Haug, H. J. & Fischer, J. E. (2004): Predicting inpatient violence in acute psychiatric wards using the Brøset-Violence-Checklist: a multicentre prospective cohort study. *Journal of Psychiatric and Mental Health Nursing*. 11, s. 422 – 427.

Almvik, R., Woods, P. & Rasmussen, K. (2000) The Broset Violence Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence*, 15 (12), s. 1284 - 1296.

Björkdahl A., Olsson D. & Palmstierna T. (2006) Nurses' short-term prediction of violence in acute psychiatric intensive care. *Acta Psychiatrica Scandinavica*, 113 (3), s. 224–229.

Bjørkly, S. & Moger, T. A. (2007) A secondstep in development of a checklist for screening risk for violence in acute psychiatric patients: evaluation of interrater reliability of the Preliminary Scheme 33. *Psychological reports*, 101 (3), s. 1145 - 1161.

Bossuyt, P. M. & Leeflang, M. M. (September 2008) Chapter 6: Developing Criteria for Including Studies. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4*. [Internett] Birmingham, The Cochrane Collaboration. Tilgjengelig fra: <http://srdta.cochrane.org> [nedlastet 2. oktober 2013]

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Moher, D., Rennie, D., de Vet, C. W. & Lijmer, J. G. (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine*, 138 (1), W-1 – 12.

Clarke, D. E., Brown, A. M. & Griffith, P. (2010) The Broset Violence Checklist: clinical utility in a secure psychiatric intensive care setting. *Journal of Psychiatric & Mental Health Nursing*, 17, s. 614 - 620.

Cornaggia, C. M., Beghi, M., Pavone, F. & Barale, F. (2011) Aggression in psychiatry wards: A systematic review. *Psychiatry Research*, 189 (1), s. 10 – 20.

Daffern, M. (2007) The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior*, 12, 116 - 130.

Deeks, J. J., Wisniewski, S., & Davenport, C. (13. September 2013) Chapter 4: Guide to the contents of a Cochrane Diagnostic Test Accuracy Protocol. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. [Internett] Birmingham, The Cochrane Collaboration. Tilgjengelig fra: <http://srdta.cochrane.org> [nedlastet 2. oktober 2013]

de Vet, H. C. W., Eisinga, A., Riphagen, I. I., Aertgeerts, B. & Pewsner D. (September 2008) Chapter 7: Searching for Studies. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4*. [Internett] Birmingham, The Cochrane Collaboration. Tilgjengelig fra: <http://srdta.cochrane.org> [nedlastet 2. oktober 2013]

Dumais, A., Larue, C., Michaud, C. & Goulet, M. H. (2012) Predictive validity and psychiatric nursing staff's perception of the clinical usefulness of the French version of the Dynamic Appraisal of Situational Aggression. *Issues in Mental Health Nursing*, 33, s. 670 - 675.

Fazel, S., Singh, J. P., Doll, H. & Grann, M. (2012) Use of risk assessment instruments to predict violence and antisocial behavior in 73 samples involving 24 827 people: systematic review and meta-analysis. [Internett] *British Medical Journal*, 2012, 345. Tilgjengelig fra: <http://www.bmj.com/content/345/bmj.e4692> [nedlastet 2. mai 2013]

Gray, N. S., Benson, R., Craig, R., Davies, H., Fitzgerald, S., Huckle, P., Maggs, R., Taylor, J., Trueman, M., Williams, T. & Snowden, R. J. (2011) The Short-Term Assessment of Risk and Treatability (START): A prospective study of inpatient behavior. *The International Journal of Forensic Mental Health*, 10, s. 305 - 313.

Griffith, J. J., Daffern, M. & Godber, T. (2013) Examination of the predictive validity of the Dynamic Appraisal of Situational Aggression in two mental health units. *International Journal of Mental Health Nursing*, 22, s. 485 - 492.

Hartvig, P., Roaldset, J., Moger, T., Ostberg, B., & Bjørkly, S. (2011). The first step in the validation of a new screen for violence risk in acute psychiatry: The inpatient context. *European Psychiatry*, 26, s. 92–99.

Helsedirektoratet (2012) *Nasjonal faglig retningslinje for utredning, behandling og oppfølging av personer med samtidig ruslidelse og psykisk lidelse – ROP-lidelser*. IS-1948. Oslo, Helsedirektoratet.

Helsedirektoratet (2013) *Nasjonal faglig retningslinje for utredning, behandling og oppfølging av psykoselidelser*. IS-1957. Oslo, Helsedirektoratet.

Higgins, J. P. T. & Green S. Red. (Mars 2011) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. [Internett] The Cochrane Collaboration, 2011. Birmingham, the Cochrane Collaboration. Tilgjengelig fra: www.cochrane-handbook.org [nedlastet 19. november 2013]

Holdsworth, N., Collis, B. & Allott, R. (1999) The development and evaluation of a brief risk screening instrument for the psychiatric inpatient setting. *Journal of psychiatric and mental health nursing*, 6, s. 43 - 52.

Jayaram, G., Samuels, J. & Konrad, S. (2012) Prediction and prevention of aggression and seclusion by early screening and comprehensive seclusion documentation. *Innovations in Clinical Neuroscience*, 9, s. 30 - 38.

Kling, R., Corbière, M., Milord, R., Morrison, J. G., Craib, K., Yassi, A., Sidebottom, C., Kidd, C., Long, V. & Saunders, S. (2006) Use of a violence risk assessment tool in an acute care hospital: effectiveness in identifying violent patients. *AAOHN journal*, 54 (11), s. 481-487.

Knutzen, M., Bjørkly, S., Bjørnstad, M., Furre, A. & Sandvik, L. (2014) *Innsamling og analyse av data om bruk av tvangsmidler og vedtak om skjerming i det psykiske helsevernet for voksne i 2012*. Kompetansesenterets Rapport 2014-2. Oslo, Kompetansesenter for sikkerhets-, fengsels- og rettspsykiatri for Helseregion Sør-Øst.

Kumar, S. & Simpson, A. I. F. (2005) Application of risk assessment for violence methods to general adult psychiatry: a selective literature review. *Australian & New Zealand Journal of Psychiatry*, 39 (5), s. 328 – 335.

Macaskill, P., Gatsonis, C., Deeks, J.J., Harbord, R. M. & Takwoingi, Y. (23. December 2010) Chapter 10: Analyzing and Presenting Results. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. [Internett] Birmingham, The Cochrane Collaboration. Tilgjengelig fra: <http://srdta.cochrane.org> [nedlastet 2. oktober 2013]

McCusker, P. J. (2007) Issues regarding the clinical use of the Classification of Violence Risk (COVR) assessment instrument. *International Journal of Offender Therapy and Comparative Criminology*, 51, s. 676 - 685.

McNiel, D. E. & Binder, R. L. (1994) Screening for risk of inpatient violence: Validation of an actuarial tool. *Law and Human Behavior*, 18, s. 579 - 586.

McNiel, D. E., Gregory, A. L., Lam, J. N., Binder, R. L. & Sullivan, G. R. (2003) Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting and Clinical Psychology*, 71, s. 945 - 953.

Mellesdal, L. (2003) Aggression on a psychiatric acute ward: a three-year prospective study. *Psychological Reports*, 92, s. 1229 - 1248.

Mellesdal, L., Mehlum, L., Wentzel-Larsen, T., Kroken, R. & Jørgensen, H. A. (2010) Suicide Risk and Acute Psychiatric Readmissions: A Prospective Cohort Study. [Internett] *Psychiatric Services*, 61 (1). Tilgjengelig fra:

<http://journals.psychiatryonline.org/article.aspx?articleid=101087> [nedlastet 25. november 2013]

Mossman, D. (1994) Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62 (4), s. 783 - 792.

