

“CADEAT”: considerations on the use of CAD (computer-aided diagnosis) in mammography

“CADEAT”: considerazioni sull’impiego di CAD (computer aided diagnosis) in mammografia

Gruppo di studio GISMa – SIRM: R. Chersevani¹ · S. Ciatto² · C. Del Favero³ · A. Frigerio⁴
L. Giordano⁵ · G. Giuseppetti⁶ · C. Naldoni⁷ · P. Panizza⁸ · M. Petrella⁹ · G. Saguatti¹⁰

¹Sezione di Studio di Senologia, Società Italiana di Radiologia Medica, Milano, Italy

²ISPO, Istituto Scientifico per la Prevenzione Oncologica, Viale A. Volta 171, 50131 Firenze, Italy

³Fondazione Don Carlo Gnocchi IRCS, S.M. Nascente, Milano, Italy

⁴Centro di Riferimento Regionale Screening Mammografico, AOU S. Giovanni Battista, Torino, Italy

⁵Unità di Epidemiologia, CPO Piemonte, AOU S. Giovanni Battista, Torino, Italy

⁶Università Politecnica delle Marche - Facoltà di Medicina, Ancona, Italy

⁷Centro di Riferimento e coordinamento screening in oncologica. Assessorato alle politiche per la salute. Regione Emilia-Romagna, Bologna, Italy

⁸Dipartimento di Radiologia Ospedale San Raffaele, Milano, Italy

⁹Programma Aziendale Screening, AUSL 2 Umbria, Perugia, Italy

¹⁰Centro di Senologia, Presidio Ospedaliero Maggiore – AUSL, Bologna, Italy

Correspondence to: S. Ciatto, Tel.: +39-055-012275, Fax: +39-055-5001623, e-mail: stefano.ciatto@gmail.com

Received: 27 May 2009 / Accepted: 26 June 2009

© Springer-Verlag 2010

Abstract

Computer-aided diagnosis (CAD) has been extensively reported to increase sensitivity by about 10% when added to a single reading while increasing recall rate by 12%, and its current use can be safely recommended in clinical practice. CAD has been suggested as a possible alternative to conventional double reading in screening. Uncontrolled comparison is consistent and suggests that CAD is comparable to double reading in incremental cancer detection rate (CAD +10.6%, double reading +9.1%) and possibly better in recall rate (CAD +12.5%, double reading +28.8%). However, controlled studies comparing single reading + CAD to conventional double reading are not consistent and on average suggest a lower cancer detection rate (−5.1%) and a lower recall rate (−9.8%) for CAD. Scientific evidence is not sufficient for a safe recommendation of single reading + CAD as a current alternative to conventional double reading.

Keywords Breast cancer · Mammography · Diagnosis · Computer aided diagnosis

Riassunto

Esiste consistente evidenza scientifica che la computer aided diagnosis (CAD), aggiunta alla lettura singola, consenta un aumento della sensibilità di circa il 10%, con un aumento del tasso di richiamo di circa il 12%: l’uso corrente di CAD nella pratica clinica è pertanto raccomandabile. La singola lettura insieme a CAD è stata suggerita come una possibile alternativa alla doppia lettura convenzionale nello screening. Confronti tra studi non controllati suggeriscono consistentemente che CAD sia comparabile alla doppia lettura quanto a tasso diagnostico incrementale di carcinoma (CAD +10,6%, doppia lettura +9,1%) e possibilmente superiore quanto a tasso di richiamo (CAD +12,5%, doppia lettura +28,8%). Al contrario, un numero limitato di studi controllati che confrontano la singola lettura+CAD con la doppia lettura non mostra risultati consistenti e in media riporta per CAD un minor tasso diagnostico incrementale di cancro (−5,1%) e un minor tasso di richiamo (−9,8%). L’evidenza scientifica disponibile non è sufficiente per raccomandare la lettura singola + CAD come alternativa corrente alla doppia lettura convenzionale.

Parole chiave Carcinoma della mammella · Mammografia · Diagnosi · Diagnosi computer assistita

Computer-aided diagnosis (CAD) has been in use as an aid to mammography reading for at least two decades. The initial available scientific evidence on CAD performance was based on uncontrolled prospective and retrospective studies. Overall evidence consistently demonstrates that adding CAD to single conventional reading allows for 10% relative incremental cancer detection rate (Table 1) at the cost of 12% incremental relative recall rate (Table 2). Based on such evidence, CAD has been recommended as a useful aid to conventional reading. CAD had limited success in Europe, as it was first available in its troublesome analogic version (high costs, need to digitise films, continuous comparison with paper printouts), allowing no reimbursement from the national health system (in the US, appropriate reimbursement allowed for a larger diffusion of the method).

