

**PGMI klassifisering
av mammografibilder ved
interne og eksterne
mamma-radiografer**



Randi Gullien

Masteroppgave
Masterprogram i helsefag
studieretning RAB fag
Institutt for samfunnsmedisinske fag
Det medisinsk- odontologiske fakultet
Universitetet i Bergen

Vår 2013

Forord.

Kvalitetssikringsarbeid er en viktig del av radiografenes arbeidsoppgaver i mammografiscreeningen i Norge. Bildekriterier og retningslinjer for klassifisering er utarbeidet i kvalitetsmanualen til det Norske Mammografiprogrammet. Likevel har jeg erfart gjennom arbeid med kvalitetssikring og klassifisering at det er rom for individuelle vurderinger. Med dette som utgangspunkt synes jeg det er interessant å se på klassifiseringsarbeid og om erfaring og arbeidssted påvirker klassifiseringsarbeidet.

Først vil jeg rette en stor takk til min veileder, Professor dr. med. Per Skaane som med sin kompetanse og støtte har vært til stor hjelp underveis.

Jeg vil rette en ekstra takk til Ph.d. kandidat Jack G. Andersen for sin uendelige tålmodighet og til min gode venninne og kollega radiograf Anne Emilie Haakull for de utallige oppmuntringene.

En takke til dere tre for et godt samarbeid gjennom mange år.

En takk til Professor dr. philos Leiv Sandvik ved for statistisk bistand.

Jeg vil også takke enhetsleder Tone Snare Berge som ga meg muligheten til å være student og gjøre dette arbeidet.

Til sist, men ikke minst, vil jeg takke min ektefelle Espen og våre to kjære barn, Martin og Kristin, for deres store tålmodighet, oppmuntringer og støtte i denne tiden.

Oslo, mai 2013.

Randi Gullien

INNHOLDSFORTEGNELSE

1. INTRODUKSJON	4
1.1 Presentasjon av problemområdet og bakgrunn for valg av tema.....	6
1.2 Sentrale teorier på området og studiens teoretiske forankring.....	7
1.3 Tidligere forskning på området.....	8
2. HENSIKT OG PROBLEMSTILLING.....	9
2.1 Hensikt.....	9
2.2 Problemstilling.....	9
3. METODE.....	10
3.1 Valg av forskningsdesign og metode.....	10
3.2 Validitet og reliabilitet.....	10
3.3 Utvalg.....	11
3.3.1 Inklusjonskriterier.....	11
3.3.2 Eksklusjonskriterier.....	11
3.3.3 PGMI radiografer.....	12
3.4 Variabler.....	12
3.5 Datainnsamling.....	13
3.6 Analyse.....	13
3.7 Litteratursøk.....	14
3.8 Etiske betraktninger.....	15
3.8.1 Etikens utvikling.....	15
3.8.2 Godkjenninger og anbefalinger.....	16
3.8.3 Etikk ved publisering.....	17
3.8.4 Egen rolle som forsker.....	18
4. RESULTATER.....	19
5. DISKUSJON.....	21
6. KONKLUSJON.....	25
7. REFERANSELISTE.....	26
8. VEDLEGG.....	29
Vedlegg 1.....	29
Vedlegg 2.....	30
Vedlegg 3.....	31
Vedlegg 4.....	32
Vedlegg 5.....	34

TITTEL: PGMI klassifisering av mammografibilder ved interne og eksterne mamma-radiografer.

FORFATTER: Randi Gullien

Studieretning RAB fag

År: 2013

SAMMENDRAG

Bakgrunn: Kvalitetsforbedringsarbeid og kvalitetssikringsarbeid er en kontinuerlig prosess for å sikre og opprettholde høy kvalitet over tid. Til dette benyttes PGMI klassifisering.

Hensikt: Finne ut om PGMI radiografene PGMI klassifiserer det samme settet med bilder forskjellig når de har forskjellige erfaringer, er interne og eksterne.

Bildekriteriene fra kvalitetsmanualen til det norske mammografiprogrammet (MP) ble brukt.

Materiale og metode: Et tilfeldig utvalg av 120 mammografiundersøkelser med fire bilder, 480 mammogrammer, fra uke 34 og 35 i 2010 ved mammografiscreeningen i Oslo ble klassifisert av tre PGMI radiografer, intern og med erfaring (A), ekstern og med erfaring (B) og intern og med liten erfaring i PGMI klassifiseringsarbeid (C). Bildene ble klassifisert etter følgende system: P=perfekt bilde=1, G=godt bilde=2, M=moderat godt bilde=3 og I=inadekvat bilde=4. Cohens kappa benyttes som statistisk metode for å beregne graden av samsvar mellom PGMI radiografene. Wilcoxon signed rank test benyttes for å analysere forskjeller i gjennomsnittsskåring mellom PGMI radiografene.

Resultater: Samsvar mellom PGMI radiografene var $\kappa = 0,40; 0,40; 0,44$ for A vs. B, B vs. C og A vs. C, respektivt og klassifiseres som litt til moderat samsvar i henhold til kappa klassifiseringen. Det beste samsvaret var mellom de to interne PGMI radiografene (A og C). Det dårligste samsvaret var mellom de to erfarne PGMI radiografene (A og B). PGMI radiografenes bruk av klassifiseringen varierte fra $p < 0,001$ for A vs. B og A vs. C, mens $p = 0,224$ for B vs. C for alle bildene og $p = 0,993$ for B vs. C for MLO bildene.

Konklusjon: Resultater indikerer en betydelig variasjon i klassifiseringen mellom PGMI radiografene, og at erfaring i PGMI klassifisering er av mindre betydning. Studien viser at det er behov for mer samkjøring og standardisert bruk av PGMI og et behov for revisjon og gjennomgang av kvalitetsmanualen.

Nøkkelord: PGMI, mammografi, mammografiscreening, kvalitetssikring, radiografi.

TITLE:

PGMI classification of mediolateral-oblique screening mammograms: Agreement among mammography radiographers.

SUMMARY

Introduction: Inter-observer agreement among three mammography radiographers were evaluated using the PGMI classifications recommended in the quality assurance manual of the Norwegian Breast Cancer Screening Program (NBCSP). The three radiographers, both internal and external, had different experience in PGMI classification.

Methods: Three radiographers, one internal experienced (A), one external experienced (B) and one internal inexperienced (C) in PGMI classification, independently classified MLO images from 120 examinations using the national guidelines for PGMI classification. The examinations were randomly selected from week 34 and 35 in 2010. The PGMI classifications were transformed into a numerical value: P=perfect=1; G=good=2; M=moderate good= 3; and I=inadequate=4. Cohen's Kappa statistic was used to measure inter-observer agreement.

Results: The kappa values were fair to moderate. Agreement of A vs. B, B vs. C, A vs. C were $\kappa = 0.34, 0.44$ and 0.45 , respectively. The highest agreement was between the two internal radiographers (A vs. C) and the lowest between the two experienced PGMI radiographers (A vs. B).

Conclusion: Different experiences, subjective opinions, and local variations regarding interpretation may influence the PGMI classification despite national guidelines. Our results indicate a lack of consequent use of the PGMI classification system. Training, examinations and exercise may be necessary to ensure uniform use of the PGMI classification.

Keywords: PGMI, mammography, mammograms, breast imaging, quality assurance, radiography.

1. INTRODUKSJON

1.1 Presentasjon av problemområdet og bakgrunn for valg av tema

Stortinget i Norge har vedtatt et landsdekkende mammografiscreening program, Mammografi Programmet (MP) der alle kvinner i alderen 50-69 år med fast bopel i Norge blir invitert til mammografiundersøkelse hvert annet år. Helse- og omsorgsdepartementet har det overordnede økonomiske ansvar for MP. Helsedirektoratet har det overordnede faglige ansvaret for MP og dette er delegert til Kreftregisteret. Kreftregisteret har ansvaret for planlegging, gjennomføring og evaluering. Nasjonal Rådgivningsgruppe er en faggruppe med bred faglig kompetanse innenfor brystdiagnostikk, som gir innspill til ledelsen av MP. Kreftregisteret har i samarbeid med Nasjonal Rådgivningsgruppe utarbeidet en kvalitetsmanual for de ulike faggruppene som er involvert i MP (1). Kvalitetsmanualen er et dokument som skal kvalitetssikre arbeidet i MP og ble første gang utgitt i 1996 (1). I denne studie benyttes den reviderte utgaven fra 2003 (1). Kvalitetsmanualen gjelder for alle yrkesgrupper i MP og inneholder et eget kapittel for radiografene, kapittel 5. Dette kapittelet inneholder krav og retningslinjer til radiografenes arbeid, bl.a. kvalitetssikringsarbeidet, kvalitetsvurderinger og klassifisering av mammografibildene (1). Radiografenes kvalitetsklassifisering av mammografibildene er ingen diagnostisk vurdering, men en kvalitetsklassifisering hvor mammografibilder blir målt opp mot en standard. Dette klassifiseringsarbeidet kalles PGMI og standarden bildene blir målt opp mot er definert som gullstandarden i MP (1). Kvalitetsmanualen tar utgangspunkt i den Europeiske kvalitetsmanualen, European Guidelines og den britiske og den irske kvalitetsmanualen, men er tilpasset norsk forhold (2-4).

PGMI klassifiseringen inndeles i P for et perfekt mammografibilde, G for et godt mammografibilde, M for et moderat godt mammografibilde og I for et inadekvat mammografi bilde. I henhold til retningslinjene i kvalitetsmanualen skal det utføres PGMI klassifiseringen av mammografibilder ved alle brystdiagnostisk sentre (BDS) i Norge (1). PGMI klassifiseringen blir utført av dedikerte radiografer ved hvert BDS og disse radiografene blir kalt PGMI radiografer.

Kvalitetsforbedringsarbeid og kvalitetssikringsarbeid er en kontinuerlig prosess for å sikre og opprettholde høy kvalitet over tid. Radiografene som tar mammografibildene på screeningen, screeningradiografene, får i henhold til retningslinjene i kvalitetsmanualen individuelle tilbakemeldinger fra PGMI radiografen

flere ganger i året (1). På denne måten kan screeningradiografene ha fokus på kvalitet og kvalitetsforbedringer av eget arbeid og måle sitt eget arbeide mot de nasjonale kravene. Det er screeningradiografenes oppgave å ta optimale bilder, og de må tilstrebe optimal posisjoneringen av brystet og optimalisere de tekniske innstillinger av apparatet slik at best mulig bilder blir tatt i hver enkelt projeksjon. I tillegg er det radiografenes ansvar kontinuerlig å tilegne seg kunnskap og tiltak for optimalisering av eget arbeid. For at PGMI resultatene for hver enkelt screeningradiograf og for at avdelingens PGMI resultat skal gi et korrekt bilde av kvalitetsnivået, er det svært viktig at PGMI radiografen gjør en mest mulig objektiv PGMI klassifisering. Ved en objektiv klassifisering av bildene kan resultatene sammenlignes mellom screeningradiografene på avdelingen, mot andre BDS og mot de nasjonale kravene. De målrettede forbedringstiltakene som screeningradiografene og avdelingen iverksetter tar vanligvis utgangspunkt i PGMI resultatene, slik at dårlig kvalitet kan forbedres og god kvalitet videreføres. Det er derfor viktig at det er samsvar mellom PGMI radiografenes vurderinger.

1.2 Sentrale teorier på området og studiens teoretiske forankring

Regelmessig mammografi reduserer dødeligheten av brystkreft (5-7). Forutsetningen for reduksjon av dødelighet er at kvinnene møter regelmessig til mammografiscreeningen, slik at brystkreft kan oppdages i et tidlig stadium. Mammografibildene er grunnlaget for å kunne stille en korrekt radiologisk diagnose. Det er flere faktorer som påvirker bildekvaliteten negativt og som kan gjøre bildetakingen og bildetolkningen vanskelig. Årsaker til dette kan være for eksempel posisjoneringsfeil, apparattekniske aspekter, eksponeringsfeil, detektorfeil, bildekontrast eller kvinnens anatomi (8-11). Mammografiscreeningen i Oslo (MiO) har digitale mammografiapparater og derfor utgår film, kassetter og fremkalling som mulig feilkilde til dårlig bildekvalitet.

