

Fontys Paramedische Hogeschool

Medisch Beeldvormende en Radiotherapeutische Technieken

“Ik zie, ik zie wat jij niet ziet”

Een aanbeveling voor een norm voor het maximaal aantal uur dat een screeningsradioloog per dag mammogrammen mag beoordelen.

Onderzoeksverslag

Linda Kuiper
2189653

Begeleider:
Rob Vogels

Opdrachtgever:
Ellen Paap

9 juni 2015



Voorwoord

Deze literatuurstudie is aangeboden door het Landelijk Referentiecentrum voor Bevolkingsonderzoek (LRCB) te Nijmegen. Dit onderzoek is uitgevoerd in het kader van de afstudeerfase van de opleiding Medisch Beeldvormende en Radiotherapeutische Technieken aan de Fontys Paramedische Hogeschool te Eindhoven.

Het doel van dit onderzoek is om een basis te leggen voor het bepalen van een norm voor het maximaal aantal uur dat een screeningsradioloog per dag mammogrammen mag beoordelen.

Graag wil ik verschillende mensen bedanken voor hun hulp bij de uitvoering van dit onderzoek. Als eerste bedank ik Ellen Paap van het LRCB voor haar goede begeleiding en feedback gedurende de afgelopen 20 weken. Iedereen werkzaam bij het LRCB wil ik bedanken voor hun enthousiasme en interesse voor dit onderzoek. Het heeft gezorgd voor veel motivatie om door te gaan.

Als tweede wil ik Rikie Deurenberg-Vos, informatiespecialist aan het Radboud UMC, bedanken voor haar begeleiding bij het creëren van een zoekstrategie. Zij heeft voor orde gezorgd in de chaos die is ontstaan bij het maken van de zoekstrategie.

Verder wil ik Rob Vogels bedanken voor de begeleiding tijdens de afstudeerfase. Niet alleen op het gebied van het onderzoek, maar ook voor het in de gaten houden van het persoonlijk functioneren gedurende de uitvoering van het onderzoek.

Als laatste wil ik mijn medestudenten Nicole van de Griend, Laura Saris en Anne Vollering bedanken voor de feedback die zij hebben gegeven op zowel mijn projectplan als onderzoeksverslag. De hoeveelheid feedback heeft ervoor gezorgd dat mijn verslag steeds een stukje beter werd.

Linda Kuiper

Utrecht, 9 juni 2015

Samenvatting

Achtergrond

Om borstkanker vroegtijdig op te sporen, worden vrouwen tussen de 50 en 75 jaar elke twee jaar uitgenodigd voor het bevolkingsonderzoek op borstkanker. De onderzoeken, mammogrammen, worden door screeningsradiologen beoordeeld. In Nederland is een minimum van 3.000 mammogrammen per jaar als norm vastgesteld om ervaring en deskundigheid te behouden. Dit minimum wordt ruim gepasseerd, waardoor de kans bestaat dat nauwkeurigheid in beoordelen omlaag gaat. In dit onderzoek wordt onderzocht wat het maximaal aantal uur is dat een screeningsradioloog per dag mammogrammen mag beoordelen.

Methode

Een literatuurstudie is uitgevoerd waarbij middels een zoekstrategie in PubMed relevante literatuur is gevonden. Gevonden literatuur is gescreend op titel, abstract en full tekst. De overgebleven studies zijn beoordeeld op methodologische kwaliteit. De belangrijkste resultaten zijn opgenomen in een data-extractietabel.

Resultaten

In dit onderzoek zijn 35 artikelen geïnccludeerd. In 25 studies wordt informatie gegeven over uitkomstmaten als sensitiviteit en het fout positieven cijfer, in relatie tot het jaarlijks volume van de radioloog. In bijvoorbeeld 11 van de 15 studies wordt beschreven dat de sensitiviteit gelijk blijft en in zes van de acht studies daalt het aantal fout positieven, wanneer het jaarlijks volume stijgt. Tien studies bevatten informatie over vermoeidheid en oogspanning. Maximaal kan 4 tot 7 uur op een dag worden gescreend voordat vermoeidheid leidt tot lagere nauwkeurigheid. Elk uur een pauze van 5 tot 10 minuten zorgt voor mindere mate van oogspanning en vermoeidheid.

Conclusie

Uit deze literatuurstudie komt een ruime range van 4 tot 7 uur naar voren als maximale aantal screeningsuren per dag. Geadviseerd wordt om elk uur, 5 tot 10 minuten een pauze te nemen. Over de uitkomstmaten in relatie tot het jaarlijks volume kan geen eenduidige uitspraak worden gedaan. Meer onderzoek is nodig om een aanbeveling te doen naar het maximale aantal screeningsuren per dag, waarin ook de mening van experts moet worden meegenomen.

Abstract

Background

For early detection of breast cancer, women between 50-75 are invited for screening every two years. The examinations, called mammograms, are reviewed by screening radiologists. In the Netherlands, screening radiologists must evaluate at least 3,000 mammograms per year. This minimum is largely exceeded, which increases the risk of lowering the accuracy of evaluation. This study investigated whether a standard can be recommended for the maximum screening hours per day a screening radiologist should evaluate mammograms.

Method

For this literature review, a search strategy is created to find relevant literature. The literature found by the search strategy is screened by title, abstract and content. The remaining studies were rated for their quality. The main results are included in a data extraction table.

Results

In this study, 35 articles were included. Twenty-five studies provide information on outcome measures, such as sensitivity and false-positive rate, in relation to the annual volume of the radiologist. If the annual volume increases, the sensitivity remains the same in 11 out of 15 studies and the false positive rate decreases in six out of eight studies. However, no clear conclusion can be extracted, because of conflicting data from these studies. Ten studies contain information about fatigue and eyestrain. After screening 4 to 7 hours a day fatigue leads to lower accuracy. Every hour a break for 5 to 10 minutes resulted in decreased eyestrain and fatigue.

Conclusion

From this study, a wide range of 4 to 7 hours is emerged as maximum screening hours per day. It is advised to take a break for 5 to 10 minutes every hour. No clear statement can be made about the outcomes in relation to the annual volume. More research is needed to make a recommendation of the maximum screening hours per day, thereby including the opinion of experts.

Inhoud

1. Inleiding	6
2. Methode	8
2.1 Database	8
2.2 Zoektermen en strategie	8
2.3 Inclusiecriteria	9
2.4 Dataselectie.....	9
2.5 Kwaliteitscriteria.....	9
2.6 Ethische paragraaf	9
3. Resultaten	11
3.1 Kwaliteitsbeoordeling.....	11
3.2 Jaarlijks volume en ervaring	12
3.3 Vermoeidheid, oogspanning en concentratie.....	14
4. Discussie en conclusie	21
Literatuur	25
Bijlage 1: Definities van afkortingen en uitkomstmaten	I
Bijlage 2: Kwaliteitsbeoordeling	II

1. Inleiding

Bij vrouwen is borstkanker de meest voorkomende vorm van kanker [1-2]. In 2013 werd bij 14.000 vrouwen in Nederland invasieve borstkanker gediagnosticeerd [2]. De kans op het krijgen van borstkanker is van 11% in 1990 gestegen naar 15% in 2010 [3-4]. Elk jaar sterven landelijk 3.200 vrouwen aan borstkanker [5]. Om borstkanker in een vroeg stadium te ontdekken is tussen 1989 en 1997 begonnen met het landelijk invoeren van het bevolkingsonderzoek naar borstkanker [5-6]. Uit verschillende studies blijkt dat wanneer kanker in een vroeg stadium wordt ontdekt de kans op herstel groter is dan in een later stadium [1,7]. Om de twee jaar worden vrouwen tussen de 50 en 75 jaar uitgenodigd om deel te nemen aan dit screeningsonderzoek, zodat borstkanker in een vroeg stadium ontdekt kan worden [6].

Tijdens het screeningsonderzoek worden twee röntgenopnames per borst gemaakt, de craniocaudale en de mediolaterale oblique opname [8-9]. Deze vier opnames worden een mammogram genoemd [8-9]. Mammografie wordt tot op heden gezien als de beste screeningsmethode om borstkanker op te sporen [7,10]. De mammogrammen worden beoordeeld door screeningsradiologen die hier speciaal voor zijn opgeleid [11]. De mammogrammen worden door twee screeningsradiologen beoordeeld om zo min mogelijk afwijkingen te missen. Het beoordelen van de mammogrammen gebeurt onafhankelijk van elkaar. Dit proces wordt double reading genoemd [11-14]. Eén van de twee radiologen is bij voorkeur een screeningsradioloog met meer dan vijf jaar ervaring [14-16].

Om als screeningsradioloog te mogen werken dient een radioloog eerst het certificaat mammascreening te behalen [17]. Het Landelijk Referentiecentrum voor Bevolkingsonderzoek (LRCB) is verantwoordelijk voor de kwaliteit van het screeningsonderzoek en zorgt voor de opleiding tot screeningsradioloog [18]. Het LRCB verzorgt een achtdaagse opleiding tot screeningsradioloog en na succesvolle afronding wordt het certificaat mammascreening uitgereikt [19]. De registratie als screeningsradioloog is vijf jaar geldig en gedurende deze jaren moet de screeningsradioloog scholingspunten halen om kennis en kundigheid te behouden [17]. Door het LRCB worden bijscholingsactiviteiten aangeboden en krijgen de screeningsradiologen door middel van een visitatie feedback op hun functioneren, het kankerdetectie- en verwijscijfer en het fout positieven cijfer [15-16, 20].

Screeningsradiologen beoordelen in een jaar tijd veel mammogrammen. In Nederland wordt een norm van minimaal 3.000 mammogrammen per jaar aangehouden om ervaring en deskundigheid te behouden [15]. Dit minimum wordt ruim gepasseerd. De variatie tussen screeningsradiologen en het aantal mammogrammen dat zij beoordelen is groot [21-23]. Screeningsradiologen in Nederland beoordelen gemiddeld 15.000 onderzoeken per jaar [Persoonlijke communicatie T. Geerste, QA bij LRCB]. Het maximale aantal onderzoeken in een jaar loopt op tot ruim 100.000 [21]. Dit zijn gegevens van de jaren 2013 tot 2015. De verschillen in het minimum dat is opgesteld is wereldwijd terug te zien.

Zo wordt in Noord-Amerika het minimum gesteld op 480 mammogrammen per jaar en in Australië 2.000 per jaar [22], terwijl dit in Europa gesteld is op 5.000 mammogrammen [14,22].

Screeningsradiologen in Nederland beoordelen op een dag 50 tot 100 of meer mammogrammen [24]. Per mammogram wordt ongeveer één tot twee minuten besteed aan beoordelingstijd [24]. In deze tijd moet een screeningsradioloog bepalen of een mammogram een afwijking bevat. In de screening is de prevalentie van een afwijking op een mammogram erg laag [25-26]. Van de 1000 mammogrammen bevatten slechts twee tot zes een carcinoom [25,27]. Vrouwen die komen voor een screening komen zonder klachten of symptomen, waardoor de kans op het vinden van een afwijking veel lager is [25-26,28]. Het beoordelen van een mammogram is hierdoor geconcentreerd werk, wat voor problemen kan zorgen [22].

Een mammogram is opgebouwd uit zwart-wittinten [24]. Het lang bekijken van een zwart-wit beeld kan effect hebben op de focus en concentratie van een radioloog [24,29]. Door lang naar een zwart-wit beeld te kijken kan oogvermoeidheid optreden, waardoor vlekken in het gezichtsveld komen en beelden minder goed te beoordelen zijn [24,29]. Niet alleen door het bekijken van een zwart-wit beeld wordt de focus en concentratie verminderd. De lengte van een werkdag en afstand tot het computerscherm hebben eveneens invloed. Zo wordt de focus en hiermee ook de diagnostische nauwkeurigheid minder na een lange werkdag en een kleine afstand tot het computerscherm [29-31]. Een ander probleem dat optreedt na een lange periode van screenen is wat onderzoekers de tevredenheid van het zoeken noemen. Na ontdekking van een afwijking wordt niet meer verder gezocht en kunnen eventuele kleinere afwijkingen worden gemist [24,32].