CAD performance, at least on the basis of uncontrolled studies, is comparable with conventional double reading (Table 3) as to incremental cancer detection rate (+10.6%

L'introduzione dell'utilizzo di computer aided diagnosis (CAD) in mammografia risale ormai ad alcuni decenni fa. Le prime evidenze scientifiche disponibili hanno riguardato studi non controllati, prospettici o retrospettivi, la cui analisi complessiva ha dimostrato in modo abbastanza consistente che CAD aggiunto alla lettura singola consente un aumento significativo della sensibilità, dell'ordine del 10% (Tabella 1), al costo di un aumento non elevato, ma statisticamente significativo, del tasso di richiamo ad approfondimento, dell'ordine del 12% (Tabella 2).

Sulla base di tale evidenza, CAD è stato consigliato come presidio utile da affiancare alla lettura singola. CAD ha avuto poco successo in Europa, sia perché all'inizio è stato disponibile solo nella versione indagativa applicabile alla mammografia analogica (costi elevati, necessità di un digitalizzatore, stampa di schemi cartacei da confrontare continuamente con le mammografie), sia perché non è stato oggetto di rimborso dai vari sistemi sanitari nazionali (SSN) (a differenza degli USA, dove un congruo rimborso

Table 1 Incremental cancer detection rate obtained by single reading + CAD

Study	Additional cancers/total cancers (%)
CSPO, Firenze (4 studies) ^a [1–4]	85/617 (+13.7%)
Freer 2001 ^b [5]	8/41 (+19.5%)
Helvie 2004 ^b [6]	1/10 (+10.0%)
Gur, 2004 ^b [7]	4/206 (+1.9%)
Khoo, 2005 [8]	2/61 (+1.3%)
Birdwell, 2005 ^b [9]	2/27 (+7.4%)
Cupples, 2005 ^b [10]	17/101 (+16.8%)
Ko 2006 ^b [11]	2/45 (+4.4%)
Morton, 2006 ^b [12]	8/105 (+7.6%)
Fenton, 2007 ^{b,c} [13]	– (+ 3.6%)
Total	129/1213 (+10.6%)

^aRetrospective study; ^bprospective study; ^ccohort comparison

Table 2 Incremental recall rate obtained by single reading + CAD

Study	Additional recalls/total recalls (%)
CSPO, Firenze (4 studies) ^a [1–4]	245/703 (+35.5%)
Freer 2001 ^b [5]	156/830 (+18.7%)
Helvie 2004 ^b [6]	34/344 (+9.8%)
Khoo, 2005 [8]	18/372 (+5.8%)
Birdwell, 2005 ^b [9]	73/887 (+8.2%)
Cupples, 2005 ^b [10]	164/2100 (+7.8%)
Ko 2006 ^b [11]	100/602 (+16.6%)
Morton, 2006 ^b [12]	191/1996 (+9.5%)
Fenton, 2007 ^{b,c} [13]	– (+ 3.0%)
Totale	981/7834 (+12.5%)

^aRetrospective study; ^bprospective study; ^ccohort comparison

Tabella 1 Tasso diagnostico aggiuntivo di carcinoma ottenuto mediante la singola lettura CAD assistita

Studio	Carcinomi aggiuntivi/totali (%)
CSPO, Firenze (4 studi) ^a [1–4]	85/617 (+13,7%)
Freer, 2001 ^b [5]	8/41 (+19,5%)
Helvie, 2004 ^b [6]	1/10 (+10,0%)
Gur, 2004 ^b [7]	4/206 (+1,9%)
Khoo, 2005 [8]	2/61 (+1,3%)
Birdwell, 2005 ^b [9]	2/27 (+7,4%)
Cupples, 2005 ^b [10]	17/101 (+16,8%)
Ko, 2006 ^b [11]	2/45 (+4,4%)
Morton, 2006 ^b [12]	8/105 (+7,6%)
Fenton, 2007 ^{b,c} [13]	– (+3,6%)
Totale	129/1213 (+10,6%)

^aStudio retrospettivo; ^bstudio prospettico; ^cconfronto tra coorti

Tabella 2 Tasso di richiamo aggiuntivo ottenuto mediante la singola lettura CAD assistita

Studio	Richiami aggiuntivi/totali (%)
CSPO, Firenze (4 studi) ^a [1–4]	245/703 (+35,5%)
Freer, 2001 ^b [5]	156/830 (+18,7%)
Helvie, 2004 ^b [6]	34/344 (+9,8%)
Khoo, 2005 [8]	18/372 (+5,8%)
Birdwell, 2005 ^b [9]	73/887 (+8,2%)
Cupples, 2005 ^b [10]	164/2100 (+7,8%)
Ko, 2006 ^b [11]	100/602 (+16,6%)
Morton, 2006 ^b [12]	191/1996 (+9,5%)
Fenton, 2007 ^{b,c} [13]	– (+3,0%)
Totale	981/7834 (+12,5%)