National Institute for Health and Clinical Excellence (NICE) (2005) Clinical Guideline 25. Violence: the short-term management of disturbed/violent behaviour in psychiatric inpatient settings and emergency departments. [Internett] Tilgjengelig fra:

<http://www.nice.org.uk/CG025NICEguideline> [nedlastet 6.april 2012]

Needham, I., Abderhalden, C., Meer, R., Dassen, T., Haug, H., Halfens, R. & Fischer, J. (2004) The effectiveness of two interventions in the management of patient violence in acute mental inpatient settings: Report on a pilot study. *Journal of Psychiatric and Mental Health Nursing*, 11, s. 595 - 601.

Nijman, H., Merckelbach, H., Evers, C., Palmstierne, T. & A Campo, J. (2002) Prediction of aggression on a locked psychiatric admissions ward. *Acta Psychiatrica Scandinavica*, 105, 390 - 395.

Ogloff, J. R. P., & Daffern, M. (2006). The Dynamic Appraisal of Situational Aggression: An Instrument to Assess Risk for Imminent Aggression in Psychiatric Inpatients. *Behavioral Sciences & the Law*, 24, s. 799 - 813.

Polit, D., & Beck, C. (2012) *Nursing research: generating and assessing evidence for nursing practice*. 9. utg. Philadelphia, Lippincott Williams & Wilkins.

Presecki, P., Jukic, M. K., Biocina, S. M., Agius, M. & Mimica, N. (2010) Assessment of risk behaviour in psychiatric patients by a Brief Risk Profile questionnaire - A pilot study. *Socijalna Psihijatrija*, 38, s 28 - 32.

QUADAS (2011) *QUADAS-2 tool* [Internett] University of Bristol. Tilgjengelig fra <http://www.bristol.ac.uk/quadas/> [Nedlastet 14. mai 2014]

Rechenmacher, J., Muller, G., Abderhalden, C. & Schulc, E. (2014) The diagnostic efficiency of the extended German Broset Violence Checklist to assess the risk of violence. *Journal of Nursing Measurement*, 22, s. 201 - 212.

Reitsma, J. B., Rutjes, A. W. S., Whiting, P., Vlassov, V. V., Leeflang, M. M. G. & Deeks, J. J. (27. October 2009) Chapter 9: Assessing methodological quality. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0*. [Internett] Birmingham, The Cochrane Collaboration. Tilgjengelig fra: <http://srdta.cochrane.org> [nedlastet 2. oktober 2013]

Singh, J. P. Grann, M. & Fazel, S. (2011) A comparative study of violence risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychological Review*, 31(3), s. 499 – 513.

Singh, J. P., Serper, M., Reinharth, J. & Fazel, (2011) Structured Assessment of Violence Risk in Schizophrenia and Other Psychiatric Disorders: A Systematic Review of the Validity, Reliability, and Item Content of 10 Available Instruments. *Schizophrenia Bulletin*, 37 (5), s. 899 - 912.

Singh, J. P., Yang, S., Mulvey, E. P. & The RAGEE Group (2015) Reporting guidance for violence risk assessment predictive validity studies: The RAGEE Statement. *Law and Human Behavior*, 39 (1), s. 15-22.

Sosial- og helsedirektoratet (2007) *Vurdering av risiko for voldelig atferd – bruk av strukturerte kliniske verktøy*. IS-9/2007. Oslo, Sosial- og helsedirektoratet.

Steinert, T. (2002) Prediction of inpatient violence. *Acta Psychiatrica Scandinavica*, 106, s. 133-141.

Watts, D. J. B., Slade, M., Holloway, F., Rosen, A. & Thornicroft, G. (2004) Clinical assessment of risk decision support (CARDS): The development and evaluation of a feasible violence risk assessment for routine psychiatric practice. *Journal of Mental Health*, 13 (6), s. 569 - 581.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M. & QUADAS-2 Group (2011) QUADAS-2: a

revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155 (8), s. 529 - 536.

Woods, P. & Ashley, C. (2007) Violence and aggression: a literature review. *Journal of Psychiatric and Mental Health Nursing*, 14, s. 652–660.

Woods, P., Ashley, C., Kayto, D. & Heusdens, C. (2008) Piloting violence and incident reporting measures on one acute mental health inpatient unit. *Issues in Mental Health Nursing*, 29, s. 455 - 469.

Yao, X., Li, Z., Arthur, D., Hu, L., An, F. R. & Cheng, G. (2014) Acceptability and psychometric properties of Broset Violence Checklist in psychiatric care settings in China. *Journal of Psychiatric & Mental Health Nursing*, 21, s. 848 - 855.

**Voldsprediksjon i psykiatriske akuttavdelinger,
en metaanalyse av validerte verktøy.**

**Prediction of violence in psychiatric acute wards;
A Meta Analysis of Validated Tools.**

Petter Jakobsen

Mastergrad i kunnskapsbasert praksis i helsefag

Avdeling for helse- og sosialfag

Høgskolen i Bergen

Artikkel planlegges publisert i Sykepleien Forskning

Antall ord: 2929

Antall figurer og tabeller: 4

Voldsprediksjon i psykiatriske akuttavdelinger, en metaanalyse av validerte verktøy.

Sammendrag:

Bakgrunn

Forekomsten av vold og aggresjon vesentlig hyppigere innen psykisk helsevern enn generelt i samfunnet, og inneliggende pasienters aggresjon blir håndtert med ulike preventive tiltak.

Samtidig er det et krav fra politikere og pasientorganisasjoner at psykisk helsevern reduserer bruken av tvang. Da er det interessant at i en studie er det vist at systematisk bruk av voldsrisikoprediksjonsverktøy reduserer både forekomsten av voldsepisoder og bruk av preventive tiltak i akuttpsykiatrien.

Hensikt

Hensikten med denne systematiske oversiktsstudien er å identifisere validerte prediksjonsverktøy, samt kartlegge hvor nøyaktige verktøyene er til å predikere vold i en akuttpsykiatrisk kontekst.

Metode

Et systematisk søk etter litteratur ble gjennomført i databasene MEDLINE, EMBASE, PsycINFO og the NCJRS Abstracts Database. Prospektive studier som validere relevante verktøy ble identifisert, og kvalitetsvurdert med sjekklisten QUADAS-2. Til slutt ble det gjennomført en metaanalyse for å sammenligne hvor nøyaktig verktøyene predikerte voldsfare.

Resultat

The Brøset Violence Checklist og The Extended Brøset Violence Checklist ble inkludert i metaanalysen. Resultatene er tilnærmet like for begge verktøyene. Prediksjonene er treffsikre ($AUC \geq 0,87$), men andelen falske positive prediksjoner er likevel vesentlig ($PPV < 15\%$).

Konklusjon

Begge verktøyene er godt egnede kliniske verktøy til fortløpende vurdering av akutte risikofaktorer for fysisk aggresjon ved innleggelser på en akuttpsykiatrisk sengepost.

Nøkkelord: Violence, Aggression, Psychiatry, Risk Assessment, Review

Prediction of violence in psychiatric acute wards; A Meta Analysis of Validated Tools.

Abstract:

Background

The occurrence of violence and aggression is significantly more frequent in mental health care than in society in general. Inpatient aggression is managed by various preventive measures, but there is a demand from politicians and patient organizations that mental health services reduce the use of coercive measures. In this context, it is interesting that one study confirms that the systematic use of tools that predict the risk of violence can reduce the frequency of violence as well as the use of preventive measures.

Aims

The purpose of this systematic study is to identify valid predictive tools and assess how accurate they predict violence in an acute psychiatric context.

Method

A systematic literature search was conducted in the databases MEDLINE, EMBASE, PsycINFO and the NCJRS Abstracts Database. Prospective studies that validate relevant tools were identified, and the checklist QUADAS-2 was used to assess their quality. Finally a meta-analysis was conducted to compare how accurately the tools predict the risk for violence.

Results

The Brøset Violence Checklist and The Extended Brøset Violence Checklist were included in the meta-analysis. The results are similar for both tools. They accurately predict the risk for violence ($AUC \geq 0.87$), but the percentage of false positive predictions is high ($PPV < 15\%$).