Hoewel in Nederland een minimum is opgesteld, is er geen maximum voor het aantal mammogrammen dat een screeningsradioloog achter elkaar mag beoordelen of hoeveel uur hij of zij zich hier per dag aan mag wijden. In dit onderzoek zal worden onderzocht in hoeverre een screeningsradioloog achter elkaar mammogrammen kan beoordelen. Om een norm te kunnen vaststellen wordt gebruik gemaakt van bestaande wetenschappelijke literatuur. Hierbij worden verschillende aspecten geanalyseerd die bij het beoordelen een rol kunnen spelen zoals oogvermoeidheid, de diagnostische nauwkeurigheid gedurende een werkdag en beeldschermwerk. Het doel van dit onderzoek is om te bepalen wat het maximaal aantal uur is dat een screeningsradioloog per dag mag besteden aan het beoordelen van mammogrammen.

2. Methode

2.1 Database

Voor de uitvoering van deze literatuurstudie is gebruik gemaakt van de database PubMed. Ruim 5.000 tijdschriften op het gebied van geneeskunde en gerelateerde vakgebieden zijn in deze database opgenomen. PubMed leent zich bovendien goed voor het creëren van een zoekstrategie [33]. Uit vooronderzoek is gebleken dat een groot aantal studies in PubMed gevonden werd en is besloten om alleen PubMed als database te gebruiken voor het doorzoeken van de literatuur.

2.2 Zoektermen en strategie

Voor het opstellen van de zoekstrategie is gebruik gemaakt van de Pearl Growing techniek [34]. Met deze techniek worden woorden of referenties uit relevante studies (de Pearls) gebruikt om meer studies te vinden. Deze woorden zijn in de zoekstrategie toegevoegd. Alle zoektermen zijn verdeeld in drie groepen om zo een structuur te creëren in de zoekstrategie. De eerste groep bevat de zoektermen radiology, radiologists, mammography, mass screening, computers en computer users. De tweede groep bevat de zoektermen image interpretation en observer performance. De laatste groep bevat zoektermen waarin de uitkomsten zijn beschreven, zoals workload, clinical competence, observer variation, eyestrain, reading volume, visual strain, visual search, diagnostic accuracy, experience en false positive reactions.

Vanuit deze zoektermen is de volgende zoekstrategie opgesteld (Tabel 1).

Tabel 1 Zoekstrategie ingevoerd in PubMed

Zoekstrategie
radiology [Mesh] OR mammography [Mesh] OR diagnostic imaging [tiab] OR diagnostic [ti] OR medical imaging [tiab] OR image [ti] OR imaging [ti] OR images [ti] OR radiologist [tiab] or radiology [tiab] or radiologists [tiab] OR computer user [tiab] OR computer users [tiab]
AND
image interpretation [tiab] OR image perception [tiab] OR image reading [tiab] OR perception [ti] OR visual perception [Mesh] OR visual [ti] OR search [ti] OR reading [ti] OR (reader [tiab] AND accuracy [tiab]) OR (reader* [tiab] AND performance [tiab]) OR (observer [ti] AND performance [tiab]) OR vision [ti] OR eye [ti] OR eyes [ti] OR interpreting [ti] OR interpretation [ti] OR interpretive [ti]
AND
Workload [Mesh] OR workload [ti] OR work load [ti] OR workday* [ti] OR volume [ti] OR diagnostic errors [Mesh] OR diagnostic errors [ti] OR false positive [ti] OR false negative [ti] OR diagnostic accuracy [ti] OR accuracy [ti] OR quality assurance, health care [Mesh] OR quality assurance [ti] OR quality [ti] OR observer variation [Mesh] OR observer [ti] OR variation [ti] OR clinical competence [Mesh] OR competence [ti] OR strain [ti] OR asthenopia [Mesh] OR fatigue [ti] OR risk factors [ti]

2.3 Inclusiecriteria

Voor dit onderzoek zijn studies gebruikt uit de Engelstalige en Nederlandstalige literatuur. Hierin zijn alleen studies meegenomen waarin radiologie, mammografie of radiologen de onderzoeksgroep zijn. Verdere criteria zijn dat er in de studies onderzoek wordt gedaan naar uitkomsten gerelateerd aan de hoeveelheid mammogrammen dat wordt beoordeeld en de nauwkeurigheid van het beoordelen van röntgenbeelden. Hierbij kan gedacht worden aan sensitiviteit, kankerdetectiecijfer en het verwijscijfer van radiologen. Daarnaast is het gebruik van computers en het effect van beelden bekijken op een computerscherm meegenomen. Dit zijn uitkomsten als oogvermoeidheid en oogspanning. De definities voor de verschillende uitkomsten zijn terug te vinden in Bijlage 1.

Tot slot zijn alleen studies van de afgelopen 15 jaar geïnccludeerd. Deze tijdslimiet is bepaald op het feit dat vanaf ongeveer 2001 de helft van de radiologie afdelingen in Nederland is overgegaan op digitale apparatuur [35]. Afgaand aan de invoering van digitale apparatuur worden röntgenbeelden op een andere manier bekeken en dat is voor het huidige onderzoek minder van belang.

2.4 Dataselectie

Voor het selecteren van relevante studies is een dataselectie doorlopen. In deze dataselectie is rekening gehouden met de inclusiecriteria die voor dit onderzoek zijn opgesteld. De eerste stap in de dataselectie is het beoordelen van de studies op titel. Als tweede stap zijn de overgebleven studies beoordeeld op het abstract. De studies zijn daarna beoordeeld op full tekst. Als laatste zijn de referenties van de studies geanalyseerd die de gehele selectie hebben doorstaan. Mochten in de referenties studies staan die relevant zijn voor het onderzoek, zijn deze meegenomen in de dataselectie. De gegevens van de geïnccludeerde studies zijn verwerkt in een data-extractietabel, waarbij een beschrijving is gegeven van het soort studiedesign, de onderzoeksgroep, de gebruikte methode, de resultaten en de uiteindelijke conclusie.

2.5 Kwaliteitscriteria

De geïnccludeerde studies zijn aan de hand van de kwaliteitseisen beoordeeld op methodologische kwaliteit (Tabel 2). De kwaliteitseisen zijn met behulp van Cochrane checklisten [36,37], de checklist van het Critical Appraisal Skills Programme [38] en het boek van Wouters et al. [39] opgesteld.

Aan de hand van de Best Evidence Synthese is aan elke studie een waarde gegeven op basis van de kwaliteitseisen [40]. Als een vraag met 'ja' is beantwoord wordt één punt toegekend. Is een vraag met 'nee' beantwoord dan zijn nul punten gegeven. De volgende conclusies zijn hieruit getrokken: 0-3 punten is zwakke kwaliteit, 4-6 punten is matige kwaliteit en 7-9 punten is sterke kwaliteit.

2.6 Ethische paragraaf

Dit onderzoek betreft een literatuurstudie, er wordt geen beroep gedaan op (proef)personen. Hierdoor hoeft het niet aan de Wet Medisch-wetenschappelijk Onderzoek (WMO) verplichting te voldoen [41].

Tabel 2 Kwaliteitseisen

Kwaliteitseisen	Voldaan Ja/Nee
1. Heeft de studie een duidelijke probleemomschrijving/doel?	
2. Is er een duidelijke onderzoeksgroep omschreven?	
3. Is de onderzoeksgroep groot genoeg om onderzoek mee te doen?*	
4. Is er een duidelijke methode omschreven?	
5. Is de periode waarin het onderzoek is gedaan representatief voor het onderzoek?**	
6. Is de studie reproduceerbaar?	
7. Zijn de resultaten duidelijk en geloofwaardig omschreven?	
8. Is de conclusie een logisch vervolg op de resultaten?	
9. Wordt er vermeld wat de beperkingen zijn van het onderzoek of welk aanvullend onderzoek er gedaan moet worden?	

**De onderzoeksgroep wordt groot genoeg bevonden als deze 100 deelnemers of meer bevat. **Als in een artikel niet duidelijk aangegeven staat in welke periode het onderzoek is uitgevoerd, worden nul punten toegekend.*

3. Resultaten

Voor dit onderzoek is middels een zoekstrategie literatuur verzameld voor het beantwoorden van de onderzoeksvraag. Op 24 maart 2015 is de zoekstrategie in PubMed ingevoerd. Hieruit zijn 1.963 studies gekomen. Twee dubbele studies zijn ontdekt, waardoor nog 1.961 studies overbleven. Voor het vinden van relevante studies is een dataselectie doorlopen. Deze selectie is schematisch weergegeven in een flowchart (Figuur 1).

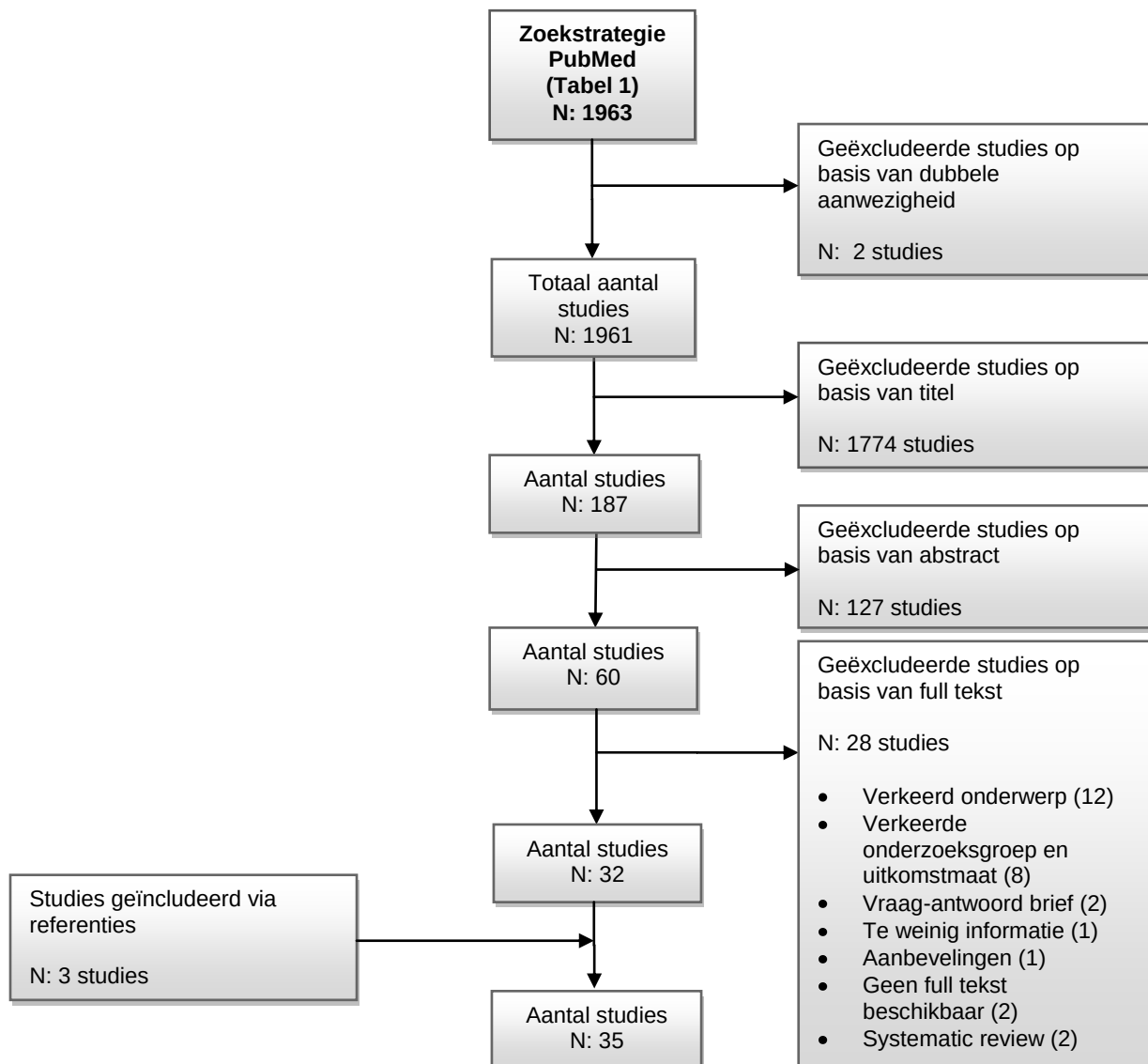
De studies zijn gescreend op titel, abstract en full tekst. Hierbij is gelet op de inclusiecriteria voor dit onderzoek. Op basis van de titel zijn 1.774 studies geëxcludeerd. De overgebleven 187 studies zijn gescreend op basis van het abstract, waarvan 127 studies zijn geëxcludeerd. De voornaamste redenen voor exclusie op titel en abstract waren de afwijkende onderwerpen en uitkomstmaten ten opzichte van de inclusiecriteria. Op basis van full tekst zijn 28 studies geëxcludeerd vanwege verschillende redenen. Van 12 studies kwam het onderwerp van de studie niet overeen met de inclusiecriteria, zoals CT-scan en therapie voor oogaandoeningen. Acht studies hadden een verkeerde uitkomstmaat of onderzoeksgroep. De uitkomstmaten beschreven niet het jaarlijks volume of nauwkeurigheid van de radiologen en twee studies hadden als onderzoeksgroep de laboranten die de mammogrammen maken. Twee studies waren een vraag-antwoord brief, één studie bevatte te weinig informatie om een oordeel uit te kunnen halen en één artikel was een aanbevelingsrapport. Van twee studies was de full tekst niet beschikbaar en kon niet op full tekst worden gescreend. Twee studies waren een literatuurstudie en zijn voor het onderzoek geëxcludeerd. Deze worden wel besproken in de discussie.