^aStudio retrospettivo; ^bstudio prospettico; ^cconfronto tra coorti

Table 3 Incremental cancer detection rate obtained by double reading

Study	Additional cancers detected/total cancers (%)
Anderson, 1994 [14]	21/170 (+12.3%)
Thurfjell, 1994 [15]	10/66 (+15.1%)
Warren, 1995 [16]	33/236 (+13.9%)
Leivo, 1999 [17]	29/261 (+11.1%)
Harvey, 2003 [18]	9/134 (+6.7%)
Kwek, 2003 [19]	7/128 (+5.4%)
Ciatto, 2005 [20]	43/670 (+6.4%)
Total	152/1,665 (+9.1%)

Tabella 3 Tasso diagnostico aggiuntivo di carcinoma ottenuto mediante la doppia lettura

Studio	Carcinomi aggiuntivi da doppia lettura/totali (%)
Anderson, 1994 [14]	21/170 (+12,3%)
Thurfjell, 1994 [15]	10/66 (+15,1%)
Warren, 1995 [16]	33/236 (+13,9%)
Leivo, 1999 [17]	29/261 (+11,1%)
Harvey, 2003 [18]	9/134 (+6,7%)
Kwek, 2003 [19]	7/128 (+5,4%)
Ciatto, 2005 [20]	43/670 (+6,4%)
Totale	152/1665 (+9,1%)

CAD vs. +9.1% double reading) and possibly better as to incremental recall rate (+12.5% CAD vs. 28.8 % double reading) (Table 4).

A smaller incremental recall rate observed with single reading + CAD versus conventional double reading may have a simple explanation. Double reading involves two radiologists with different individual diagnostic criteria and often with a different level of attention for specific morphological features (masses, parenchymal distortions, microcalcifications). Such differences in perception allow one reader to detect cancers missed by another reader and, at the same time, increase the recall rate of lesions with a specific morphological pattern. Conversely, with CAD, it is the same radiologist who adopts the same morphological

Table 4 Incremental recall rate obtained by double reading

Studio	Additional recalls/total recalls (%)
Anderson, 1994 [14]	569/1,086 (+52.3%)
Warren, 1995 [16]	1026/2,328 (+44.0%)
Harvey, 2003 [18]	371/320 (+11.45%)
Kwek, 2003 [19]	625/1,560 (+40.0%)
Ciatto, 2005 [20]	1,250/5,138 (+24.3%)
Total	3,841/13,332 (+28.8%)

ha reso la metodica appetibile sul piano economico).

Sulla base degli studi non controllati, le prestazioni di CAD, peraltro, appaiono paragonabili a quelle della doppia lettura per quanto riguarda il tasso diagnostico aggiuntivo (Tabella 3), con un 10,6% di CAD rispetto al 9,1% della doppia lettura, e possibilmente superiori per quanto riguarda l'effetto sul tasso di richiamo (Tabella 4), con un 12,5% di CAD rispetto al 28,8 % della doppia lettura.

Il minor effetto negativo di CAD, rispetto alla doppia lettura, sul tasso di richiamo è di spiegazione abbastanza facile. Nella doppia lettura i due lettori hanno ovviamente criteri diagnostici diversi e, in genere, una diversa attenzione per lesioni di morfologia diversa (opacità di massa, distorsioni, calcificazioni): proprio questa diversità di percezione consente da un lato ad un lettore di identificare alcune neoplasie che l'altro ha sottovalutato, dall'altro di richiamare preferenzialmente determinate morfologie, diverse dai richiami dell'altro lettore. Questo potenziamento del tasso di richiamo tra due lettori è maggiore di quello legato a CAD, perché in questo caso le varie entità morfologiche vengono rivalutate dallo stesso lettore, in sostanza con gli stessi criteri diagnostici. In altre parole un lettore che tende a dare poco importanza alle piccole opacità regolari, assumendole come benigne, continuerà a considerarle tali anche se CAD le evidenzia, mentre se gli associamo un altro lettore che invece dà importanza a quel tipo di alterazioni, il caso ha una maggiore probabilità di essere richiamato.

Se l'evidenza da studi non controllati suggerisce che la lettura singola integrata da CAD potrebbe essere una valida alternativa alla doppia lettura convenzionale, è indubbio che le caratteristiche degli studi in questione (selezione delle casistiche, contesti, protocolli, operatori diversi) comportino una scarsa confrontabilità, perché non concordano per una serie di fattori confondenti. Per questo motivo da tempo si auspica la conduzione di studi controllati che confrontino, a parità di casistica e di operatori, la lettura singola+CAD e la doppia lettura convenzionale. Gli studi finora disponibili in questo ambito sono riassunti nella Tabella 5.