Conclusion

Both checklists are well suited clinical tools for the continuous assessment of acute risk for physical aggression in patients when admitted to an acute psychiatric ward.

Keywords: Violence; Aggression; Psychiatry; Risk Assessment; Review

Voldsprediksjon i psykiatriske akuttavdelinger, en metaanalyse av validerte verktøy.

INTRODUKSJON

I utgangspunktet er forekomsten av vold og aggresjon vesentlig hyppigere i psykiatrien enn generelt i samfunnet. Samtidig er helsetilstanden til sykepleiere som regelmessig eksponeres for aggresjon signifikant forringet i forhold til andre helsearbeidere (1). En vesentlig del av en helsearbeiders hverdag i akuttpsykiatrien handler om å håndtere aggresjon blant inneliggende pasienter med forskjellige tiltak, alt fra personlig oppfølging av pasienter til iverksettelse av preventive tiltak (2). Preventive tiltak er gjerne bruk av tvangsmidler som belter, remmer, fastholding, isolasjon og korttidsvirkende legemidler. Men behandlingstiltaket skjerming er også mye brukt. Skjerming handler om å holde pasienten adskilt fra avdelingen, og begrense kontakten med andre til kun et fåtalls pleiepersonell. Bruk av tvang er nøye regulert av lovgivningen, men samtidig er det et krav fra politikere, helsebyråkrater og pasientorganisasjoner at psykisk helsevern generelt reduserer bruken av tvang (3). I denne sammenhengen er det relevant at en studie (4) finner at systematisk bruk av et voldsrisikoprediksjonsverktøy i akuttpsykiatrien reduserte både bruken av preventive tiltak og forekomsten av voldsepisoder. Effektene kom som et resultat av økt bevissthet omkring potensielle farer, og dermed mer fokusert bruk av preventive tiltak, men samtidig ble personalet også mer fokusert på konfliktdempende samhandling med pasientene. I retningslinjer anbefales det at psykiatriske pasienter farlighetsvurderes, men anbefalingene er enten ikke basert på et oppdatert kunnskapsgrunnlag (5, 6), eller er kun juridisk fundamenterte anbefalinger (7, 8).

Vurderinger omkring personers potensielle farlighet har normalt vært sikkerhets- og fengselspsykiatriens domene (9, 10), og fokuset har vært på langtidsprediksjon av risiko når potensielt farlige personer returnerer til samfunnet. Men i løpet av de siste 15 årene er det likevel publisert en rekke studier som har fokus på prediksjon av voldsfare innen ordinær psykiatri (10, 11). Disse prediksjonsverktøyene kan deles inn i to grupper. En gruppe er verktøy som fortløpende brukes til å vurdere de akutte risikofaktorene ved en pasient. Den andre gruppen er verktøy som identifiserer potensielt farlige pasienter, for så å selekere dem til en grundigere utredning av farlighet. Det er i tillegg to forskjellige tilnærminger til hvordan strukturerte risikovurderinger er designet (9, 10, 12). Sjekklistene klassifiserer graden av voldsfare. Samt verktøy for å gjøre en strukturert klinisk vurdering, slik at systematisk innsamlet materiale kan brukes for å støtte behandlerens konklusjoner, basert på

informert klinisk skjønn. I denne studien er målet å finne fram til gode validerte systematiske verktøy som kan benyttes til fortløpende å vurdere akutte risikofaktorer for pasienters potensiale for fysisk aggresjon ved innleggelser i akuttpsykiatrien. For dette formålet fremstår sjekklister som det mest hensiktsmessige formatet. Basert på omtale i litteraturen (10, 11) er det særskilt to verktøy som fremstår som relevant for dette prosjektet, The Brøset Violence Checklist (BVC) og The Dynamic Appraisal of Situational Aggression (DASA-IV). Begge er enkle sjekklister som det tar få minutter å administrere og tar utgangspunkt i sykepleierens observasjon av en pasient gjennom et arbeidsskift. Graden av voldsfare defineres av sumskåren til sjekklisten. Hensikten med denne studien og problemstillingen blir dermed å identifisere hvilke ulike verktøy som finnes, samt forsøke å kartlegge hvor nøyaktige de er til å forutse voldsfare i en akuttpsykiatrisk kontekst. Dette gjøres som en systematisk oversikt, ved å inkludere artikler som beskriver prediksjonsverktøy. En artikkel kan beskrive ett eller flere verktøy, og ett verktøy kan være beskrevet i flere artikler.

METODE

Et systematisk søk etter litteratur, for å få oversikten over samtlige relevante studier egnet til å besvare denne studiens forskningsspørsmål, ble gjennomført 11. mars 2014. Det ble søkt i databasene MEDLINE, EMBASE, PsycINFO og The US National Criminal Justice Reference Service Abstracts. I forkant av søket ble det utviklet en søkestrategi i tråd med anbefalingene i Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy (DTA håndboken) (13). Søkestrategien var som følgende: (violen* or aggress* or attack* or assault*) AND (risk* or assess* or predict* or tool* or instrument* or measur* or scale* or test* or survey* or checklist* or questionnaire* or appraisal* or sensitivity or specificity) AND (mental* or bipolar* or schizo* or psychos* or personality* or manic* or psychiatric*) AND (psychiatric adj (inpatient* or hospital* or admission* or readmission*)) Den 28. oktober 2015 ble det i tillegg gjort et oppdateringssøk i de samme databasene, et enkelt søk hvor det kun ble søkt på navnene til verktøyene identifisert gjennom inklusjonsprosessen, både i fulltekst og forkortet form. Totalt 2462 referanser ble identifisert gjennom begge søkene, etter at dubletter var slettet.

De identifiserte referansene ble så vurdert i forhold til et sett av inklusjonskriterier. Prosessen ble gjennomført i tråd med anbefalingene i Cochrane Handbook for Systematic Reviews of Interventions (14). Inklusjonskriteriene var at studien måtte validere et verktøy for risikovurdering av inneliggende voksne akuttpsykiatriske pasienter, med et tidsperspektiv for

vurderingen på maksimum 24 timer. Samtidig måtte verktøyet være egnet for bruk av sykepleiere, og være hensiktsmessig til sin intenderte bruk, altså en form for sjekklister. I tillegg måtte studien ha en prospektiv design, grunnet at retrospektive studier er betraktet som vesentlig metodologisk svakere enn prospektive (15). Studier gjort utelukkende på en sikkerhetspsykiatrisk populasjon, samt studier om barn og ungdom ble ekskludert. I tillegg ble kun referanser på engelsk og skandinaviske språk vurdert for inklusjon. Totalt ble 7 artikler som omtalte 3 ulike voldsprediksjonsverktøy inkludert i prosessen. Inklusjonsprosessen er vist som et flytdiagram i figur 1.

Sjekklisten QUADAS-2 (17, 18) ble benyttet til å kvalitetsvurdere de inkluderte studiene, og studiene som sjekklister vurderte til å være av lav kvalitet ble ekskludert fra metaanalysen. Formålet med kvalitetsvurderingen er å vurdere om de inkluderte studiene er valide (16). QUADAS-2 tar utgangspunkt i en definisjon av kvalitet som et forhold mellom fraværet av potensielle systematiske feil i studien, og anvendbarheten av resultatene i forhold til det relevante forskningsspørsmålet (18). Men QUADAS-2 er primært utviklet for å vurdere studier om diagnostiske verktøys nøyaktighet, og derfor har det vært visse utfordringer ved å benytte den i kvalitetsvurderingen. I studier om diagnostiske verktøy blir det studerte verktøyet vurdert i forhold til en objektiv referansestandard (16). Den eventuelle fremtidige hendelsen som benyttes som referansestandard innen voldsprediksjon, er ikke objektiv. Siden en vesentlig del av en helsearbeiders hverdag i akuttpsykiatrien handler om å håndtere aggresjon blant inneliggende pasienter, er det ikke urimelig å anta at andelen falske positive prediksjoner delvis er et resultat av godt klinisk arbeid og preventive tiltak (19, 20). Etisk sett vil det være svært betenkelig å ikke iverksette tiltak ovenfor en potensielt farlig pasient, bare for å unngå en økning i falske positive prediksjoner. Tilsvarende ville det være uetisk om personalet var blindet for resultatet av en prediksjon, da slik kunnskap krever preventiv handling (20, 21, 22). Konsekvensen av dette blir at referansestandard dimensjonen i QUADAS-2 ikke vektlegges i kvalitetsvurderingen av studiene. Kvalitetsvurderingen ble utført av to personer uavhengig av hverandre, og uenigheter ble diskutert og løst. Fire studier ble ekskludert i prosessen. En oppsummering av kvalitetsvurderingen presenteres i resultatdelen.