Mammografibilder som blir tatt i MP er standardisert til frontbilde (cranio-caudal, CC) og skråbilde (mediolateral - oblique, MLO) av hvert bryst (1). Det er standardiserte projeksjonskriteriene for både CC og MLO bildene (1). Disse kriteriene er nedfelt i Kvalitetsmanualen og legges til grunn i PGMI klassifisering av bildene (Vedlegg 1, s. 29). MP's indikator for kvalitet er at minimum 75 % av bildene skal være perfekte eller gode, P + G. Ved PGMI klassifisering er kvalitetskravet ytterligere spesifisert for de resterende 25 %: minst 22 % skal være moderat gode bilde og maksimum 3 % skal være inadekvate bilder. Disse kvalitetskravene er et

minimumskrav i PGMI klassifiseringen og brukes som gullstandard i MP (1). Den enkelte screeningradiograf og avdelingen blir på denne måten kvalitetsmålt opp mot gullstandarden.

PGMI klassifisering skal være en mest mulig objektiv klassifisering som følger kriteriene. Det vil være et problem dersom det ikke er samsvar i PGMI klassifiseringen og de forskjellige PGMI radiografene klassifiserer bildene forskjellig til tross for retningslinjene. Spuur et. al. konkluderer i sine studier at det er mangel på samsvar mellom PGMI radiografene og at det er viktig å redusere den subjektive vurderingen i klassifiseringsarbeidet (10;12). En norsk studie fant at det var store variasjoner mellom norske PGMI radiografer (9).

Tidligere forskning viser at mammografi og tidlig oppdagelse av brystkreft reduserer dødeligheten (5;7). Som et ledd i tidlig oppdagelse av brystkreft er det viktig at kvinnene får en god undersøkelse med mammografibilder av optimal kvalitet.

1.3 Tidligere forskning på området

Tidligere PGMI studier ved BDS, OUS, har vært presentert på European Congress of Radiology (ECR) (13-15). Disse studiene viser at avdelingens totale PGMI resultat ligger over kravene i Kvalitetsmanualen, men at det var individuelle kvalitetsforskjeller blant screeningradiografene ved MiO. Disse forskjellene viste seg å være uavhengig av screeningradiografenes arbeidserfaring med mammografi og resultatene varierte i løpet av prosjektperioden. Studiet viste at enkelte feil og mangler ved bildene gjentok seg hyppigere enn andre og at det var et forbedringspotensial i gruppen moderat gode MLO bilder.

Det er publisert artikler om temaene omstillings- og læringsinteresse, vurdering av forskjellige klassifiserings systemer, kriteriene i PGMI klassifiseringen og inter-observatør variabilitet mellom PGMI radiografer og mellom radiologer (9;10;12;16-18). Tidligere forskning har vist at det er en direkte sammenheng mellom bildekvalitet og cancer deteksjon. Optimal bildekvalitet gir en tidligere oppdagelse av cancer og færre intervall cancer (19). Taplin et. al. fant i sin studie at bildekvalitet og posisjonering hadde signifikant effekt på cancer deteksjonsraten (11). Ifølge Basset et. al. var dårlig posisjonering hovedgrunnen til at amerikanske brystsentre ikke fikk akkreditering etter de amerikanske retningslinjene i American College of Radiology (ACR) mammografiprogram (8).

Klassifisering av mammografibilder er ikke standardisert internasjonalt. Mange steder i Europa bruker PGMI klassifiseringen, mens Australia i stor grad har adoptert

og tilpasset PGMI klassifiseringen til EAR hvor EAR står for excelent, acceptable og repeat (18). En norsk PGMI klassifiseringsstudie viser store forskjeller i klassifisering av bilder mellom lokale PGMI radiografer, et ekspertpanel med PGMI radiografer og imellom disse to gruppene (9).

2. HENSIKT OG PROBLEMSTILLING

2.1 Hensikt

Hensikten med studien er å se om PGMI radiografene PGMI klassifiserer det samme settet med bilder forskjellig når de har forskjellige erfaringer og kommer fra forskjellige BDS. Den individuelle tilbakemeldingen screeningradiograf får og avdelingens resultater må være objektive og korrekt, og dermed til å stole på. PGMI resultatene skal være uavhengig av hvilken PGMI radiograf som klassifiserer bildene. Studien skal evaluere i hvilken grad tre PGMI radiografer med forskjellig erfaring i PGMI klassifiseringsarbeid og som er interne og eksterne, samsvarer når de klassifiserer kvalitet i et utvalg av mammografibilder uavhengig av hverandre. Bildene skal klassifiseres etter de gitte PGMI kriterier som er nedfelt som gullstandard i Kvalitetsmanualen (1).

Kvalitetssikring og kvalitetsforbedringsarbeid er en kontinuerlig prosess hvor arbeidet med optimalisering og å opprettholde høy kvalitet på mammografibildene og iverksette forbedringstiltak, er med på å heve kompetanse til radiografene. Oslo universitetssykehus (OUS) har fire hovedfokus i sin strategiplan: Strategi 2013 – 2018, hvor av det ene hovedmålet er: ”Være et fremragende universitetssykehus, en lærende og skapende organisasjon”. Videre står det blant annet: ”OUS har ansvar for å tilby spesialisert diagnostikk og behandling av høy kvalitet. Vi skal kvalitetssikre behandlingstilbudet gjennom systematisk registrering av behandlingsresultater i medisinske kvalitetsregistre” (20). Denne studien er ett av mange ledd i arbeidet for å nå disse målene.

2.2 Problemstilling

Hvor godt er samsvaret mellom (i hvilken grad samsvarer) PGMI radiografenes kvalitetsklassifisering av mammografibilder tatt ved MiO når de har forskjellig erfaring i PGMI klassifisering og er interne og eksterne? En PGMI klassifisering skal være en objektiv klassifisering som følger standardiserte kriterier uten rom for individuelle oppfatninger. I denne studien er det tre PGMI radiografer som utfører klassifiseringen.

To er interne og en av disse er erfaren og en er uerfaren PGMI radiograf. Den siste er en erfaren og ekstern PGMI radiograf.

3. METODE

3.1 Valg av forskningsdesign og metode

Studien har et tverrsnittsdesign og er basert på et utvalg av mammografibilder av 120 kvinner tatt ved mammografiscreeningen i Oslo og tre PGMI radiografer som klassifiserer hvert bilde i forhold til PGMI kriteriene.

I denne studien skal resultatene fra PGMI klassifiseringen systematiseres, registreres og analyseres statistiske. Det er valgt å se på PGMI resultatene i to grupper, se på alle bildene samlet, CC og MLO under ett, og tilsvarende for MLO bildene. Det blir gjort beregninger for å finne graden av samsvar mellom PGMI radiografene ved hjelp av inter-rater reliabilitet Cohens kappa analyser (21-23). Videre bruktes Wilcoxon signed rank test for å analysere forskjeller i gjennomsnittsskåring mellom PGMI radiografene (21;22).

3.2 Validitet og reliabilitet

Validitet står for relevans, overførbarhet og kunnskapens gyldighet og om metoden måler det den er tenkt å måle (21;24;25). Med validitet menes hvor gyldig resultatene av studien er og om gyldigheten er overførbar til andre BDS som bruker de samme PGMI klassifiseringskriteriene og måleskalaen (21;24;25). De innsamlede dataene må være relevante i forhold til problemstillingen, være valide. Bevisst eller ubevisst, ut fra egeninteresse eller mangelfull kunnskap kan jeg som forsker påvirke den interne og eksterne validiteten ved å unnlate, mistolke, misforstå, tilside sette viktig informasjon både i innsamlingsfasen og i registreringsfasen (21;24;25).

Reliabilitet betyr pålitelighet og er forbundet med målesikkerheten i resultatene og knyttet til den klassifiseringsmetoden som er brukt. I denne studien er målemetoden PGMI klassifisering hvor PGMI er brukt som måleskala og er omgjort til en numerisk skala fra 1 til 4. PGMI klassifiseringen er nedfelt i kvalitetsmanualen for MP og ved å bruke den nasjonalt standardiserte metoden, gullstandarden for bildeklassifisering, styrker det studiens reliabilitet. PGMI klassifiseringen er etterprøvable og reliabiliteten styrkes ved at tre PGMI radiografer klassifiserer det samme settet med bilder uavhengig av hverandre. Graden av samsvar mellom PGMI radiografene vil påvirke studiens reliabilitet.

Validiteten og reliabiliteten kan påvirkes av de to interne PGMI radiografene som kan være forutinntatte når de klassifiserer kollegaers bilder, være påvirket av en lokal trend i PGMI klassifiseringen og latt dette prege klassifiseringsarbeidet. En ekstern PGMI radiografen i tillegg til de to interne PGMI radiografene vil styrke studiens validitet og reliabilitet Dalland (24).

3.3 Utvalg

Prosjektleder har gjort et tilfeldig utvalg av 120 mammografiundersøkelser med fire bilder, totalt 480 mammografibilder. Undersøkelsene ble tatt i uke 34 eller uke 35 2010 ved mammografiscreeningen i Oslo, og PGMI klassifiseringen ble utført i ukene 38 til 40 i 2010. Alle mammografiundersøkelsene var digitale og utført på Selenia Dimensions (Hologic, Inc; Bedford, MA, USA). Undersøkelsene ble hentet fra det digitale bildearkivet, picture archiving and communication system (PACS) og lagret på en dedikert arbeidsstasjon for mammografi. Prosjektleder laget egen arbeidsliste på arbeidsstasjonen slik at PGMI radiografene enkelt kunne finne mammografibildene som skulle klassifiseres. Tidligere undersøkelser ble ikke brukt til sammenligning. PGMI klassifiseringen ble ført på eget registreringsskjema (Vedlegg 2, s. 30). Prosjektleder kontrollerte at PGMI radiografene arbeidet uavhengig av hverandre ved å tildele dem arbeidstid på forskjellige dager. Det ble ikke tatt hensyn til hvilke screeningradiografer som hadde utført undersøkelsene.

3.3.1 Inklusjonskriterier

- Mammografiundersøkelser av kvinner tatt ved MiO.
- Undersøkelser som består av fire mammografibilder (to bilder av hvert bryst, CC og MLO).

3.3.2 Eksklusjonskriterier

- Kvinner med implantater i brystene ble ekskludert fordi det er en egen prosedyre for mammografi av kvinner med implantater i brystene. Denne prosedyren innebærer flere bilder og svært liten kompresjon av brystene.
- Undersøkelser hvor fremmedelementer som f. eks. pacemakere er med på bildene ble ekskludert. Fremmedelementer vil gjøre at det kan være vanskelig å posisjonere og komprimere brystet optimalt. Det vil være vanskelig å foreta en PGMI klassifisering når deler av brystet kan være dekket av fremmedelementet.

- Undersøkelser med flere enn fire bilder ble ekskludert fordi kvalitetsmanualen ikke er tydelig i sin anbefaling for undersøkelser med flere bilder i en projeksjon.

3.3.3 PGMI radiografer

De tre radiografer som deltok i studien hadde forskjellig bakgrunn (Tabell I). En erfaren PGMI radiograf defineres som en radiograf med minst fire års erfaring i å ta mammografibilder og minst to års erfaring med å bruke PGMI klassifiseringen på mammografibilder. Med en uerfaren PGMI radiograf menes en radiograf med minst fire års erfaring i å ta mammografibilder og mindre enn to års erfaring med å praktisere PGMI klassifisering på mammografibilder. En erfaren og ekstern PGMI radiograf arbeidet opprinnelig ved et annet BDS, men startet sitt arbeid ved MiO 2 uker før prosjektstart. Denne PGMI radiografen defineres som en ekstern.

Tabell I.