I risultati di questi studi sono a prima vista piuttosto diffusi quanto a incremento del tasso diagnostico di carcinomi e di richiami. La differenza del tasso diagnostico di carcinoma basato sulla somma delle osservazioni è ovvia-

Tabella 4 Tasso di richiamo aggiuntivo ottenuto mediante la doppia lettura

Studio	Richiami aggiuntivi da doppia lettura (%)
Anderson, 1994 [14]	569/1086 (+52,3%)
Warren, 1995 [16]	1026/2328 (+44,0%)
Harvey, 2003 [18]	371/320 (+11,45%)
Kwek, 2003 [19]	625/1560 (+40,0%)
Ciatto, 2005 [20]	1250/5138 (+24,3%)
Totale	3841/13332 (+28,8%)

Table 5 Controlled studies comparing conventional double reading and single reading + computer-aided design (CAD)

Reading	Double	Single + CAD	Difference for CAD
True positives (cancers)			
Khoo, 2005 ^a [8]	61	57	-6.6%
Karssemeijer, 2003 ^b [21]	57	53	-7.5%
Georgian-Smith, 2007 ^a [22]	15	13	-15.1%
Gromet, 2008 ^{a,c} [23]	530	499	-5.8%
Gilbert (CADET), 2008 ^a [24]	199	198	-0.5%
Total	862	820	-5.1%
False positives (recalls)			
Khoo, 2005 ^a [8]	352	373	+5.9%
Georgian-Smith, 2007 ^a [22]	459	406	-13.0%
Gromet, 2008 ^{a,c} [23]	14,138	12,594	-10.9%
Gilbert (CADET), 2008 ^a [24]	947	1,100	+14.7%
Total	15,896	14,473	-9.8%

^aProspective; ^bretrospective; ^cfirst reading by radiographers

Tabella 5 Studi controllati che confrontano doppia lettura convenzionale e lettura singola+CAD

Lettura	Doppia	Singola+CAD	Differenza per CAD
Veri positivi (carcinoma)			
Khoo, 2005 ^a [8]	61	57	-6,6%
Karssemeijer, 2003 ^b [21]	57	53	-7,5%
Georgian-Smith, 2007 ^a [22]	15	13	-15,1%
Gromet, 2008 ^{a,c} [23]	530	499	-5,8%
Gilbert (CADET), 2008 ^a [24]	199	198	-0,5%
Totale	862	820	-5,1%
Falsi positivi (richiami)			
Khoo, 2005 ^a [8]	352	373	+5,9%
Georgian-Smith, 2007 ^a [22]	459	406	-13,0%
Gromet, 2008 ^{a,c} [23]	14138	12594	-10,9%
Gilbert (CADET), 2008 ^a [24]	947	1100	+14,7%
Totale	15896	14473	-9,8%

^aProspettici; ^bretrospettivi; ^cprima lettura da tecnici di radiologia

criteria for suspicion and re-analyses areas selected by CAD for review, and this translates to a lower incremental recall rate. In other words, a reader discounting small regular opacities, assumed as benign, will insist in discounting them, even if marked by CAD, whereas another reader with a special concern for missing cancers depicted as regular opacities is more likely to recall them, and possibly detect additional cancers.

Although the evidence from uncontrolled studies suggests single reading combined with CAD as a valid alternative to conventional double reading, no doubt these studies are hardly comparable due to differences in design (case selection, setting, diagnostic protocol, different operators), implying possible confounders. For this reason, controlled studies (same series, same readers) have been advocated for a long time to accurately compare single reading + CAD and double reading. Available studies to date are summarised in Table 5.

mente discutibile, trattandosi di studi del tutto diversi come disegno, scenario, numerosità di casistica e operatori, ma, con questa limitazione, sembra indicare che le due modalità non differiscano di molto, in media con una minore sensibilità (-5,1%) per la lettura singola+CAD, e differiscano maggiormente, ma questa volta a favore della lettura singola+CAD, quanto a tasso di richiamo (-9,8%).

La discrepanza dei risultati tra il confronto degli studi non controllati di lettura singola+CAD e doppia lettura e gli studi controllati può spiegarsi con il fatto che tutti questi studi riguardano casistiche, lettori e disegni di studio diversi: tutte queste variabili possono influenzare in modo importante i risultati. Anche gli studi controllati presentano le stesse variabili: una menzione speciale merita il recente studio *Computed-Aided Detection Evaluation Trial (CADET)* [25], che ha probabilmente il disegno di studio più robusto e affidabile, che riporta una sostanziale equivalenza della doppia lettura e della lettura singola+CAD, e in

Results are inhomogeneous at first glance, both for incremental detection and recall rate. A simple estimate of average results may be criticised for the intrinsic differences among studies; however, with such a limitation, it suggests no major differences between the two modalities (-5.1% in detection rate and -9.8% in recall rate for single reading + CAD).