Utgangspunktet for å gjennomføre en metaanalyse med troverdige resultater, er at det kun analyseres studier av god metodologisk kvalitet, som konkret omhandler problemstillingen (23). Formålet med analysen er å fremskaffe informasjon om hvor god diagnostisk tester er til å skille mellom pasienter med en sykdomstilstand og de uten. Dette gjøres ved å vurdere

nøyaktigheten, eller treffsikkerheten, til en test i forhold til to faktorer. Sensitiviteten som er et estimat som uttrykker sannsynligheten for at testen vil identifisere en person som faktisk har tilstanden som positiv, og spesifisiteten som uttrykker sannsynligheten for at testen vil identifisere en person som ikke har tilstanden som negativ. Metoden skiller kun mellom om sykdomstilstanden er til stede eller ikke, og en definert grenseterskel for sykdom er en forutsetning for å kunne beregne estimatene (23). I tillegg rapporterer studier resultater i en Receiver Operating kurve (ROC), en grafisk modell som viser forholdet mellom sensitivitet og spesifisitet over alle grensetersklene. Nøyaktigheten til en test estimeres fra ROC som areal under kurven (AUC). En AUC verdi på 1 tilsvarer en 100 % treffsikker test, mens verdien 0,5 innebærer at testen ikke predikerer bedre enn tilfeldighet (15). Estimaten som trolig gir mest mening for personer uten statistisk kompetanse, er de positive og negative prediksjonsverdiene. Disse estimatene uttrykker prosentvis hvor stor sannsynlighet det er for at resultatet fra en prediksjon faktisk er korrekt (15, 23). Det er også viktig å teste for heterogenitet, eller graden av ulikhet mellom studiene som skal inngå i analysen. Er det heterogenitet, altså at forskjellen mellom studiene er for stor, anbefales ikke en metaanalyse for materialet (15).

RESULTATER

Som en konsekvens av kvalitetsvurderingen ble fire av syv inkluderte studier ekskludert fra metaanalysen. En oppsummert oversikt over resultatene vises i tabell 1. Tre studier ekskluderes fra metaanalysen grunnet bruk av en referansestandard som ikke har overføringsverdi til å besvare denne studiens forskningsspørsmål. To av studiene (22, 24) ekskluderes fordi det benyttes både verbal og fysisk aggresjon som referansestandard, og i tillegg er studiepopulasjonen i en av studiene (24), vurdert som ikke å være utført på en akuttpsykiatrisk pasientgruppe. En studie (25) ekskluderes fordi referansestandard som benyttes er både fysisk aggresjon og kasting av gjenstander. Den fjerde studien (26) utgår fordi den ikke er utført på en akuttpsykiatrisk pasientgruppe.

Tre studier om to forskjellige voldsprediksjonsverktøy ble inkludert til metaanalysen etter kvalitetsvurderingen, og en av artiklene (19) beskriver begge verktøyene. Verktøyene er The Brøset Violence Checklist (BVC) og The Extended Brøset Violence Checklist (BVC-CN). BVC er en sjekkliste med seks variabler, forvirret, irritabel, brautende atferd, verbale trusler, fysiske trusler og angrep på gjenstander (21). Variablene skåres som enten fraværende [0] eller tilstede [1], og sumskåren av sjekklisten klassifiserer graden av voldsfare. Null

innebærer liten voldsfare, mens en og to indikerer moderat voldsfare. Sumskårer over tre klassifiseres som høy voldsfare, og samtidig anbefales det iverksettelse av preventive tiltak. BVC-CN er BVC utvidet med en VAS-skala (20). VAS-skalaen, et mål for personalets subjektive syn, går fra ingen risiko (0) til veldig høy risiko (6), og resultatet summeres med sumskåren fra BVC. En totalsum på under syv vurderes som liten sannsynlighet for fysisk aggresjon, mellom syv og ni vurderes som moderat risiko, og over ni vurderes som høy sannsynlighet for fysisk utagering. Ingen av artiklene (24, 26) om verktøyet DASA-IV, som omtales i innledningsdelen, passerte kvalitetsvurderingen. Dette er beklagelig, siden DASA-IV i tillegg til å predikere voldsfare, også er et hjelpemiddel til klinisk håndtering av voldsrisikoen, slik at voldsepisoder ikke oppstår (27).

Det var ikke hensiktsmessig å gjennomføre en statistisk heterogenitetstest på materialet. Årsaken er at det var for få studier til å estimere mål på heterogenitet statistisk. Det ble likevel besluttet å gjennomføre en metaanalyse, fordi samtlige studier ble vurdert til å være av god metodologisk kvalitet i kvalitetsvurderingen. Det er dermed liten sannsynlighet for klinisk heterogenitet i materialet, siden studiene er utført på akuttpsykiatriske avdelinger, og pasientene ble fortløpende inkludert ved innleggelse fra avdelingens opptaksområde. Det er også liten sannsynlighet for metodologisk heterogenitet, siden verktøyene er benyttet med lik grenseterskel, samt med rimelig lik tidshorisont for prediksjonen (15). At det er noe varians i variablene alder, spredning i alder og kjønn mellom studiene (se tabell 2), er ikke av betydning for kvaliteten på voldsprediksjoner (28). Samlet sett er det grunnlag for å anta at eventuelle forskjellene i resultatene mellom studiene kan forklares med tilfeldigheter.

Det kombinerte Forest Plotet i tabell 3 viser at testene generelt er gode til å identifisere personene som ikke vil utøve vold, fordi verdiene for spesifisitet ligger samlet nær 1. Selv om det er større grad av variasjon mellom studiene i forhold til testenenes sensitivitet, så viser tabell 3 at testene generelt sett er mindre gode til å identifisere individene som faktisk vil utøve vold, altså at andelen falske positive prediksjoner er betydelig. Dette må sees i lys av at det å predikere noe alltid vil medføre en viss risiko for at det predikeres feil, og at det derfor må tas en beslutning om hvilken type feil som er minst ønskelig å gjøre (29). Fastsetting av en bestemt grenseterskel for sykdom er å gjøre et slikt valg.

Sammenligningen av BVC og BVC-CN i tabell 2 viser at testene predikerer likt og med en høy grad av nøyaktighet på tvers av alle grensetersklene. En AUC verdi på over 0,80 er betraktet som et godt resultat (11). Begge testene er også svært gode til å identifisere de sanne

negative tilfellene (NPV). Den positive prediksjonsverdien (PPV) viser at BVC-CN predikerer med lavest andel falske positive, men at det likevel er 85 % sannsynlighet for at en positiv prediksjon er feil. Men samtidig viser estimatet for sensitivitet at det er størst sannsynlighet for at en positiv prediksjon fra BVC vil være en sann positiv. Resultatene er altså motstridene om hvor treffsikker testene er på å identifisere de sanne positive tilfellene. Forklaringen er å finne i utregningen av estimatene. For mens sensitiviteten estimeres som et forhold mellom sanne positive og falske negative, så estimeres PPV som et forhold mellom sanne positive og falske positive. Youden's Index kan benyttes til å oppklare paradokset, siden dette statistiske estimatet forholder seg likeverdig til begge feilkildene, altså de falske positive og de falske negative. Youden's Index estimerer et tall mellom null og en, og nærmest 1 tilsier færrest feilkilder (23). Resultatene viser en svak tendens til at BVC er verktøyet som predikerer med størst treffsikkerhet.