In totaal zijn 32 studies geïnccludeerd. Naar aanleiding van de referenties zijn nog drie studies geïnccludeerd, waardoor 35 studies zijn geïnccludeerd [7,9,22,23,29-31,42-69].

3.1 Kwaliteitsbeoordeling

Elke studie is beoordeeld op methodologische kwaliteit. In de data-extractietabellen (Tabel 4 en 5) is per studie de kwaliteitsbeoordeling opgenomen. De volledige beoordeling is terug te vinden in Bijlage 2. Drie studies scoren als een zwak artikel en hebben een lage bewijslast [47,61,63]. Deze studies zijn editorials en scoren alleen punten in de criteria over de probleemstelling, conclusie en beperkingen. De overige 32 studies scoren hoge punten variërend van zeven tot negen punten. Vijf studies scoren zeven punten [29-31,49,52], 12 studies scoren acht punten [22,42,44-46,52,54,64-68] en 15 studies scoren negen punten [7,9,23,43,48,51,53,55-60,62,69]. Zestien studies hadden een kleinere onderzoeksgroep dan 100 deelnemers en in negen studies was de interventietijd niet duidelijk beschreven.

De geïncludeerde studies zijn opgedeeld in twee groepen, gebaseerd op het onderwerp van de studie. De eerste groep bevat studies over het aantal mammogrammen dat screeningsradiologen per jaar beoordelen (jaarlijks volume), de werkervaring en de invloed van beide op de uitkomstmaten en de diagnostische nauwkeurigheid. De tweede groep bevat studies over vermoeidheid en de oogspanning na lange werkdagen door (screenings)radiologen en door gebruik van computers. Alle gegevens van de studies zijn schematisch weergegeven in twee data-extractietabellen (Tabel 4 en 5).



Figuur 1: Schematische weergave dataselectie.

3.2 Jaarlijks volume en ervaring

In totaal zijn er 25 studies waarin het jaarlijks volume wordt vergeleken met de nauwkeurigheid van de beoordeling door de (screenings)radiologen [7,9,22,23,42-62]. In Tabel 3 wordt in het kort de bevindingen uit deze studies weergegeven. In de studies worden de volgende uitkomstmaten onderzocht: de sensitiviteit, specificiteit, het fout positieven cijfer, het kankerdetectiecijfer, het

verwijscijfer van de screeningsradiologen en de positief voorspellende waarde (zie Tabel 4 en Bijlage 1 voor de definities).

In acht van de 25 studies hebben radiologen een jaarlijks volume hoger dan 5.000 onderzoeken per jaar [23,45-46,49,52,54,58-59]. De overige studies hebben een marge van jaarlijks 480 – 5.000 onderzoeken per jaar. In zeven studies worden testsets gebruikt, waarmee de resultaten zijn verzameld [9,48,52-54,59-60]. De overige studies hebben gedurende een bepaalde periode gegevens verzameld van de (screenings)radiologen in een werkelijke screeningssetting.

In 16 studies is een uitspraak gedaan over de sensitiviteit en specificiteit. In 11 studies is beschreven dat de sensitiviteit gelijk blijft bij elk jaarlijks volume [7,9,23,43,46,49,51,53-54,58,59]. Daartegenover beweren vier studies dat de sensitiviteit omhoog gaat, als het jaarlijks volume van radiologen eveneens omhoog gaat of als er door de radiologen meer mammogrammen worden bekeken die zij hebben doorverwezen voor verder onderzoek [42,44,48,56]. Voor wat betreft de specificiteit beschrijven drie studies dat er geen relatie is tussen het jaarlijks volume en de specificiteit [7,9,46]. Daarentegen beschreven Barlow et al. [42] dat de specificiteit omlaag gaat, zodra het jaarlijks volume omhoog gaat. Echter beschreven vier andere studies dat de specificiteit omhoog gaat als het jaarlijks volume hoger wordt [52-54,56]. Van deze vier studies geven Reed et al. [54] en Smith-Bindman et al. [56] specifiek aan dat deze stijging een stabilisatie bereikt vanaf respectievelijk 2.000 en 4.000 onderzoeken per jaar.

In acht studies wordt het fout positieven cijfer beschreven [23,43-44,49,51,56-58]. Twee van de acht studies concluderen dat het jaarlijks volume geen invloed heeft op het fout positieven cijfer [51,57]. De overige zes studies beschreven dat een hoger volume gepaard gaat met een lager fout positieven cijfer [23,43-44,49,56,58]. Buist et al. [44] en Haneuse et al. [49] geven hierbij specifiek een stabilisatie aan van respectievelijk 2.000 en 4.000 onderzoeken per jaar.

Over de invloed van het jaarlijks volume ten opzichte van het kankerdetectiecijfer doen tien studies een uitspraak [22-23,43-45,48-50,55-56]. Volgens vijf studies wordt het kankerdetectiecijfer beter als het jaarlijks volume omhoog gaat [23,44,48,50,55]. Hiertegenover staan twee studies die aangeven dat het detectiecijfer omlaag gaat bij een hoger jaarlijks volume [45,49] en drie studies waarbij wordt beschreven dat het jaarlijks volume geen invloed heeft op het kankerdetectiecijfer [22,43,56].

In drie studies wordt het verwijscijfer beschreven [43,45-46]. Buist et al. [43] en Cornford et al. [45] geven de conclusie dat een hoger jaarlijks volume zorgt voor een lager verwijscijfer, terwijl Duncan et al. [46] beschrijven dat het jaarlijks volume geen invloed heeft op het verwijscijfer. In twee studies wordt de positief voorspellende waarde meegenomen in het onderzoek [22,45]. Volgens Coldman et al. [22] gaat de positief voorspellende waarde omhoog als het volume groter wordt, waarbij stabiliteit is vanaf een volume van 2.000 onderzoeken per jaar. Daartegenover concludeert Cornford et al. [45] dat het jaarlijks volume geen invloed heeft op de positief voorspellende waarde.

Tabel 3 Samenvatting van relatie tussen de uitkomstmaten en een hoger jaarlijks volume van een radioloog.

Uitkomstmaat	Hoger*	Gelijk	Lager	Totaal aantal studies**
Sensitiviteit	4	11	-	15 / 25
Specificiteit	4	3	1	8 / 25
Fout positieven cijfer	-	2	6	8 / 25
Kankerdetectiecijfer	5	3	2	10 / 25
Verwijscijfer	-	1	2	3 / 25
Positief voorspellende waarde	1	1	-	2 / 25

*Het aantal studies waarin de uitkomstmaat omhoog of omlaag gaat of gelijk blijft bij een hoger wordend jaarlijks volume. **Totaal aantal studies waarin deze uitkomstmaat wordt besproken.

Een extra bevinding in dit onderzoek is de relatie tussen het aantal jaren ervaring van de radioloog, het jaarlijks volume en de uitkomstmaten. In 14 studies wordt hierover een uitspraak gedaan [7,9,41,45-46,50-51,53-54,56-57,59,61-62]. In vijf studies wordt geen bewijs gevonden over een relatie tussen de ervaring van een radioloog en de uitkomstmaten [9,45-46,56,59]. De overige studies geven verschillende en tegenstrijdige conclusies over de aanwezigheid van een relatie, waarbij de werkervaring zowel een positieve als negatieve invloed heeft op de uitkomstmaten. Hier kan zodoende geen duidelijk uitspraak over worden gedaan.

3.3 Vermoeidheid, oogspanning en concentratie

In tien studies worden de uitkomsten vermoeidheid, oogspanning en concentratie onderzocht en het effect dat zij hebben op de prestatie van (screenings)radiologen en professionele computergebruikers [29-31,63-69].

In zeven studies wordt de oogspanning gemeten [29-31,64,67-69]. In deze studies wordt beschreven dat de oogspanning omhoog gaat gedurende een paar uur beoordelen. Radiologen hebben gedurende een werkdag meer moeite om te kunnen focussen. Uit twee studies blijkt dat radiologen na een werkdag een hogere oogspanning hebben dan de radiologen in opleiding [29,31]. Een algemene indicator voor een hoge oogspanning zijn lange werkuren. In verschillende studies wordt aanbevolen om pauzes te nemen. In drie studies wordt aanbevolen om elk uur een pauze te nemen [66,68-69]. Deze pauzes zijn van korte duur en duren ongeveer vijf tot tien minuten. Een goede nachtrust blijkt eveneens een positief effect te hebben op zowel oogspanning als de vermoeidheid [65].

Vijf studies doen onderzoek naar vermoeidheid. Na een werkdag of een paar uur beoordelen is de vermoeidheid hoger in vergelijking met het begin van een werkdag [29-31,64-65]. De tijd tussen de eerste en laatste metingen van vermoeidheid varieert tussen de 4 en 7 uur. Uit de studie van Maeda et al. [65] blijkt dat na 4 uur beoordelen per dag de vermoeidheid al toeslaat en Vertinsky et al. [69] concludeert dat 7 uur beoordelen per dag het maximum is. De andere drie studies variëren hier tussenin. Reed et al. [66] beschreven dat één beoordelingssessie gemiddeld 2 uur duurt, waarin ongeveer 200 onderzoeken worden beoordeeld. In één sessie wordt de aandacht en concentratie van een al radioloog minder [66].