The discrepancy between the evidence provided by uncontrolled comparison of single reading + CAD and double reading may be explained by the fact that the comparative studies dealt with different series, different readers and different study designs, and all these variables are likely to heavily influence incremental detection rates. Controlled studies may also be affected by the same biases. The recently published Computer-Aided Detection Evaluation Trial (CADET) deserves special mention [25] as the study with probably the most robust unbiased design, which reports a substantial equivalence of single reading + CAD and conventional double reading as to cancer detection rate, thus confirming the results of uncontrolled comparisons. However, it is evident that further controlled studies are needed before a final word may be said on this topic.

Considering that several CAD algorithms are presently marketed without proper validation of their efficacy or a careful analysis of their advantages and disadvantages, the latter being not immediately perceived in current use, the joint study group of the Gruppo Italiano di Studio per lo Screening Mammografico (GISMa) and Società Italiana di Radiologia Medica (SIRM) Senology Study Section thought it worthwhile to publish these considerations and recommendations on the use of CAD in current mammography practice. Considerations and recommendations are addressed to radiologists planning to use CAD but also to health service managers who may be interested in the possible advantages of the method. Overall, existing evidence on CAD use in mammography allows for a number of considerations and recommendations:

1. Scenario of use: CAD helps identify mammographic abnormalities, indicating the need for re-evaluation. CAD may thus be equally employed in a clinical setting, conventionally using single reading; or in screening, currently adopting double reading.
2. "Caveat" for CAD algorithm evaluation; Performance of different CAD algorithms in commerce is not always supported by published scientific validation. Validation should be ideally based on the assessment of incremental cancer detection rate and recall rate to diagnostic assessment, obtained through double-blind reading of current mammography activity or of a ad hoc predefined case series [1–4, 25]. Comparative studies of different CAD algorithms tested on the same series [1, 26] show that algorithms may differ to a substantial

questo sembra confermare quanto suggerito dal confronto di studi non controllati. È peraltro evidente che sono necessari altri studi controllati per dire una parola definitiva sull'argomento.

In considerazione del fatto che diversi algoritmi CAD vengono attualmente commercializzati senza una rigorosa verifica di efficacia e/o una attenta disamina dei pro e dei contro della metodica, tenuto conto che la corretta comprensione di questi ultimi non è sempre immediata, il gruppo di studio congiunto Gruppo Italiano di Studio per lo Screening Mammografico (GISMa) e sezione di Senologia della Società Italiana di Radiologia Medica (SIRM) ha ritenuto opportuno riportare queste considerazioni, precisazioni e raccomandazioni rispetto al possibile uso di CAD nella pratica clinica, ad oggi. Tali raccomandazioni e osservazioni sono rivolte sia ai radiologi che vogliono prendere in considerazione l'uso di CAD, sia agli amministratori sanitari interessati ai possibili vantaggi della metodica. Complessivamente l'evidenza disponibile sull'uso di CAD in mammografia consente di fare alcune considerazioni, precisazioni e raccomandazioni:

1. *Scenario di uso. CAD aiuta l'identificazione di anomalie mammografiche proponendole per una più attenta valutazione. In tal senso CAD può essere impiegato sia a supporto della mammografia clinica, che impiega correntemente una lettura singola, che della mammografia di screening, che impiega correntemente la doppia lettura.*
2. *Caveat relativo alla validazione dell'algoritmo. La performance dei diversi algoritmi in commercio non è sempre supportata da una verifica scientifica pubblicata in letteratura. La verifica in genere si basa sulla determinazione del tasso incrementale diagnostico di carcinoma e di richiamo ad approfondimento, verificata con metodica cieca su casistica corrente o su set di validazione approntati ad hoc [1–4, 25]. Studi comparativi di algoritmi CAD testati sulla stesa casistica [1, 26] dimostrano come gli algoritmi possano differire in modo rilevante. Ogni algoritmo CAD oggetto di commercializzazione dovrebbe, prima di essere impiegato nella pratica clinica, essere testato in maniera congrua, ad esempio consentendo un confronto affidabile di performance con altri algoritmi CAD di performance nota.*
3. *Caveat relativo alle modalità d'uso di CAD. I sistemi CAD di cui è nota la performance in mammografia sono in genere dotati di bassa specificità. Essi identificano un eccesso di regioni di interesse (ROI) proponendole alla rivalutazione (poco meno di una ROI per proiezione). Il radiologo è quindi tenuto necessariamente, per evitare un intollerabile eccesso di approfondimenti diagnostici, a trascurare, come non rilevanti la maggioranza delle ROI, concentrandosi su una minoranza di esse. Tale processo è intrinseco all'uso di CAD e riguarda sia lo screening che l'uso clinico, in quanto entrambi, nono-*