DISKUSJON

Samtlige inkluderte studier i metaanalysen viser den samme tendensen til at testene generelt er gode til å identifisere personene som ikke utøver vold, men at testene er mindre gode til å identifisere individene som faktisk vil utøve vold. Til tross for forskjeller mellom de ulike studiene og verktøyene er andelen falske positive prediksjoner generelt rimelig stor, knapt mer enn 1 av 10 prediksjoner er korrekte i forhold til å identifisere en fremtidig voldsutøver. Umiddelbart fremstår akkurat dette forholdet som en stor svakhet ved testene. At testene generelt er gode til å identifisere personene som ikke utøver vold, er nødvendigvis ikke formildende, for som det fremgår av figur 4 (Forest Plot) er forekomsten av voldsepisoder svært få i forhold til det totale antallet prediksjoner. Men for prediksjon av sjeldne forekomne tilfeller er det uunngåelig at andelen falske positive blir vesentlig (12). Likevel er den høye andelen falske positive etisk betenkelig, i hvert fall om pasienter utsettes for unødvendig tvangsbruk som en konsekvens av prediksjonen. Men nå er det ikke nødvendigvis slik at samtlige potensielt farlige pasienter blir usatt for ulike tvangstiltak i preventivt hensyn. Mye handler om helsepersonellens bevissthet omkring risikoen, samt den personlige oppfølgingen og samhandlingen med pasientene. Likevel bør det være en viss bekymring omkring den høye andelen falske positive prediksjoner. For mange pasienter kan det oppleves som stigmatiserende å bli identifisert som potensiell voldsutøver. Men samtidig kan det ikke utelukkes at det er en sterk sammenheng mellom andelen falske positive, og personalets kliniske arbeid. Antakelsen bekreftes av en systematisk oversikt (1), som finner at kvaliteten

og fagligheten i interaksjonen mellom pasient og sykepleier er av stor betydning for aggresjonsnivået i en psykiatrisk sengepost. Tilsvarende forhold er også omtalt tidligere, i forhold til hvordan bruk av fremtidig voldshandling som referansestandard kan være problematisk. For det er ikke urimelig å anta at klinisk arbeid, samt gjennomføring av relevante preventive tiltak øker andelen falske positive prediksjoner. Dette forholdet kan illustreres med et teoretisk eksempel, bare for å spissformulere poenget. Sett at man har det perfekte voldsrisikoprediksjonsverktøyet, og prøver det ut ved en perfekt avdeling med usedvanlige dyktige og erfarne sykepleiere. En avdeling hvor pasientene håndteres slik at fysisk vold aldri forekommer. Resultatet ville blitt et verktøy med spesifisitet i toppskiktet. Mens den positive prediksjonsverdien ville vist et verktøy som predikerte med 0 % sannsynligheten for at et positivt svar på testen tilsvarte en sann positiv prediksjon. Dermed er det vist at en stor andel falske positive prediksjoner nødvendigvis ikke er problematisk. Og samtidig at andelen falske positive er et resultat av tre forhold, at det predikeres sjeldne forekomne tilfeller, et bevisst valg om hvilken type feil som bør unngås, samt kvaliteten på det kliniske arbeidet. Estimaten for kvaliteten på prediksjonene må også sees i lys av at de er basert på en statistisk metode som forutsetter en objektiv referansestandard. Og fremtidig vold er ikke en objektiv referansestandard.

Jeg har så langt vist at selv om begge verktøyene er langt fra perfekte, så kan man til en viss grad stole på dem, og begge prediksjonsverktøyene, BVC og BVC-CN, fremstår som gode hjelpemidler i klinisk psykiatrisk arbeid. Dog ikke for å selektere pasientene direkte til preventive tiltak, men som et systematisk hjelpemiddel å støtte seg på i klinikken. Det konkluderes også tilsvarende om verktøyet BVC-CN (20), at den største gevinsten med verktøyet er funksjonen som rådgivende hjelpemiddel i den kliniske beslutningsprosessen, samt som et nyttig redskap for sykepleierne i vurderinger omkring vold og aggresjon. Andre finner at bruk av et voldsrisikoprediksjonsverktøy i akuttpsykiatrien reduserte både forekomsten av preventive tiltak og voldsepisoder (4). Resultatet forklares med økt bevissthet omkring potensielle farer, og dermed mer fokusert bruk av preventive tiltak. I tillegg ble personalet mer fokusert på konfliktdempende samhandling med pasientene. Med et slikt utgangspunkt er det beklagelig at studiene om prediksjonsverktøyet DASA-IV ikke passerte kvalitetsvurderingen, og derfor ikke inngikk i metaanalysen. For et verktøy som foruten å predikere voldsfare, også har i seg å være et hjelpemiddel til klinisk håndtering av voldsrisiko, fremstår definitivt som et verktøy med potensiell stor klinisk relevans.

KONKLUSJON

Med utgangspunkt i resultatene fra metaanalysen kan det ikke trekkes noen konkret konklusjon om hvilket verktøy, BVC eller BVC-CN, som er best egnet for å predikere fremtidige voldsepisoder. Men gjennom diskusjonen, med støtte i resultatene, er det vist at begge verktøyene er godt egnede kliniske verktøy til fortløpende vurdering av akutte risikofaktorer for fysisk aggresjon ved innleggelser i akuttpsykiatrien. I tillegg fremstår verktøyene å ha en klinisk nyttefunksjon utover bare voldsprediksjonen, ved at personalet blir mer fokusert på konfliktdempende samhandling med pasientene.

Etiske vurderingar

Det eksisterer ingen interessekonflikter for denne forfatteren, verken i form av økonomiske eller ansettelsesmessige bindinger.

Takksigelser

Flere personer må takkes for å ha bidratt i arbeidet med artikkelen. Gunhild Austrheim for hjelp til utviklingen av en søkestrategi. Jill Bjarke for å ha vært den uavhengige andrepersonen i inklusjons- og kvalitetsvurderingsprosessen. Rolf Gjestad for hans statistiske ekspertise. Tilslutt vil jeg takke veilederen for dette prosjektet, Geir Smedslund ved Kunnskapssenteret, for gode råd og et fruktbart samarbeid.

REFERANSER

1. **Cornaggia, C. M., Beghi, M., Pavone, F. & Barale, F.** Aggression in psychiatry wards: A systematic review. *Psychiatry Research* 2010;189 (1):10–20.
2. **Björkdahl A., Olsson D. & Palmstierna T.** Nurses' short-term prediction of violence in acute psychiatric intensive care. *Acta Psychiatrica Scandinavica* 2006;113:224–229.
3. **Knutzen, M., Bjørkly, S., Bjørnstad, M., Furre, A. & Sandvik, L.** Innsamling og analyse av data om bruk av tvangsmidler og vedtak om skjerming i det psykiske helsevernet for voksne i 2012. Kompetansesenterets Rapport 2014-2. Oslo. 2014.