PGMI radiografenes erfaring og tilhørighet

PGMI radiograf	Tilhørighet	Erfaring i PGMI	Erfaring i mammografi
A	Intern	Ja	Ja
B	Ekstern	Ja	Ja
C	Intern	Nei	Ja

3.4 Variabler

Denne studien er basert på variabelen ”bildekvalitet” som kan benyttes til å klassifisere kvaliteten av et mammografibilde (Tabell II). Denne variabelen skåres på en skala fra 1 til 4, hvor hver tallverdi representerer en PGMI klassifiseringsskår for hvert enkelt bilde.

Tabell II.

PGMI klassifiseringen omgjort til tallverdi og kvalitetskrav.

	Tallverdi	PGMI krav
Perfekt bilde	1	P + G ≥ 75 %
Godt bilde	2	
Moderat godt bilde	3	M < 22 %
Inadekvat bilde	4	I < 3 %

3.5 Datainnsamling

Studien er godkjent av enhetsleder og anbefalt av forsknings- og undervisningsavdelingen ved OUS (FASUUS), og har fått tilråding fra personvernombudet og Regional komité for medisinsk og helsefaglig forskningsetikk, Sør- Øst Norge (REK). Studien er en del av en større forskningsstudie som heter: Oslo Tomosynthesis Screening Trial (OTST). OTST studien har den nødvendige tilråding fra forskningsutvalget, personvernombudet og REK (Vedlegg 3 s. 31, vedlegg 4, s. 32 og vedlegg 5, s. 34).

Forskningsdata fra studien skal oppbevares på egen server. En godkjenning av FASUUS innebærer en tilgang til å få opprettet en egen database ved sykehusets forskningsserver slik at aidentifiserte personopplysninger kan leges inn på egen filen. Tilgang til denne serveren forutsetter at studien er formalisert og godkjent av de nødvendige instanser. Studien får eget lagringsområde og det er prosjektansvarlig som bestemmer hvem som skal ha tilgang til denne datafilen. Ved å benytte denne serveren vil kravene til informasjonssikkerhet i henhold til Datatilsynets krav være dekket.

En egen arbeidsliste med de aktuelle undersøkelsene ble laget på en av arbeidsstasjonene på avdelingen. PGMI radiografene hadde et introduksjonsmøte med prosjektleder ved prosjektstart. På møtet ble registreringsskjema, kopi av kvalitetsmanualens bildekriterier og kodeliste delt ut gått gjennom (Vedlegg 1, s. 29 og vedlegg 2, s.30). På denne måten fikk PGMI radiografene den samme informasjonen og det var mulighet for å stille spørsmål slik at uklarheter kunne bli oppklart. De tre PGMI radiografene fikk tildelt hvert sitt tidspunkt for klassifiseringsarbeidet ved arbeidsstasjonen slik at de utførte PGMI klassifiseringen uavhengig av hverandre. Ved PGMI klassifisering av bildene var kvinnens undersøkelsesdato, fødselsår og initialer kjent for PGMI radiografene. Innsamlet data ble aidentifisert og lagret på forskerserveren slik at verken utøvende radiografs navn eller kvinnenens navn er tilgjengelig.

3.6 Analyse

I denne studien er det valgt 120 undersøkelser, 240 CC bilder og 240 MLO bilder, til sammen 480 mammografibilder. Cohens kappa er valgt som statistisk metode for å beregne graden av samsvar mellom PGMI radiografene. (21-23). Kappa verdiene, κ , sier noe om graden av samsvar, men dette samsvaret sier ikke noe om i hvilken grad PGMI radiografene klassifiserer riktig eller feil, godt eller dårlig, dvs. om begge klassifiserer feil eller riktig. En kappa verdi viser hvor godt eller dårlig et samsvar er.

Kappa graderes mellom null og en. En kappa på 1 indikerer et perfekt samsvar og en kappa på null indikerer ikke samsvar (22). Kappa verdiene graderes som vist i tabell III (Tabell III) (22).

Tabell III.

Cohens kappa gradering.

Kappa, κ , verdi	Styrke av enighet/ samsvar
< 0,20	Dårlig
0,20 – 0,40	Litt
0,41 - 0,60	Moderat
0,61 – 0,80	God
0,81 – 1,00	Svært god

I denne studien utarbeides det kappa verdier for 4-punkt graderingsskala, κ 4. Skalaen for PGMI skår er en ikke-lineær skala hvor det er større forskjell mellom et godt bilde og et moderat bilde, enn mellom et perfekt bilde og et godt bilde. Det er aksept for at det ikke er mulig bare å ta perfekte bilder, og i kvalitetsmanualen er kravet til perfekte og gode bilder slått sammen, P + G, og minst 75 %. Det utarbeides derfor kappa verdier for en kollabert 4-punkt graderingsskala til en 3 nivå graderingsskala, κ 3. I den kollaberte skalaen er skår for perfekte bilder, P, og skår for gode bilder, G, slått sammen, mens skår for moderate og inadekvate bilder er uendret.

Wilcoxon signed rank test er en ikke-parametrisk test som kan brukes til å sammenligne gjennomsnittlig PGMI klassifiseringsverdier på de samme mammografibildene parvis mellom PGMI radiografene. Klassifiseringsskalaen er fra 1 til 4 og hvor 1 er best og skår for et perfekt bilde, mens 4 er dårligst og er skår for et inadekvat bilde.

PGMI skårene ble analysert statistisk ved å bruke SPSS statistikkprogram (SPSS SmartViewer version 15.0 for Windows XP, SPSS Inc., Chicago, Illinois).

3.7 Litteratursøk

Det ble benyttet flere fremgangsmåter for å finne litteratur til oppgaven. Som ansatt ved OUS har jeg tilgang til medisinsk biblioteks databaser og bøker og til bibliotekstjenester. Ved å søke i de såkalte "åpne" databasene eller web som er tilgjengelig for alle med internett kan man få treff som ikke alltid er kvalitetssikret og leseren må kritisk vurdere stoffet, bl.a. ved å vurdere om stoffet er lagt ut for å påvirke

eller formidle faktisk informasjon, finne ut hvem som har opphavet (organisasjon eller privatperson), hvem som er ansvarlig for nettsiden og om siden er oppdatert eller foreldet. Via biblioteket ved OUS har jeg tilgang til ”lukkede” databaser. Disse databasene kan kreve lisenser eller tilgang via arbeidsplass som betaler lisensen eller tilgang som registrert helsepersonell.

Forskjellige søkedatabaser har litt forskjellig søkestrategi og forskjellige søkeord, nøkkelord eller emneord. Derfor er det viktig å finne de beste søkeordene slik at søkene kan bli mest mulig effektive og Medical Subject Headings (MeSH) terminologien brukes til dette (26). MeSH er National Library of Medicine i USAs synonymvokabular i medisin og består av termer som navngir og beskriver den hierarkiske strukturen som tillater søking på forskjellige nivåer i databaser. Dette er spesielt utviklet for databasene Medline og PubMed, men kan også brukes i flere andre databaser (26). Et MeSH søkeord brukes til å beskrive hva artikkelen omhandler og det er koblet synonymer til dette søkerordet. Dette sikrer at forskjellige stavemåter, språk, entalls- og flertallsendelser blir tatt hensyn til.

Det har vært utført søk etter litteratur i databasene PubMed, Chinal, Ovid og Embase. om PGMI, inter-variabilitet, inter-observer og studier som omhandler PGMI. Søkene ble gjort delvis alene og delvis sammen med medisinsk bibliotekar på medisinsk bibliotek ved OUS. Søkerord som ble brukt var: quality control, PGMI, guidelines, interobserver, intervariability, mammography, radiographer og kombinasjoner av disse. Søkene har vært i databasene på engelsk og på skandinaviske språk. Ved å benytte seg av bibliotekartjenesten ved OUS kvalitetssikres litteratursøkene.

3.8 Etiske betraktninger

3.8.1 Etikkens utvikling

Det medisinske og helsefaglige etiske syn i forskning har endret seg gjennom tidene. Hippokrates (ca. 466 – 377 f. Kr.) regnes som grunnlegger av profesjonsetikken og har hatt stor betydning for utvikling av etikk innenfor medisin (27). Hippokrates fokuserte på at moralske prinsipper og etisk forsvarlighet skulle komme den syke til gode i tillegg til det medisinske håndverk. Fra den Hippokratiske ed til de etiske komiteene har medisinsk etikk hatt en lang og til tider vanskelig utvikling (27). Det har endret seg fra nesten å være fraværende og nærmest å kunne benytte nesten hvem som helst til hvilket som helst forsøk, til at forskning blir regulert, overvåket og etisk vurdert

(27). Tidligere var det vanlig å utføre forskning og eksperimenter på fattige eller sosialt svake, og at hensynet til forskningen gikk foran hensynet til individene. De nåværende etiske prinsipper har sin bakgrunn i Nürnberg-koden som kom som følge etter oppgjøret med krigsforbrytere etter den andre verdenskrig i Nürnberg, Nürnberg-prosessen. Det ble da diskutert og satt spørsmål ved de etiske vurderingene ved forskningen. Som et resultat kom man frem til Nürnberg-koden hvor det står at ingen kan fraskrive seg sitt personlige ansvar for egne handlinger uansett om det er gitt instruks eller ordre (28). Nürnberg-koden påpeker at forskning bare skal foregå på friske, voksne som kan avgi et frivillig og informert samtykke. På bakgrunn av dette behandler Helsinki- deklarasjonen medisinsk etikk, krav til vitenskapelige fremgangsmåter og personalets kvalifikasjoner. Denne deklarasjonen ble utarbeidet av Verdens legeförening, World Medical Association, og vedtatt på generalforsamlingen i 1964 (29). Helsinki- deklarasjonen fra 1964 inneholder viktige prinsipper innenfor medisinsk forskning og er senere revidert flere ganger, siste gang i 2008 (29). I Norge ble det undervist i etikk i sosialmedisin på 1950- tallet og utover 1960 og -70 tallet ble det undervist i medisinsk etikk ved de medisinske fakultetene og regionale komiteene for forskningsetikk ble opprettet (27). I motsetning til det man valgte i utlandet, valgte man i Norge en tverrfaglig sammensetning av de etiske komiteene, hvor de medisinskfaglige forskere er i mindretall (27;30). At medisinsk forskning har en medisinsk- og helsefaglig etisk forankring og er vurdert av en tverrfaglig etisk komité oppleves som en positiv forankring av den medisinske forskningen (27).

Forskningsetikkloven i Norge innebærer at all forskning som innebærer forsøk og forskning på mennesker må søke om tillatelse (31). Med personvern menes den rett alle individer har for å verne om sin private sfære og til å forvalte sannheten om seg selv (32). Personvern må ikke forveksles med helse, miljø og sikkerhetshensyn (HMS). Vanligvis kommer ikke HMS inn på personvern, for HMS dreier seg om andre deler av menneskenes sikkerhet. Personvern er i prinsippet alt som kommer inn av all informasjon som kan knyttes til det enkelte individ. Forvaltningen av denne informasjonen kommer inn under Personopplysningsloven (POL) med tilhørende Personopplysningsforskrift (32). Alle skal kunne føle seg trygge på at opplysninger om en selv ikke kan misbrukes av helsevesen, myndigheter eller andre.

3.8.2 Godkjenninger og anbefalinger

Det er viktig å sikre at regler og retningslinjer for medisinsk forskning blir fulgt og at forskningsprosjektet er godkjent og fått de tilrådning fra de nødvendige instanser

før prosjektet starter (Vedlegg 3, s. 31, vedlegg 4, s. 32, og vedlegg 5, s. 34). Denne studien er inkludert i disse søknadene.

Det er viktig at prosjektansvarlige forsker og alle i forskningsteamet kjenner de forskningsetiske prinsipper som gjelder for all forskning. Det viser at studien tar sikte på en høy etisk standard og at etiske prinsipper blir ivaretatt både i innsamlingen av data, under bearbeidingen og ved publiseringer. Alle prosjektdeltagere og forskeren er fast ansatt ved sykehuset og er utdannet helsearbeidere og har ved ansettelse underskrevet taushetsløfte. Det kan likevel være på sin plass med påminnelse om personvern og taushetsplikt til forskningsteamet.

