Linea guida pubblicata nel Sistema Nazionale Linee Guida

Roma, 23 dicembre 2025

LINEE GUIDA PER LO SCREENING E LA DIAGNOSI DEL TUMORE DELLA MAMMELLA (*ADOLOPMENT* LINEE GUIDA EUROPEE)
Capitolo sull'uso della Intelligenza Artificiale

Società Scientifiche coinvolte

- AIE Associazione Italiana di Epidemiologia
- AIFM Associazione Italiana di Fisica Medica
- AIOM Associazione Italiana di Oncologia Medica
- AIRO Associazione Italiana di Radioterapia e Oncologia Clinica
- FASTeR/AITERS Federazione delle Associazioni Scientifiche dei Tecnici di Radiologia
- GISMa Gruppo Italiano Screening Mammografico
- SIAPEC-IAP Società Italiana di Anatomia Patologica e Citologia Diagnostica (Divisione italiana della International Academy of Pathology)
- SICPRE Società Italiana di chirurgia plastica ricostruttiva ed estetica
- SIRM Società Italiana di Radiologia Medica ed Interventistica
- SITI Società Italiana d'Igiene, Medicina Preventiva e Sanità Pubblica

Comitato Tecnico Scientifico (CTS)

1. AIE, Salvatore Scondotto, Università degli Studi di Enna "Kore". Professore Associato Igiene generale e applicata
2. AIFM, Vittorio Cannatà, Responsabile del Servizio di Fisica Sanitaria presso l'IRCCS Ospedale Pediatrico Bambino Gesù di Roma; Trianni Annalisa Fisico, specializzato in fisica sanitaria, Direttore UO di fisica Sanitaria dell'Azienda Provinciale per i servizi sanitari di Trento (Apss)
3. AIOM, Franceso Perrone, Direttore Dipartimento Ricerca Traslazionale, Unità Sperimentazioni Cliniche. Istituto Nazionale Tumori - IRCCS Fondazione Pascale - Napoli
4. AIRO, Carlotta Becherini, Oncologa radioterapista, Azienda Ospedaliero Univarsitaria Careggi - Firenze
5. FASTER/AITERS, Pacifici Stefano, Tecnico Sanitario di radiologia medica (TSRM), Unicamillus International Medical University – Roma
6. GISMA, Stefano Ferretti, Patologo, Dip. Medicina Traslazionale e per la Romagna, Università di Ferrara Registro Tumori della Regione Emilia-Romagna, Unità funzionale Azienda USL Ferrara
7. SIAPEC, Castellano Isabella Anatomo Patologo, Dipartimento di Scienze Mediche Università di Torino
8. SICPRE, Maurizio Ressa, Direttore Struttura Complessa di Chirurgia Plastica IRCCS "Giovanni Paolo II" Istituto dei Tumori di Bari Chirurgo plastico, Ospedale Fatebenefratelli Isola Tiberina Roma
9. SIRM, Federica Pediconi, Dipartimento di Scienze Radiologiche, Oncologiche ed Anatomo-Patologiche, Radiologia Diagnostica ed Interventistica della Mammella - Sapienza Università di Roma
10. SITI, Torri Emanuele, Igienista, presso Provincia autonoma di Trento

Gruppo di Lavoro metodologico

1. Battisti Francesca, Igienista, Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
2. Benato Cortecci Mattia, Department of Social and Political Sciences, CERGAS-SDA and Dondena, Università Bocconi, Milano (ERT)
3. Deandrea Silvia, Epidemiologa, ATS della Provincia di Pavia, DG Welfare Regione Lombardia
4. Fattore Giovanni, Department of Social and Political Sciences, CERGAS-SDA and Dondena, Università Bocconi, Milano (ERT)
5. Giordano Livia, Epidemiologa, CPO Piemonte Città della Salute e della Scienza, TORINO
6. Giorgi Rossi Paolo, Epidemiologo, Azienda Unità Sanitaria Locale - IRCCS di Reggio Emilia
7. Mantellini Paola, Oncologa, Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
8. Morgano Gian Paolo, European Commission, Joint Research Centre (JRC), Ispra, Italy
9. Paci Eugenio, Epidemiologo, Firenze
10. Rossi Martina, Segreteria Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
11. Francesco Venturelli, Epidemiologo, Azienda Unità Sanitaria Locale - IRCCS di Reggio Emilia
12. Zappa Marco, Epidemiologo, Firenze

Membri del panel

Catia Angiolini	Oncologa, Azienda Ospedaliero Universitaria Careggi, Firenze
Adriana Bonifacino	Fond. Fondazione IncontraDonna
Laura Bonvicini	Epidemiologa, Azienda Unità Sanitaria Locale-IRCCS di Reggio Emilia
Beniamino Brancato	Radiologo, Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
Lauro Bucchi	Epidemiologo, Registro Tumori dell'Emilia-Romagna, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) Dino Amadori, Meldola, Forlì
Eva Carnesciali	Tecnico Sanitario Radiologia Medica, Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
Francesca Caumo	Radiologa, Istituto Oncologico Veneto IOV – IRCCS, Padova, Italia
Silvia Deandrea	Epidemiologa, Co-chair metodologico, - ATS della Provincia di Pavia - DG Welfare Regione Lombardia
Flori Degrassi	Presidente ANDOS Nazionale
Livia Giordano	Epidemiologa, CPO Piemonte Città della Salute e della Scienza, TORINO
Paolo Giorgi Rossi	Epidemiologo, Azienda Unità Sanitaria Locale-IRCCS di Reggio Emilia
Paola Golinelli	Fisica medica, Ausl Modena
Francesco Gongolo	Medico legale, Responsabile SOSD screening e malattie cronico degenerative Azienda Sanitaria Universitaria Friuli Centrale
Stafania Gori	Direttore Oncologia Medica - Direttore Dipartimento Oncologico IRCCS Ospedale Sacro Cuore Don Calabria, Negrar di Valpolicella
Carlo Magliocca	Chirurgo plastico, Ospedale Fatebenefratelli – Isola Tiberina, Roma
Icro Meattini	Oncologo radioterapista, Azienda Ospedaliero Universitaria Careggi - Firenze
William Mantovani	Direttore UO Epidemiologia Clinica e Valutativa - Trento.
Walter Mazzucco	Professore Ordinario di Igiene e Medicina Preventiva dell'Università di Palermo, Dirigente Medico presso l'Azienda Ospedaliera Universitaria Policlinico "Paolo Giaccone" - Palermo
Staefania Montemezzi	Radiologo, Direttore U.O.C. di Radiologia BT, Direttore DAI Patologia e Diagnostica - Azienda Ospedaliera Universitaria Integrata - Verona
Marco Moschetta	Radiologo, Responsabile U.O.S.D. Radiodiagnostica ad Indirizzo Senologico "Aldo Moro", Coordinatore Clinico Breast Care Unit, Azienda Ospedaliero Universitaria Consorziale Policlinico Bari
Paola Mosconi	Biologa, Istituto di Ricerche Farmacologiche Mario Negri IRCCS – Milano

Andrea Nitrosi	Responsabile Struttura Semplice Fisica della Diagnostica per Immagini Responsabile del Programma Management Integrato per lo Sviluppo Tecnologico in Ambito del Sistema Ris/Pacs - AUSL di Reggio Emilia
Eugenio Paci	Epidemiologo, Firenze
Laura Pandis	Medico chirurgo**,** Responsabile della Qualità, UOC Chirurgia Plastica, Azienda Ospedaliera, Università di Padova
Francesca Pietribiasi	Patologo, ASL TO5 - Ospedale S. Croce, Moncalieri (TO)
Francesco Regini	Radiologo, SC Screening e Prevenzione Secondaria - Istituto per lo studio, la prevenzione e la rete oncologica (ISPRO) Firenze
Elena Ricci	Assegnista di ricerca Scuola Superiore Sant'Anna di Pisa; Professoressa a contratto, Università Europea di Roma
Gioele Santucci	Tecnico Sanitario Radiologia Medica, Università degli Studi di Bologna, Ausl Romagna
Francesco Sardanelli	Radiologo, Lega Italiana per la Lotta contro i Tumori Milano Monza Brianza, Milano
Priscilla Sassoli de' Bianchi	Biologa specializzata in Statistica Sanitaria Coordinatore regionale Screening, Regione Emilia-Romagna
Cristian Scatena	Patologo, Università di Pisa ed Azienda Ospedaliero Universitaria Pisana
Carlo Senore	Epidemiologo, CPO Piemonte Città della Salute e della Scienza, Torino
Margherita Serra	Chirurgo, Policlinico S. Orsola- Malpighi, Dip. della Salute della donna, del bambino e dell'adolescente
Daniela Severi	Tecnico Sanitario Radiologia Medica, Ausl Romagna
Francesco Venturelli	Epidemiologo, Azienda Unità Sanitaria Locale-IRCCS di Reggio Emilia
Marco Zappa	Epidemiologo, Firenze

Al fine di facilitare il processo di votazione si è costituito un **Panel mobile**, ovvero un sottogruppo di esperti già appartenenti al Panel che ha espresso interesse alla partecipazione attiva alle plenarie e conseguenti votazioni. Il panel è stato creato su adesione volontaria dei seguenti esperti in materia: Catia Angiolini, Adriana Bonifacino, Laura Bonvicini, Beniamino Brancato, Lauro Bucchi, Eva Carnesciali, Francesca Caumo, Silvia Deandera, Flori Degrassi, Paolo Giorgi Rossi, Paola Golinelli, Francesco Gongolo, Carlo Magliocca, William Mantovani, Walter Mazzucco, Stefania Montemezzi, Marco Moschetta, Paola Mosconi, Andrea Nitrosi, Eugenio Paci, Laura Pandis, Loredana Pau, Francesca Pietribiasi, Francesco Regini, Elena Ricci, Gioele Santucci, Francesco Sardanelli, Priscilla Sassoli de' Bianchi, Cristian Scatena, Carlo Senore, Margherita Serra, Daniela Severi, Francesco Venturelli, Marco Zappa.