4. **Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H. J., & Fischer, J.** (2008) Structured risk assessment and violence in acute psychiatric wards: randomized controlled trial. *The British Journal of Psychiatry* 2008;193:44-50.
5. **National Institute for Health and Clinical Excellence.** Clinical Guideline 25. Violence: the short-term management of disturbed/violent behaviour in psychiatric inpatient settings and emergency departments. 2005. Tilgjengelig fra: <http://www.nice.org.uk/CG025NICEguideline> (nedlastet 6.april 2012).
6. **Helsedirektoratet.** Nasjonal faglig retningslinje for utredning, behandling og oppfølging av psykoselidelser. IS-1957. Oslo, Helsedirektoratet. 2013.
7. **Sosial- og helsedirektoratet.** Vurdering av risiko for voldelig atferd – bruk av strukturerte kliniske verktøy. IS-9/2007. Oslo, Sosial- og helsedirektoratet. 2007.
8. **Helsedirektoratet.** Nasjonal faglig retningslinje for utredning, behandling og oppfølging av personer med samtidig ruslidelse og psykisk lidelse – ROP-lidelser. IS-1948. Oslo, Helsedirektoratet. 2012.
9. **Woods, P. & Ashley, C.** Violence and aggression: a literature review. *Journal of Psychiatric and Mental Health Nursing* 2007;14:652–660.
10. **Hartvig, P., Roaldset, J., Moger, T., Ostberg, B., & Bjørkly, S.** The first step in the validation of a new screen for violence risk in acute psychiatry: The inpatient context. *European Psychiatry* 2011;26:92–99.
11. **Daffern, M.** The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior* 2007;12:116-130.
12. **Singh, J. P., Serper, M., Reinharth, J. & Fazel,** (2011) Structured Assessment of Violence Risk in Schizophrenia and Other Psychiatric Disorders: A Systematic Review of the Validity, Reliability, and Item Content of 10 Available Instruments. *Schizophrenia Bulletin* 2011;37:899-912.
13. **de Vet, H. C. W., Eisinga, A., Riphagen, I. I., Aertgeerts, B. & Pewsner D.** Chapter 7: Searching for Studies. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. *Cochrane Handbook for*

Systematic Reviews of Diagnostic Test Accuracy Version 0.4. The Cochrane Collaboration, Birmingham. 2008. Tilgjengelig fra: <http://srdata.cochrane.org> (nedlastet 2. oktober 2013).

14. **Higgins, J. P. T. & Green S. Red.** Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration, Birmingham. 2011. Tilgjengelig fra: www.cochrane-handbook.org (nedlastet 19. november 2013).

15. **Polit, D., & Beck, C.** Nursing research: generating and assessing evidence for nursing practice. 9. utg. Philadelphia, Lippincott Williams & Wilkins. 2012.

16. **Reitsma, J. B., Rutjes, A. W. S., Whiting, P., Vlassov, V. V., Leeflang, M. M. G. & Deeks, J. J.** Chapter 9: Assessing methodological quality. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. red. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, Birmingham. 2009. Tilgjengelig fra: <http://srdata.cochrane.org> (nedlastet 2. oktober 2013).

17. **QUADAS.** QUADAS-2 tool. University of Bristol. 2011. Tilgjengelig fra <http://www.bristol.ac.uk/quadas/> (Nedlastet 14. mai 2014).

18. **Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M. & QUADAS-2 Group.** QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;155:529-536.

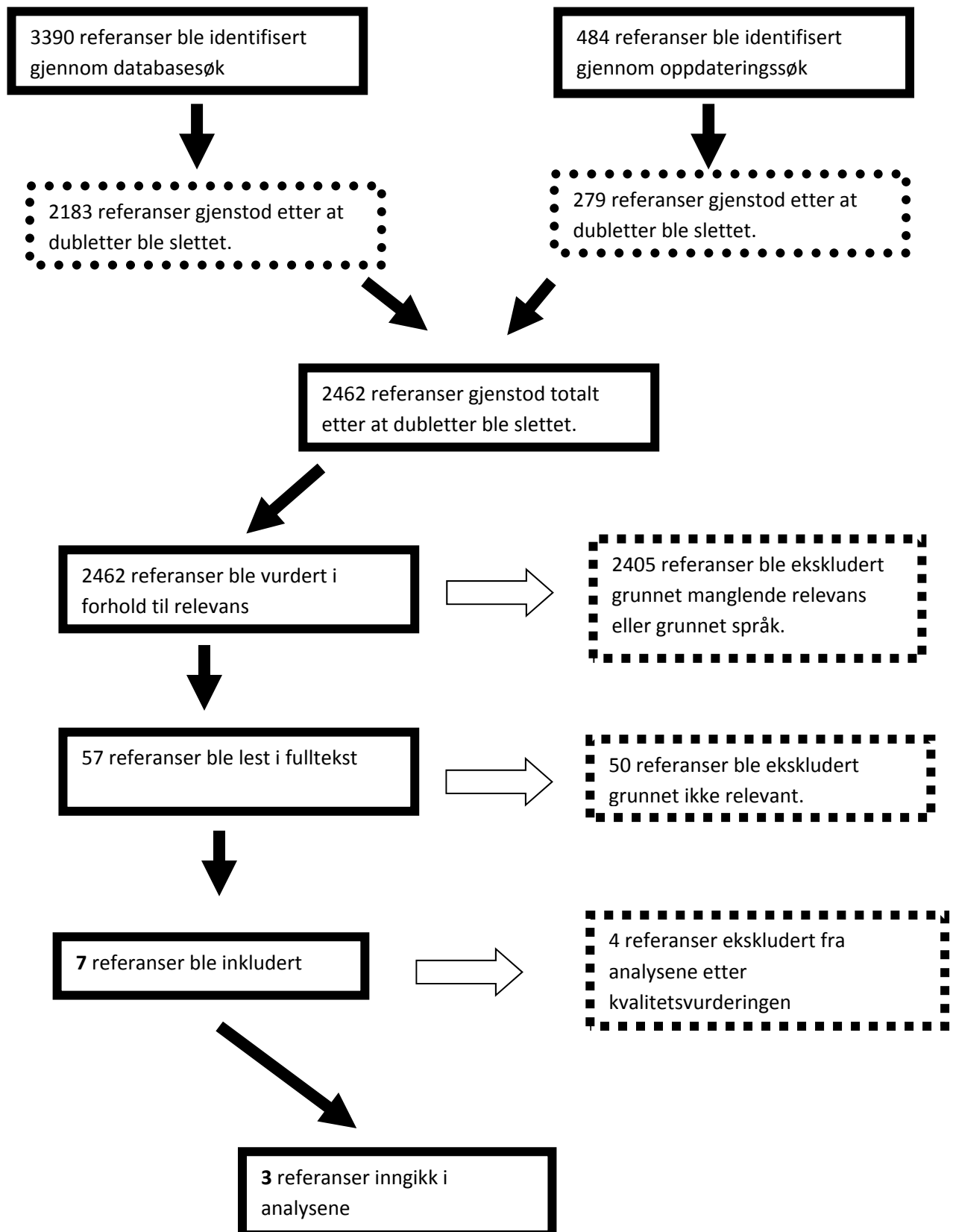
19. **Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H. J., & Fischer, J.** Predicting inpatient violence using an extended version of the Broset-Violence-Checklist: instrument development and clinical application. *BMC Psychiatry* 2006;6. Tilgjengelig fra: <http://www.biomedcentral.com/1471-244X/6/17> (nedlastet 09. sept. 2013)

20. **Rechenmacher, J., Muller, G., Abderhalden, C. & Schulc, E.** The diagnostic efficiency of the extended German Broset Violence Checklist to assess the risk of violence. *Journal of Nursing Measurement* 2014;22:201-212.

21. **Abderhalden, C., Needham, I., Miserez, B., Almvik, R., Dassen, T., Haug, H. J. & Fischer, J. E.** Predicting inpatient violence in acute psychiatric wards using the Brøset-Violence-Checklist: a multicentre prospective cohort study. *Journal of Psychiatric and Mental Health Nursing* 2004;11:422–427.