Forskningsdata fra studien skal lagres på eget lagringsområde ved sykehusets forskningsserver. Ved å benytte forskningsserveren vil kravene til informasjonssikkerhet i henhold til Datatilsynets krav være dekket. Tilgang til forskningsserveren er en del av søknaden til personvernombudet ved OUS.

I denne studien skal ikke bare kvinnene som får sine bilder klassifisert bli sikret anonymitet, men også hver enkelte radiograf som utfører undersøkelsene, slik at det ikke er mulig å spore tilbake til verken kvinne eller radiograf. For å kunne gi en individuell tilbakemelding til hver enkelt radiograf må PGMI radiografene ha tilgang til hvilken radiograf som har tatt hvilke bilder. Deretter skal navnet kodes etter gitte retningslinjer for forskning.

Det er til enhver tid forskers ansvar at dette gjennomføres i hele forskningsprosessen, både i innsamlingen av data, under bearbeidingen og ved publiseringer, selv om godkjenning og tilrådning foreligger er. Gjennom gode rutiner og et godt planlagt prosjekt er faren for at sensitiv informasjon kan komme på avveie eller ikke håndteres korrekt, redusert.

3.8.3 Etikk ved publisering

Det finnes etiske retningslinjer for publisering av vitenskapelige artikler i medisinske tidsskrifter. Vancouver-konvensjonen skal sikre at medisinske tidsskrifter holder en høy etisk standard. Retningslinjene er utarbeidet av International Committee of Medical Journal Editors og de fleste tidsskrifter benytter disse retningslinjene (33). Dette er viktige momenter i de etiske betraktninger fordi forskningen vanligvis ender i en skriftlig publikasjon eller på annen måte publiseres. I enhver publikasjon må det sikres anonymitet til pasienter og kollegaer og de følge internasjonale retningslinjene for publisering må følges (33).

3.8.4 Egen rolle som forsker

Som student ved Universitetet i Bergen, UiB, er det viktig at jeg setter meg inn i og følger de etiske retningslinjene for UiB i tillegg til de allmenngyldige regler og retningslinjene (34). Jeg har tidligere avlagt eksamen ved UiB og ved Høgskolen i Bergen, HiB, og disse besvarelsene inneholder bl.a. avsnitt om litteratursøk, metodikk og etikk. Dette er ikke publisert materiale. Jeg har ikke kopiert disse besvarelsene, men føler likevel at jeg til tider har formulert på samme måte og forholdsvis likt som i disse besvarelsene. Dette er nok fordi det er min måte å formulere meg på og fordi jeg ikke er rutinert i skrivearbeid. Jeg mener at jeg bruker kildehenvisninger på en korrekt måte, slik at det er mulig å gå til de kilder jeg har oppgitt.

På er generelt grunnlag stilles det i dag krav til objektivitet, forskningskunnskap og systematikk til forskning. Forskeren skal bruke kritisk refleksjon, bidra med kunnskapsutvikling som kan etterprøves og deles, og med ambisjon om overførbarhet av resultatene ut over den sammenheng fra den enkelte studie (25). Forskning er ikke bare ny banebrytende forskning og med oppsiktsvekkende funn, men også studier som bekrefter tidligere funn.

Før studien starter er det viktig at jeg som forsker og prosjektleder har satt meg inn i de nødvendige lover og regler og at alle godkjenninger og anbefalinger er innhentet før studiet starter. De etiske aspektene og personvern hensyn i studien vil dermed være ivaretatt. Det er viktig at jeg erkjenner mine begrensninger i kunnskap om forskningsmetodikk, litteratursøk og statistikk og søker råd og veiledning fra personer med nødvendig kompetanse. Det er viktig at alle data blir korrekt behandlet og at jeg som forsker ikke lar forutinntatte holdninger påvirke studien. Jeg må være bevisst på mine egne meninger, oppfatninger og synspunkter og ikke la dette legge føringer og påvirke klassifiseringsarbeidet til PGMI radiografene. Dette kan være vanskelig fordi min nærhet til studien og hyppig kontakt med PGMI radiografene kan gjøre det vanskelig med kritisk refleksjon (25). Objektivitet er også viktig når statistiske metoder skal brukes. Resultatet av studien kan være annerledes enn forventet og jeg må være forberedt på at resultatet kan rokke ved mine forutinntatte meninger og erfaringer (25). Som forsker bør jeg lese litteraturen kritisk. Interessante funn i litteratur kan vise seg og ikke være overførbart til dette studiet fordi det kan dreie seg om klassifisering innenfor andre fagområder. Likevel kan litteraturen gi nye innfallsvinkler og momenter som kan være relevante og viktige.

Dokumentasjon av resultatene i studien er viktig. Som forsker har jeg ansvar for muligheten til å etterprøve og kontrollere de data som er innsamlet.

Jeg må håndtere informasjon, datamaterialet og resultatene på en slik måte at resultater fra studien kan komme pasienter og kollegaer til gode. Jeg må som forsker være ydmyk og være mitt ansvar bevisst.

4. RESULTATER

Hensikten med studien var å analysere graden av samsvar mellom PGMI radiografenes PGMI klassifisering. Kappa verdiene på 4 punkt skalaen for alle bildene, CC og MLO, sett samlet, viste at samsvaret var litt til moderat samsvar med variasjonsbredde på 0,40 – 0,44. Det beste samsvaret viste seg å være mellom de to interne PGMI radiografene, A vs. C, med kappa 0,44. Det var noe mindre samsvar mellom de to erfarne PGMI radiografene, A vs. B, med på kappa 0,40, og mellom den erfarne eksterne PGMI radiografen, B, vs. den interne uerfarne PGMI radiografen, C, med en kappa på 0,40 (Tabell IV).

Kappa verdiene for samsvar på den kollaberte 4 punkt skalaen til en 3 punkt skala for CC og MLO bildene, derimot, viste moderat samsvar med variasjonsbredde på 0,47 – 0,61. Det beste samsvaret var mellom de to interne PGMI radiografene, A vs. C, med en kappa verdi på 0,61. Det var mindre samsvar mellom de to erfarne PGMI radiografene, A vs. B, med en kappa verdi på 0,53. Det laveste eller dårligste samsvaret var mellom den erfarne eksterne PGMI radiografen, B, vs. den interne uerfarne PGMI radiografen, C, med en kappa på 0,48 (Tabell IV).

Tabell IV.

Kappa verdi for samsvar mellom PGMI radiografene for CC og MLO bildene vurdert samlet.

PGMI radiograf	Kappa verdi for alle bildene, CC og MLO sett samlet.	
	B	C
A	$\kappa 4 = 0,40$ $\kappa 3 = 0,53$	$\kappa 4 = 0,44$ $\kappa 3 = 0,61$
B		$\kappa 4 = 0,40$ $\kappa 3 = 0,47$

$\kappa 4$ = kappa verdi for en 4-punkters skala for PGMI klassifisering.
 $\kappa 3$ = kappa verdi for en 3-nivå skala av den kollaberte 4-punkters gradering av PGMI klassifiseringen.

Kappa verdiene på 4 punkt skalaen for alle MLO bildene viste litt til moderat samsvar med variasjonsbredde mellom 0,34 – 0,45. Det beste samsvaret mellom de to interne radiografene, A vs. C med en kappa verdi på 0,45, til tross for deres forskjellige erfaringer i PGMI klassifiserings arbeid. Det var dårligste samsvar var mellom de to PGMI radiografene som begge har er faring i både mammografi og klassifiseringsarbeid, A vs. B med en kappa verdi på 0,34. Samsvaret mellom B vs. C var kappa 0,44 og tilsvare et moderat samsvar. Resultatene viser at samsvaret varierte mellom litt og moderat samsvar for MLO resultatene (Tabell V).

Kappa verdien for samsvar for den kollaberte 4 punkt skalaen til en 3 punkt skala for alle MLO bildene viste et moderat samsvar med variasjonsbredde på 0,46 – 0,57. Det beste samsvaret er mellom de to interne PGMI radiografene, A vs. C, med kappa verdi på 0,57. Det var mindre samsvar mellom den erfarne eksterne PGMI radiografen, B, vs. den interne uerfarne PGMI radiografen, C, med en kappa på 0,49. Det laveste eller dårligste samsvaret var mellom de to erfarne PGMI radiografene, A vs. B, med på kappa 0,46 (Tabell V).

Tabell V.

Kappa verdi for samsvar mellom PGMI radiografene for MLO bildene.

PGMI radiograf	Kappa verdier for PGMI slår for alle MLO bildene	
	B	C
A	$\kappa 4 = 0,34$ $\kappa 3 = 0,46$	$\kappa 4 = 0,45$ $\kappa 3 = 0,57$
B		$\kappa 4 = 0,44$ $\kappa 3 = 0,49$
$\kappa 4$ = kappa verdi for en 4-punktters skala for PGMI klassifisering. $\kappa 3$ = kappa verdi for en 3-nivås skala av den kollaberte 4-punktters gradering av PGMI klassifiseringen.		

Wilcoxon signed rank test ble brukt som metode for å se på forskjellen mellom PGMI radiografenes bruk av klassifiseringsskalaen. For alle bildene, CC og MLO, sett samlet, var $p < 0,001$ for A vs. B; $p < 0,001$ for A vs. C og $p = 0,224$ for B vs. C. For alle MLO bildene var $p < 0,001$ for A vs. C; $p < 0,001$ for A vs. B og $p = 0,993$ for B vs. C. Disse resultatene viser den samme tendensen for alle bildene sett under ett (CC og MLO) og for MLO bildene sett for seg. Resultatene indikerer at den erfarne

eksterne, B, og den interne uerfarne, C, brukte PGMI klassifiseringsskalaen forholdsvis likt og at det var større forskjeller og variasjoner i bruk av PGMI klassifiseringsskalaen mellom de to erfarne PGMI radiografene, A og B, og mellom de to interne PGMI radiografene, A og C.

5. DISKUSJON

MP har i sin kvalitetsmanual klare retningslinjer for PGMI klassifiseringsarbeidet, kriterier for bildekvalitet og en gullstandard for PGMI klassifiseringskriterier (1). En PGMI klassifisering skal være en objektiv klassifisering av mammografibildene hvor kriteriene blir brukt. PGMI klassifiseringer bør derfor være uavhengig av PGMI radiografenes arbeidssted og deres erfaring i klassifiseringsarbeidet.

Imidlertid viser resultatene fra denne studien at det ikke er så godt samsvar som forventet. Noe overraskende er det at samsvaret mellom PGMI radiografene hovedsakelig ikke var bedre enn moderat. Det eneste samsvaret som kunne defineres som godt var for CC og MLO bildene sett samlet. Dette samsvaret var mellom de to interne PGMI radiografene, A og C, med en kollabert kappa verdien var: $\kappa_3 = 0,61$. Dette er i nedre sjikt for hva som er definert som et godt samsvar, for nedre grense er $\kappa = 0,60$. Det ville vært naturlig å tro at erfaring i PGMI klassifisering og det å være vant med å bruke kriteriene i kvalitetsmanualen var viktig. Derfor er det overraskende at det dårligste samsvaret var mellom de to erfarne PGMI radiografene, A og B, med bare litt samsvar ($\kappa = 0,34$).

En objektiv klassifisering er strengt å forholde seg til oppgitte kriterier og gi klassifiserings skår etter dette. Ved å bruke Wilcoxon signed rank test vises det at PGMI radiografene brukte klassifiseringsskalaen noe forskjellig. Den eksterne erfarne, B, og den interne uerfarne, C, brukte klassifiseringsskalaen forholdsvis likt, med en p verdi $p < 0,224$ for CC og MLO samlet og $p < 0,993$ for MLO bildene. Resultatene sier ikke om PGMI radiografene B og C er strengere eller mildere i sin klassifisering enn radiograf A, men at de brukte klassifiseringsskalaen likere enn de gjør i forhold til PGMI radiograf A, og at de ikke spriker i sine klassifiseringer i forhold til hverandre. De to erfarne PGMI radiografene bruker klassifiseringsskalaen forskjellig, $p < 0,005$ og det kan det tyde på at de har utviklet sine egne normer for hva de aksepterer eller ikke. De kan ha dannet seg egne oppfatninger om klassifiseringsarbeidet gjennom erfaringene sine. De erfarne kan være umotiverte og være lei av PGMI arbeid som de

har arbeidet med i lang tid og derfor være ukonsentrerte, unøyaktige og akseptere synlige mangler eller feil ved bildene. Alle PGMI radiografene kan være påvirket og preget av sine radiograf- og radiologkollegaer på arbeidsplassen, og lokale forhold og hva det enkelte BDS fokuserer på og vektlegger.