Revisori esterni

Daniela Bernardi, Radiologa

Simona Vecchi, Metodologa

Definizione acronimi utilizzati nel documento

CERGAS: Centre for Research on Health and Social Care Management, SDA Bocconi

CNEC: Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure

CTS: Comitato Tecnico Scientifico

DBT: Tomosintesi mammaria digitale (Digital Breast Tomosynthesis)

DM: Mammografia digitale (Digital Mammography)

ECIBC: The European Commission Initiative on Breast Cancer

EtD: The GRADE Evidence to Decision frameworks

ERT: Evidence Review Team

FASO: Federazione delle Associazioni degli Screening Oncologici

GDG: The Guideline Development Group

GRADE: Grading of Recommendations, Assessment, Development, and Evaluation (The GRADE method)

IA: Intelligenza Artificiale

ONS: Osservatorio Nazionale Screening

PICO: Quesito clinico formulato secondo il framework Popolazione, Intervento, Comparatore, Outcomes

PRU: PICO Responsible Unit

SNLG: Sistema Nazionale Linee Guida

SINOSSI RACCOMANDAZIONI UTILIZZO DELL'INTELLIGENZA ARTIFICIALE

Quesito 1. *La singola lettura umana supportata dall'intelligenza artificiale vs. la doppia lettura umana senza supporto dell'intelligenza artificiale dovrebbe essere utilizzata per leggere le mammografie eseguite con mammografia digitale (2DFFDM) o tomosintesi digitale del seno per la diagnosi precoce del cancro della mammella nei programmi di screening mammografico?*

Raccomandazione: Il panel italiano adotta la raccomandazione del Gruppo di Sviluppo delle Linee Guida (GDG) ECIBC e **suggerisce di non utilizzare la singola lettura umana supportata dall'intelligenza artificiale** per leggere le mammografie eseguite con mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro della mammella nei programmi di screening mammografico.

(Conditional recommendation against the intervention, very low certainty of evidence; raccomandazione su condizione contro l'intervento, grado di certezza delle evidenze molto basso.)

Quesito 2. *La doppia lettura umana supportata dall'intelligenza artificiale vs. la doppia lettura umana senza supporto dell'intelligenza artificiale dovrebbe essere utilizzata per leggere le mammografie eseguite con mammografia digitale (2DFFDM) o tomosintesi digitale del seno per la diagnosi precoce del cancro della mammella nei programmi di screening mammografico?*

Raccomandazione: Il panel italiano adotta la raccomandazione del Gruppo di Sviluppo delle Linee Guida (GDG) ECIBC e **suggerisce di utilizzare la doppia lettura umana supportata dall'intelligenza artificiale** (con consenso o arbitrato in caso di letture discordanti) invece che la doppia lettura umana senza supporto dell'intelligenza artificiale (con consenso o arbitrato in caso di letture discordanti) per leggere le mammografie eseguite con mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro della mammella nei programmi di screening mammografico.

(Conditional recommendation for the intervention, very low certainty of evidence; raccomandazione su condizione a favore dell'intervento, grado di certezza delle evidenze molto basso.)

Quesito 3. *Una strategia di triage basata sull'intelligenza artificiale per assegnare a singola o doppia lettura umana ciascuna mammografia invece che la doppia lettura umana per tutte le mammografie dovrebbe essere utilizzata nello screening mammografico basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico?*

Raccomandazione Per donne asintomatiche con un rischio medio di cancro della mammella, il panel italiano **suggerisce di utilizzare una strategia di triage basata sull'intelligenza artificiale per assegnare a singola o doppia lettura umana** ciascuna mammografia invece che la doppia lettura umana per tutte le mammografie nello screening mammografico basata su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico.

(Conditional recommendation for the intervention, moderate certainty of evidence; raccomandazione su condizione a favore dell'intervento, grado di certezza delle evidenze moderato.)

Per la giustificazione della raccomandazione e le considerazioni per la sua implementazione si veda la sezione Conclusions dell'Evidence-to-Decision Framework.

	Adattato	Adottato	PICO de novo
Quesito 1		**X**	
Quesito 2		**X**	
Quesito 3			**X**

Per i due quesiti adottati direttamente dalle linee guida europee non sono stati prodotti EtD arricchiti dei dati di contesto.

METODOLOGIA ADOLOPMENT RACCOMANDAZIONI UTILIZZO DELL'INTELLIGENZA ARTIFICIALE

Sintesi processo e fonti informative

Nell'ambito del processo di adattamento Italiano delle raccomandazioni ECIBC sull'uso dell'intelligenza artificiale, il CTS ha deciso di adottare le due raccomandazioni votate a luglio 2023 dal GDG ECIBC e di sviluppare una terza raccomandazione a partire da un PICO definito *de novo*, alla luce delle prime due raccomandazioni e delle evidenze pubblicate successivamente a luglio 2023.

Il PICO definito *de novo* ed il razionale per lo sviluppo della terza raccomandazione sono riportati nell'introduzione dell'*Evidence-to-Decision Table*.

A supporto dello sviluppo della raccomandazione, l'Evidence Review Team ha svolto una revisione sistematica della letteratura, una valutazione economica ed è stata condotta una survey su accettabilità e fattibilità dell'introduzione dell'IA nei programmi di screening mammografico. Nell'ambito delle attività della PRU e del panel sono stati inoltre identificati studi non inclusi nella revisione sistematica perché non in linea con i criteri di eleggibilità, ritenuti informativi ed integrati nell'EtD a supporto di considerazioni aggiuntive. Nel capitolo 3, paragrafi 3.5-3.8, è riportata la metodologia utilizzata per la produzione delle evidenze riportate nell'EtD.

Indice

METODI DI LAVORO PER LE LINEE GUIDA ITALIANE PER LO SCREENING E LA DIAGNOSI DEL CANCRO DELLA MAMMELLA (ADOLOPMENT DELLE LINEE GUIDA EUROPEE) 9
1. Introduzione e Razionale 9
2. Obiettivo 9
3. Metodologia 10
4. Revisione esterna 24
5. Applicabilità 24
6. Aggiornamento della LG: tempi e modalità 24
7. Conflitti di interesse editoriale 24
8. Bibliografia 25
9. Regole per il panel 30
EVIDENCE TO DECISION FRAMEWORK - QUESITO 3 33
ASSESSMENT 34
SUMMARY OF JUDGEMENTS 46
TYPE OF RECOMMENDATION 47
CONCLUSIONS 47
Bibliografia 52
APPENDICI 54

METODI DI LAVORO PER LE LINEE GUIDA ITALIANE PER LO SCREENING E LA DIAGNOSI DEL CANCRO DELLA MAMMELLA (ADOLOPMENT DELLE LINEE GUIDA EUROPEE)

1. Introduzione e Razionale

Le Linee guida europee sullo screening mammografico (Perry 2006) esistono da oltre 20 anni e forniscono agli Stati Membri le indicazioni per l'organizzazione dei programmi di screening, come richiesto dalla Raccomandazione del Consiglio dell'Unione europea del 2 dicembre 2003. A partire dal 2016, le raccomandazioni (European Breast Cancer Guidelines on Screening and Diagnosis – in seguito Linee Guida Europee) (ECIBC 2019) sono sviluppate con il metodo GRADE (Grading of Recommendations Assessment, Development and Evaluation) e il framework GRADE Evidence to Decision (EtD) (Alonso-Coello P 2016), lo stesso richiesto dal Sistema Nazionale Linee Guida (SNLG) per l'accettazione di linee guida nel sistema previsto dalla legge 24/2017 ("Legge Gelli").

Quindi, considerata la disponibilità di linee guida di riferimento dal punto di vista istituzionale e comunitario che presentano anche le caratteristiche richieste per l'inclusione nel sistema italiano, e la contestuale assenza di documenti analoghi sviluppati a livello nazionale, si è ritenuto utile attivare un processo di adozione e adattamento attraverso una procedura di adolopment (Schünemann 2017), così come raccomandato dal Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure (CNEC) (CNEC 2019).

Nei precedenti capitoli, il panel per l'adozione e adattamento delle linee guida sullo screening mammografico, sulla base delle evidenze elaborate per le linee guida europee, suggerisce che la mammografia sia offerta alle donne a partire dai 45 anni fino ai 74 anni di età. La raccomandazione è forte per le donne tra i 50 ed i 69 anni mentre nelle due fasce estreme (45-49, 70-74) è una raccomandazione 'con riserva', cioè il panel ha concluso che gli effetti desiderabili del fare una mammografia in queste fasce di età superano probabilmente gli effetti indesiderabili, ma esiste ancora un margine di incertezza.

2. Obiettivo

Il progetto si pone l'obiettivo finale di migliorare la qualità dello screening mammografico organizzato in Italia favorendo l'erogazione di attività clinica in accordo alle migliori evidenze. Per ottenere questo risultato è necessario favorire l'adozione e l'implementazione delle raccomandazioni delle linee guida nei programmi di screening italiani, sia a livello di politiche regionali sia a livello di prassi nelle realtà locali, attraverso la diffusione istituzionale di raccomandazioni adattate o adottate

dalle Linee Guida Europee, o sviluppate de-novo con la stessa metodologia, in coerenza con l'assetto metodologico e istituzionale di SNLG.

Utilizzatori target:

- Ministero della Salute
- Regioni
- Conferenza Stato-Regioni
- Coordinamento interregionale di prevenzione
- Aziende Sanitarie e Ospedaliere Universitarie
- Associazioni di cittadini e pazienti.

3. Metodologia

3.1 La governance del processo

Il promotore dell'iniziativa è l'Osservatorio Nazionale Screening (ONS), la società scientifica referente è il Gruppo Italiano Screening Mammografico (GISMa). Il processo di adolopment si avvale di un Comitato Tecnico Scientifico (CTS) composto dalle seguenti società scientifiche:

- AIE Associazione Italiana di Epidemiologia
- AIFM Associazione Italiana di Fisica Medica
- AIOM Associazione Italiana di Oncologia Medica
- AIRO Associazione Italiana di Radioterapia e Oncologia Clinica
- FASTeR Federazione delle Associazioni Scientifiche dei Tecnici di Radiologia
- SIAPEC-IAP Società Italiana di Anatomia Patologica e Citologia Diagnostica (Divisione italiana della International Academy of Pathology)
- SICi Società Italiana di Citologia
- SICPRE Società Italiana di chirurgia plastica ricostruttiva ed estetica
- SIRM Società Italiana di Radiologia Medica ed Interventistica
- SITI Società Italiana di Igiene Medicina Preventiva e Sanità Pubblica

La composizione del CTS include tutte le società scientifiche accreditate dal Ministero, la cui area è interessata dalle raccomandazioni delle Linee Guida Europee, e che hanno aderito all'invito trasmesso ai rispettivi presidenti dall'ONS.

Ciascuna società scientifica ha nominato un proprio rappresentante nel CTS. Il CTS, in linea con quanto richiesto da SNLG, ha contribuito alla definizione dell'ambito di interesse (*scope)* delle linee guida italiane e in funzione di questo ha nominato i membri del Panel, di cui supervisionerà le attività.