Tabel 4 Data-extractietabel: Studies over volume, ervaring en nauwkeurigheid

Auteur, jaar [Ref]	Studie design	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal radiologen	Aantal mammo-grammen				
Barlow et al, 2004 [42]	Kwantitatief (survey/observatie-neel)	124	500 – 5000 p.j.	Vragenlijst over persoonskarakteristieken en data verzameling van 1 jaar mammogrammen.	- Hoger volume: sensitiviteit omhoog (P=0,004) en specificiteit omlaag (P=0,002). - Meer ervaring/ouder worden: sensitiviteit gaat omlaag en specificiteit gaat omhoog. - Recall ratio: hoger bij jongere, minder ervaren radiologen en hoger volume.	Nauwkeurigheid van interpretatie staat in relatie tot aantal jaren ervaring of jaarlijks volume.	8
Beam et al, 2003 [9]	Kwantitatief (survey/observatie-neel)	110	Gemiddeld 1912 p.j.	148 mammogrammen worden beoordeeld en een vragenlijst over persoonlijke en werkgegevens.	- Sterke variatie in sensitiviteit en specificiteit. - Volume of ervaring in mammografie heeft geen invloed op nauwkeurigheid, wel aantal jaren na afstuderen. Hoe langer het aantal jaren, hoe lager de nauwkeurigheid.	Radiologen met meer recente training hebben een betere nauwkeurigheid.	9
Buist et al, 2011 [43]	Kwantitatief (survey/observatie-neel)	120	1000-5000 p.j.	Vragenlijst over volume, berekening van screening focus, sensitiviteit, FPR en CDR.	- Gemiddeld is de sensitiviteit 85,2%, FPR 9,1% en CDR 4,2 op de 1000 mammogrammen. - Radiologen met het grootste volume per jaar hebben een lage screening focus en zijn over het algemeen vrouwen, ouder en hebben training gehad in het beoordelen van mammogrammen. - Sensitiviteit gaat omlaag bij een hoger screeningspercentage - Radiologen met een screening focus van >90% hebben een lagere sensitiviteit Radiologen met lage screening focus hebben hogere FPR. - CDR verschilde niet per volume. - Hoogste FPR en CDR bij radiologen met focus lager dan 80%.	- Hoger volume zorgt voor betere interpretatievermogen - Hoger volume gaat gepaard met lage FPR en RR, zonder dat de sensitiviteit of CDR omlaag gaat.	9
Buist et al, 2014 [44]	Kwantitatief (survey/observatie-neel)	96	-	Vragenlijst over demografische gegevens en jaarlijks volume. Sensitiviteit, FPR en CDR zijn gemeten.	- Gemiddeld is de sensitiviteit 85,3%, FPR 9,1% en CDR 4,1 op de 1000 mammogrammen. - Als radiologen hun eigen recall mammogrammen ook diagnostisch bekijken, gaat de sensitiviteit (P=0,039), FPR (P=0,004) en CDR (P=0,001) omhoog. - FPR gaat omlaag als het totale jaarlijks volume groter wordt, waarbij een stabilisatie optreedt bij een volume van 2000 per jaar.	Radiologen die ook hun eigen recall mammogrammen bekijken hebben een hogere sensitiviteit, CDR en FPR.	8
Carney et al, 2014 [60]	Kwantitatief (survey/observatie-neel)	119	Minimaal 50 p.w.	Radiologen beoordelen een testset (109 onderzoeken), hogere prevalentie. Beoordelingstijd wordt vastgelegd. Vragenlijst over demografische en klinische gegevens.	64% heeft meer dan 10 jaar ervaring. - Hoe langer een onderzoek met kanker werd bekeken, hoe groter de kans was op een terecht positieve beoordeling. - Langere beoordelingstijd zorgt voor een kleine stijging in sensitiviteit, maar verlaging in specificiteit. - Geen relatie tussen aantal jaren ervaring en duur van beoordelen.	Langere interpretatietijd zorgt voor een kleine stijging in sensitiviteit, maar een grotere kans op hogere FPR.	9
Coldman et al, 2006 [22]	Kwantitatief (survey/observatie-neel)	304	480 - ≥5000 p.j.	Over een periode van 3 jaar gegevens verzameld over screening volume, CDR en PPV.	- Het jaarlijks volume heeft geen effect op de CDR. - PVV gaat omhoog als het volume groter wordt, maar wordt stabiel bij een jaarlijks volume van 2000. - Gemiddeld gaat PVW 34% omhoog tot een volume van 2000 in vergelijking met een volume van 480-699 mammogrammen.	Volume heeft geen effect op CDR, wel op PVW.	8
Cornford et al, 2011 [45]	Kwantitatief (survey/observatie-neel)	37 (waarvan 16 laboranten)	Gemiddeld 6000 p.j.	Over een periode van 3 jaar gegevens verzameld over CDR, RR, PVW en gemiste kankers Verdeling in groepen o.b.v. jaarlijks volume.	- Geen bewijs dat ervaring invloed heeft op de uitkomsten. - Geen verschil gemiste kankers of PVW tussen de groepen - Gemiddelde CDR lager bij hoog volume (>8000) dan alle andere groepen bij elkaar. - Laagste RR in de hoge volume groep.	Hoog volume heeft geen invloed op gemiste kanker of PVW, maar zorgt wel voor lager CDR.	8

Auteur, jaar [Ref]	Studie design	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal radiologen	Aantal mammo-grammen				
Duncan et al, 2011 [46]	Kwantitatief (survey/observatie-neel)	37	<5000 - >8000 p.j.	Gegevens verzameld over periode van 3 jaar over RR, CDR, gemiste kankers. Verdeling in groepen op basis van jaarlijks volume.	- Er is geen relatie gevonden tussen het volume en de sensitiviteit of specificiteit. - Gemiddeld voor alle volumes: sensitiviteit 94%, specificiteit 95% en RR is 5%. - Het aantal jaren ervaring heeft geen invloed op sensitiviteit of specificiteit. - Geen verschillen in sensitiviteit en RR tussen de groepen.	Dit onderzoek is een reactie op Corford et al en levert geen bewijs dat er slechter wordt gepresteerd bij hoge volumes (>8000).	8
Elmore et al, 2002 [47]	Editorial	-	-	-	- Reflectie op het artikel van Esserman et al, 2002. - Zowel sensitiviteit als specificiteit gaat omhoog bij hogere volumes, al zijn de verschillen in specificiteit heel klein. - Het gebruik van testset is gevaarlijk, want dit is niet representatief voor de echte wereld.	-	3
Elmore et al, 2003 [61]	Editorial	-	-	-	- Het artikel van Beam et al geeft aanleiding tot vragen over de definities van een ervaren radioloog. - Er wordt afgevraagd of het gebruik van testset representatief is voor de echte wereld.	-	3
Elmore et al, 2009 [7]	Kwantitatief (survey/observatie-neel)	205	Gemiddeld 3131 p.j.	Vragenlijst over demografische en werkgerelateerde aspecten Gegevens verzameld over RR, FPR, PVW, CDR en sensitiviteit.	- Gemiddeld is RR 9,3%, FPR 8,9%, PVW 4,0% en sensitiviteit is 83,8%. - Geen relatie gevonden tussen volume en nauwkeurigheid. - Groot verschil in sensitiviteit onder radiologen met dezelfde FPR. - Radiologen met training hebben hogere sensitiviteit en FPR en is de enige factor met invloed op nauwkeurigheid. - Hogere RR en FPR bij vrouwelijke radiologen, radiologen met training en radiologen met <10 jaar ervaring.	Radiologen met training hebben hogere sensitiviteit en nauwkeurigheid, maar ook hogere RR en FPR.	9
Esserman et al, 2002 [48]	Kwantitatief (survey/observatie-neel)	59 U.S. 194 U.K.	<100 - >300 p.m.	Radiologen moeten 60 onderzoeken beoordelen (testset), hogere prevalentie. Verschil tussen UK en US radiologen. Sensitiviteit, specificiteit en totale nauwkeurigheid zijn gemeten.	- Gemiddelde sensitiviteit bij lage volumes US 70,3%, hoge volumes US 77,0% en hoge volumes UK 79,3%. - Gemiddelde specificiteit bij lage volumes US 83,6%, hoge volumes US en UK beide 88,0%. - Sensitiviteit is hoger bij hogere volumes. - Specificiteit van hoge volumes van US kwamen overeen met hoge volumes van UK. - Hoe hoger het volume, hoe beter de nauwkeurigheid. - Verschil in CDR: lage volume s US 71,5%, hoge volumes US 78,6% en hoge volumes UK 83,5%.	Hoe hoger het volume, hoe beter de nauwkeurigheid. (overeenkomst met Kan et al.).	9
Haneuse et al, 2012 [49]	Kwantitatief (survey/observatie-neel)	107	480-6000 p.j.	Gemaakte vragenlijst over demografische aspecten en ervaring in mammografie Metingen naar sensitiviteit, FPR en CDR.	- Radiologen met hoge volumes zijn over het algemeen vrouwen. - Sensitiviteit is bij alle jaarlijkse volumes gelijk . - CDR en FPR gaan omlaag als het jaarlijks volume groter wordt dan 4000, tot 4000 gaan ze omhoog.	Sensitiviteit blijft gelijk, maar CDR en FPR gaan omlaag als het volume groter wordt dan 4000 mammogrammen per jaar.	7

Auteur, jaar [Ref]	Studie design	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal radiologen	Aantal mammo-grammen				
Kan et al, 2000 [50]	Kwantitatief (survey/observatie-neel)	35	1000-5199 p.j.	Radiologen met min. 3 jaar ervaring zijn op basis van volume in groepen verdeeld. Gegevens verzameld over CDR en AIR.	- Het jaarlijks volume is niet gerelateerd aan aantal jaren ervaring - AIR is het laagst bij volumes van 3000-3999 per jaar. - CDR wordt hoger als volume ook groter wordt. - Radiologen met minder dan 2000 onderzoeken per jaar hebben een hogere AIR en een lagere CDR. - Radiologen met 4000 of meer hebben een hoge AIR, maar ook een hoge CDR.	Volgens dit onderzoek moet het jaarlijks volume een range hebben van 2000-4000 onderzoeken per jaar, hierin worden de beste resultaten behaald.	8
Miglioretti et al, 2007 [51]	Kwantitatief (survey/observatie-neel)	123	500->2000 p.j.	Vragenlijst over demografische en klinische aspecten. Gegevens verzameld over sensitiviteit en FPR.	- Gemiddeld is de sensitiviteit 79% en FPR 4,3%. - Volume heeft geen invloed op sensitiviteit of FPR. - Minder dan 10 jaar ervaring zorgt voor hogere FPR, maar sensitiviteit blijft gelijk.	Meer jaren ervaring zorgt voor een lagere FPR.	9
Miglioretti et al, 2009 [62]	Kwantitatief (survey/observatie-neel)	231	-	Vragenlijst over demografische aspecten. Gegevens verzameld over RR, FPR, PVW en sensitiviteit. Vergelijk in training en geen training.	- Met training: sensitiviteit wordt beter naarmate er meer ervaring is. Geen relatie met RR en FPR. - Zonder training: geen verband tussen ervaring en sensitiviteit. Naarmate er meer ervaring is, gaat de RR en FPR omlaag en de PVW omhoog.	Radiologen zonder training hebben een grotere leercurve dan radiologen met training.	9
Molins et al, 2008 [52]	Kwantitatief (survey/observatie-neel)	28	1900-13.000 p.j.	200 onderzoeken zijn beoordeeld (testset), hoge prevalentie. Verdeeld in groepen op basis van wel of niet regelmatig beoordelen. Vragenlijst over volume, ervaring.	- Gemiddeld is sensitiviteit 84% en specificiteit 64%. - Nauwkeurigheid is 56% t.o.v. 66% waarbij radiologen die regelmatig mammogrammen beoordelen hoger scoren. - Radiologen met hoger volume hebben hogere specificiteit en nauwkeurigheid. - Feedback zorgt voor lagere sensitiviteit en hogere specificiteit.	Radiologen met hoger volume hebben betere nauwkeurigheid en specificiteit.	7
Rawasdesch et al, 2013 [53]	Kwantitatief (survey/observatie-neel)	116	Gemiddeld 4000 p.j.	60 testonderzoeken, locatie sensitiviteit en specificiteit zijn gemeten, demografische aspecten zijn verzameld en radiologen zijn op basis van volume in groepen verdeeld.	- Sensitiviteit verschilde tussen de groepen niet. - Specificiteit is gerelateerd aan jaren ervaring en volume. - Radiologen met een jaarlijks volume van 5000 of meer presteren beter naarmate hun ervaring groter wordt. - Specificiteit wordt beter als volume groter wordt.	Het detecteren van een normaal onderzoek is belangrijk en is groter bij een hoger volume.	9
Reed et al, 2010 [54]	Kwantitatief (survey/observatie-neel)	69	<1000-20.000 p.j.	50 testonderzoeken, sensitiviteit en specificiteit zijn gemeten. Radiologen zijn op basis van jaarlijks volume in groepen verdeeld.	- Er is een positieve relatie tussen de prestatie van radiologen en het aantal jaren ervaring ($P=0,0001$) en aantal uur interpreteren per week ($P=0,003$). - Radiologen met volume van 2000 en meer scoren beter dan lagere volumes ($P=0,04$). - Specificiteit is vanaf 2000 onderzoeken stabiel.	Er is een sterke relatie tussen het volume en prestatie van radiologen, maar is van vele factoren afhankelijk.	8
Rickard et al, 2005 [55]	Kwantitatief (survey/observatie-neel)	133	270->5000 p.j.	Gedurende 2 jaar gegevens verzameld over CDR. Radiologen zijn in groepen verdeeld op basis van jaarlijks volume.	- CDR varieerde van 0-14,1 per 1000 onderzoeken. Grootste verschillen in lage volumes. - Gemiddeld CDR van 3,8 per 1000 onderzoeken. - CDR ging omhoog als het volume ook omhoog ging, maar werd stabiel na een jaarlijks volume van 1400 mammogrammen.	CDR is bij lage volume laag, maar wordt beter bij hogere volumes.	9
Smith-Bindman, 2005 [56]	Kwantitatief (survey/observatie-neel)	209	835-5000 p.j.	Gegevens verzameld over een periode van 5 jaar over jaarlijks volume, sensitiviteit, CDR en FPR.	- Gemiddeld is sensitiviteit 77% en FPR 10%. - Jaren ervaring en volume hebben geen relatie met sensitiviteit. - Specificiteit wordt beter tot een volume van 4000 per jaar. - FPR ging omlaag met 50% als radioloog ouder wordt, als het volume omhoog gaat en aantal jaren na afstuderen omhoog gaat. - CDR blijft bij elk volume gelijk.	Meest ervaren radiologen hebben lage FPR en een betere nauwkeurigheid.	9