Uavhengig av erfaring og arbeidssted så er man forskjellige som personer. Noen er mer nøyaktige enn andre, noen aksepterer mer, er "snillere", enn andre og noen aksepterer mindre, er "strengere", enn andre. Resultatene kan påvirkes ved at en "snill" PGMI radiografene kan ha en tendens til hyppigere klassifisere bildene med bedre skår enn en "streng" PGMI radiograf. Motsatt kan det være at en "streng" PGMI radiograf hyppigere klassifiserer lavere enn en "snill" PGMI radiograf. PGMI radiografer kan være mer objektive enn andre og klassifiserer bilder i henhold til kriteriene med påfølgende skår, uten tanke for om det er en god eller dårlig skår. Andre kan derimot føle det ubehagelig og kvie seg for å gi dårlig skår hyppig. Undersøkelsene var ikke blindet for radiografnavn og det var mulig å se hvem som hadde tatt bildene. Det kan påvirke de interne PGMI radiografene som klassifiserer kollegaers arbeid. Alle PGMI radiografene kan ubevisst være strengere eller snillere enn de ellers ville vært ved en ordinær PGMI klassifisering fordi de deltar i en studie. Deres motivasjoner og interesse for å delta kan derfor påvirke klassifiseringen. De kan bevisst eller ubevisst ha et ønske om å hjelpe prosjektleder, de kan føle det positivt å bli spurt, få et avbrekk i arbeidsdagen, føle seg presset, ha en faglig interesse eller ha andre grunner for å delta som PGMI radiograf. Fordi de er deltagere i et prosjekt kan de være opptatt av å bruke kriteriene i kvalitetsmanualen korrekt og de kan være motivert eller umotivert for arbeidet.

PGMI klassifiseringen er et tidkrevende arbeid og det er viktig at PGMI radiografene får avsatt nødvendig arbeidstid og tilgang til arbeidsstasjoner. I en travel arbeidshverdag kan PGMI klassifiseringen bli nedprioritert i forhold til andre arbeidsoppgaver. I denne studien fikk PGMI radiografene tilgang til arbeidsstasjonen når de ikke var i bruk av andre, slik at de ikke skulle bli forstyrret i arbeidet eller føle press til å arbeide raskere enn de ønsket. PGMI klassifiseringen ble utført delvis i og delvis etter arbeidstid. Det er derfor vanskelig å si om hvorvidt bruk av privat tid påvirket hvor raskt de utførte PGMI klassifiseringen og om dermed indirekte påvirke resultatet. Det ble ikke registrert tidsbruk til klassifiseringen.

Det finnes ingen krav, kurs eller utdanning for å bli en PGMI radiograf, hverken lokalt eller nasjonalt. Det kreves heller ingen test med oppfølginger i etterkant eller annen form for kvalitetssikring av PGMI radiografene for å sikre at bruken av kriteriene

er korrekte og objektive. Det er ingen krav til innrapporteringer av PGMI resultater fra BDS' ene til Kreftregisteret. Det er ingen sentral ekspertgruppe i regi av Kreftregisteret, som har et overordnet ansvar for å sikre kvaliteten på PGMI radiografene. Det er heller ingen innsending av bilder til Kreftregisteret for klassifisering eller utveksling av bilder for klassifisering mellom fylkene. Det er ikke organisert et faglig nasjonalt nettverk mellom PGMI radiografene ved de forskjellige BDS. PGMI radiografene arbeider derfor "alene" uten et faglig kollegialt nettverk som kan være med å forhindre lokale variasjoner og kvalitetssikre PGMI radiografenes arbeid i hvert BDS. Sporadisk kontakt mellom PGMI radiografene ved forskjellige BDS finnes naturligvis, men det er personavhengig og ikke i organisert form.

Til tross for at kvalitetsmanualen fastslår detaljerte kriterier for hver enkelt projeksjon i PGMI klassifiseringen, indikerer resultatene i denne studien at det likevel er subjektive oppfatninger blant PGMI radiografene. En årsak til dette kan være hvordan lett asymmetri og asymmetri er beskrevet for CC bildene og for MLO bildene. Når f. eks. to MLO bilder hver for seg blir klassifisert som gode bilder, men har en asymmetri, så er det uklart beskrevet i Kvalitetsmanualen om begge bildene skal klassifiseres som moderate bilder eller om ett av bildene skal klassifiseres som godt og det andre som moderat. Det er derfor mulig at PGMI radiografene har forskjellige oppfatninger om dette og det kan medvirke til subjektive oppfatninger og lokale variasjoner.

Man skulle tro at erfaring i PGMI klassifisering var viktig, og det var derfor overraskende at det dårligste samsvaret var mellom de to erfarne PGMI radiografene og at det beste samsvaret var mellom de to interne radiografene. Det kan indikere at de individuelle vurderingene er påvirket av avdelingens praksis. Det kan derfor være behov for en mer detaljerte kriterier i Kvalitetsmanualen. Denne studien viser at det kan tyde på at det er behov for kurs, opplæring og trening i PGMI klassifiseringsarbeid. Et nasjonalt nettverk for PGMI radiografer kan være nyttig, likeledes kan innrapporteringer til Kreftregisteret, utveksling av bilder mellom BDS, og klassifisering av hverandres bilder være en mulighet for å sikre objektiv klassifisering av bildene og godt samsvar mellom PGMI radiografene uavhengig av erfaring og arbeidssted. En stillingstittel som PGMI radiograf ved alle BDS kan vurderes og bør være knyttet til obligatoriske kurs eller utdanning.

Blant de 120 valgte screeningundersøkelsene kan det være bilder som er svært vanskelige å klassifisere, selv for et ekspertpanel, og det kan påvirke resultatene.

Kvalitetssikring og forbedringsarbeid ved de enkelte BDS tar utgangspunkt i de individuelle PGMI resultatene og i avdelingens totale PGMI resultat. For at forbedringstiltakene skal være hensiktsmessige, målrettede og være med på å bidra til forbedringer og god kvalitet, er det viktig at PGMI resultatene er objektive og er til å stole på. Ved gal eller mangelfullt utført PGMI klassifisering kan de forbedringstiltakene som velges, være feil og påvirke screeningradiografene, og dermed bildekvaliteten, negativt. Tiltak som bidrar til å opprettholde gode rutiner og god kvalitet på bildene er også påvirket av en objektiv PGMI klassifisering slik at riktige tiltak blir iverksatt. Det er viktig at det utføres PGMI klassifiseringer etter at det er iverksatt forbedringstiltak og at denne PGMI klassifiseringen også er objektiv slik at man korrekt kan sammenligne resultatene med forutgående PGMI skår. På den måten er det mulig å følge kvalitetssikringsarbeidet over tid.

En svakhet ved studien er at bildene ikke var PGMI klassifisert av erfarne PGMI radiografer i konsensus i forkant eller i etterkant av studien som en fasit og at studieresultatene dermed kunne ha blitt målt opp mot en fasit. Et konsensusmøte med de tre PGMI radiografene i etterkant av studien og hvor alle bildene ble gått gjennom og en skår kunne vært brukt som fasitsvar. For å måle intra-variabiliteten eller hvor likt hver enkelt PGMI radiograf PGMI klassifiserer bilder, måtte de samme bildene blitt PGMI klassifisert på nytt 8 uker eller senere, etter den første klassifiseringen, for å sikre at PGMI radiografen ikke husker sin egen klassifisering av bildene. En studie med flere PGMI radiografer fra flere fylker ville vært av interesse for å se om tendensen er tilsvarende ved andre BDS.

Bildene som ble klassifisert var ikke anonymisert på pasient- eller radiografnavn. Pasientnavn kom automatisk på bildene og for å få frem radiografnavnet må man utføre en enkel tilleggs operasjon på arbeidsstasjonen. Det er usikkert om PGMI radiografene tok seg tid til dette og derfor er det usikkert i om og i hvilken grad dette kan ha påvirket resultatene. Sannsynligvis vil det bare påvirket de interne PGMI radiografenes klassifisering som kjenner radiografenes signatur.

En videre begrensning i studien er antall PGMI radiografer. Tre PGMI radiografer, to interne og en ekstern, er et for lite antall PGMI radiografer til å kunne trekke generelle konklusjoner. Studiens resultater indikerer derimot en trend som det kan være av interesse å studere nærmere. Et større antall PGMI radiografer og fra flere fylker ville styrket studiens validitet og reliabilitet.

6. KONKLUSJON

Resultater fra denne studien indikerer at det er betydelig variasjon i klassifiseringen av mammografibilder mellom PGMI radiografene, og at erfaring i PGMI klassifisering er av mindre betydning. En revisjon og gjennomgang av kvalitetsmanualens kriterier for PGMI kan derfor være aktuelt. Studien viser at det er behov for mer samkjøring og standardisert bruk av PGMI. Likeledes bør det vurderes om Kreftregisteret skal ha et overordnet ansvar for innrapportering av PGMI resultater og ha en jevnlig kvalitetsvurdering av PGMI radiografer. Et faglig nettverk mellom PGMI radiografene kan være nyttig. Det kan tenkes at kurs og utdanning for PGMI radiografer i tillegg til en stillingstittel som PGMI radiograf samt tid til PGMI klassifiseringsarbeidet kan være viktig. Objektivitet i PGMI klassifiseringen bør tilstrebes slik at resultatene i minst mulig grad er påvirket av subjektive oppfatninger og lokale variasjoner.

7. REFERANSELISTE

- (1) Mammografiprogrammet: kvalitetsmanual. Oslo: Kreftregisteret; 2003.
- (2) UK. NHSBSP 63: Quality assurance guidelines for mammography: Including radiographic quality control. NHS Cancer Screening Programmes; 2006.
- (3) Guidelines for Quality Assurance in Mammography Screening. 3. edition ed. The National Cancer Screening Service Board, Ireland; 2008.
- (4) Guidelines for Quality Assurance in Mammography Screening, 2008.
<http://www.breastcheck.ie/sites/default/files/bcheck/documents/bc-qa-guidelines.pdf>
- (5) Tabár L, Vitak B, Chen TH-H, Yen AM-F, Cohen A, Tot T, et al. Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality during 3 Decades. *Radiology* 2011 Sep 1;260(3):658-63.
- (6) Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001 May 1;91(9):1724-31.
- (7) van Schoor G., Moss SM, Otten JD, Donders R, Paap E, den Heeten GJ, et al. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011 Mar 15;104(6):910-4.
- (8) Bassett LW, Farria DM, Bansal S, Farquhar MA, Wilcox PA, Feig SA. Reasons for failure of a mammography unit at clinical image review in the American College of Radiology Mammography Accreditation Program. *Radiology* 2000 Jun;215(3):698-702.
- (9) Hofvind S, Vee B, Sørnum R, Hauge M, Ertzaas AK. Quality assurance of mammograms in the Norwegian Breast Cancer Screening Program. *Radiography* 2009 Mar;1(1):22-9.
- (10) Spuur K, Poulos A. Evaluation of the pectoral muscle in mammography images: The Australian experience. *Eur J Radiography* 2009 Mar;1(1):12-21.
- (11) Taplin SH, Rutter CM, Finder C, Mandelson MT, Houn F, White E. Screening mammography: clinical image quality and the risk of interval breast cancer. *AJR Am J Roentgenol* 2002 Apr;178(4):797-803.
- (12) Spuur K, Hung WT, Poulos A, Rickard M. Mammography image quality: Model for predicting compliance with posterior nipple line criterion. *Eur J Radiol* 2010 Jul 9.
- (13) Gullien R, Haakull AE, Andersen JG, Rostad E, Skaane P. PGMI classification of full-field digital mammograms in the Norwegian population-based breast cancer screening program according to the radiographers experience. *Eur Radiol* 2008;18: (Suppl. 1): 262.
- (14) Gullien R, Rostad E, Haakull AE, Andersen JG, Schou-Bredal I, Strømme H. Are there improvements in PGMI results after implementations of different