Il Panel è il gruppo consultivo indipendente, multidisciplinare e multiprofessionale, composto da esperti di contenuto e membri laici, che ha lo scopo principale di sviluppare le raccomandazioni della linea guida finale. Le regole di lavoro del panel sono riportate in dettaglio in un documento specifico.

Il CTS e il Panel si avvalgono del supporto di un gruppo di coordinamento ed indirizzo, composto da rappresentanti dell'ONS, del GISMa ed esperti coinvolti nello sviluppo delle Linee Guida Europee, che ha la finalità di coordinare ed organizzare le sessioni di lavoro, così come di contribuire alla redazione dei documenti finali, dei verbali e delle varie rendicontazioni. È cura del CTS, supportato dal gruppo di coordinamento e indirizzo, definire le regole a cui il Panel deve attenersi nel percorso di adolopment monitorandone l'aderenza.

3.2 Ambito di interesse delle linee guida (scope)

Le Linee Guida Europee, cioè le linee guida di partenza per il processo di adolopment, sono composte da 74 domande cliniche per un numero simile di raccomandazioni che coprono le seguenti aree/processi del percorso di diagnosi precoce del cancro della mammella:

- Modalità organizzative dei programmi di screening
- Test di primo livello (es. tipo di esame, frequenza, fascia di età)
- Test di secondo livello
- Diagnosi
- Stadiazione preoperatoria
- Modalità di informazione/invito ai programmi di screening e comunicazione dei risultati
- Formazione degli operatori

Dal punto di vista della prospettiva, le Linee Guida Europee adottano sia quella individuale sia quella di popolazione. Le popolazioni di interesse sono rappresentate dalle donne eleggibili per lo screening mammografico e le donne sottoposte ad approfondimento diagnostico, sia per sintomi e/o

segni, per invio dopo una valutazione del rischio o per richiamo da primo livello di screening. Sono esclusi gli individui di sesso maschile, le ricorrenze loco-regionali dopo diagnosi di cancro, il cancro metastatico, il percorso delle donne ad alto rischio, la prevenzione primaria se non come intervento all'interno di un programma di screening. Le linee guida includono sia il contesto pubblico, sia il contesto privato.

Il CTS si è confrontato rispetto all'ambito di interesse (*scope*) delle linee guida candidate all'adolopment per valutare se mantenere lo stesso interesse per le linee guida nazionali oppure evidenziare eventuali aree carenti o, al contrario, non prioritarie.

Le conclusioni rispetto all'ambito di interesse delle linee guida italiane sono:

- si conferma la priorità per tutti i processi correlati alla diagnosi e allo screening del tumore della mammella.

Nei paragrafi successivi saranno presentate le modalità per l'inclusione delle raccomandazioni su screening e diagnosi.

3.3 Tempi e organizzazione del lavoro

Tutti i PICO delle Linee Guida Europee verranno considerati per l'adozione o adattamento (v. paragrafo 6). Il CTS ha condiviso l'idea di affrontare inizialmente l'argomento fasce d'età ed intervalli e successivamente la personalizzazione dello screening. A seguire verranno valutate le raccomandazioni relative al secondo livello, valutazione e stadiazione preoperatoria e aspetti organizzativi, comunicazione e formazione.

La formulazione di PICO *de novo* sugli argomenti non trattati dalle Linee Guida Europee potrà procedere parallelamente, ma la revisione della letteratura e il percorso di formulazione delle raccomandazioni sarà successivo all'adolopment. Il CTS valuterà la proposta di uno o più PICO inerenti al percorso di diagnosi precoce delle neoplasie della mammella per le donne portatrici di protesi (estetica o ricostruttiva) dopo avere valutato i contenuti delle linee di indirizzo prodotte dal progetto FOCUS ON di Senonetwork.

3.4 Temi che ricadono nell'ambito di interesse (scope) delle Linee Guida Europee

Adozione o adattamento delle raccomandazioni

Le raccomandazioni delle *Linee Guida Europee* possono essere "adottate" oppure "adattate":

- adozione: la direzione e la forza della raccomandazione non sono modificate. Nel caso di raccomandazioni "a condizione" (*conditional recommendations*), in cui le condizioni per preferire un'opzione siano di carattere contestuale (disponibilità di risorse, costo efficacia, impatto organizzativo) il panel ed il CTS possono declinare quali siano le attuali condizioni

del contesto italiano; modifiche sostanziali nel contesto potrebbero portare ad un adattamento della raccomandazione (v. punto successivo)

- adattamento: sono modificate la direzione e/o la forza della raccomandazione.

Il CTS ritiene che le raccomandazioni coerenti sia con la legislazione italiana, sia con le pratiche correnti. È quindi improbabile che ci siano discrepanze nei giudizi di fattibilità e accettabilità, abbiano come esito finale l'adozione. Per questo motivo, il CTS sottopone al Panel i diversi PICO accompagnati da una valutazione preliminare non vincolante degli stessi secondo i criteri di coerenza sopra citati e dando una delle due possibili indicazioni: adozione oppure adattamento.

Ruolo e funzioni della PICO Responsible Unit (PRU)

La PICO responsible unit (unità responsabile di un PICO – PRU) è un gruppo di panelisti che include un metodologo panelista, un membro laico del panel e altri 2/3 panelisti con profilo clinico inerente ai PICO in oggetto o che abbiamo partecipato allo sviluppo dei PICO delle Linee Guida Europee. Possono inoltre partecipare anche i presidenti (chair) o i vicepresidenti (vice chair). I membri di un PRU possono essere candidati dal gruppo di coordinamento o candidarsi spontaneamente.

L'attività del PRU si concretizza nella compilazione dell'EtD per la raccomandazione italiana (finestra "adolopment" in GRADEPro inclusi i giudizi sui criteri) che viene inviato al panel per il pre-voto (almeno due settimane prima della plenaria) e nella presentazione del razionale e sintesi delle valutazioni durante la plenaria stessa.

Metodi per giungere alla formulazione della nuova raccomandazione: adattamento

In questo caso, il CTS sottopone al Panel uno o più PICO con l'indicazione di ripercorrere l'EtD in funzione di un adattamento al contesto italiano.

1. I membri del PRU prendono visione degli EtD delle raccomandazioni europee riportate nel sito ECIBC e chiedono tutti i chiarimenti che ritengono necessari in teleconferenza con i membri dell'ECIBC che hanno contribuito alla stesura delle raccomandaazioni europee, che fanno anche parte del panel italiano, e che hanno partecipato alla loro stesura, indipendentemente che siano essi stessi membri del PRU o meno.
2. Uno o più membri del PRU effettuano una ricognizione preliminare delle evidenze supplementari e le presentano durante una seconda teleconferenza del PRU, durante la quale il gruppo le valuta e si esprime seguendo il diagramma di flusso 1 per le evidenze recenti relative a effetti desiderabili e indesiderabili (desirable e undesirable effects) e il diagramma

di flusso 2 per le evidenze sugli elementi di contesto (costi, costo-efficacia, accettabilità, fattibilità, valori e preferenze)

3. Sulla base delle decisioni prese nel passaggio numero 2, il PRU procede alla stesura dell'EtD italiano per il pre-voto italiano e formula un nuovo giudizio sui criteri per i PICO che non hanno richiesto un passaggio con il CTS secondo quanto previsto nelle flow-chart. Questa fase può svolgersi interamente in autonomia se non sussistono temi particolari che richiedano un ulteriore confronto del PRU in teleconferenza.
4. Il PRU invia al panel almeno due settimane prima del meeting virtuale la bozza dell'EtD per l'effettuazione di un pre-voto del giudizio sui criteri tramite la funzione Panel Voice del software GradePro.
5. Un membro del PRU introduce razionale e sintesi delle valutazioni all'inizio del meeting del panel.
6. Il panel, guidato dai presidenti, procede nella formulazione della raccomandazione secondo le procedure standard indicate dal metodo GRADE (Alonso-Coello P 2016).

Come indicato nei diagrammi di flusso riportati in coda al documento, due casi richiedono un confronto con il CTS prima della prosecuzione della stesura dell'EtD italiano:

- Evidenze su effetti desiderabili e indesiderabili provenienti da studi recenti che possono modificare in modo significativo il giudizio del panel italiano: in questo caso, il CTS valuterà se richiedere una revisione sistematica completa di aggiornamento e valutazione delle evidenze, raccordandosi in questo senso con l'aggiornamento della raccomandazione europea organizzato dal JRC.
- Evidenze di contesto che possono modificare in modo significativo una raccomandazione in adozione: in questo caso, il CTS valuta la mozione del PRU rispetto alla propria valutazione di adozione e può modificare o meno il mandato al PRU verso l'adattamento.

Questi casi possono richiedere un posticipo della discussione del PICO in oggetto rispetto a quanto calendarizzato.

Il PRU può inoltre proporre al panel una riformulazione del PICO qualora non venga più considerato utile o pertinente alle necessità di decisioni cliniche. La riformulazione del PICO obbliga a una nuova revisione sistematica e dunque deve essere sottoposta al processo di prioritizzazione. Un PICO può essere riformulato quando, ad esempio, cambia leggermente la popolazione (es. una

raccomandazione fatta per le sole micro-calcificazioni può essere estesa a micro e micro più lesione o viceversa), oppure l'intervento o il comparatore sono leggermente ridefiniti (es. tomo con o senza 2D sintetica può diventare tomo con 2D sintetica) perché la letteratura ha portato a definire il contesto e il razionale del PICO in modo leggermente differente.

Metodi per giungere alla formulazione della nuova raccomandazione: adozione

In questo caso, il CTS sottopone al Panel uno o più PICO con l'indicazione di adottare la raccomandazione delle *Linee Guida Europee*.

1. Si costituisce un PRU che può essere di numerosità inferiore a quello previsto per le raccomandazioni in adattamento
2. Il PRU acquisisce il materiale caricato su GRADEPro durante il processo di sviluppo delle Linee Guida Europee e prepara una bozza della raccomandazione e delle considerazioni (implementazione, monitoraggio, sottogruppi, ricerca) in italiano.

Non viene richiesto un voto espresso su GRADEPro. Durante questa fase di pre-analisi delle raccomandazioni, i membri del panel possono segnalare la necessità di adattare la raccomandazione e la richiesta viene trasmessa al CTS. Se il CTS accetta la mozione, la discussione del PICO viene rimandata ad una seduta successiva e viene nominato un PRU completo.

3.5 Revisione sistematica della letteratura

3.5.1 Metodi

Quesito clinico

Il quesito clinico prioritarizzato dal CTS è il seguente: dovrebbe una strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana ciascuna mammografia invece che la doppia lettura umana per tutte le mammografie essere utilizzato nello screening mammografico basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico?