22. **Yao, X., Li, Z., Arthur, D., Hu, L., An, F. R. & Cheng, G.** Acceptability and psychometric properties of Broset Violence Checklist in psychiatric care settings in China. *Journal of Psychiatric & Mental Health Nursing* 2014;21:848-855.
23. **Macaskill, P., Gatsonis, C., Deeks, J.J., Harbord, R. M. & Takwoingi, Y.** Chapter 10: Analyzing and Presenting Results. I: Deeks, J. J., Bossuyt, P. M. & Gatsonis, C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*. The Cochrane Collaboration, Birmingham. 2010. Tilgjengelig fra: <http://srdta.cochrane.org> (nedlastet 2. oktober 2013).
24. **Griffith, J. J., Daffern, M. & Godber, T.** Examination of the predictive validity of the Dynamic Appraisal of Situational Aggression in two mental health units. *International Journal of Mental Health Nursing* 2013;22:485-492.
25. **Almvik, R., Woods, P. & Rasmussen, K.** The Broset Violence Checklist: Sensitivity, specificity, and interrater reliability. *Journal of Interpersonal Violence* 2000;15:1284-1296.
26. **Dumais, A., Larue, C., Michaud, C. & Goulet, M. H.** Predictive validity and psychiatric nursing staff's perception of the clinical usefulness of the French version of the Dynamic Appraisal of Situational Aggression. *Issues in Mental Health Nursing* 2012;33:670-675.
27. **Ogloff, J. R. P., & Daffern, M.** The Dynamic Appraisal of Situational Aggression: An Instrument to Assess Risk for Imminent Aggression in Psychiatric Inpatients. *Behavioral Sciences & the Law* 2006;24:799-813.
28. **Singh, J. P. Grann, M. & Fazel, S.** A comparative study of violence risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychological Review* 2011;31:499–513.
29. **Mossman, D.** (1994) Assessing predictions of violence: being accurate about accuracy. *Journal of Consulting and Clinical Psychology* 1994;62:783-792.

Fig. 1:



Tabell 1:

	<u>Risk of Bias</u>				<u>Applicability Concerns</u>		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Abderhalden 2004	+	+	?	+	+	+	+
Abderhalden 2006	+	+	?	+	+	+	+
Almvik 2000	+	+	?	+	+	+	-
Dumais 2012	+	+	?	+	-	+	+
Griffith 2013	+	+	?	+	-	+	-
Rechenmacher 2014	+	+	?	+	+	+	+
Yao 2013	+	+	?	+	+	+	-

 High	 Unclear	 Low
---	--	--

(Review Manager 5.3)

Tabell 2:

Descriptive Statistics						
Select		N	Minimum	Maximum	Mean	Std. Deviation
1	Sensitivitet	2	.64	.68	.6593	.02321
	Spesifisitet	2	.90	.94	.9221	.02454
	Youden	2	.58	.58	.5814	.00133
	PPV	2	.11	.11	.1124	.00179
	NPV	2	.99	1.00	.9946	.00140
	LRplus	2	7.09	10.62	8.8545	2.49129
	LRneg	2	.36	.38	.3693	.01534
	AUC	2	.86	.88	.8700	.01414
	Age	2	36.40	39.90	38.1500	2.47487
	Age_SD	2	13.00	13.30	13.1500	.21213
	Male	2	.61	.61	.6100	.00000
	Valid N (listwise)	2				
	2	Sensitivitet	2	.54	.59	.5644
Spesifisitet		2	.92	.97	.9439	.03333
Youden		2	.46	.56	.5083	.06705
PPV		2	.11	.19	.1472	.05367
NPV		2	.99	.99	.9929	.00257
LRplus		2	6.79	18.10	12.4449	7.99974
LRneg		2	.43	.50	.4624	.05205
AUC		2	.83	.93	.8805	.07142
Age		2	36.40	47.04	41.7200	7.52362
Age_SD		2	13.30	17.39	15.3450	2.89207
Male		2	.50	.61	.5530	.08061
Valid N (listwise)		2				

1 = BVC, 2 = BVC-CN

(IBM SPSS Statistics 22.0.0.2)

Tabell 3:

BVC

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Abderhalden 2004	9	72	5	1117	0.64 [0.35, 0.87]	0.94 [0.92, 0.95]		
Abderhalden 2006	25	195	12	1852	0.68 [0.50, 0.82]	0.90 [0.89, 0.92]		

BVC-CN

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Abderhalden 2006	20	163	17	1884	0.54 [0.37, 0.71]	0.92 [0.91, 0.93]		
Rechenmacher 2014	10	44	7	1310	0.59 [0.33, 0.82]	0.97 [0.96, 0.98]		

(Review Manager 5.3)

VEDLEGG I: Tabell 1**Database(s): Embase 1974 to 2014 April 08**

Search Strategy: #	Searches	Results
1	exp aggression/68064	
2	exp violence/ 100060	
3	(violen* or aggress* or attack* or assault*).tw.	333292
4	1 or 2 or 3	425946
5	exp "danger, risk, safety and related phenomena"/	1807981
6	checklist/	7578
7	exp questionnaire/	390979
8	prediction/	216554
9	"sensitivity and specificity"/	195757
10	(risk* or assess* or predict* or tool* or instrument* or measur* or scale* or test* or survey* or checklist* or questionnaire* or appraisal* or sensitivity or specificity).tw.	8568103
11	5 or 6 or 7 or 8 or 9 or 10	9171362
12	4 and 11	212338
13	exp mental disease/	1518686
14	(mental* or bipolar* or schizo* or psychos* or personality* or manic* or psychiatric*).tw.	709585
15	13 or 14	1788325
16	12 and 15	57986
17	(psychiatr* adj3 (inpatient* or hospitali* or admission* or readmission*)).tw.	13077
18	16 and 17	1197

VEDLEGG II: Tabell 2**Database(s): Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present**

Search Strategy: #	Searches	Results
1	Violence/	24134
2	exp Aggression/	28161
3	(violen* or aggress* or attack* or assault*).tw.	261268
4	1 or 2 or 3	279368
5	exp Checklist/	2124
6	Questionnaires/	295091
7	exp risk/	819675
8	exp "Predictive Value of Tests"/	139698
9	exp "Sensitivity and Specificity"/	409219
10	(risk* or assess* or predict* or tool* or instrument* or measur* or scale* or test* or survey* or checklist* or questionnaire* or appraisal* or sensitivity or specificity).tw.	7068042
11	5 or 6 or 7 or 8 or 9 or 10	7396894
12	4 and 11	135141
13	exp Mental Disorders/	936504
14	(mental* or bipolar* or schizo* or psychos* or personality* or manic* or psychiatric*).tw.	558815
15	13 or 14	1205832
16	12 and 15	27049
17	(psychiatr* adj3 (inpatient* or hospitali* or admission* or readmission*)).tw.	10194
18	16 and 17	740

VEDLEGG III: Tabell 3**Database(s): PsycINFO 1967 to April Week 1 2014**

Search Strategy: #	Searches	Results
1	exp patient violence/	1025
2	Violence/	22539
3	exp aggressive behavior/	113666
4	aggressiveness/3377	
5	(violen* or aggress* or attack* or assault*).tw.	138373
6	1 or 2 or 3 or 4 or 5	184369
7	exp risk management/	3461
8	risk assessment/	8825
9	exp measurement/	249518
10	(risk* or assess* or predict* or tool* or instrument* or measur* or scale* or test* or survey* or checklist* or questionnaire* or appraisal* or sensitivity or specificity).tw.	1658811
11	7 or 8 or 9 or 10	1681310
12	6 and 11	97894
13	exp mental disorders/	425666
14	(mental* or bipolar* or schizo* or psychos* or personality* or manic* or psychiatric*).tw.	625135
15	13 or 14	816330
16	12 and 15	32525
17	exp psychiatric hospitalization/	8696
18	(psychiatr* adj3 (inpatient* or hospitali* or admission* or readmission*)).tw.	16183
19	17 or 18	20893
20	16 and 19	1405

VEDLEGG IV: Tabell 4**Search the NCJRS Abstracts Database**

The NCJRS Abstracts Database contains summaries of the more than 215,000 criminal justice, juvenile justice, and substance abuse resources housed in the NCJRS Library collection.