- improvement initiatives? Follow-up project (ECR 2008). Eur Radiol 2009;19: (Suppl. 1): 292.
- (15) Gullien R, Andersen JG, Haakull AE. Identifying the most common deviations in mediolateral-oblique (MLO) mammograms classified as "Moderate" before and after implementation of improvement initiatives. Insight Imaging 2010;1: (Suppl. 1): 18.
 - (16) Gur D, Abrams GS, Chough DM, Ganott MA, Hakim CM, Perrin RL, et al. Digital Breast Tomosynthesis: Observer Performance Study. Am J Roentgenol 2009 Aug 1;193(2):586-91.
 - (17) Moran S, Warren-Forward H. Assessment of the willingness of radiographers in mammography to accept new responsibilities in role extension: Part one ΓÇô Quantitative analysis. Radiography 2011 Nov;17(4):270-4.
 - (18) Moreira C, Svoboda K, Poulos A, Taylor R, Page A, Rickard M. Comparison of the validity and reliability of two image classification systems for the assessment of mammogram quality. J Med Screen 2005;12(1):38-42.
 - (19) Feig SA. Image quality of screening mammography: effect on clinical outcome. AJR Am J Roentgenol 2002 Apr;178(4):805-7.
 - (20) OUS. OUS Strategi 2013 - 2018. OUS Strategi 2013-2018 2013 Available from: URL: <http://www.oslo-universitetssykehus.no/SiteCollectionDocuments/Om%20oss/Mål%20og%20strategier/Strategi-OUS-2013-2018.pdf>
 - (21) Aalen OO, Frigessi A. Statistiske metoder i medisin og helsefag. Oslo: Gyldendal akademisk; 2006.
 - (22) Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.
 - (23) Cohen J. Nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;(70, 213).
 - (24) Dalland O. Metode og oppgaveskriving for studenter. Oslo: Gyldendal akademisk; 2012.
 - (25) Malterud K. Kvalitative metoder i medisinsk forskning: en innføring. Oslo: Universitetsforl.; 2011.
 - (26) U.S.National Library of Medicine. MeSH. MeSH 2013 Available from: URL: <http://www.nlm.nih.gov/>
 - (27) Ruyter KW, Solbakk JH, FÅrde R. Medisinsk og helsefaglig etikk. Oslo: Gyldendal akademisk; 2007.
 - (28) Vancouverkonvensjonen, Etisk kommite. Vancouverkonvensjonen 2013 Available from: URL: <http://www.etikkom.no/FBIB/Praktisk/Lover-og-retningslinjer/Vancouverkonvensjonen/>

- (29) Etikkomiteen. Helsinkideklarasjonen. Helsinkideklarasjonen 2010 Available from: URL: <http://www.etikkom.no/Forskningsetikk/Etiske-retningslinjer/Medisin-og-helse/Helsinki-deklarasjonen/>
- (30) REK. REK helseforskning. https://helseforskning.etikkom.no/ikbViewer/page/forside?_ikbLanguageCode=n 2013
- (31) Forskningsetiske komiteer. Forskningsetikkloven. Forskningsetikkloven 2007 Available from: URL: <http://www.etikkom.no/FBIB/Praktisk/Lover-og-retningslinjer/Forskningsetikkloven/>
- (32) Datatilsynet personvernloven. Datatilsynet 2013 January 2 Available from: URL: <http://www.datatilsynet.no/>
- (33) Vancouverkonvensjonen, ICMJE. International Committee of Medical Journal Editors, ICMJE 2013 January 2 Available from: URL: www.icmje.org
- (34) Universitetet i Bergen. Etikk i utdanningen, UiB. <http://www.uib.no/utdanning/om-aa-studere/etikk-i-utdanningen> 2013

8. VEDLEGG

Vedlegg 1.

Kvalitetsmanualen bildekriterier.

<p>P: Perfekte bilder</p> <ol style="list-style-type: none">1. Samtlige projeksjonskriterier er oppfylt2. Korrekt merking av bildet3. Optimal eksponering4. Optimal kompresjon5. Fravær av bevegelse6. Ingen hudfolder7. Symmetriske bilder	<p>M: Moderat gode bilder</p> <ol style="list-style-type: none">1. Korrekt merking av bildet2. Optimal eksponering3. Optimal kompresjon4. Fravær av bevegelser <p>Mangler:</p> <ol style="list-style-type: none">5. Pektoralismuskel ikke til brystvortenivå6. Brystvorten ikke i profil7. Submammære vinkel ikke fremstilt8. Hudfolder, store, ikke gjennomeksponert.9. Ikke symmetri
<p>G: Gode bilder</p> <ol style="list-style-type: none">1. Samtlige projeksjonskriterier er oppfylt2. Korrekt merking av bildet3. Optimal eksponering4. Optimal kompresjon5. Fravær av bevegelse <p>*) Mangler:</p> <ol style="list-style-type: none">6. Små gjennomsiktige hudfolder/ rynker7. Lett asymmetriske bilder	<p>I: Inadekvate bilder</p> <p>Mangler:</p> <ol style="list-style-type: none">1. Deler av brystet ikke fremstilt2. Feil eksponering3. Inadekvat kompresjon4. Bevegelsesuskarphet5. Bilder er ikke navnet.

Krav til PGMI:

$P + G \geq 75 \%$

$M < 22 \%$

$I < 3 \%$

Ref: Kvalitetsmanualen, kapittel 5.7.3, tabell 5.

Vedlegg 2.

PGMI registreringsskjema.

Radiograf navn:

	Us.dato	F. år	Initialer	R MLO skår	L MLO skår	R CC skår	L CC skår
1.							
2.							
3.							
4.							
5.							
6.							
7.							
8.							
9.							
10.							
11.							
12.							
13.							
14.							
15.							
16.							
17.							
18.							
19.							
20.							
21.							
22.							
23.							
24.							
25.							

Vedlegg 3.

Godkjenning fra forskningsutvalget OUS.

---Opprinnelig melding----

Fra: Faleide Evi
Sendt: 18. januar 2010 12:43
Til: Skaane Per
Emne: Registrering av forskingsprosjekter i FASUUS

Til orientering:

Prosjekt med tittel: "Tomosyntese studien. Tomosynthese og CAD som supplement til konvensjonell digital mammografi for påvisning av forandringer i det kvinnelige brystet. " er registrert i FASUUS med projektnr. 1460.

For å få tilgang til forskningsserver må prosjektet ha tilrådning fra personvernombudet. Når tilrådingen er mottatt, kontakt Nihal D.Perera (pnih@ullevål.no) for tilgang.

Hilsen Evi Faleide

Forsknings- og utdanningsavdelingen
Oslo universitetssykehus, Ullevål

IKKE SENSITIVT INNHOLD

Vedlegg 4.

Godkjenning fra personvernombudet.

Oslo universitetssykehus HF

PERSONVERNOMBUDETS UTTALELSE OM FORSKNINGSPROSJEKT

Til: Per Skaane

Kopi: Randi Gullen

Fra: Personvernombudet

Saksbehandler: Heidi Thorstensen

Dato: 19/4-2010

Offentlighet: Ikke unntatt offentlighet

Sak: Personvernombudets uttalelse om innsamling og databehandling av personopplysninger i forskningsstudie

Saksnummer/
Personvernnummer: 1460

Personvernombudets uttalelse til innsamling og databehandling av personopplysninger for forskning i prosjektet ” Tomosyntese studien”

Viser til innsendt melding om behandling av personopplysninger / helseopplysninger. Det følgende er et formelt svar på meldingen. Forutsetningene nedenfor må være oppfylt før rekruttering til studien kan starte.

Personvernombudet har vurdert det til at den planlagte databehandlingen av personopplysninger / helseopplysninger tilfredsstillende de krav som stilles i helse- og personvernlovgivningen.

Personvernombudet har ingen innvendinger til at den planlagte databehandlingen av personopplysninger / helseopplysninger kan igangsettes under forutsetning av følgende:

1. Behandling av personopplysningene / helseopplysninger i studien skjer i samsvar med og innenfor det formål som er oppgitt i meldingen (se vedlagte meldeskjema).

2. Studien må vurderes og godkjennes av Regional komité for medisinsk og helsefaglig forskningsetikk (REK), og eventuelle merknader må følges. Kopi av anbefaling fra personvernombudet vedlegges søknaden til REK. REKs uttalelse sendes i retur til undertegnede.
3. Vedlagte samtykke benyttes. Eventuelle endringer til dette som berører formålet, utvalget inkluderte eller databehandlingen må forevises personvernombudet før de tas i bruk.
4. Data lagres som oppgitt i meldingen.
5. Kodeliste som kobler aidentifiserte data med personopplysninger lagres på papir, nedlåst på prosjektleders avlåste kontor. Listen lagres fysisk adskilt fra andre registerdata.
6. Data slettes eller anonymiseres senest 31/12-2015. Kodelisten makuleres innen denne tid.
7. Dersom formålet, utvalget av inkluderte eller databehandlingen endres må personvernombudet gis forhåndsinformasjon om dette.
8. Kontaktperson for prosjektet skal hvert tredje år sende personvernombudet ny melding som bekrefter at databehandlingen skjer i overensstemmelse med opprinnelig formål og helseregisterlovens regler.

Studien er registrert i sykehusets offentlig tilgjengelig database over forsknings- og kvalitetsstudier.

Lykke til med studien!

Med vennlig hilsen

Heidi Thorstensen
personvernombud/IKT-sikkerhetssjef
Kompetansesenter for personvern og sikkerhet
Oslo universitetssykehus HF
www.uus.no/personvern

Vedlegg 5.

Godkjenning fra regional komité for medisinsk og helsefaglig forskningsetikk Sør-Øst.



UNIVERSITETET I OSLO
DET MEDISINSKE FAKULTET

Vedlegg 5.
Godkjenning fra REK.

Overlege Per Skaane
Oslo universitetssykehus
Brystdiagnostisk senter
Kirkevn. 166
0407 Oslo

**Regional komité for medisinsk og helsefaglig
forskningsetikk Sør-Øst D (REK Sør-Øst D)**
Postboks 1130 Blindern
NO-0318 Oslo

Telefon: 22 85 05 93
Telefaks: 22 85 05 90

E-post: i.m.middelthon@medisin.uio.no
Nettadresse: www.etikkom.no

Dato: 16.02.10
Deres ref.:
Vår ref.: 2010/144

Digital brysttomosyntese (DBT): Snittbilder i brystdiagnostikken

Vi viser til søknad av 04.01.10 for det ovenfor nevnte forskningsprosjekt.

Prosjektleder er professor dr. med. Per Skaane.

Forskningsansvarlig er Oslo universitetssykehus, avd. Ullevål ved øverste administrative ledelse.

Prosjekttema:

Tomosyntese (DBT), en 3-dimensjonal rekonstruksjon av tynne snittbilder gjennom brystet, som et supplement til mammografi. Dette forventes å bedre brystkreftdiagnostikken, både ved å påvise flere krefttilfeller og å redusere falske positive funn i screening. Computer-assistert deteksjon (CAD) er en annen avansert teknikk ved digital mammografi som analyserer bilder og markerer suspekter forandringer. Formålet med studien er i fase 1 å evaluere tomosyntese hos kvinner til utredning for symptomer, høyrisikokvinner eller suspekter mammografifunn. I fase 2 vil DBT og CAD bli anvendt hos asymptotiske kvinner som møter til mammografi-screening for å bedre resultatene.

Vedtak:

Komiteen har vurdert søknaden og godkjenner prosjektet med hjemmel i helseforskningsloven § 10. Det knytter seg imidlertid vilkår til godkjenningen som må oppfylles før prosjektet settes i gang.