Per poter impostare la strategia di ricerca della letteratura e definire i criteri di eleggibilità degli studi, il quesito clinico è stato riformulato secondo il framework PICO, come segue:

- **Popolazione**: Mammografie di screening acquisite mediante mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico.
- **Intervento**: Strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana
- **Comparatore**: Doppia lettura umana

- **Esiti**: Cancri Intervallo (Interval Cancer Rate); Tasso di identificazione dei tumori (Cancer detection rate); Tasso di richiamo (Recall rate); Tasso di falsi positivi (False positive rate); Valore Predittivo Positivo (Positive predictive value); Carcinomi duttali in situ (Ductal carcinoma in situ); Riduzione del carico di lavoro (Workload reduction).

Strategia di ricerca e database

Per garantire la massima completezza della ricerca bibliografica, l'Evidence Review Team ha condotto una revisione sistematica della letteratura, successivamente discussa con l'intera PRU, prendendo in considerazione il periodo compreso tra gennaio 2022 e marzo 2025 (data di primo lancio della ricerca: Novembre 2024; data di aggiornamento della ricerca Marzo 2025). L'avvio della ricerca bibliografica dal 2022 è stato motivato dalla conclusione nel 2023 della precedente revisione del GDG ECIBC (ECIBC, 2023) includendo un anno di sovrapposizione al fine di garantire la completezza. L'analisi è stata effettuata sui database PubMed, Scopus e Ovid (Journals@Ovid Full Text), e le pubblicazioni sono state considerate eleggibili esclusivamente qualora il titolo o l'abstract contenessero le parole chiave riportate in Appendice 1.

Criteri di eleggibilità

Sono stati considerati eleggibili gli studi la cui popolazione, intervento, comparatore e outcome fossero in linea con il PICO prioritarizzato. Non sono stati posti limiti al disegno di studio. Sono stati inclusi articoli pubblicati in lingua inglese o italiano. Pubblicazioni in preprint non sono stati inclusi.

Processo di selezione degli studi ed estrazione dati

Per l'identificazione e la rimozione dei duplicati ed il processo di selezione degli studi è stato utilizzato Rayyan, un software specificamente sviluppato per la gestione delle systematic reviews. La selezione degli studi per titolo e abstract e la selezione dei full text è stata condotta in doppio dai membri dell'Evidence Review Team (ERT). I risultati del processo di selezione degli studi sono stati riportati utilizzando il PRISMA flow diagram. (Figura 1) (Page 2021)

Da ciascuno studi incluso sono state estratte informazioni relative a primo autore, anno, paese di conduzione, setting, disegno di studio, caratteristiche della popolazione inclusa, strategia di triage utilizzata, software IA utilizzato ed esiti.

Valutazione della qualità degli studi inclusi

La valutazione del rischio di bias dei due studi inclusi dalla revisione sistematica della letteratura è stata condotta in doppio utilizzando gli strumenti RoB 2 per gli studi riguardanti il MASAI trial (Lång 2023; Hernström 2025) e ROBINS-I per lo studio di Lauritzen et al. (Lauritzen 2024)

Metodologia di sintesi dei risultati

È stata condotta una sintesi narrativa delle caratteristiche degli studi inclusi. È stato deciso di non effettuare una meta-analisi e di confrontare gli esiti degli studi randomizzati e non randomizzati riportando le stime per

ciascun outcome prioritarizzato nelle summary of findings table, includendo le stime di effetti relativi ed effetti assoluti. La scelta di non condurre una meta-analisi, oltre alla preferenza a non combinare i risultati di studi randomizzati e non-randomizzati, deriva dal fatto che i due studi adottano soglie diverse di malignancy score per distingurere le mammografie da inviare a doppia lettura da quelle da sottoporre a singola lettura umana. Questo determina una eterogeneità spiegata negli outcome recall rate e riduzione del workload. Si è preferito fornire il range degli effetti, dando conto di questa eterogeneità, piuttosto che calcolare una stima pooled con intervallo di confidenza ampio.

Valutazione del grado di certezza delle evidenze

La valutazione del grado di certezza delle evidenze è stata effettuata per singolo outcome utilizzando il metodo GRADE (Schünemann 2017) raccomandato dal manuale metodologico del SNGL (CNEC 2019).

3.5.2 Risultati revisione sistematica della letteratura

La ricerca ha permesso di recuperare complessivamente 3.991 pubblicazioni, di cui 1.403 da PubMed, 2.520 da Scopus e 68 da Ovid.

Il processo di selezione degli studi ha portato all'eliminazione di 1.307 duplicati (1.265 individuati automaticamente da Rayyan e 42 rilevati manualmente dall'Evidence Review Team), riducendo il numero di pubblicazioni da analizzare a 2.684. Successivamente, la selezione degli studi per titolo e abstract, condotta in doppio dall'Evidence Review Team, ha portato all'esclusione di 2.578 pubblicazioni, di cui 5 escluse in quanto non recuperabili, lasciando 106 studi per la valutazione dei full-text. Di questi, 103 sono stati esclusi, concludendo il processo con l'inclusione di 3 pubblicazioni finali relative a due studi. La Figura 1 riporta il PRISMA Flow Diagram (Page 2021), che descrive l'intero processo di identificazione e selezione degli studi.

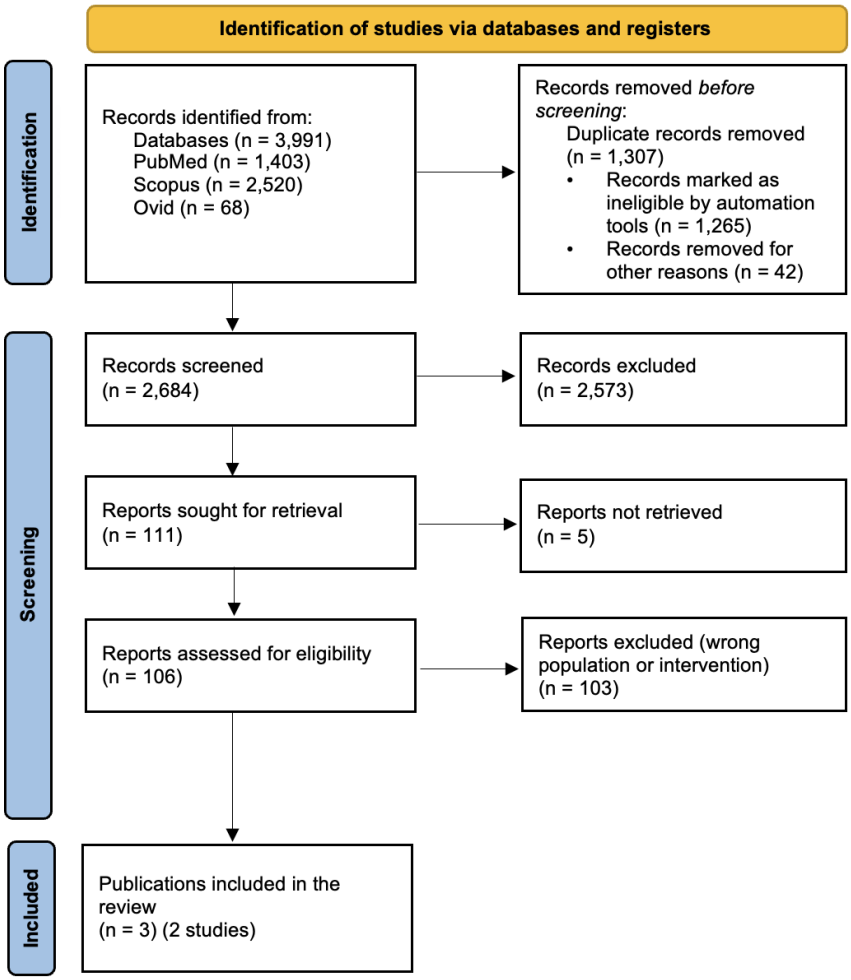


Figura. 1: Flow diagram PRISMA (Page 2021)

Caratteristiche degli studi inclusi

Le tre pubblicazioni analizzate riguardano due studi, le cui caratteristiche principali sono illustrate nella tabella 1 e approfondite nei paragrafi successivi. Entrambi gli studi inclusi hanno riportato evidenze su tutti gli esiti di interesse: tasso di rilevazione dei tumori, tasso di richiamo, tasso di falsi positivi, valore predittivo positivo, carcinoma duttale in situ, riduzione del carico di lavoro.

STUDI	SETTING	DISEGNO DI STUDIO	CAMPIONE	SOFTWARE (IA) UTILIZZATO
MASAI TRIAL – (Lång 2023; Hernström 2025).	Svezia; Centri: Malmö, Lund, Landskrona, Trelleborg; screening aprile 2021–dicembre 2022.	Studio randomizzato, controllato, a gruppi paralleli, di non inferiorità, in singolo cieco.	53.043 (screening con IA) vs 52.872 (doppia lettura senza IA). 40-74 anni.	Transpara (v.1.7.0) ScreenPoint Medical.
Lauritzen 2024	Danimarca; Regione di Copenaghen; screening: ottobre 2020–novembre 2021 (pre-IA) e novembre 2021–ottobre 2022 (post-IA)	Studio pre/post.	60.751 (prima dell'IA), 58.246 (dopo l'IA). 50-70 anni.	Transpara (v.1.7.1) ScreenPoint Medical.

Tabella 1. Caratteristiche degli studi inclusi nella revisione sistematica.

Lauritzen et al. 2024: Lo studio non randomizzato pre/post Lauritzen et al. (Lauritzen 2024), *"Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer",* condotto nel programma di screening mammografico della Regione di Copenhagen, confronta le prestazioni e le differenze nel carico di lavoro prima e dopo l'implementazione di un sistema di intelligenza artificiale (Transpara v.1.7.1, ScreenPoint Medical). Il campione include donne di età compresa tra 50 e 69 anni, sottoposte a screening biennale. Prima dell'introduzione del software di IA, tutte le mammografie sono sottoposte a doppia lettura indipendente con arbitraggio; successivamente, gli esami classificati dall'IA come probabilmente normali (score ≤5, poi ≤7) vengono letti singolarmente da un radiologo senior, mentre gli altri vengono sottoposti a doppia lettura con supporto decisionale dell'IA.