- extent. Any CAD algorithm proposed for commerce should be properly tested before clinical use. Ideally, comparison of a new algorithm with already validated algorithms with regard to diagnostic accuracy should be current standard.
3. “Caveat” for CAD use modality: CAD systems with known performance rates usually have a low specificity. An excess of regions of interest (ROI) is identified and proposed for review (at least one ROI per view). Radiologists are thus obliged to discount the large majority of ROIs to avoid unacceptable recall rates, indicating the need for diagnostic assessment only in a minority of ROIs. Such a discounting process is intrinsic and obligatory with CAD use, both in a clinical or screening setting, although screening is better “ruled” with reference standards than is clinical practice, as in both scenarios excess false positive recalls and unnecessary assessment have to be avoided.
 4. “Caveat” for expected CAD results: The impact of CAD on diagnostic accuracy varies from one radiologist to another. Studies of CAD performance [1–4, 27, 28] demonstrated that even well-trained radiologists with considerable experience may react in different ways to CAD use, the magnitude of CAD influence on diagnosis being quite variable. For this reason, it is of paramount importance that radiologists do not automatically assume the same average advantages of CAD use reported in the literature: each reader should test his or her own CAD-related incremental detection and recall rate. A different impact of CAD is expected on individual radiologists and teams of radiologists. When CAD is adopted in a department, results should be checked to assure the procedure is convenient. In the event that CAD shows poor local performance (too few extra cancers detected with an excess of extra recalls), it should not be used in current practice, being of no benefit for the patients
 - 5 “Caveat” for CAD use by unexperienced radiologists: As discounting most CAD-generated ROIs is unavoidable, this may be difficult for radiologists of limited experience in mammography reading, who tend to adopt a low baseline threshold for suspicion and a high recall rate, as they feel less confident with the methodology. Thus, CAD use might translate into an even higher recall rate. CAD should be used only by experienced radiologists. Although it is not easy to agree on a definition of “experienced”, a number of at least 5,000 mammograms read every year is currently indicated by the EC, the Ministry of Health, and GISMa as a reference standard for “good experience” and could be used a reference also for CAD use.
 6. Medicolegal “caveat” relative to ROI discounting: Let us suppose that one radiologist is sued for diagnostic delay *stante la mammografia clinica non sia “normata” come lo screening, devono evitare l’eccesso di falsi positivi e di accertamenti (anche invasivi) inutili.*
4. *Caveat relativo ai risultati attesi con l’uso di CAD. L’impatto di CAD sulle prestazioni diagnostiche varia da radiologo a radiologo. Gli studi che hanno verificato l’effetto di CAD sulla diagnosi [1–4, 27, 28] riportano correntemente come anche radiologi esperti reagiscono diversamente a CAD, facendosi influenzare da esso in diversa misura. Per questo motivo è opportuno che il radiologo che usa CAD non assuma che ne avrà il vantaggio medio riportato in letteratura, ma verifichi, dopo una determinato periodo d’uso, il reale impatto di CAD sul tasso diagnostico di carcinoma e di approfondimenti diagnostici nella propria esperienza individuale. Tali considerazioni non sono valide solo sul piano individuale ma anche di struttura. Una unità operativa (UO) che adottasse l’uso di CAD dovrebbe verificarne i risultati per considerare la procedura realmente conveniente. Ove infatti non si verificasse che l’uso di CAD è vantaggioso (ad es. identifica troppo pochi carcinomi aggiuntivi rispetto all’eccesso di approfondimenti che genera), tale uso dovrebbe essere abbandonato, per motivi etici, in quanto non utile, anzi dannoso, per il paziente.*
 5. *Caveat relativo all’uso di CAD da parte di radiologi poco esperti. Definito come inevitabile questo processo di selezione delle ROI generate da CAD, è evidente che un radiologo di limitata esperienza in mammografia (e per questo già tendente ad eccedere negli approfondimenti per la minore confidenza con il metodo) potrebbe risentire negativamente dell’uso di CAD, che lo porterebbe facilmente ad un tasso di falsi positivi, e relativi approfondimenti inutili, eccessivo. È quindi auspicabile che CAD sia usato solo da radiologi esperti in mammografia. Ferma restando la difficoltà di definire l’esperienza con un parametro numerico, il limite minimo di 5000 mammografie lette ogni anno, indicato dalle raccomandazioni della Comunità Europea (CE), del Ministero della Salute, e del GISMa, come indicativo di “buona esperienza” potrebbe essere assunto anche come riferimento per l’uso di CAD.*
 6. *Caveat medico-legale relativo alla censura delle ROI. Se una lesione di cui si sostiene la mancata identificazione per imperizia era stata identificata come ROI, questo può essere una aggravante dell’errore del radiologo in un contenzioso medico-legale? Come indicato al punto 3, il radiologo che usa CAD ha l’obbligo della censura della grande maggioranza delle ROI per garantire una buona pratica diagnostica. Quindi il fatto che abbia trascurato una ROI non può essere preso come aggravante di un supposto errore diagnostico, ma l’eventuale omissione verrà valutata in base alla revisione del radiogramma incriminato, possibilmente con modalità cieca [29],*

due to misperception of a mammographic abnormality, which had been selected as an ROI by CAD. Should ROI discounting be assumed as an indicator of diagnostic error? As discussed Caveat 3, CAD use involves the obligation for the radiologist to discount most ROIs to avoid unacceptable recall rates. Thus, ROI discounting cannot be assumed per se as a proof of diagnostic error: diagnostic error will be evaluated only on review of the mammogram, possibly using a blind review methodology [29], and without CAD use. Should ROIs be retrieved, exactly the same algorithm in use at the time of the supposed diagnostic error should be employed. As algorithms are often updated, ROIs are not necessarily the same with a new algorithm release, and ROIs are not currently saved and stored.