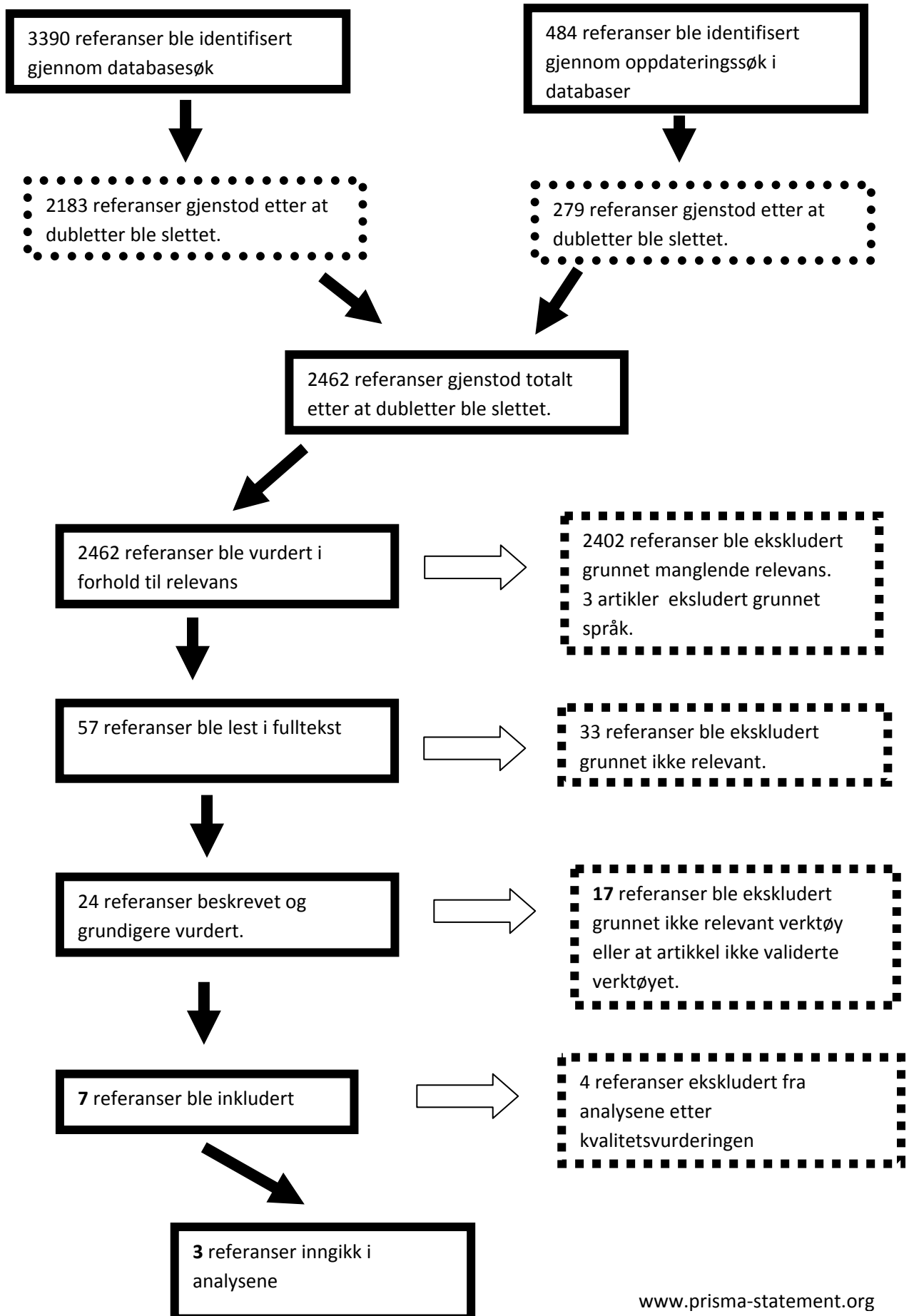
Tell us how you use the NCJRS Abstracts Database. Send us your [Feedback](#).

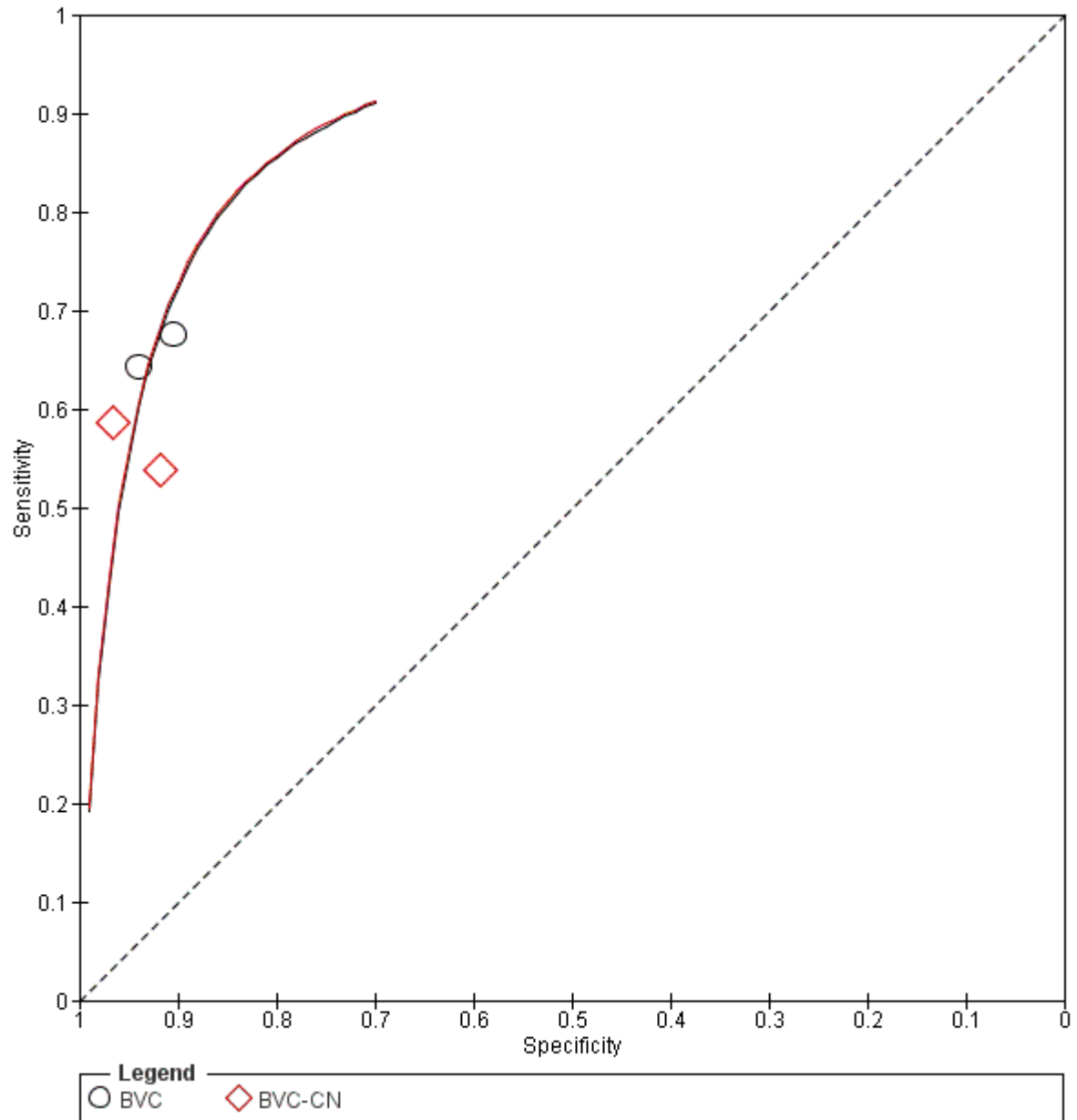
Choose a search type All Any Phrase and enter your search below

Title	<input type="text"/>
Author	<input type="text"/>
Journal Name	<input type="text"/>
NCJ Number Numbers Only	<input type="text"/>
General Search	<input type="text" value="risk violen* inpatient*"/>
Language	<input type="text" value="ALL"/>
Date Range	<input type="text"/> - <input type="text"/>
	<small>Start Date End Date (Enter date as mm/yyyy for example: 10/2002.)</small>
	<input type="button" value="Search"/> <input type="button" value="Clear"/>

Resultat: 48

VEDLEGG V: Figur 1: PRISMA FLYTEDIAGRAM



VEDLEGG VII: *Figur 5: SROC kurve*

(Review Manager 5.3)