Auteur, jaar [Ref]	Studie design	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal radiologen	Aantal mammo-grammen				
Tan et al, 2006 [57]	Kwantitatief (survey/observatie-neel)	1067	<200->560 p.j.	Gegevens verzameld over de FPR en de karakteristieken van de radiologen. Verdeeld in groepen op basis van leeftijd en volume.	- Gemiddeld is FPR 6,3%. - Radiologen die jong zijn, vrouw zijn en recente training hebben gehad hebben een hogere FPR. - Radiologen met een langere tijd na afstuderen hebben een lagere FPR. - Volume heeft geen invloed op de FPR.	Vrouwen en recent getrainde radiologen hebben een hoge FPR en ervaren radiologen een lage FPR (overeenkomst met Barlow en Smith).	9
Thèberge et al, 2005 [23]	Kwantitatief (survey/observatie-neel)	275	307-7386 p.j.	Over periode van 2 jaar zijn gegevens verzameld over het FPR en CDR.	- Volume van radioloog heeft geen invloed op sensitiviteit. - FPR gaat omlaag als volume van radioloog omhoog gaat (P=0,001). - CDR is hoger bij volume vanaf 4000 per jaar dan bij <2000.	Bij hoger volume gaat FPR omlaag en CDR omhoog.	9
Thèberge et al, 2013 [58]	Kwantitatief (survey/observatie-neel)	340	500-6000 p.j.	Gedurende 6 jaar gegevens verzameld over volume, FPR en sensitiviteit.	- Gemiddeld is sensitiviteit 86,7% en FRP is 9,3%. - Sensitiviteit bleef gelijk naarmate volume hoger werd. - Nauwkeurigheid wordt beter als volume omhoog gaat (P=0,0005) en heeft beste verbetering bij >3000. - Radiologen die minimaal 4000 onderzoeken per jaar beoordelen hebben een verbetering in nauwkeurigheid van 32% in vergelijking met een jaarlijks volume van 500 tot 1000 onderzoeken. - FPR ging omlaag als volume hoger werd (P=0,001).	Nauwkeurigheid wordt beter als volume omhoog gaat.	9
Timmers et al, 2014 [59]	Kwantitatief (survey/observatie-neel)	112	1000-60.000 p.j.	60 onderzoeken, (testset) hoge prevalentie. Gegevens verzameld over sensitiviteit en specificiteit. Vragenlijst over karakteristieken.	- Er is geen relatie gevonden tussen de sensitiviteit en de karakteristieken van de radiologen (jaarlijks volume, leeftijd, werkervaring in mammografie). 50% beoordeeld meer dan 10.000 mammogrammen per jaar. - Per week wordt er gemiddeld 8 uur (max. 40 uur) besteed aan het beoordelen van mammogrammen.	Geen relatie tussen sensitiviteit en karakteristieken.	9

P value = P, geeft de mate aan van een significant verschil tussen twee groepen of condities, waarbij $P < 0,5$ significant is; **RR** = Recall Rate of het verwijscijfer, alle onderzoeken die positief zijn beoordeeld ten opzichte van alle screeningsonderzoeken; **Sensitiviteit** = aantal positieven onderzoeken onder vrouwen die na 1 jaar de diagnose borstkanker hebben gekregen; **Specificiteit** = aantal negatieve onderzoeken onder vrouwen die na 1 jaar geen borstkanker bleken te hebben; **Screening focus** = percentage screening volume van het totale volume per jaar; **FPR** = fout positieven ratio, aantal positieven beoordelingen onder vrouwen die geen kanker bleken te hebben; **CDR** = kanker detectie ratio, aantal onderzoeken per 1.000 onderzoeken met kanker; **PVW** = positief voorspellende waarde, percentage van aantal positieve mammogrammen die ook daadwerkelijk borstkanker bevatten; **AIR** = abnormale interpretatie ratio of het verwijscijfer, zie definitie recall rate; **p.j.** = per jaar; **p.m.** = per maand; **p.w.** = per week.

Tabel 5 Data-extractietabel: Studies over vermoeidheid en oogspanning na lange werkdagen en computergebruik

Auteur, Jaar [Ref]	Studiedesign	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal deelnemers	Aantal uur/ aantal cases				
Cawson, 2009 [63]	Editorial	-	-	-	- Het artikel van Reed et al 2009 wordt besproken. - De lage prevalentie kan een negatief effect hebben op concentratie en aandacht van de lezer en kan leiden tot fouten. - In Australië wordt gemiddeld 12 uur per week gescreend. - Radiologen in artikel van Reed et al. ronden een leessessie van 50-200 studie af in 1-3 uur.	-	3
Ikushima et al, 2013 [64]	Kwantitatief (survey/ monitor)	17 radiologen	2,99 uur (1,42-5,00)	CFF metingen, 2x voormiddag en 2x namiddag. SSQ vragenlijst.	- Oogspanning ging omhoog als er langer de tijd werd genomen om een afbeelding te beoordelen. - Mentale vermoeidheid ging omlaag als omringende verlichting omhoog ging en als het aantal beoordeelde afbeeldingen omlaag ging. - Radiologen met bril hebben meer last van oogspanning, gevolgd door radiologen met contactlenzen.	Omringende verlichting moet omhoog, liever lenzen gebruiken i.p.v. een bril en het verlagen van aantal afbeeldingen, zorgt voor minder vermoeidheid bij radiologen, zowel mentaal als visueel.	8
Krupinski et al, 2009 [31]	Kwantitatief (monitor)	6 (3 radiologen/ 3 assistenten)	5,73 uur	Aan begin en einde van de werkdag metingen van de ogen op verschillende afstanden, van 20,25,66,40,50 cm en 61,91,122,152,183 cm van het oog af.	- Voor alle gemeten afstanden geldt dat lezen op kleine afstand slechter wordt. - Slechter in staat om te focussen op zowel dichte als verre afstanden. - Radiologen scoorden slechter dan assistenten.	Na een dag diagnostisch lezen treed visuele vermoeidheid op en is meer inspanning vereist om te kunnen focussen.	7
Krupinski et al, 2010 [29]	Kwantitatief (survey/ monitor)	20 (10 radiologen/ 10 assistenten)	60 cases	Oogspanning in de ochtend en namiddag gemeten. SOFI en SSQ vragenlijsten.	- Reading time: vroeg gemiddeld 52,1 sec en laat 51,5 sec. - Oogspanning is na een werkdag verhoogd, radiologen meer dan assistenten. - Op basis van SOFI blijkt vermoeidheid groter aan het eind van de dag. - Nauwkeurigheid gaat omlaag, verschil van 4%.	Na een werkdag hebben radiologen een verminderde mogelijkheid om te focussen, meer symptomen van oogspanning en vermoeidheid en een verminderde nauwkeurigheid.	7
Krupinski et al, 2012 [30]	Kwantitatief (survey/ monitor)	22 (11 radiologen/ 11 assistenten)	110 cases	Ogen gemeten aan het begin van de werkdag en aan het eind van de werkdag. SOFI vragenlijst is ingevuld.	- Oogspanning is in na een werkdag hoger. - Na een werkdag wordt er hoger gescoord bij de SOFI lijst op vermoeidheid. Assistenten scoren hoger dan radiologen. - Nauwkeurigheid is bij assistenten na een werkdag verlaagd, bij radiologen blijft dit hetzelfde.	Deelnemers hebben verhoogde symptomen van vermoeidheid en oogspanning. Assistenten hebben dit erger dan radiologen en bij hen is de nauwkeurigheid ook verlaagd.	7
Maeda et al, 2011 [65]	Kwantitatief (survey/ monitor)	7 radiologen (6 mannen/ 1 vrouw)	4 uur	CFFF elke dag, een maand lang, gemeten voor en na een beoordelingssessie. Vragenlijst over slaapgedrag, scorelijst vermoeidheid VAS.	- Na een leessessie gaat de vermoeidheid (subjectief) omhoog en gaat de CFFF omlaag. - Er is geen relatie gevonden tussen deze factoren. - Wel een relatie tussen slaapgedrag en CFFF, hoe korter de nacht ervoor geslapen is, hoe lager de CFFF blijkt te zijn.	Na 4 uur is de CFFF verlaagd en een hogere mate van vermoeidheid. Goede nachtrust zorgt voor beter visueel functioneren.	8

Auteur, jaar [Ref]	Studiedesign	Populatie (N=)		Methode	Resultaten	Conclusie	Kwaliteits beoordeling
		Aantal deelnemers	Aantal uur / aantal cases				
Reed et al, 2009 [66]	Kwantitatief (survey)	83 radiologen	4 uur	Vragenlijst over werkgerelateerde aspecten, zoals het jaarlijks volume, ervaring, werktijd en mate van concentratie.	- Gemiddeld wordt er 2 min naar een foto gekeken. - Er wordt 2 uur gedaan over één sessie, deze bevat 200-300 onderzoeken. - Geen relatie tussen interpretatietijd en volume. - Concentratie en aandacht verminderen gedurende een sessie. - Nemen van pauze is beste manier om concentratie en aandacht te verbeteren.	Er moet onderzoek worden gedaan naar beste omstandigheden en aantal mammogrammen voor optimale concentratie. Nemen van een pauze van 5-15 min na 45 min of 100 onderzoeken.	8
Safdar et al, 2009 [67]	Kwantitatief (survey/ monitor)	48 radiologen (33 mannen/ 14 vrouwen)	-	Gedurende een jaar elke dag 3 metingen van de ogen. Leeskaart op afstand van 81,3 cm. Vragenlijst over ogen.	- Vrouwen hadden een oogscherpte van 20/16 en mannen van 20/15. - Contactlenzen hadden geen effect. - In de ochtend was de scherpte lager dan in de (na)middag, maar is een klein verschil. - Gemiddelde scherpte van 20/15, maar laagste meting 20/30 en hoogste meting 20/10. - Er is geen relatie tussen aantal uren slaap en oogscherpte gevonden.	Hoewel er een verschil in oogscherpte was in de ochtend en de (na)middag, is dit een te klein verschil om van enige betekenis te zijn. De waardes zijn normaal.	8
Toomingas et al, 2014 [68]	Kwantitatief (survey)	1246 (498 mannen/ 785 vrouwen)	-	Aan het begin is een vragenlijst ingevuld die geldt als basis. Gedurende een jaar elke maand dezelfde vragenlijst ingevuld.	- Als vrouw heb je een hoger risico om problemen aan je ogen te krijgen door langdurig computergebruik. 3 uur of meer achter de computer zorgt voor meer problemen aan je ogen, dit geldt ook voor 1,5 intensief computergebruik. - Slechte verlichting of weerspiegeling in het scherm zorgen ook voor oogproblemen.	Professionele computer gebruikers hebben verhoogde kansen op oogproblemen. Regelmatige pauzes en verbeterde werkcondities zullen de risico's verlagen.	8
Verinsky et al, 2005 [69]	Kwantitatief (survey)	380 radiologen	-	Vragenlijst via het internet. Meten van oogspanning door een 5-puntenschaal in de vragenlijst.	- Oogspanning werd niet beïnvloed door bekijken van digitale of analoge beelden. - Sterke indicatoren voor oogspanning zijn: vrouwelijk geslacht, jonge leeftijd, lange werkuren. - Oogspanning werd hoger als langer dan 6 uur werd beoordeeld. - Elk uur een korte pauze zorgt voor lagere oogspanning. - Gebruik lenzen heeft geen relatie met oogspanning.	Niet meer dan 7 uur per dag beoordelen van radiologische beelden. Elk uur een korte pauze nemen zorgt voor gunstig effect op de oogspanning.	9

CFF/CFFF = Critical (Flicker) Fusion Frequency, de frequentie waarbij een knipperend licht of beeld overgaat naar een continue lichtbron. Het wordt gebruikt als maat voor de vermoeidheid, een lage C(F)FF is een hoge mate van vermoeidheid; **SSQ** = Simulator Sickness Questionnaire, zeven vragen over de oogspanning en vermoeidheid; **SOFI** = Swedish Occupational Fatigue Inventory, bevat 20 expressies over de fysieke aspecten van vermoeidheid; **VAS** = Visual Analog Score, schaalverdeling voor de mate van vermoeidheid (subjectief).