7. “Caveat” for the health service managers: When combined with single reading currently used for clinical mammography, CAD is expected to increase cancer detection rate with a similar relative increase in recall rates. Overall (Tables 1 and 2), the ratio between additional detected cancers and additional recalls is expected in the range of 1:10. On this basis, CAD use is associated with an increase of service costs (licence, additional diagnostic assessment). In other words, clinical diagnosis, with the use of CAD, is more complex and costly and requires a greater workload, although this is counterbalanced by earlier diagnosis of a limited number of cancers. The cost-effectiveness of CAD might be different in a screening setting if single-reading + CAD is used as an alternative to double reading. Such an alternative implies sparing one radiologist’s workload and reducing recall rates, though at the cost of reduced sensitivity (around 5% less). Such calculations are just indicative as they are based on average values drawn from rather variable numbers in the literature. It is more reasonable to say that in the absence of sound scientific evidence that single reading + CAD is as effective as conventional double reading, CAD use is not recommended as an alternative to double reading. If used, potential diagnostic errors should be judged more severely, as recommended policies (such as double reading) for maximising screening efficacy were not implemented. Managers of screening programmes adopting such a protocol would be also sued for incompetence and would share part of the responsibility of the diagnostic error. In other words, conventional double reading remains obligatory in mammography screening. CAD may possibly be used in addition to double reading.

senza CAD. In ogni caso, l’algoritmo CAD nella indicazione delle ROI non è necessariamente riproducibile nelle nuove versioni dello stesso prodotto commerciale. Gli algoritmi CAD, infatti, sono spesso soggetti a modifiche migliorative della performance (in genere della specificità) e possono, pertanto, differire da un anno all’altro. Poiché le ROI non vengono normalmente archiviate, ove si volesse verificare retrospettivamente una ROI, deve essere impiegato lo stesso algoritmo in uso al tempo dell’esame contestato, che potrebbe differire dall’algoritmo corrente.

7. *Caveat per gli amministratori rispetto all’uso di CAD. Se combinato alla lettura singola, di comune uso in pratica clinica, CAD dovrebbe aumentare il tasso diagnostico di carcinoma al costo di un aumento comparabile di approfondimenti diagnostici. In linea di massima (Tabelle 1 e 2) il rapporto tra carcinomi e approfondimenti aggiuntivi dovrebbe essere dell’ordine di 1:10. Sul piano dei costi di esercizio, pertanto, l’uso di CAD comporta un aumento dei costi, legato ai costi intrinseci di CAD (licenza) e al costo degli approfondimenti diagnostici aggiuntivi generati da CAD. In altre parole, la diagnostica senologica clinica, con l’uso di CAD, diviene più complessa e costosa e richiede più tempo/lavoro, anche se questo è compensato da un aumento della precocità diagnostica in un limitato numero di carcinomi. Diversamente interessante potrebbe essere l’impiego di CAD in pratica di screening, ove si usi CAD come alternativa alla doppia lettura. Una simile scelta porterebbe indubbiamente al risparmio di un radiologo lettore, apparentemente con una riduzione dei richiami, ma al prezzo di una riduzione della sensibilità con perdita di un 5% di carcinomi. È opportuno ricordare che tali calcoli sono piuttosto grossolani in quanto basati su valutazioni di medie dei risultati della letteratura, di per sé poco affidabili. È più ragionevole affermare che, al momento, in mancanza di evidenza scientifica di una reale equivalenza tra lettura singola+CAD e doppia lettura, l’uso di CAD in alternativa alla doppia lettura non è raccomandato. Ove lo si usasse a tal fine, eventuali errori diagnostici sarebbero aggravati dal fatto di non aver adottato quei presidi (appunto la doppia lettura) raccomandati per garantire l’efficacia dello screening. L’amministrazione responsabile del programma di screening che avesse adottato tale protocollo sarebbe, pertanto, chiamata a corredo in caso di dovuto indennizzo. In altre parole, la doppia lettura resta una procedura d’obbligo nello screening. L’uso di CAD può essere adottato come impiego aggiuntivo.*

Conflict of interest statement The authors declare that they have no conflict of interest to the publication of this article.