I tillegg til vilkår som fremkommer av dette vedtaket er tillatelsen gitt under forutsetning av at prosjektet gjennomføres slik det er beskrevet i søknaden, protokollen og de bestemmelser som følger av helseforskningsloven med forskrifter.

Vilkår vedrørende informasjonsskrivet:

Da det ikke er aktuelt å innhente stedfortredende samtykke i denne studien må dette tas ut av samtykkeerklæringen.

Vilkår vedrørende informasjonssikkerhet:

Komiteen forutsetter at data lagres på et nettverk som er godkjent for dette og under kontroll av prosjektets forskningsansvarlige institusjon. Forskningsprosjektets data skal oppbevares forsvarlig, se personopplysningsforskriften kapittel 2, og Helsedirektoratets veileder for «Personvern og informasjonssikkerhet i forskningsprosjekter innenfor helse- og

omsorgssektoren», <http://www.norsk-helsenett.no/informasjonsikkerhet/bransjenormen/Personvern%20og%20informasjonssikkerhet%20i%20forskningsprosjekter%20v1.pdf>

Tillatelsen gjelder til 31.12.2016. Av dokumentasjonshensyn skal opplysningene likevel bevares inntil 31.12.2021. Opplysningene skal lagres aidentifisert, dvs. adskilt i en nøkkel- og en opplysningsfil. Opplysningene skal deretter anonymiseres eller slettes.

Prosjektet skal sende sluttmelding til REK Sør-Øst D senest 31.06.2017.

Komiteens vedtak kan påklages til Den nasjonale forskningsetiske komité for medisin og helsefag, jf. forvaltningsloven 28 flg. Eventuell klage kan sendes til REK Sør-Øst D. Klagefristen er tre uker fra mottak av dette brevet.

REK har gått over til elektronisk saksbehandling og fått ny saksportal: <http://helseforskning.etikkom.no>. Vi ber om at svar på merknader og henvendelser til REK sendes inn via denne portalen eller på epost: post@helseforskning.etikkom.no. Vennligst oppgi REKs saksnummer.

Med vennlig hilsen

Stein A. Evensen (sign.)
Professor dr.med.
leder

Ingrid Middelthon
seniorrådgiver

Kopi:
Oslo universitetssykehus, v/adm.dir. Siri Hatlen

Randi Gullien
Breast Imaging Centre
Department of Medical Imaging and Intervention
Radiology and Nuclear medicine
Oslo University Hospital
Kirkeveien 166
NO – 0407 Oslo
Norway

To Editor of Radiography.

**Re: PGMI classification of mediolateral-oblique screening mammograms:
Agreement among mammography radiographers.**

I hereby confirm that this manuscript is not and has not been under consideration for publication elsewhere, but a part of this work has been published as an oral presentation at European Congress of Radiology (ECR) in 2011.

There are no conflicts of interest.

This manuscript is a part of my master degree work.

On the following pages, please find the manuscript.

I hope you accept the article. Please contact me if you have any questions or if you want some modifications / improvements.

Sincerely,

Randi Gullien.

Title

PGMI classification of mediolateral-oblique screening mammograms: Agreement among mammography radiographers.

Randi Gullien

Author details

Randi Gullien, Senior Radiographer, Breast Imaging Centre, Department of Medical Imaging and Intervention, Radiology and Nuclear medicine, Oslo University Hospital, Norway

Original research

Corresponding authors:

Senior radiographer Randi Gullien
Breast Imaging Centre
Department of Medical Imaging and Intervention
Radiology and Nukleaer medicine
Oslo University Hospital
Kirkeveien 166
N - 0407 Oslo / Norway
Phone: + 47 - 23 01 65 22
Fax: + 47 - 23 01 65 35
E-Mail: uxraul@ous-hf.no

Title

PGMI classification of mediolateral-oblique screening mammograms: Agreement among mammography radiographers.

Introduction

Inter-observer agreement among three mammography radiographers were evaluated using the PGMI classifications recommended in the quality assurance manual of the Norwegian Breast Cancer Screening Program (NBCSP). The three radiographers, both internal and external, had different experience in PGMI classification.

Methods

Three radiographers, one internal experienced (A), one external experienced (B) and one internal inexperienced (C) in PGMI classification, independently classified MLO images from 120 examinations using the national guidelines for PGMI classification. The examinations were randomly selected from week 34 and 35 in 2010. The PGMI classifications were transformed into a numerical value: P=perfect=1; G=good=2; M=moderate good= 3; and I=inadequate=4. Cohen's Kappa statistic was used to measure inter-observer agreement.

Results

The kappa values were fair to moderate. Agreement of A vs. B, B vs. C, A vs. C were $\kappa = 0.34, 0.44$ and 0.45 , respectively. The highest agreement was between the two internal radiographers (A vs. C) and the lowest between the two experienced PGMI radiographers (A vs. B).

Conclusion

Different experiences, subjective opinions, and local variations regarding interpretation may influence the PGMI classification despite national guidelines. Our results indicate a lack of consequent use of the PGMI classification system. Training, examinations and exercise may be necessary to ensure uniform use of the PGMI classification.

Keywords:

PGMI, mammography, breast imaging, quality assurance, radiography, kappa statistic.

Introduction

Breast cancer is the most common cancer among women worldwide and early detection of (small) cancers is very important to reduce the mortality (1-3).

Mammography is the most widely used examination for detecting breast cancer. High quality mammograms are a basic element to improve the breast cancer detection rate, to reduce interval cancers and to reduce mortality. Several key components are mandatory for high quality mammograms, including technical aspect related to the mammography unit as well as the radiographer's positioning of the patient, compression of the breast, and the selection of correct exposure (4;5).

The Norwegian Breast Cancer Screening Program (NBCSP) is a governmentally organized population-based mammography-screening program and is administrated by the Cancer Registry in Norway. The program started in 1996 and was nationwide in 2005. NBCSP invites biannually women aged 50 – 69 years which are resident in Norway (6). All counties in Norway have dedicated mammography-screening centres and breast-imaging centre (BIC). The BIC is responsible for the follow-up performance of the mammography-screening examinations.

To avoid mammograms of low quality, guidelines have been developed nationally and internationally (6-8). In Norway, the NBCSP has developed a quality assurance manual for all professions within the mammography-screening program (6). This quality assurance manual is based on European Guidelines and the guidelines for the National Health Service Breast Screening Program (NHSBSP) (7;8). One of the chapters in the Norwegian quality assurance manual is dedicated to the radiographers and all parts of the radiographers work. One section describes specific criteria and recommendations regarding the projection and to the quality of the mammograms and contains the images criteria for both cranio-caudal (CC) and mediolateral-oblique (MLO) images. The NBCSP decided to use the mammography classification system called PGMI classification system (6). PGMI is an acronym where P is a perfect mammogram, G is a good mammogram, M is a moderately good mammogram and I is an inadequate mammogram. The quality assurance manual states that P and G images together should be equal or better than 75 %, at least 22 % of the images should be of moderately good quality and not more than 3 % should be inadequate images (Table 1). The PGMI classification system is not a diagnostic evaluation of the mammograms, but

a quality classification of each of the standard mammograms. Dedicated radiographers at each breast-imaging centre, called PGMI radiographers, perform the PGMI classifications of screening mammograms at their own centre. Despite guidelines, the classification system may be variable, subjective and interpreted locally, making comparison of PGMI classification important (9;10).

The aim of this study is to analyse any differences between the PGMI radiographers in their PGMI classification work when they independently assess the same set of screening mammograms. Previous studies have pointed out that there is a lack of agreement between radiographers when classifying mammograms despite guidelines with specific criteria for the classification work (9;10). Subjective opinions and different knowledge and experience in the classification work might influence the PGMI results. This article will focus on agreement between PGMI radiographers when assessing MLO images. MLO projection is the projection that captures most of the breasts tissue and therefore it is very important in order to have high quality on the MLO views.

Methods

This study is part of a large-scaled mammography-screening project where all participants gave informed consent. The Regional Committees for Medical and Health Research Ethics approved the study. All examinations were performed at the mammography-screening centre in Oslo and full-field digital mammographic examinations (FFDM) were performed on Selenia Dimensions (Hologic, Inc; Bedford, MA, USA).

A total of 120 randomly selected mammographic examinations were selected. These examinations had been carried out in week 34 and 35 of 2010, and consisted of CC and MLO views of both breasts. The PGMI classification criteria of the Quality Assurance Manual for the NBCSP were used for classifying mammograms (Table 2) (6). The PGMI categories were transformed into a numerical value, where P is a score of 1, G a score of 2, M a score of 3, and I a score of 4. Including only classification score from the MLO images, a total of 240 images were analysed.

Examinations with more than four images and examinations of women with breast implants or a foreign substance as pacemakers were excluded from the study.

PGMI criteria of MLO views

The MLO projection is the image that visualises and captures most of the breast tissue. Criteria used to assess the positioning of the breast are related to adequate visualization of the pectoral muscle, the inframammary angle (sub mammary angle), the nipple in profile, and spreading the glandular tissue (Table 3). The pectoral muscle is an area of importance on a well-positioned mammogram although the amount and orientation can be discussed (11). Positioning failure is mostly depending on the performing radiographer (4;5). A perfect positioning of the breasts requires a highly skilled performance by the radiographers. Perfect and proper positioning of the breast occasionally requires modification of technique due to the woman's habitus. To perform such modifications depends on the radiographers knowledge, training and experience in mammography screening. The Norwegian quality assurance manual states that mammograms classified as perfect must contain correct labelling and exposure, optimal compression of the breast, visualisation of the inframammary angle, the pectoral muscle to the nipple level, the nipple in profile, and no skin folds (Figure 1). There must be symmetric images when comparing with the contralateral MLO mammogram (Table 2 and 3).

PGMI radiographers

Three screening mammography radiographers, called PGMI radiographers, carried out the PGMI interpretations. One PGMI radiographer (A) was internal with long experience in mammography screening work and in PGMI classification work. The external PGMI radiographer (B) had long experience in both mammography screening and classification work. She worked in a screening centre in another county, but started to work at our screening centre two week before the study started. She was considered to be an external PGMI radiographer. The third PGMI radiographer (C) was inexperienced in PGMI classification, but was experienced in mammographic examinations. PGMI radiographer A and C came from the same centre.

The classification was carried out on dedicated PACS workstations with 5K screens and specific work list. All three PGMI radiographers used the same workstation for the classification work. The PGMI radiographers reviewed and classified the MLO images for each exam using the recommendation of the NBCSP. The classification

score for each image was recorded on an onsite form (Table 4). Previous examinations were not used for comparison.

In order to reduce bias between the PGMI radiographers, the classification was done independently on given date and time on different workdays. The classification work was carried out during week 38 to week 40, 2010.

Data analysis

To measure strength of agreement, Cohen's Kappa, κ , were used. Cohens Kappa grading is shown in table 5 (Table 5) (12). The kappa scores were estimated on the 4-point rating scale for PGMI classifications as well as on a 3-level scale of the collapsed 4-point scale. For this purpose the PGMI score for P ($P = 1$) and for Good ($G = 2$) was merged into one category. Kappa values were evaluated for all MLO views and for right and for left MLO views separately.

The PGMI scores were analysed using the SPSS spread sheet (SPSS SmartViewer version 15.0 for Windows XP, SPSS Inc., Chicago, Illinois).

Results

Each PGMI radiographer classified the 240 MLO images using the criteria of the NBCSP guidelines in the quality assurance manual (6). The completed forms were returned to project leader and coded in a spreadsheet. It was not discovered any errors or any inadequate forms when coding the results.

Kappa value on the 4-point rating scale for all MLO views was fair to moderate (range 0.34 – 0.45) (Table 6). The kappa value for the 3-level scale for all MLO views was moderate (range 0.46 – 0.57) (Table 6).

The MLO images of right breast (R-MLO) and left breasts (L-MLO) were classified independently. Kappa value on the 4-point rating scale for the R-MLO views was fair to moderate (range 0.35 – 0.46) (Table 7). The kappa value for the 3-level scale for the R-MLO views was moderate (range 0.42 – 0.49) (Table 7). The kappa value for the 4-point rating scale for the L-MLO views was fair to moderate (range 0.33 – 0.50) (Table 8). The kappa value for the 3-level scale for the L-MLO views was moderate to good (range 0.50 – 0.64) (Table 8).