MASAI trial - (Lång 2023; Hernström 2025): Il trial MASAI è uno studio randomizzato, controllato, a gruppi paralleli, di non inferiorità e in singolo cieco, condotto all'interno del programma nazionale di screening in Svezia. Le partecipanti, donne tra 40 e 74 anni, reclutate in quattro centri di screening, sono state randomizzate a screening supportato dall'IA (Transpara v.1.7.0) o a doppia lettura standard senza IA.
Nel gruppo intervento, l'IA attribuisce a ciascun esame uno score di rischio compreso tra 1 e 10H,

in base al quale viene definita la modalità di lettura. Gli esami con punteggio compreso tra 1 e 9 vengono sottoposti a singola lettura da parte di un radiologo con supporto dell'IA, mentre quelli con punteggio pari a 10 sono esaminati attraverso doppia lettura radiologica, anch'essa assistita dall'IA. Gli esami con uno score di rischio uguale a 10H (rischio estremamente elevato) vengono sottoposti a doppia lettura radiologica con supporto dell'IA. In quest'ultimo caso i radiologi sono istruiti a procedere al richiamo in quasi tutti i casi, salvo chiara evidenza di falso positivo.

Valutazione di qualità degli studi inclusi e del grado di certezza delle evidenze.

Il rischio di bias del trial randomizzato MASAI (Lång 2023; Hernström 2025) è stato considerato non serio per tutti gli esiti prioritarizzati. Il rischio di bias dello studio non randomizzato, pre/post (Lauritzen 2024) è stato valutato come serio per possibile confondimento sugli esiti: Tasso di identificazione dei tumori (Cancer detection rate), Valore Predittivo Positivo (Positive predictive value); Carcinomi duttali in situ (Ductal carcinoma in situ). Tale giudizio è dovuto a differenze tra il periodo pre-e post-introduzione dell'IA in termini di tempo dall'ultimo round di screening e di storia di screening delle donne arruolate. Per i restanti esiti si è ritenuto che il rischio di bias delle stime fosse non serio.
La valutazione del grado di certezza delle evidenze è stata effettuata per ciascun esito e valutando i criteri previsti dal metodo GRADE (Schünemann 2017) in doppio e validata tramite confronto tra i membri dell'ERT. I risultati della valutazione sono stati riportati nella tabella summary of findings in Appendice 3.
Sintesi dei risultati

La sintesi dei risultati degli studi inclusi per ciascun outcome prioritarizzato è riportata nella tabella summary of findings in Appendice 3 e nella tabella Evidence-to-Decision (EtD).

3.6 Integrazione delle evidenze per considerazioni aggiuntive

Su suggerimento dei membri della PRU, ulteriori pubblicazioni, sebbene non perfettamente aderenti al PICO definito per la revisione sistematica, sono state considerate per fornire ulteriori informazioni ed incluse nelle considerazioni aggiuntive nell'EtD (Dembrower 2023; Brancato 2025; Sharma 2023; Hovda 2025; Bergan 2024; Oberije 2025; Larsen 2022; Byng 2022; Lång 2021; Graewingholt 2021; Wanders 2022; Chang 2025; Pesapane 2024; Ozcan 2025; Holen 2024; Pearce 2025; de Vries 2021; Tang 2024). Tra queste, uno studio prospettico svedese (Dembrower 2023), in cui l'IA ha sostituito la seconda lettura utilizzando il software Lunit INSIGHT MMG ed uno studio retrospettivo italiano più piccolo (Brancato 2025) che ha utilizzato lo stesso strumento di IA (Lunit INSIGHT MMG) sono stati suggeriti dalla PRU per informare eventuali considerazioni aggiuntive su tasso di richiamo e tasso di indentificazione. Uno studio restrospettivo (Sharma 2023) in cui l'IA (utilizzando lo strumento di IA "MIA") ha sostituito la seconda lettura, multicentrico e condotto su mammografie acquisite da quattro fornitori diversi di apparecchiature mammografiche, è stato utilizzato per informare considrazioni aggiuntive in merito alla generalizzazione delle performance dei software AI in diversi contesti. Le pubblicazioni incluse nelle considerazioni aggiuntive per i domini dell'EtD "effetti desiderabili"

relativamente alla sensibilità per Cancri Intervallo sono state identificate a seguito di un aggiornamento rapido della revisione dei due PICO adottati dall'ECIBC che, avendo un approccio diagnostico, includevano anche studi di confronto di accuratezza diagnostica. (Hovda 2025, Bergan 2024; Oberije 2025; Larsen 2022; Byng 2022; Lång 2021; Graewingholt 2021; Wanders 2022) Una recente pubblicazione che conferma i risultati di quelle già incluse su suggerimento della PRU è stata inclusa nella fase di revisione dell'EtD da parte del panel, successiva alla plenaria (Yu 2025). Sei studi sono stati considerati per fornire informazioni sull'accettabilità dell'IA da parte delle donne nel contesto italiano (Pesapane 2024) ed internazionale (Ozcan 2025; Holen 2024; Pearce 2025; de Vries 2021; Tang 2024).

3.7 Analisi economica e di Costo-Efficacia

L'inclusione di sistemi di intelligenza artificiale nello screening mammografico comporta costi diretti legati non solo all'acquisizione del software, ma anche costi, non inclusi in quest'analisi, legati all'integrazione del software con le infrastrutture esistenti, ai test di interoperabilità e alle attività di validazione necessarie a garantirne l'efficacia. A questi si aggiungono i costi, non inclusi in quest'analisi, per la formazione dei radiologi e per le esercitazioni di simulazione, nonché per il monitoraggio continuo. È importante sottolineare che le risorse necessarie possono variare in maniera significativa tra le diverse organizzazioni del SSN. Le stime presentate in questa sezione mirano a garantire un'applicabilità generale a differenti scenari di implementazione. L'analisi è condotta dalla prospettiva del SSN e ipotizza che il prezzo per lettura includa tutti i costi connessi all'utilizzo di applicazioni di intelligenza artificiale nella lettura delle mammografie intesi come costi addizionali a regime.

Il costo per lettura con l'ausilio di un software di IA è stato stimato da un recente elaborato PoliS-Lombardia (Università di Milano) (https://www.myecole.it/ooc/wp-content/uploads/2024/11/PW_Santicchia-Sonia.pdf) a €2,5. Questa stima si riferisce ad un prezzo per un singolo centro e quindi, verosimilmente, risulta sovrastimata nel contesto di gare d'appalto competitive con volumi di attività più elevati. Si è pertanto adottata una stima di €2 per lettura, ritenuta più realistica ma ancora conservativa. Tale valore risulta in linea con le evidenze riportate dal NICE (NICE 2021), che ha stimato un costo compreso tra £0,60 e £3,00 per lettura, al netto dell'IVA. Si tratta in ogni caso di costi emergenti, che richiedono l'impiego di risorse aggiuntive.

Questi costi sono compensati dalla significativa riduzione del tempo di lettura da parte dei radiologi. Secondo una simulazione recente effettuata al Policlinico di Milano, il costo per ora di lettura (inclusi tempi preparatori e tempi morti fisiologici) è pari a €63 per ora. In base a una media ponderata della retribuzione lorda (RAL) di radiologi senior e junior, questo corrisponde a €4,2 per lettura, nell'ipotesi che vengano svolte 15 letture all'ora.

Assumendo la stessa riduzione del numero di letture osservata in Hernström et al., (Hernström 2025) (44,2%), l'intervento permetterebbe di risparmiare 913 letture ogni 1.000 donne sottoposte a screening. Questo

comporterebbe un risparmio di €1.835 per 1.000 donne. Per un centro che esegue 15.000 mammografie all'anno, ciò si tradurrebbe in un risparmio di €27.520.

Dall'analisi di sensibilità è emerso che l'intervento rimane cost-saving fino ad un costo per esame dell'IA pari a € 3,83.

I dati di efficacia dei due studi suggeriscono un effetto positivo o almeno neutro dal punto di vista dell'efficacia complessiva. Allo stesso tempo entrambi gli studi mostrano una riduzione significativa del numero di letture dei radiologi. La riduzione delle seconde letture può inoltre ridurre il tempo di refertazione delle mammografie e liberare tempo ai radiologi per i secondi livelli, riducendo i tempi d'attesa per le risposte negative e per il prosieguo del percorso diagnostico/terapeutico.

Nonostante quindi gli oneri iniziali, l'intervento può generare risparmi attraverso la riduzione del carico di lavoro dei radiologi e il miglioramento dell'efficienza organizzativa complessiva.

Come anticipato nell'introduzione di questa sezione, le risorse richieste possono variare significativamente tra le diverse organizzazioni del SSN. Alcune dispongono già delle infrastrutture tecnologiche necessarie per una rapida implementazione (es. PACS e RIS), mentre altre potrebbero necessitare di investimenti più consistenti per acquisire e integrare tali sistemi. I prezzi futuri delle soluzioni di IA sono difficili da prevedere, ma è probabile che in una fase iniziale le aziende mantengano prezzi contenuti per acquisire quote di mercato. Più critiche per il SSN potrebbero risultare le strategie di pricing nel medio periodo, anche in relazione al rischio di situazioni di monopolio o lock-in da parte delle imprese.

3.8 Evidenze a supporto della valutazione di accettabilità e fattibilità

Per quanto riguarda i domini di accettabilità e fattibilità, non è stata condotta una revisione sistematica della letteratura. Le evidenze qui riportate provengono dai due studi inclusi (Hernström et al., 2025; Lauritzen et al., 2024), da una survey condotta tra 305 professionisti attivi nei centri di screening italiani per esplorare le attitudini verso l'impiego dell'IA, da uno studio di Pesapane et al. (2024) sulle percezioni delle pazienti discusso con la PRU e da cinque ulteriori studi utilizzati come riferimento per un confronto internazionale (Ozcan 2025; Holen 2024; Pearce 2025; de Vries 2021; Tang 2024).

Survey sull'utilizzo dell'intelligenza artificiale nei programmi di screening italiani:

La survey è stata promossa dalla Federazione delle Associazioni degli Screening Oncologici (FASO) e dal centro di ricerca CERGAS (Centre for Research on Health and Social Care Management, SDA Bocconi. La survey è stata condotta nel rispetto della normativa italiana vigente e non ha richiesto approvazione etica, poiché rivolta a professionisti delle reti dell'Osservatorio Nazionale Screening (ONS) e FASO nell'ambito delle loro funzioni istituzionali. I risultati sono stati raccolti tramite i referenti regionali screening che hanno coinvolto i referenti locali. La survey è stata creata tramite Google Form, si è ispirata al framework NASSS (Greenhalgh et al., 2017) ed è stata pretestata per assicurarne chiarezza e validità.

L'obiettivo principale era quello di indagare il livello di penetrazione delle applicazioni di intelligenza artificiale nei programmi di screening oncologici italiani, analizzandone le potenzialità e le principali barriere all'implementazione.