References/Bibliografia

- Ciatto S, Ambrogetti D, Bonardi R et al (2004) Comparison of two commercial systems for computer aided detection (CAD) of breast cancer by screening mammography. *Radiol Med* 107:480–488
- Ciatto S, Rosselli Del Turco M, Burke P et al (2003) Comparison of standard and double reading and computer aided detection (CAD) of interval cancers at prior negative screening mammograms blind review. *Br J Cancer* 89:1645–1649
- Ciatto S, Rosselli Del Turco M, Risso G et al (2003) Comparison of standard reading and computer aided diagnosis (CAD) on a national proficiency test of screening mammography. *Eur J Radiol* 45:135–138
- Ciatto S, Brancato B, Rosselli Del Turco M et al (2003) Comparison of standard reading and computer aided diagnosis (CAD) on a proficiency test of screening mammography. *Radiol Med* 106:59–65
- Freer TW, Ulissey MJ (2001) Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 220:781–786
- Helvie MA, Hadjiiski L, Makariou E et al (2004) Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: pilot clinical trial. *Radiology* 231:208–214
- Gur D, Sumkin JH, Rockette HE et al (2004) Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst* 96:185–190
- Khoo LA, Taylor P, Given-Wilson RM (2005) Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. *Radiology* 237:444–449
- Birdwell RL, Bandodkar P, Ikeda DM (2005) Computer-aided detection with screening mammography in a university hospital setting. *Radiology* 236:451–457
- Cupples TE, Cunningham JE, Reynolds JC (2005) Impact of computer-aided detection in a regional screening mammography program. *AJR Am J Roentgenol* 185:944–950
- Ko JM, Nicholas MJ, Mendel JB, Slanetz PJ (2006) Prospective assessment of computeraided detection in interpretation of screening mammography. *AJR Am J Roentgenol* 187:1483–1491
- Morton MJ, Whaley DH, Brandt KR, Amrami KK (2006) Screening mammograms: interpretation with computer-aided detection – prospective evaluation. *Radiology* 239:375–383
- Fenton JJ, Taplin SH, Carney PA et al (2007) Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 356:1399–1409
- Anderson ED, Muir BB, Walsh JS, Kirkpatrick AE (1994) The efficacy of double reading mammograms in breast screening. *Clin Radiol* 49:248–251
- Thurfjell EL, Lerneval KA, Taube AA (1994) Benefit of independent double reading in a population based mammography screening program. *Radiology* 19:241–244
- Warren RM, Duffy SW (1995) Comparison of single reading with double reading of mammograms, and change in effectiveness with experience. *Br J Radiol* 68:958–962
- Leivo T, Salminen T, Sintonen H et al (1999) Incremental cost-effectiveness of double-reading mammograms. *Breast Cancer Res Treat* 54:261–267
- Harvey SC, Geller B, Oppenheimer RG et al (2003) Increase in cancer detection and recall rates with independent double interpretation of screening mammography. *AJR Am J Roentgenol* 180:1461–1467
- Kwek BH, Lau TN, Ng FC, Gao F (2003) Non-consensual double reading in the Singapore Breast Screening Project: benefits and limitations. *Ann Acad Med Singapore* 32:438–441
- Ciatto S, Ambrogetti D, Bonardi R et al (2005) Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme. *J Med Screen* 12:103–106
- Karssemeijer N, Otten JD, Verbeek AL et al (2003) Computer-aided detection versus independent double reading of masses on mammograms. *Radiology* 227:192–200
- Georgian-Smith D, Moore RH, Halpern E et al (2007) Blinded comparison of computer-aided detection with human second reading in screening mammography. *AJR Am J Roentgenol* 189:1135–1141
- Gromet M (2008) Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol* 190:854–859
- Gilbert FJ, Astley SM, Gillan MG et al; CADET II Group (2008) Single reading with computer-aided detection for screening mammography. *N Engl J Med* 359:1675–1684
- Ciatto S, Brancato B, Rosselli Del Turco M et al (2003) Comparison of standard reading and computer aided diagnosis (CAD) on a proficiency test of screening mammography. *Radiol Med* 106:59–65
- Ciatto S, Cascio D, Fauci F et al (2009) Computer-assisted diagnosis (CAD) in mammography: comparison of diagnostic accuracy of a new algorithm (Cyclopus, Medcad) with two commercial systems. *Radiol Med* 114:626–635
- Bazzocchi M, Mazzarella F, Del Frate C et al (2007) CAD systems for mammography: a real opportunity? A review of the literature. *Radiol Med* 112:329–353
- Bazzocchi M, Facecchia I, Zuiani C et al (2001) Application of a computer-aided detection (CAD) system to digitalized mammograms for identifying microcalcifications. *Radiol Med* 101:334–340
- Fratello F, Federici A, Ghirardini A et al; Gruppo di lavoro SIRM-GISMa (2008) Programma di screening mammografico: errori interpretativi supposti nella lettura della mammografia e cancri di intervallo. Ministero della Salute, Direzione Generale della Prevenzione Sanitaria, Gennaio 2008