4. Discussie en conclusie

In dit onderzoek is aan de hand van bestaande wetenschappelijk literatuur onderzocht of kan worden vastgesteld wat het maximaal aantal uur is dat een screeningsradioloog per dag mammogrammen mag beoordelen. Hierbij is de invloed van het jaarlijks volume van de radioloog en het effect van vermoeidheid en oogspanning op de nauwkeurigheid in het beoordelen van mammogrammen onderzocht. In totaal zijn 35 studies geïnccludeerd. Deze studies zijn beoordeeld op methodologische kwaliteit. Op drie studies na, scoren alle studies een goede kwaliteit. Uit 25 studies worden tegenstrijdige conclusies verkregen over het effect van het jaarlijks volume van radiologen op de verschillende uitkomstmaten. Zo geven bijvoorbeeld 11 van de 16 studies aan dat een hoger volume geen effect heeft op de sensitiviteit, maar dit wordt niet door alle studies ondervonden. Deze tegenstrijdigheden gelden eveneens voor de specificiteit, kankerdetectiecijfer, het fout positieven cijfer, de positief voorspellende waarde en het verwijscijfer. Tien studies geven informatie over de vermoeidheid en oogspanning gedurende een werkdag. Lange werkuren zorgen voor een toename in oogspanning en vermoeidheid, waardoor de nauwkeurigheid in het beoordelen van mammogrammen gedurende de werkdag omlaag gaat. In dit onderzoek komt naar voren dat maximaal 4 tot 7 uur op een dag leidt tot toename in oogspanning en vermoeidheid. Elk uur een pauze van 5 tot 10 minuten leidt tot minimale oogspanning en vermoeidheid.

De geïnccludeerde studies zijn in verschillende landen uitgevoerd, zoals de Verenigde Staten (US), Australië en het Verenigd Koninkrijk (UK). Het minimum voor het jaarlijks volume verschilt tussen deze landen [14,22]. Australië hanteert een minimum van 2.000 onderzoeken per jaar en de UK van 5.000 onderzoeken per jaar [14-15,22]. Dit is vergelijkbaar met het minimum van 3.000 onderzoeken per jaar in Nederland [14-15,22]. Afgaande op deze aantallen lijken de studies uit Australië en de UK het meest toepasbaar op de Nederlandse screeningssituatie. Dit geldt niet voor de studies uit de US waar een veel lager minimum wordt gehanteerd. Esserman et al. [48] en Suleiman et al. [70] maken een vergelijking tussen screeningsradiologen uit respectievelijk de US en de UK en Australië en de US. Zij concluderen dat radiologen met een hoog jaarlijks volume beter presteren dan radiologen met een laag jaarlijks volume. Deze vergelijking werd door Elmore et al. [47] aangemoedigd, omdat zo duidelijk wordt waarom verschillen in jaarlijks volume bestaan. In een literatuurstudie van Moss et al. [71] wordt voor dit verschil als reden gegeven dat de prestatie en nauwkeurigheid in beoordelen van de radiologen uit de US meer gerelateerd is aan het aantal jaar ervaring dan aan het jaarlijks volume. In huidig onderzoek hebben de radiologen in de meeste studies een volume van maximaal 5.000 onderzoeken per jaar. In een aantal studies stijgt dit aantal tot wel 20.000 of 60.000 onderzoeken per jaar [54,59]. In Nederland zijn er radiologen die veel onderzoeken per jaar beoordelen, oplopend tot ruim 100.000 onderzoeken [21, Persoonlijke communicatie T. Geerste, QA bij LRCB]. De meeste studies in dit onderzoek hebben een jaarlijks volume dat niet overeenkomt met dat van Nederland. De literatuurstudie van Moss et al. [71] beschrijft echter dat weinig bewijs is gevonden voor een relatie tussen het jaarlijks volume en de nauwkeurigheid van het beoordelen bij volumes hoger dan 5.000

onderzoeken per jaar. In huidig onderzoek geven een aantal studies aan dat vanaf een volume van ongeveer 3.000 onderzoeken per jaar de nauwkeurigheid hetzelfde blijft in vergelijking met lagere volumes [49-50,53-56,58]. Hoewel de studies uitgevoerd in Australië en de UK het dichtst bij de Nederlandse screeningssetting lijken te komen, is het lastig om met deze gegevens een vergelijking tussen de landen te maken.

Verschillende studies in dit onderzoek geven aan dat na een aantal uur vermoeidheid optreedt gedurende een werkdag. Echter is er een spreiding in het maximaal aantal uur dat in deze studies worden aan bevolen. Zo geeft Maeda et al. [65] een maximum van 4 uur en Vertinsky et al. [69] een maximum van 7 uur. Dit geeft een aanzienlijk verschil aan van 3 uur gedurende één werkdag. Hoewel duidelijk wordt dat gedurende een werkdag vermoeidheid optreedt en dit de diagnostische nauwkeurigheid verlaagd, is het maximaal aantal uur dat gewerkt kan worden op één dag niet eenduidig. Voor meer duidelijkheid hierover in de Nederlandse screeningssetting zou een advies op basis van expert opinions geformuleerd kunnen worden. Daarentegen zijn de studies wel duidelijk in de aanbeveling voor het nemen van pauzes gedurende een werkdag. Om vermoeidheid en oogspanning te verminderen wordt aanbevolen om ieder uur een pauze te nemen van 5 tot 10 minuten. In een literatuurstudie van Kagadis et al. [72] wordt aangegeven dat het nemen van een pauze na invoering van de digitalisering in de radiologie belangrijker is geworden. Voorheen konden radiologen een pauze nemen omdat nieuwe onderzoeken aan de lichtkasten moesten worden bevestigd [72]. Na de digitalisering zijn deze pauzes weggefallen, waardoor het belangrijk is om deze pauzes zelf in te plannen op een werkdag. In de studies van Kagadis et al. [72] en Vertinsky et al. [69] worden de aanbevelingen gedaan om de 20-20-20 regel aan te houden, bekend in de computerliteratuur. Hierbij wordt elke 20 minuten, 20 voet (6 meter) ver weggekeken gedurende 20 seconden, waardoor de oogspieren kunnen ontspannen en er minder kans is op een verhoogde oogspanning en vermoeidheid [72]. Screeningsradiologen dienen bewust te zijn van het feit dat de oogspanning en vermoeidheid verminderd kunnen worden door het regelmatig nemen van pauzes gedurende de werkdag.

In dit onderzoek hebben zeven studies testsets gebruikt om tot hun resultaten te komen [9,48,52-54,59-60]. Elmore et al. [47,61] bediscussiëren in hun studies het effect van testsets. In de testsets is de prevalentie veel hoger dan in een werkelijke screeningssetting. Zoals Elmore et al. [61] aangeven zijn deze testsets geen realistische weergave van de werkelijkheid. Recent is onderzoek gedaan naar het gebruik van testsets. Soh et al. [73] beschreven dat de sensitiviteit in testsets enigszins hoger is dan in een klinische situatie, maar dat testsets zeker gebruikt kunnen worden. In de studie van Gur et al. [74] wordt echter beschreven dat de sensitiviteit in een klinische situatie hoger is dan in een testset. Volgens hen is een testset niet representatief voor de werkelijke screeningssetting. De gemeten uitkomsten in huidig onderzoek zouden kunnen verschillen met de werkelijke screeningssetting. Hoewel de voorkeur uitgaat naar studies uitgevoerd in de werkelijke screeningssetting [47,61,74], wordt het gebruik van testsets niet geheel afgeraden [73].

Dit onderzoek bevat een aantal sterke punten. Voorafgaand aan het uitvoeren van dit onderzoek heeft een uitgebreid vooronderzoek plaatsgevonden, waarin is nagedacht over de vele aspecten waarmee een screeningsradioloog te maken heeft tijdens het beoordelen van mammogrammen. Gebaseerd op dit vooronderzoek is een uitgebreide zoekstrategie gecreëerd, waardoor de conclusie voor dit onderzoek onderbouwd wordt door een ruim aantal studies. Door de uitgebreide zoekstrategie kan een onderscheid worden gemaakt in de studies die zijn geïnccludeerd. Gekozen is om de onderzoeksvraag van twee kanten te benaderen. In dit onderzoek is onderscheid gemaakt tussen het aantal onderzoeken dat een screeningsradioloog per jaar beoordeeld en de vermoeidheid gedurende een werkdag. Zoals uit de studies naar voren komt, wordt de vermoeidheid gedurende een werkdag groter, waardoor focus en concentratie verminderd. Vermoeidheid is hierdoor een goede indicator voor het bepalen van het aantal uur dat een screeningsradioloog per dag mag werken. Het jaarlijks volume van een radioloog is een optelsom van het maximaal aantal onderzoeken dat een radioloog op een dag beoordeeld. Hoewel al vele studies zijn gedaan naar het minimale aantal onderzoeken dat een screeningsradioloog per jaar mag beoordelen, is een maximum nog niet opgesteld. Dit onderzoek is het eerste onderzoek dat hier een norm voor probeert op te stellen. Doordat hier nog geen informatie over bekend is, kan dit onderzoek de eerste aanzet zijn om hier verder onderzoek naar te doen.

Desondanks bevat dit onderzoek eveneens een aantal beperkingen. De zoekstrategie is opgebouwd naar aanleiding van de gevonden Pearl-artikelen. Mogelijk is hierdoor de zoekstrategie te specifiek gericht is op de Pearl-artikelen en dus andere relevante studies niet gevonden zijn. Een voorbeeld hiervan het is het artikel van Smith-Bindman [56]. Deze studie is in dit onderzoek meegenomen middels de referenties, maar is niet door middel van de zoekstrategie gevonden. Ten tweede zijn de kwaliteitseisen voor het bepalen van de methodologische kwaliteit door de onderzoeker zelf opgesteld. Deze lijst is echter niet gevalideerd in de praktijk. De kans bestaat dat de kwaliteitsbeoordeling hierdoor niet geheel betrouwbaar is. Verder is dit onderzoek door één onderzoeker uitgevoerd, waardoor de kans bestaat dat er sprake is van subjectiviteit gedurende het uitvoeren van het onderzoek. Een laatste beperking van dit onderzoek is dat er geen studies zijn gevonden die zijn uitgevoerd in de Nederlandse screeningssetting. De geïnccludeerde studies zijn voornamelijk uitgevoerd in de UK, US en Australië en zijn lastig te vergelijken met Nederland door een verschil in het jaarlijks volume van radiologen.

Op basis van deze literatuurstudie worden de volgende aanbevelingen gedaan voor een verder vervolgonderzoek. Vanuit de resultaten kan geen duidelijke conclusie worden gegeven over het effect van het jaarlijks volume van een screeningsradioloog op de diagnostische nauwkeurigheid. Meer onderzoek is hiernaar nodig om te bepalen in welke mate het jaarlijks volume invloed heeft. Het is van belang dat het onderzoek in de Nederlandse screening wordt uitgevoerd, aangezien in huidig onderzoek veelal studies zijn gebruikt waarin het jaarlijks volume lager is dan in Nederland. Eveneens zal verder onderzoek gedaan moeten worden naar de invloed van de werkervaring van screeningsradiologen op de diagnostische nauwkeurigheid. Uit dit onderzoek blijkt dat de werkervaring een relatie heeft met de diagnostische nauwkeurigheid. Door de vele tegenstrijdige

conclusies wordt echter niet duidelijk wat deze relatie is. Meer onderzoek is nodig zijn naar het aantal uur dat een radioloog op een dag mag beoordelen. Uit dit onderzoek komt naar voren dat er een range is van 4 tot 7 uur. Hierop kan echter geen concrete norm worden vastgesteld, aangezien deze range tamelijk groot is.

In dit onderzoek is onderzoek gedaan naar de fysieke vermoeidheid van radiologen gedurende een werkdag. In een studie van Krupinski en Reiner [75] wordt echter beschreven dat ook cognitieve vermoeidheid een grote oorzaak is van het maken van fouten bij het stellen van een diagnose. Verder onderzoek is nodig om te bepalen in welke mate cognitieve vermoeidheid invloed heeft op de diagnostische nauwkeurigheid van een radioloog gedurende een werkdag.