Il questionario si articola in due sezioni: una generale, destinata ai professionisti di centri screening erogatori delle prestazioni di primo e secondo livello nel processo di screening oncologico (screening LEA mammografico, colorettale e cervicale, nuovi screening prostata e polmone, screening HCV, segreterie organizzative) e una specifica dedicata allo studio dell'utilizzo delle applicazioni di intelligenza artificiale nello screening per il tumore della mammella.

La survey ha raccolto 305 risposte, con la rappresentanza di tutte le regioni italiane ad eccezione del Molise e della Provincia Autonoma di Bolzano.

I risultati evidenziano un forte interesse verso le potenzialità dell'IA nei processi di screening (74,7%), in particolare per quanto riguarda gli algoritmi di classificazione del rischio e il supporto alla diagnostica offline.

Per contro, le principali barriere all'introduzione dei sistemi di IA nella pratica clinica sono identificate dall'80% del campione nella mancanza di linee guida ufficiali, nell'insufficienza di risorse economiche per sostenere gli investimenti necessari e nella carenza di un'infrastruttura IT adeguata.

Allo stato attuale, l'adozione effettiva di sistemi di IA da parte del campione risulta limitata: solo il 18% ha dichiarato di utilizzare tali soluzioni, di cui il 9% nell'ambito delle attività di screening, organizzato e opportunistico, ed il restante 9% per finalità di ricerca. Infine, la survey ha esplorato il livello di accordo sull'implementazione di un modello ispirato al trial MASAI (Lång 2023; Hernström 2025), evidenziando una valutazione complessivamente positiva in termini di volontà di sperimentazione, desiderabilità e fattibilità.

3.9 Formato previsto delle linee guida finali

Le linee guida saranno disponibili attraverso la pagina di consultazione delle linee guida di SNLG e sul sito dell'Osservatorio Nazionale Screening. Inoltre, le società scientifiche facenti parte del CTS dovranno segnalarle sui loro siti.

3.10 Modalità di comunicazione ad interim e finale

Una volta completato un gruppo di raccomandazioni che costituisca un corpus sensato e di senso compiuto di raccomandazioni, questo viene inviato a SNLG per la valutazione e l'approvazione. I metodi e i risultati delle raccomandazioni saranno anche resi disponibili sotto forma di pubblicazione scientifica nazionale o internazionale.

3.11 Prospettive per lo sviluppo di un "percorso donna"

Un limite dell'approccio delle linee guida "per PICO" è la frammentarietà delle raccomandazioni, che appaiono come indicazioni scollegate e non permettono di delineare un percorso che la donna e i professionisti possano seguire. Il CTS ha discusso della possibilità di ampliare l'ambito di attività del panel

al di là dell'adolopment dell'ECIBC fino ad arrivare alla proposta di un PDTA nazionale per la prevenzione e la diagnosi del tumore della mammella nelle donne a medio rischio. L'obiettivo sarebbe fornire alla donna e ai portatori di interesse raccomandazioni basate sulla migliore evidenza disponibile, su quali interventi diagnostici o di prevenzione possono essere adottati con l'obiettivo di prevenire o diagnosticare precocemente il tumore della mammella in base alla propria fascia di età. Questa progettualità è stata ritenuta di interesse e sarà ridiscussa ed eventualmente oggetto di un aggiornamento di questo documento metodologico al termine del percorso di adolopment delle Linee Guida Europee.

4. Revisione esterna

Il documento che include gli EtD approvati e revisionati dal Panel è stato inviato ad esperti esterni per una revisione del contenuto e dell'interpretazione delle prove a supporto delle raccomandazioni. Sono stati identificati due revisori le cui competenze coprono gli ambiti principali di interesse delle raccomandazioni sottoposte a revisione: un revisore con profilo diagnostico-clinico (radiologa) e un'epidemiologa esperta di sanità pubblica e di metodologia di sviluppo delle linee guida. I membri del panel estensori della versione finale degli EtD hanno prodotto un riscontro punto per punto delle osservazioni dei revisori, che hanno dato luogo a modifiche del testo quando ritenute opportune dal Panel. Tutti gli esperti hanno dichiarato il loro conflitto di interessi.

5. Applicabilità

Le raccomandazioni di questa LG sono state elaborate e formulate per rendere direttamente applicabili alla popolazione italiana le evidenze scientifiche sulla diagnosi precoce del cancro della mammella. Il Panel segnala la presenza di fattori potenzialmente ostacolanti quali la persistenza di percorsi di tipo opportunistico e spontaneo per sottoutilizzo dei percorsi organizzati per scarsa conoscenza o problematiche organizzative. Questa LG si propone come primo documento condiviso da tutte le società scientifiche coinvolte nel percorso verso una maggiore standardizzazione e appropriatezza del percorso di diagnosi precoce in Italia.

6. Aggiornamento della LG: tempi e modalità

In considerazione della continua evoluzione delle conoscenze medico scientifiche e della conseguente disponibilità della letteratura d'interesse, l'aggiornamento del documento è previsto in concordanza con l'aggiornamento delle linee guida europee, descritto nel documento ECIBC Guidelines updating strategy workflow (https://healthcare-quality.jrc.ec.europa.eu/ecibc/methodologies/guidelines-updating).

7. Conflitti di interesse editoriale

I membri del panel hanno sottoscritto una dichiarazione su eventuali conflitti di interesse. I membri del panel si astengono dalla votazione della forza della raccomandazione nei seguenti casi:

- Quando fanno parte dell'*authorship* di uno o più lavori considerati per la raccomandazione.
- Quando hanno ricevuto finanziamenti diretti o indiretti da aziende titolari dell'intervento che si sta prendendo in esame.

8. Bibliografia

- Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016;353:i2016. Published 2016 Jun 28. doi:10.1136/bmj.i2016
- Bergan MB, Larsen M, Moshina N, et al. AI performance by mammographic density in a retrospective cohort study of 99,489 participants in BreastScreen Norway. Eur Radiol. 2024;34(10):6298-6308. doi:10.1007/s00330-024-10681-z
- Brancato B, Magni V, Saieva C, et al. AI-supported approaches for mammography single and double reading: A controlled multireader study. Eur J Radiol. 2025;187:112101. doi:10.1016/j.ejrad.2025.112101
- Byng D, Strauch B, Gnas L, et al. AI-based prevention of interval cancers in a national mammography screening program. Eur J Radiol. 2022;152:110321. doi:10.1016/j.ejrad.2022.110321
- CNEC (Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure) Manuale metodologico per la produzione di linee guida di pratica clinica. 2019 https://snlg.iss.it/wp-content/uploads/2019/04/MM_v1.3.2_apr_2019.pdf
- Chang YW, Ryu JK, An JK, et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study. Nat Commun. 2025;16(1):2248. Published 2025 Mar 6. doi:10.1038/s41467-025-57469-3
- de Vries CF, Morrissey BE, Duggan D, Staff RT, Lip G. Screening participants' attitudes to the introduction of artificial intelligence in breast screening. J Med Screen. 2021;28(3):221-222. doi:10.1177/09691413211001405
- Dembrower K, Crippa A, Colón E, Eklund M, Strand F; ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health. 2023;5(10):e703-e711. doi:10.1016/S2589-7500(23)00153-X
- ECIBC. European Commission Initiative on Breast Cancer. Recommendations for the European Breast Cancer Guidelines, 2019. Available at: https://ecibc.jrc.ec.europa.eu/recommendations
- ECIBC, European guidelines on breast cancer screening and diagnosis. Use of artificial intelligence. Published on Jult 2023. Available at: https://cancer-screening-and-care.jrc.ec.europa.eu/en/ecibc/european-breast-cancer-guidelines?topic=64&usertype=60&updatef2=0
- Graewingholt A, Rossi PG. Retrospective analysis of the effect on interval cancer rate of adding an artificial intelligence algorithm to the reading process for two-dimensional full-field digital mammography. J Med Screen. 2021;28(3):369-371. doi:10.1177/0969141320988049

- Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025;7(3):e175-e183. doi:10.1016/S2589-7500(24)00267-X
- Holen ÅS, Martiniussen MA, Bergan MB, Moshina N, Hovda T, Hofvind S. Women's attitudes and perspectives on the use of artificial intelligence in the assessment of screening mammograms. Eur J Radiol. 2024;175:111431. doi:10.1016/j.ejrad.2024.111431
- Hovda T, Larsen M, Bergan MB, Gjesvik J, Akslen LA, Hofvind S. Retrospective evaluation of a CE-marked AI system, including 1,017,208 mammography screening examinations. Eur Radiol. 2025;35(9):5685-5694. doi:10.1007/s00330-025-11521-4
- Lång K, Josefsson V, Larsson AM, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. 2023;24(8):936-944. doi:10.1016/S1470-2045(23)00298-X
- Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I. Can artificial intelligence reduce the interval cancer rate in mammography screening?. Eur Radiol. 2021;31(8):5940-5947. doi:10.1007/s00330-021-07686-3
- Larsen M, Aglen CF, Lee CI, et al. Artificial Intelligence Evaluation of 122 969 Mammography Examinations from a Population-based Screening Program. Radiology. 2022;303(3):502-511. doi:10.1148/radiol.212381
- Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. Radiology. 2024;311(3):e232479. doi:10.1148/radiol.232479
- NICE 2021, https://www.nice.org.uk/advice/mib242/resources/artificial-intelligence-in-mammography-pdf-2285965629587653.2021
- Oberije CJG, Currie R, Leaver A, et al. Assessing artificial intelligence in breast screening with stratified results on 306 839 mammograms across geographic regions, age, breast density and ethnicity: A Retrospective Investigation Evaluating Screening (ARIES) study. BMJ Health Care Inform. 2025;32(1):e101318. Published 2025 May 14. doi:10.1136/bmjhci-2024-101318
- Ozcan BB, Dogan BE, Xi Y, Knippa EE. Patient Perception of Artificial Intelligence Use in Interpretation of Screening Mammograms: A Survey Study. Radiol Imaging Cancer. 2025;7(3):e240290. doi:10.1148/rycan.240290
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. Published 2021 Mar 29. doi:10.1136/bmj.n71
- Pearce A, Carter S, Frazer HM, et al. Implementing artificial intelligence in breast cancer screening: Women's preferences. Cancer. 2025;131(9):e35859. doi:10.1002/cncr.35859
- Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, 2006. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth Edition. European Breast Cancer Network

(EBCN), Lyon. Luxembourg: Office for Official Publications of the European Communities, 2006. ISBN 92-79-01258-4

- Pesapane F, Giambersio E, Capetti B, et al. Patients' Perceptions and Attitudes to the Use of Artificial Intelligence in Breast Cancer Diagnosis: A Narrative Review. Life (Basel). 2024;14(4):454. Published 2024 Mar 29. doi:10.3390/life14040454
- Schünemann HJ, Wiercioch W, Brozek J, et al. GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. J Clin Epidemiol. 2017;81:101-110. doi:10.1016/j.jclinepi.2016.09.009
- Sharma N, Ng AY, James JJ, et al. Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms. BMC Cancer. 2023;23(1):460. Published 2023 May 19. doi:10.1186/s12885-023-10890-7
- Tang JS, Frazer HM, Kunicki K, Basnayake P, Omori M, Lippey J. Australian healthcare workers' views on artificial intelligence in BreastScreen: Results of a mixed method survey study. Prev Med Rep. 2024;48:102917. Published 2024 Oct 28. doi:10.1016/j.pmedr.2024.102917
- Wanders AJT, Mees W, Bun PAM, et al. Interval Cancer Detection Using a Neural Network and Breast Density in Women with Negative Screening Mammograms. Radiology. 2022;303(2):269-275. doi:10.1148/radiol.210832
- Yu TT, Hoyt AC, Joines MM, et al. Mammographic classification of interval breast cancers and artificial intelligence performance. J Natl Cancer Inst. 2025;117(8):1627-1638. doi:10.1093/jnci/djaf103.