Regardless the use of the 4-point rating scale or the 3-level scale, the highest agreement, moderate to good agreement, was between the two internal PGMI

radiographers, A and C. The lowest agreement, fair to moderate agreement, was between the two experienced PGMI radiographers, A and B. There was a moderate agreement between the external experienced PGMI radiographer, B, and the internal inexperienced PGMI radiographer, C. The result from this study indicates that PGMI classification varies among the PGMI radiographers despite experience.

Discussion

The Norwegian quality assurance manual has recommendations and requirements for the quality of the mammographic examinations and for the radiographers. There are in addition criteria for the use of PGMI classification and recommendations for classification score (6). Based on this, it was expected that experience in classification work would be of importance and that the highest agreement might be expected between the two experienced PGMI radiographers (A and B). Surprisingly, the lowest agreement, fair to moderate agreement was found between those radiographers. This indicates that experience in PGMI classification work is of less value than expected. It might also be a difference in the understanding in how to classify mammograms and that different local variation has been developed in the two counties. Local trends might explain why the highest agreement was found between the two internal radiographers, A and C, despite their difference of experience in PGMI classification work. It seems to be possible that individual opinion and interpretation and local trend in classification might be of more importance than suggested.

Factors, which may affect the result of the PGMI classification, include how the classification system is used and how it is interpreted. Furthermore, the training and education of the PGMI radiographers are important. There is no postgraduate training for the radiographers to work within mammography screening program and there is no organized education or training for the PGMI radiographers in the classification work. There is no national network group of PGMI radiographers although that might be useful. A possible reason for the low agreement might be a lack of concise and detailed instruction for the classification systems in the quality assurance manual that leads to inter-operator variability. This study might be an input in the discussion of the use of PGMI as quality assessment, and it seems to be useful to require a revision of the criteria and the use of PGMI as a quality assessment tool to avoid inter-observer variability. It should be discussed if it might be useful to include even more detailed

criteria's, guidelines and training and evaluation of the PGMI radiographers in the classification work.

There are some limitations of this study. A small number of images were included and it would be preferable if even more mammograms had been PGMI classified. It would be desirable and useful with more than three PGMI radiographers included in the study to classify the images. The two internal PGMI radiographers classifies their colleagues mammograms and it might influence the classification. The mammograms in this study were randomly selected and therefore it is not certain they contained images from all classification groups.

Conclusion

Inter-observer agreement of the PGMI classification in this study varied between fair and good according to the Kappa statistic. The agreement was highest between the two internal PGMI radiographers and lowest between the two experienced PGMI radiographers. This indicates that experience in PGMI classification work is of less importance. There is a need for the PGMI classification to have a balanced and consistent quality in the classification work, since there is room for individual assessments and local variations despite the guidelines. An audit and review of the quality assurance manual criteria PGMI may be applicable, and it is obviously a need for more detailed criteria. It is possible that the responsibility for reporting PGMI quality results regularly should be a part of the ongoing quality assurance work of a mammography screening program. A network among PGMI radiographers nationwide might be useful. It is possible that courses and education as PGMI radiographers might be of importance. It is of uttermost importance that the PGMI classification results in a minimum of subjective and local variations.

Acknowledgements

The author gratefully acknowledges Per Skaane and Jack G. Andersen at the Department of Radiology at Oslo University Hospital, for their invaluable support and contribution. I am in debt to Anne Emilie Haakull, Eline Rostad and Robin Lee Hammond the same department for excellent support and to Leiv Sandvik at Centre of Clinical Research at Oslo University Hospital for important statistical support.

Reference List

- (1) Tabár L, Vitak B, Chen TH-H, Yen AM-F, Cohen A, Tot T, et al. Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality during 3 Decades. *Radiology* 2011 Sep 1;260(3):658-63.
- (2) Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001 May 1;91(9):1724-31.
- (3) van Schoor G., Moss SM, Otten JD, Donders R, Paap E, den Heeten GJ, et al. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011 Mar 15;104(6):910-4.
- (4) Bassett LW, Farria DM, Bansal S, Farquhar MA, Wilcox PA, Feig SA. Reasons for failure of a mammography unit at clinical image review in the American College of Radiology Mammography Accreditation Program. *Radiology* 2000 Jun;215(3):698-702.
- (5) Taplin SH, Rutter CM, Finder C, Mandelson MT, Houn F, White E. Screening mammography: clinical image quality and the risk of interval breast cancer. *AJR Am J Roentgenol* 2002 Apr;178(4):797-803.
- (6) Mammografiprogrammet: kvalitetsmanual. Oslo: Kreftregisteret; 2003.
- (7) UK. NHSBSP 63: Quality assurance guidelines for mammography: Including radiographic quality control. NHS Cancer Screening Programmes; 2006.
- (8) Guidelines for Quality Assurance in Mammography Screening. 3. edition ed. The National Cancer Screening Service Board, Ireland; 2008.
- (9) Hofvind S, Vee B, Sørnum R, Hauge M, Ertzaas AK. Quality assurance of mammograms in the Norwegian Breast Cancer Screening Program. *Radiography* 2009 Mar;1(1):22-9.
- (10) Moreira C, Svoboda K, Poulos A, Taylor R, Page A, Rickard M. Comparison of the validity and reliability of two image classification systems for the assessment of mammogram quality. *J Med Screen* 2005;12(1):38-42.
- (11) Bentley K, Poulos A, Rickard M. Mammography image quality: Analysis of evaluation criteria using pectoral muscle presentation. *Radiography* 2008 Aug;14(3):189-94.
- (12) Cohen J. Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;(70, 213.).

Table 1.
PGMI quality assessments for NBCSP.

Category	Quality assurance
P = Perfect	$P + G \geq 75 \%$ $M < 22 \%$ $I < 3 \%$
G = Good	
M = Moderately good	
I = Inadequate	

Table 2.

Overall PGMI classification criteria.

<p>P: Perfect</p> <p>All projection criteria fulfilled</p> <p>Correct labelling</p> <p>Optimal exposure</p> <p>Optimal compression</p> <p>Absence of movement</p> <p>Non translucent skin folds</p> <p>Symmetric images</p>	<p>M: Moderately good</p> <p>Correct labelling</p> <p>Optimal exposure</p> <p>Optimal compression</p> <p>Absences of movement</p> <p>Missing:</p> <p>Pectoral muscle not to nipple level</p> <p>Nipple not in profile</p> <p>Submammære vinkel ikke fremstilt</p> <p>Translucent skin folds</p> <p>Asymmetrical positioning</p>
<p>G: Good</p> <p>Correct labelling</p> <p>Optimal exposure</p> <p>Optimal compression</p> <p>Absences of movement</p> <p>Missing:</p> <p>Small translucent skin folds</p> <p>Slight asymmetrical positioning</p>	<p>I: Inadequate</p> <p>Missing:</p> <p>Incorrect labelling</p> <p>Exposure failure</p> <p>Inadequate compression</p> <p>Motion blur</p> <p>Parts of breast missing</p>

Table 3.

Criteria's for the MLO projections.

MLO projection criteria
Nipple in profile.
Pectoral muscle to the nipple level.
The inframammary area (Sub mammary area) is to be stretched from chest wall without non-translucent and overshadowing skin folds and perform a 90-degree angle with chest wall.
The breast exterior contour should be seen.
The women breast should be positioned so lymph nodes in axillary area will be visualized with a relaxed and well-compressed pectoral muscle.
Good compression.
Symmetry.
Non-translucent skin folds.

Table 5.

Cohens kappa.

Kappa (κ) value	Strength of agreement
< 0,20	Poor
0,20 – 0,40	Fair
0,41 - 0,60	Moderate
0,61 – 0,80	Good
0,81 – 1,00	Very good

Table 6.

Kappa score according to the agreement for the MLO views.

PGMI radiographer	PGMI radiographer	
	B	C
A	$\kappa 4 = 0.34$ $\kappa 3 = 0.46$	$\kappa 4 = 0.45$ $\kappa 3 = 0.57$
B		$\kappa 4 = 0.44$ $\kappa 3 = 0.49$
<p>$\kappa 4$ = kappa value on the 4–point rating scale for PGMI classification.</p> <p>$\kappa 3$ = kappa value on the 3–level scale of the collapsed 4–point rating scale for PGMI classification.</p>		

Table 7.

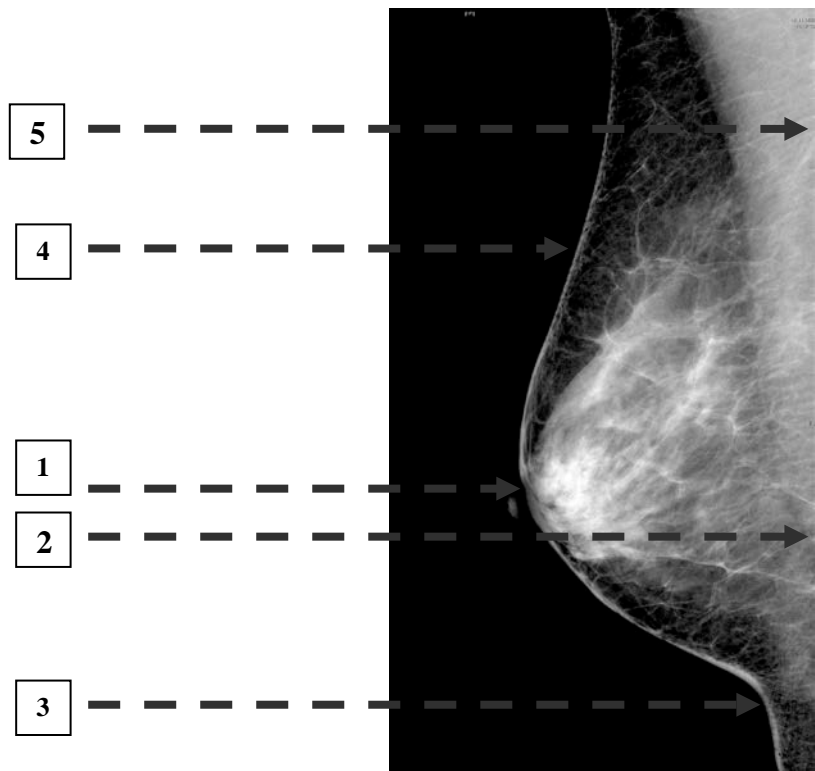
Kappa score according to the agreement for right MLO views (R-MLO).

PGMI radiographer	PGMI radiographer	
	B	C
A	$\kappa 4 = 0.35$ $\kappa 3 = 0.42$	$\kappa 4 = 0.40$ $\kappa 3 = 0.49$
B		$\kappa 4 = 0.46$ $\kappa 3 = 0.46$
$\kappa 4$ = kappa value on the 4–point rating scale for PGMI classification. $\kappa 3$ = kappa value on the 3–level scale of the collapsed 4-point rating scale for PGMI classification.		

Table 8.

Kappa score according to the agreement for the left MLO views (L-MLO).

PGMI radiographer	PGMI radiographer	
	B	C
A	$\kappa 4 = 0.33$ $\kappa 3 = 0.50$	$\kappa 4 = 0.50$ $\kappa 3 = 0.64$
B		$\kappa 4 = 0.43$ $\kappa 3 = 0.52$
<p>$\kappa 4$ = kappa value on the 4–point rating scale for PGMI classification. $\kappa 3$ = kappa value on the 3–level scale of the collapsed 4-point rating scale for PGMI classification.</p>		



MLO projection criteria
Nipple in profile.
Pectoral muscle to the nipple level.
The inframammary area (Sub mammary area) is to be stretched from chest wall without non-translucent and overshadowing skin folds and perform a 90-degree angle with chest wall.
The breast exterior contour should be seen.
The women breast should be positioned so lymph nodes in axillary area will be visualized with a relaxed and well-compressed pectoral muscle.
Good compression.
Symmetry.
Non-translucent skin folds.

Figure 1.

Perfect right MLO view with classification criteria.