Eveneens kan verder onderzoek gedaan worden naar de prestaties van radiologen gedurende een beoordelingssessie. Taylor-Phillips et al. [76] hebben hier recent een literatuurstudie naar gedaan. In deze studie is geconcludeerd dat er tijdens een beoordelingssessie veranderingen zijn in gedrag en prestatie van een (screenings)radioloog [76]. Meer onderzoek moet gedaan worden naar wat deze veranderingen precies inhouden, waarbij meer deelnemers verzameld moeten worden en ook de werkervaring van de radioloog als factor van invloed moet worden meegenomen [76].

Op basis van deze literatuurstudie kan geconcludeerd worden dat er niet duidelijk kan worden vastgesteld wat het maximaal aantal uur per dag is dat een screeningsradioloog mag beoordelen. Uit dit onderzoek komt naar voren dat gedurende een werkdag van 4 tot 7 uur, vermoeidheid en oogspanning verhoogd wordt, wat leidt tot een slechtere nauwkeurigheid in het beoordelen van mammogrammen. Door het nemen van pauzes, elk uur gedurende 5 tot 10 minuten, wordt de concentratie verbeterd en de vermoeidheid en oogspanning verminderd. Verder is in dit onderzoek het effect van het jaarlijks volume op de nauwkeurigheid in beoordelen onderzocht. Vele studies doen een uitspraak over het effect van het jaarlijks volume op de verschillende uitkomstmaten. Door de tegenstrijdige conclusies uit deze studies kan geen eenduidige conclusie worden getrokken. Om een aanbeveling te kunnen doen naar het maximaal aantal uur dat op een dag mag worden gescreend is meer onderzoek nodig en kan gebruik worden gemaakt van meningen van experts. In een vervolg onderzoek zullen verschillende aspecten als werkervaring en het aantal onderzoeken dat op één dag door een screeningsradioloog kan worden beoordeeld moeten worden meegenomen.

Literatuur

1. Fracheboud J, Sankatsing VDV, Ripping TM, Broeders MJM, Otten JDM, Ineveld BM van, et al. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland 1990 – 2011/2012. LECB, 2014 Jan.
2. Integraal Kankercentrum Nederland. Meest voorkomende soorten kanker [Internet]. 2014. [Cited 19 Feb 2015]. Available from: <http://www.cijfersoverkanker.nl/meest-voorkomende-soorten-52.html>.
3. Waal D van der, Verbeek AL, Ripping TM, Broeders MJ. Breast Cancer Diagnosis and death in the Netherlands: a changing burden. The European Journal of Public Health. 2014 Jun 27.
4. Borstkankervereniging Nederland. Risicofactoren [Internet]. [Updated 2015 Jan 4; cited 2015 Feb 24]. Available from: http://www.borstkanker.nl/wat_is_borstkanker_risicofactoren.
5. Rijksinstituut voor Volksgezondheid en Milieu. Kans op borstkanker [Internet]. [Cited 2015 Feb 24]. Available from: http://www.rivm.nl/Onderwerpen/B/Bevolkingsonderzoek_borstkanker/Wat_is_borstkanker/Kans_op_borstkanker.
6. Gezondheidsraad: Het nut van bevolkingsonderzoek naar borstkanker [Internet] Den Haag: Gezondheidsraad, 2002; publicatie nr. 2002/03. [Cited 25-01-2015]. Available from: <http://www.gezondheidsraad.nl/sites/default/files/0203N.pdf>.
7. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. Radiology. 2009;253(3):641-651.
8. Hensen J, Jaarsveld-Bolman K, Dam T, Dol-Jansen J, Geers S. Radiologie Techniek en onderzoek. Eerste druk. Amsterdam: Reed Business; 2011.
9. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. Journal of the National Cancer Institute. 2003;95(4):282-290.
10. Tria Tirona M, Breast Cancer Screening Update. American Family Physician, 2013;87(4):274-278.
11. Jong N de. Uitvoeringskader bevolkingsonderzoek Borstkanker. Rijksinstituut voor Volksgezondheid en Milieu (RIVM) 2014.
12. Gezondheidsraad. Wet bevolkingsonderzoek: digitale screeningsmammografie [Internet]. Den Haag: Gezondheidsraad, 2006; publicatie nr. 2006/05WBO. [Cited 25-01-2015]. Available from: <http://www.gezondheidsraad.nl/sites/default/files/200605WBO.pdf>.
13. Gezondheidsraad. Bevolkingsonderzoek naar borstkanker: verwachtingen en ontwikkelingen. Den Haag; Gezondheidsraad, 2014; publicatie nr. 2014/01.
14. Perry N, Broeders M, Wolf C de, Törnberg S, Holland R, Karsa L von. European Guideline for quality assurance in breast cancer screening and diagnosis, fourth edition. European Communities, 2006.
15. Landelijk Referentiecentrum voor bevolkingsonderzoek. Kwaliteitsregister screeningsradiologen in het bevolkingsonderzoek op borstkanker in Nederland [Internet]. [Update 15-10-2010; Cited 15-02-2015]. Available from: http://www.lrcb.nl/uploads/userfiles/files/kwaliteitsregister_screeningsradiologen.pdf.

16. Landelijk Referentiecentrum voor Bevolkingsonderzoek. Visitatieprotocol [Internet]. Nijmegen; April 2013 [Cited 28-03-2015]. Available from: http://www.lrcb.nl/uploads/userfiles/files/visitatieprotocol%20LRCB_versie%20april%202013_DEFINITIEF.pdf.
17. Landelijk Referentiecentrum voor Bevolkingsonderzoek. Screeningradiologen [Internet]. 2011 [Cited 28-03-2015]. Available from: <http://www.lrcb.nl/nl/page/screeningsradiologen>.
18. Landelijk Referentiecentrum voor Bevolkingsonderzoek. Audit [Internet]. 2011 [Cited 28-03-2015]. Available from: <http://www.lrcb.nl/nl/page/audit>.
19. Landelijk Referentiecentrum voor Bevolkingsonderzoek. Achtdaagse Opleiding [Internet]. 2011 [Cited 28-03-2015]. Available from: <http://www.lrcb.nl/nl/page/achtdaagse-opleiding>.
20. Landelijk Referentiecentrum voor Bevolkingsonderzoek. Bij- en nascholing [Internet]. 2011 [Cited 28-03-2015]. Available from: <http://www.lrcb.nl/nl/page/bij-en-nascholing>.
21. Otten JDM, Karssemeijer N, Hendriks JHCL, Groenewoud JH, Fracheboud J, Verbeek ALM et al. Effect of Recall Rate on Earlier Screen Detection of Breast Cancer Based on the Dutch Performance Indicators. *Journal of the National Cancer Institute*. 2005;97(10):748-754.
22. Coldman AJ, Major D, Doyle GP, D'yachkova Y, Phillips N, Onysko J, et al. Organized breast screening programs in Canada: effect of radiologist reading volumes on outcomes. *Radiology*. 2006;238(3):809-815.
23. Théberge I, Hébert-Croteau N, Langlois A, Major D, Brisson J. Volume of screening mammography and performance in the Quebec population-based Breast Cancer Screening Program. *Canadian Medical Association Journal*. 2005;172(2):195-199.
24. Samei E, Krupinski EA. *The handbook of Medical Image Perception and Techniques*. New York: Cambridge University Press; 2010.
25. Gur D, Sumkin JH, Rockette HE, Ganott M, Hakim C, Hardesty L, et al. Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System. *Journal of the National Cancer Institute*. 2004;96(3):185-190.
26. Cain MS, Biggs AT, Darling EF, Mitroff SR. A little bit of history repeating: splitting up multiple-target visual searches decreases second-target miss errors. *Journal of Experimental Psychology*. 2014;20(2):112-125.
27. Biggs AT, Adamo SH, Mitroff SR. Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica*. 2014;152:158-165.
28. Buntinx F. Screening versus diagnostiek: complexe problemen [Internet]. [Published April 2009, Cited 17-03-2015]. Available from: <http://www.henw.org/archief/volledig/id2992-screening-versus-diagnostiek-complexe-problemen.html>.
29. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long Radiology workdays Reduce Detection and Accommodation Accuracy. *Journal of the American College of Radiology*. 2010; 7(9):698-704.
30. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Madsen MT, Kramer DJ. Do Long Radiology Workdays Impact Nodule Detection in Dynamic CT Interpretation? *Journal of the American College of Radiology*. 2012;9(3):191-198.
31. Krupinski EA, Berbaum S. Measurement of Visual Strain in Radiologists. *Academic Radiology*. 2009; 16(8):947-950.

32. Berbaum KS. Satisfaction of search experiments in advanced imaging. Proc. SPIE 8291, Human Vision and Electronic Imaging XVII. 2012 Feb 9.
33. Walaeus Bibliotheek. Handleiding Pubmed [Internet]. [Update Mei 2012, Cited 07-03-2015]. Available from: <https://www.lumc.nl/sub/1060/att/908061147582339/1205160105042341.pdf>.
34. Ramer SL. Site-ation pearl growing: methods and librarianship history and theory. Journal of the Medical Library Association. 2005; 93(3):397-400.
35. Zoetelief J, Faulkner K. Equipment requirements and specification for digital and interventional radiology. Radiation Protection Dosimetry. 2001;94(1-2):43-48.
36. Dutch Cochrane Centre. Beoordeling van de kwaliteit van een onderzoek naar de waarde van een diagnostische test [Internet]. [Cited: 12-03-2015]. Available from: <http://dcc.cochrane.org/sites/dcc.cochrane.org/files/uploads/4.2%20Checklist%20Beoordeling%20diagnostisch%20onderzoek.pdf>.
37. Dutch Cochrane Centre. Beoordeling van de kwaliteit van een artikel over prognose [Internet]. [Cited 12-03-2015]. Available from: <http://dcc.cochrane.org/sites/dcc.cochrane.org/files/uploads/4.3%20Checklist%20Beoordeling%20van%20de%20kwaliteit%20van%20een%20artikel%20over%20prognose.pdf>.
38. Critical Appraisal Skills Programme. 12 questions to help you make sense of a cohort study [Internet]. [Cited 04-03-2015]. Available from: <http://calder.med.miami.edu/portals/ebmfiles/UM%20CASP%20Case-Controls%20Assessment%20Tool.pdf>.
39. Wouters E, Zaal Y van, Bruijning J. Praktijkgericht onderzoek in de (para)medische zorg. Tweede druk. Bussum: Uitgeverij Coutinho; 2015.
40. KNGF Evidence Based Products. A.9.3. Datasynthese, Aanbevelingen voor de interventies. [Internet]. [Cited 11-03-2015]. Available from: <https://www.fysionet-evidencebased.nl/index.php/richtlijnen/richtlijnen/beroerte-2014/verantwoording-en-toelichting/a-inleiding/a-9-3-datasynthese>.
41. Centrale Commissie Mensgebonden Onderzoek. Uw onderzoek, WMO plichtig of niet? [Internet]. [Cited 11-03-2015]. Available from: <http://www.ccmo.nl/nl/uw-onderzoek-wmo-pliktig-of-niet>.
42. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. Journal of the National Cancer Institute. 2004;96(24):1840-50.
43. Buist DS, Anderson ML, Haneuse SJ, Sickles EA, Smith RA, Carney PA, et al. Influence of annual interpretive volume on screening mammography performance in the United States. Radiology. 2011;259(1):72-84.
44. Buist DS, Anderson ML, Smith RA, Carney PA, Miglioretti DL, Monsees BS, et al. Effect of radiologists' diagnostic work-up volume on interpretive performance. Radiology. 2014;273(2):351-64.
45. Cornford E, Reed J, Murphy A, Bennett R, Evans A. Optimal screening mammography reading volumes; evidence from real life in the East Midlands region of the NHS Breast Screening Programme. Clinical radiology. 2011;66(2):103-7.
46. Duncan KA, Scott NW. Is film-reading performance related to the number of films read? The Scottish experience. Clinical radiology. 2011;66(2):99-102.
47. Elmore JG, Carney PA. Does practice make perfect when interpreting mammography? Journal of the National Cancer Institute. 2002;94(5):321-3.