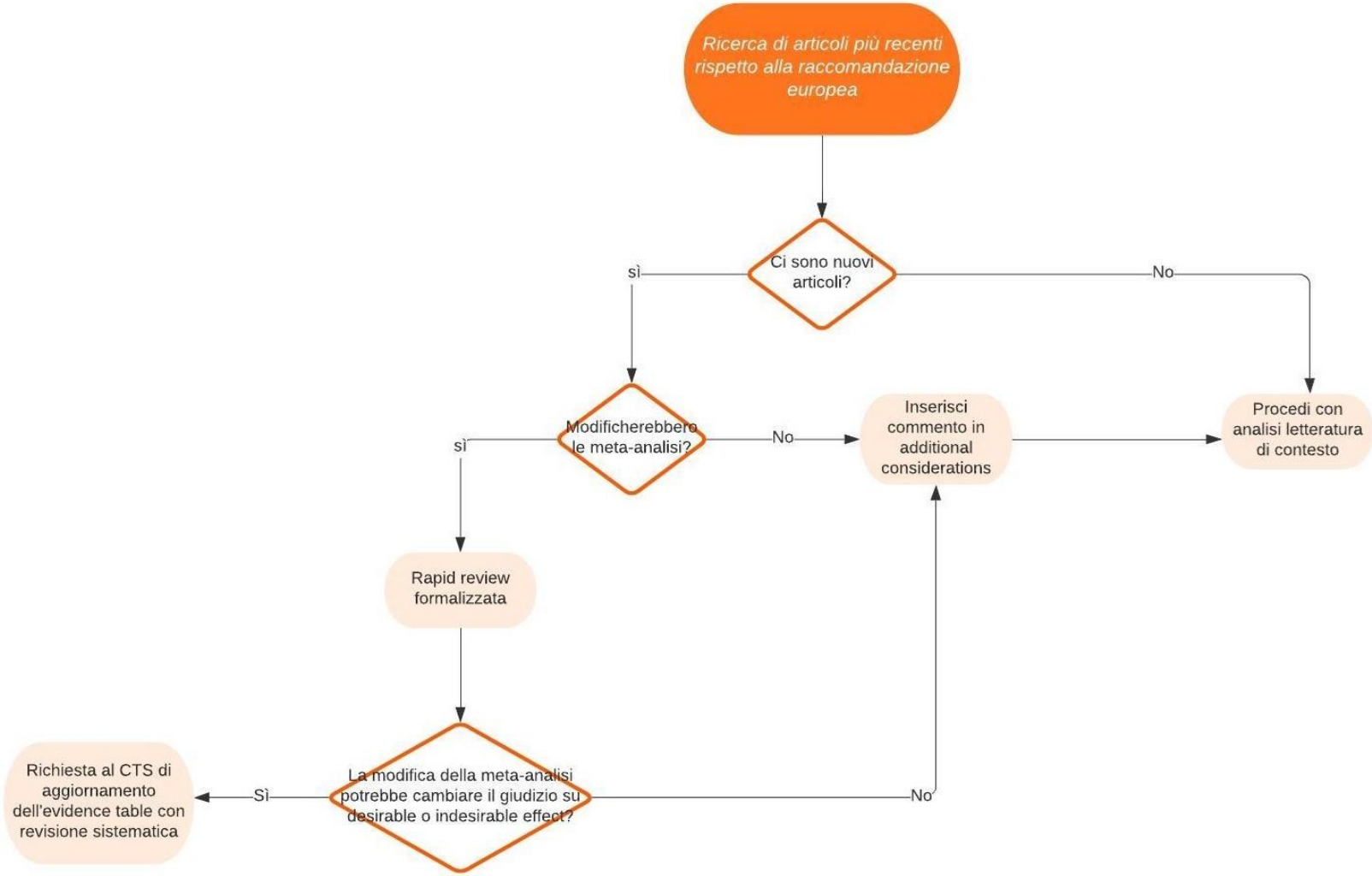


Diagramma di flusso 1: effetti desiderabili e indesiderabili

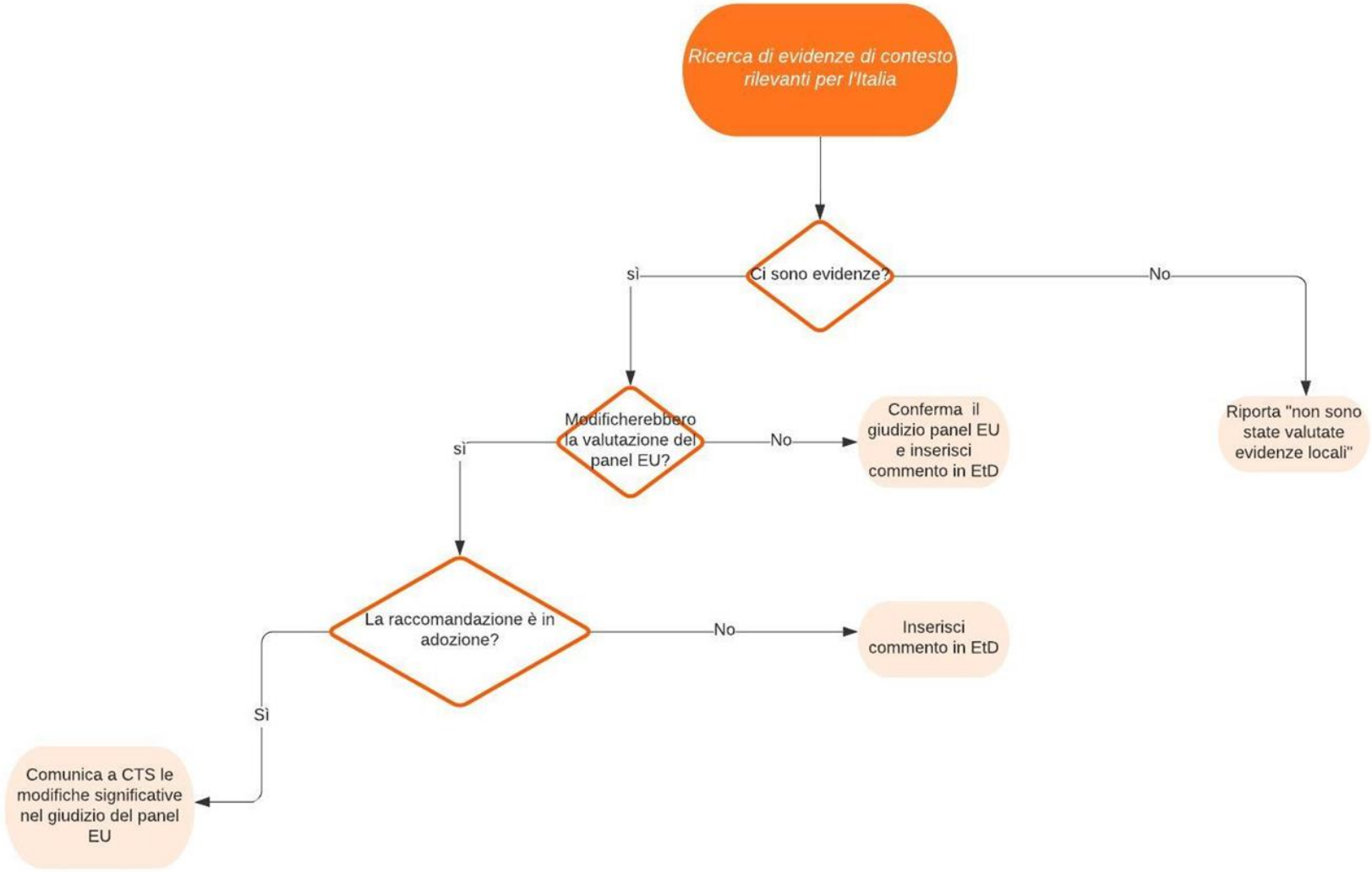


Diagramma di flusso 2: Elementi di contesto

9. Regole per il panel

9.1 Presidente (Chair)

Presidente clinico: Francesco Sardanelli

Presidente metodologico: Francesco Venturelli

Presidente epidemiologico: Carlo Senore

Dal verbale dell'11/11/2020 (Allegato 1) e verbale del 14/02/2023 (Allegato 2). In seguito al rinnovo dei presidenti delle Società Scientifiche anche i membri del CTS e del Panel sono stati in parte rinnovati nel corso dell'aprile 2025. Tutti i Presidenti delle Società scientifiche hanno dato il loro assenso alla partecipazione o hanno nominato un loro sostituto.

l panel prevede tre tipologie di presidente: clinico, epidemiologico e metodologico. I tre soggetti individuati rappresentano competenze, rispettivamente, in ambito diagnostico del cancro della mammella, in ambito screenologico e in metodologia di sviluppo delle linee guida con GRADE.

I presidenti sono scelti tra i membri del panel da parte del CTS su proposta del gruppo di coordinamento e indirizzo e sono eletti nel corso della prima riunione del panel.

Ciascuna plenaria del panel è condotta da uno dei presidenti, a seconda dell'ordine del giorno.

9.2 Modalità di sostituzione dei presidenti e dei membri del panel dimissionari

In caso di dimissioni di un presidente, si ripete la procedura di elezione dettagliata al punto 1 per identificare un sostituto.

9.3 Membri associati/supplementari

Il CTS può individuare delle figure professionali non rappresentate all'interno del panel (ad esempio economista, antropologo, etc.) per ottenere contributi su tematiche specifiche. A queste professionalità saranno applicate le procedure relative al conflitto di interesse.