48. Esserman L, Cowley H, Eberle C, Kirkpatrick A, Chang S, Berbaum K, et al. Improving the accuracy of mammography: volume and outcome relationships. *Journal of the National Cancer Institute*. 2002;94(5):369-75.
49. Haneuse S, Buist DS, Miglioretti DL, Anderson ML, Carney PA, Onega T, et al. Mammographic interpretive volume and diagnostic mammogram interpretation performance in community practice. *Radiology*. 2012;262(1):69-79.
50. Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology*. 2000;215(2):563-7.
51. Miglioretti DL, Smith-Bindman R, Abraham L, Brenner RJ, Carney PA, Bowles EJ, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *Journal of the National Cancer Institute*. 2007;99(24):1854-63.
52. Molins E, Macia F, Ferrer F, Maristany MT, Castells X. Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC health services research*. 2008;8:91.
53. Rawashdeh MA, Lee WB, Bourne RM, Ryan EA, Pietrzyk MW, Reed WM, et al. Markers of good performance in mammography depend on number of annual readings. *Radiology*. 2013;269(1):61-7.
54. Reed WM, Lee WB, Cawson JN, Brennan PC. Malignancy detection in digital mammograms: important reader characteristics and required case numbers. *Academic radiology*. 2010;17(11):1409-13.
55. Rickard M, Taylor R, Page A, Estoesta J. Cancer detection and mammogram volume of radiologists in a population-based screening programme. *Breast (Edinburgh, Scotland)*. 2006;15(1):39-43.
56. Smith-Bindman R, Chu P, Miglioretti DL, Quale C, Rosenberg RD, Cutter G, et al. Physician predictors of mammographic accuracy. *Journal of the National Cancer Institute*. 2005;97(5):358-67.
57. Tan A, Freeman DH, Jr., Goodwin JS, Freeman JL. Variation in false-positive rates of mammography reading among 1067 radiologists: a population-based assessment. *Breast cancer research and treatment*. 2006;100(3):309-18.
58. Theberge I, Chang SL, Vandal N, Daigle JM, Guertin MH, Pelletier E, et al. Radiologist interpretive volume and breast cancer screening accuracy in a Canadian organized screening program. *Journal of the National Cancer Institute*. 2014;106(3):djt461.
59. Timmers JM, Verbeek AL, Pijnappel RM, Broeders MJ, den Heeten GJ. Experiences with a self-test for Dutch breast screening radiologists: lessons learnt. *European radiology*. 2014;24(2):294-304.
60. Carney PA, Bogart TA, Geller BM, Haneuse S, Kerlikowske K, Buist DS, et al. Association between time spent interpreting, level of confidence, and accuracy of screening mammography. *AJR American journal of roentgenology*. 2012;198(4):970-8.
61. Elmore JG, Miglioretti DL, Carney PA. Does practice make perfect when interpreting mammography? Part II. *Journal of the National Cancer Institute*. 2003;95(4):250-2.
62. Miglioretti DL, Gard CC, Carney PA, Onega TL, Buist DS, Sickles EA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology*. 2009;253(3):632-40.

63. Cawson J. Reader practice in mammography screen reading in Australia. *Journal of medical imaging and radiation oncology*. 2009;53(6):514-5.
64. Ikushima Y, Yabuuchi H, Morishita J, Honda H. Analysis of dominant factors affecting fatigue caused by soft-copy reading. *Academic radiology*. 2013;20(11):1448-56.
65. Maeda E, Yoshikawa T, Hayashi N, Akai H, Hanaoka S, Sasaki H, et al. Radiology reading-caused fatigue and measurement of eye strain with critical flicker fusion frequency. *Japanese journal of radiology*. 2011;29(7):483-7.
66. Reed W, Poulos A, Rickard M, Brennan P. Reader practice in mammography screen reporting in Australia. *Journal of medical imaging and radiation oncology*. 2009;53(6):530-7.
67. Safdar NM, Siddiqui KM, Qureshi F, Mirza MK, Knight N, Nagy P, et al. Vision and quality in the digital imaging environment: how much does the visual acuity of radiologists vary at an intermediate distance? *AJR American journal of roentgenology*. 2009;192(6):W335-40.
68. Toomingas A, Hagberg M, Heiden M, Richter H, Westergren KE, Tornqvist EW. Risk factors, incidence and persistence of symptoms from the eyes among professional computer users. *Work* (Reading, Mass). 2014;47(3):291-301.
69. Vertinsky T, Forster B. Prevalence of eye strain among radiologists: influence of viewing variables on symptoms. *AJR American journal of roentgenology*. 2005;184(2):681-6.
70. Suleiman WI, Lewis SJ, Georgian-Smith D, Evanoff MG, McEntee MF. Number of mammography cases read per year is a strong predictor of sensitivity. *Journal of Medical Imaging*. 2014 SPIE; 1(1).
71. Moss SM, Blanks RG, Bennett RL. Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature. *Clinical Radiology*. 2005;60:623-626.
72. Kagadis GC, Walz-Flannigan A, Krupinski EA, Nagy PG, Katsanos K, Diamantopoulos A, et al. Medical imaging displays and their use in image interpretation. *Radiographics : a review publication of the Radiological Society of North America, Inc*. 2013;33(1):275-90.
73. Soh BP, Lee W, McEntee MF, Kench PL, Reed WM, Heard R, et al. Screening Mammography: Test Set Data Can Reasonably Describe Actual Clinical Reporting. *Radiology*. 2013;268(1):46-53.
74. Gur D, Bandos AI, Cohen CS, Hakim CM, Hardesty LA, Ganott MA, et al. The "Laboratory" Effect: Comparing Radiologists' Performance and Variability during Prospective Clinical and Laboratory Mammography Interpretations. *Radiology*. 2008;249(1):47-53.
75. Krupinski EA, Reiner BI. Real-time occupational Stress and Fatigue Measurement in Medical Imaging Practice. *Journal of Digital Imaging*. 2012;25:319-324.
76. Taylor-Phillips S, Elze MC, Krupinski EA, Dennick K, Gale AG, Clarke A, et al. Retrospective review of the drop in observer detection performance over time in lesion-enriched experimental studies. *Journal of Digital Imaging*. 2015;28(1):32-40.

Bijlage 1: Definities van afkortingen en uitkomstmaten

Tabel 6 Definities van afkortingen en uitkomstmaten

Uitkomst	Definitie
CDR	Kanker detectie ratio, aantal onderzoeken per 1.000 onderzoeken met kanker.
CFF/CFFF	Critical (Flicker) Fusion Frequency, de frequentie waarbij een knipperend licht of beeld overgaat naar een continue lichtbron. Het wordt gebruikt als maat voor de vermoeidheid, een lage C(F)FF is een hoge mate van vermoeidheid.
FPR	Fout positieven ratio, aantal positieven beoordelingen onder vrouwen die geen kanker bleken te hebben.
Jaarlijks volume	Het totaal aantal onderzoeken dat een radioloog in een jaar tijd heeft beoordeeld. Optelsom van het maximaal aantal onderzoek dat een radioloog op een dag beoordeeld.
P value	P, geeft de mate aan van een significant verschil tussen twee groepen of condities, waarbij $P < 0,5$ significant is.
PVW	Positief voorspellende waarde, percentage van aantal positieve mammogrammen die ook daadwerkelijk borstkanker bevatten.
RR / AIR	Recall rate of het verwijscijfer, alle onderzoeken die positief zijn beoordeeld ten opzichte van alle screeningsonderzoeken.
Screening focus	Percentage screening volume van het totale volume per jaar.
Sensitiviteit	Aantal positieven onderzoeken onder vrouwen die na 1 jaar de diagnose borstkanker hebben gekregen.
SOFI	Swedish Occupational Fatigue Inventory, bevat 20 expressies over de fysieke aspecten van vermoeidheid.
Specificiteit	Aantal negatieve onderzoeken onder vrouwen die na 1 jaar geen borstkanker bleken te hebben.
SSQ	Simulator Sickness Questionnaire, zeven vragen over de oogspanning en vermoeidheid.
VAS	Visual Analog Score, schaalverdeling voor de mate van vermoeidheid (subjectief).

Bijlage 2: Kwaliteitsbeoordeling

Op basis van verschillende gevalideerde checklisten is een kwaliteitslijst opgesteld [36-39]. Alle studies zijn op methodologische kwaliteit beoordeeld.

Tabel 7 Methodologische kwaliteitsbeoordeling

	Duidelijke probleemstelling?	Duidelijke onderzoeksgroep?	Is onderzoeksgroep groot genoeg?*	Duidelijke methode?	Interventietijd representatief?*	Is de studie reproduceerbaar?	Zijn resultaten gelooftwaardig en duidelijk?	Is de conclusie logisch?	Zijn beperkingen vermeld?	Totale score
	1.	2.	3.	4.	5.	6.	7.	8.	9.	
Barlow et al, 2011 [42]	1	1	1	1	0	1	1	1	1	8
Beam et al, 2003 [9]	1	1	1	1	1	1	1	1	1	9
Buist et al, 2011 [43]	1	1	1	1	1	1	1	1	1	9
Buist et al, 2014 [44]	1	1	0	1	1	1	1	1	1	8
Carney et al, 2012 [60]	1	1	1	1	1	1	1	1	1	9
Cawson 2009 [63]	1	0	0	0	0	0	0	1	1	3
Coldman et al, 2006 [22]	1	1	1	0	1	1	1	1	1	8
Cornford et al, 2011 [45]	1	1	0	1	1	1	1	1	1	8
Duncan et al, 2011 [46]	1	1	0	1	1	1	1	1	1	8
Elmore et al, 2002 [47]	1	0	0	0	0	0	0	1	1	3
Elmore et al, 2003 [61]	1	0	0	0	0	0	0	1	1	3
Elmore et al, 2009 [7]	1	1	1	1	1	1	1	1	1	9
Esserman et al, 2002 [48]	1	1	1	1	1	1	1	1	1	9
Haneuse et al, 2012 [49]	1	1	1	1	1	1	0	0	1	7
Ikushima et al, 2013 [64]	1	1	0	1	1	1	1	1	1	8
Kan et al, 2000 [50]	1	1	0	1	1	1	1	1	1	8
Krupinski et al, 2009 [31]	1	1	0	1	0	1	1	1	1	7
Krupinski et al, 2010 [29]	1	1	0	1	0	1	1	1	1	7
Krupinski et al, 2012 [30]	1	1	0	1	0	1	1	1	1	7
Maeda et al, 2011 [65]	1	1	0	1	1	1	1	1	1	8
Miglioretti et al, 2007 [51]	1	1	1	1	1	1	1	1	1	9
Miglioretti et al, 2009 [62]	1	1	1	1	1	1	1	1	1	9
Molins et al, 2008 [52]	1	1	0	1	0	1	1	1	1	7
Rawasdesht et al, 2013 [53]	1	1	1	1	1	1	1	1	1	9
Reed et al, 2009 [66]	1	1	0	1	1	1	1	1	1	8
Reed et al, 2010 [54]	1	1	0	1	1	1	1	1	1	8
Rickard et al, 2006 [55]	1	1	1	1	1	1	1	1	1	9
Safdar et al, 2009 [67]	1	1	0	1	1	1	1	1	1	8
Smith-Bindman et al, 2005 [56]	1	1	1	1	1	1	1	1	1	9
Tan et al, 2006 [57]	1	1	1	1	1	1	1	1	1	9
Theberge et al, 2005 [23]	1	1	1	1	1	1	1	1	1	9
Theberger et al, 2014 [58]	1	1	1	1	1	1	1	1	1	9
Timmers et al, 2014 [59]	1	1	1	1	1	1	1	1	1	9
Toomingas et al, 2014 [68]	1	1	1	1	0	1	1	1	1	8
Vertinsky et al, 2005 [69]	1	1	1	1	1	1	1	1	1	9

1 = deze vraag wordt met ja beantwoord en krijgt één punt. 0 = deze vraag wordt met nee beantwoord of is niet te achterhalen en krijgt nul punten. Totaal = 0-3 punten is zwakke kwaliteit, 4-6 punten is matige kwaliteit, 7-9 punten is sterke kwaliteit. *De onderzoeksgroep wordt groot genoeg bevonden als deze 100 deelnemers of meer bevat. **Als in een studie niet duidelijk aangegeven staat in welke periode het onderzoek is uitgevoerd, worden nul punten toegekend.