9.4 Osservatori

Non è prevista la partecipazione di osservatori.

9.5 PICO Responsible Unit - PRU

Il panel può avviare lavori attraverso PRU dedicato ad un singolo PICO o a un gruppo di PICO collegati fra loro. I PRU devono vedere integrate le varie funzioni e competenze, in particolare devono prevedere la partecipazione di chi ha condotto le revisioni e le eventuali integrazioni alla letteratura e gli esperti del panel. In ogni PRU, ci sarà un rapporteur che si farà carico della presentazione delle conclusioni della PRU in plenaria e che, nel caso non siano presenti all'interno della PRU, si rapporterà con i presidenti ed i vicepresidenti.

9.6 Agenda dei meeting

L'ordine del giorno dei meeting del panel è proposto dall'Osservatorio Nazionale Screening, in accordo con le indicazioni dello Steering Committee.

9.7 Convocazione dei meeting

La convocazione dei meeting è a cura della segreteria ONS. In considerazione anche della particolare situazione sanitaria, saranno inizialmente programmate sessioni in videoconferenza che potranno essere successivamente affiancate o sostituite, se le condizioni lo permettono, da riunioni in presenza.

9.8 Preparazione dei meeting

La prima seduta del panel prevede la candidatura e nomina dei presidenti. Segue quindi una sessione di formazione e a seguire la votazione del primo PICO. A partire dalla seduta successiva, per ogni sessione di plenaria (che può avere in agenda uno o più PICO) la segreteria ONS richiede ai panelisti di indicare l'interesse per la tematica in discussione e di garantire la presenza alla plenaria con effettuazione del pre-voto tramite la funzione Panel Voice del software GradePro nei tempi richiesti. Il CTS insieme al gruppo di coordinamento valuta se il numero di panelisti che hanno dato conferma di adesione per la data può plausibilmente permettere di includere tutte le professionalità ritenute indispensabili dalla specifica tematica dei PICO e una pluralità di prospettive. Se le condizioni non sono verificate si procede a una cooptazione attiva di ulteriori membri del panel.

9.9 Regole di deliberazione del panel

Il panel effettua l'attività propria di formulazione delle raccomandazioni in accordo alle regole del metodo GRADE-Evidence to Decision, delineate nelle pubblicazioni corrispondenti. Per la direzione della raccomandazione e per il giudizio sui criteri dell'EtD è sufficiente la maggioranza semplice, mentre per la forza della raccomandazione è necessaria una maggioranza qualificata (80%).

Perché la votazione sia valida è necessario che sia presente il 50%+1 dei panelisti che hanno dato l'adesione per il PICO in discussione, sia per le votazioni in plenaria sia per le votazioni in pre-voto con il software.

Il Panel è tenuto a deliberare su tutte le raccomandazioni, sia quelle in adattamento che quelle in adozione. La deliberazione sulla raccomandazione in adozione avviene esclusivamente attraverso la funzione di voto del software GRADEPro, tranne nel caso di raccomandazione finale per cui ha espresso disaccordo almeno il 20% dei panelisti. In questo caso, si ritiene il voto elettronico non valido e si procede a votazione in plenaria.

9.10 Verbali delle sedute

I contenuti sono rappresentati dai risultati dell'EtD, con l'aggiunta di eventuali altre decisioni extra EdT. Il verbale è redatto dalla segreteria ONS.

9.11 Protezione dati personali

I dati personali dei membri del panel sono conservati a cura della segreteria ONS in modalità sicura.

9.12 Gestione del conflitto di interesse

Il CTS valuta i moduli di conflitto compilati dai tutti i componenti del panel e assegna, seguendo i principi definiti dal Guidelines International Network (Schunemann et al, 2015) ad ogni dichiarazione uno di tre livelli di potenziale conflitto:

1. minimo o insignificante;
2. potenzialmente rilevante;
3. rilevante.

Ad ogni livello corrispondono le misure da intraprendere secondo la tabella riportata:

La valutazione del conflitto di interesse è specifica per gruppi di PICO collegati fra loro. Si userà il modulo proposto dal Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure dell'Istituto Superiore di Sanità (appendice 5 del Manuale metodologico per la produzione di linee guida di pratica clinica https://snlg.iss.it/wp-content/uploads/2019/04/MM_v1.3.2_apr_2019.pdf). (CNEC 2019)

L'analisi dei COI è stata effettuata nel mese di Aprile 2025 tramite consultazione del CTS via mail. Per nessun membro del panel è stata inibita la partecipazione allo sviluppo delle raccomandazioni.

9.13 Regole sulla confidenzialità

Si richiede al panel di mantenere la confidenzialità sui contenuti delle sedute fino al momento della pubblicazione (release) della raccomandazione. Saranno comunque messi a disposizione sul sito ONS i documenti principali, il cronoprogramma e un aggiornamento puntuale su quali PICO sono stati già trattati.

9.14 Riferimenti

- CNEC (Centro Nazionale per l'Eccellenza Clinica, la Qualità e la Sicurezza delle Cure) Manuale metodologico per la produzione di linee guida di pratica clinica. 2019 https://snlg.iss.it/wp-content/uploads/2019/04/MM_v1.3.2_apr_2019.pdf
- Schünemann HJ, Al-Ansary LA, Forland F, et al. Guidelines International Network: Principles for Disclosure of Interests and Management of Conflicts in Guidelines. Ann Intern Med. 2015;163(7):548-553. doi:10.7326/M14-1885.

EVIDENCE TO DECISION FRAMEWORK - QUESITO 3

Dovrebbe una strategia di triage basato sull'intelligenza artificiale (IA) per assegnare a singola o doppia lettura umana ciascuna mammografia invece che la doppia lettura umana per tutte le mammografie essere utilizzato nello screening mammografico basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico?	
POPOLAZIONE:	Mammografie di screening acquisite mediante mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico.
INTERVENTO:	Strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana
COMPARATORE:	Doppia lettura umana
ESITI PRINCIPALI:	Cancri Intervallo (Interval Cancer Rate); Tasso di identificazione dei tumori (Cancer detection rate); Tasso di richiamo (Recall rate); Tasso di falsi positivi (False positive rate); Valore Predittivo Positivo (Positive predictive value); Carcinomi duttali in situ (Ductal carcinoma in situ); Riduzione del carico di lavoro (Workload reduction).
SETTING:	Italia
PROSPETTIVA:	Popolazione (Sistema Sanitario Nazionale)
RAZIONALE:	Il tumore al seno è il tumore più comune tra le donne, non solo nell'UE, ma in tutto il mondo. Le statistiche del Sistema Informativo Europeo sul Cancro (ECIS) stimano oltre 350.000 nuovi casi di tumore al seno diagnosticati nel 2020 solo nell'UE-27, pari al 29% di tutte le nuove diagnosi di tumore nelle donne. La pubblicazione "I Numeri del Cancro in Italia 2024" riporta una stima di 53.065 nuove diagnosi di tumore al seno tra le donne nel 2024, pari al 30% delle nuove diagnosi di tumore nelle donne. (https://www.aiom.it/i-numeri-del-cancro-in-italia/) I programmi di screening svolgono un ruolo cruciale nella diagnosi precoce del tumore al seno; possono aumentare le possibilità di sopravvivenza e ridurre la mortalità specifica per malattia. La mammografia rimane il metodo migliore per individuare il tumore al seno in fase precoce. Il Guideline Development Group (GDG) dell'Eureopean Commission Initiative on Brest Cancer (ECIBC) raccomanda attualmente lo screening mammografico utilizzando un approccio a doppia lettura, poiché ne migliora la sensibilità e probabilmente l'efficacia. In molti programmi la disponibilità di radiologi è limitata e ciò rende difficile sostenere programmi di screening, creando ritardi negli inviti e nei tempi di refertazione. La riduzione delle doppie letture potrebbe in parte risolvere il problema della scarsità di radiologi. È stato dimostrato che l'intelligenza artificiale (IA) può essere potenzialmente utile per supportare, integrare o sostituire il lavoro umano. Inoltre, l'IA potrebbe superare alcune limitazioni come l'affaticamento e la variabilità tra letture, oltre a ridurre il carico di lavoro. La possibilità di ridurre le seconde letture, tramite un triage effettuato sulla base dell'intelligenza artificiale, potrebbe avere anche il vantaggio di ridurre i costi, la complessità di implementazione e i tempi di attesa. Il triage per la seconda lettura è stato proposto in letteratura con modalità leggermente differenti, variando: la soglia di score utilizzata per il triage (per attribuire a una o due letture umane); le modalità di interazione tra IA e lettura del radiologo (cioè se lettura cieca rispetto alle informazioni elaborate dall'IA, parzialmente informata dall'IA o completamente informata, e le modalità di arbitrato e i livelli di discordanza a cui applicarlo); la presenza di livelli di score per il quale si prevede l'invio ad approfondimenti a prescindere dall'esito delle letture umane. In questo PICO sono inclusi interventi indipendentemente dalla soglia adottata, dalle modalità di lettura del radiologo, dalle modalità di applicazione dell'arbitrato e dalla presenza di piccoli gruppi per i quali il richiamo è indipendente dalla lettura umana (Appendice 2).
CONFLITTO DI INTERESSI:	Gestione del conflitto di interesse (CoI): Le dichiarazioni di interesse di tutti i membri del panel sono state valutate dal CTS e gestite dall'ONS seguendo la procedura adottata per il Sistema Nazionale Linee Guida. Per nessun membro del panel è stata inibita la partecipazione allo sviluppo delle raccomandazioni.

ASSESSMENT

Problem Is the problem a priority?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ○ Probably yes ● Yes ○ Varies ○ Don't know	Il Guideline Development Group dell'ECIBC raccomanda attualmente lo screening mammografico utilizzando un approccio a doppia lettura, poiché ne migliora la sensibilità e probabilmente l'efficacia. In molti programmi la disponibilità di radiologi è limitata e ciò rende difficile sostenere programmi di screening, creando ritardi negli inviti e nei tempi di refertazione. La riduzione delle doppie letture potrebbe in parte risolvere il problema della scarsità di radiologi. L'intelligenza artificiale (IA) può essere potenzialmente utile per supportare, integrare o sostituire il lavoro umano. Inoltre, l'IA potrebbe superare alcune limitazioni come l'affaticamento e la diagnosi soggettiva, oltre a ridurre il carico di lavoro. La possibilità di ridurre il numero di seconde letture, tramite un triage effettuato sulla base dell'intelligenza artificiale, potrebbe avere il vantaggio di ridurre i costi, la complessità di implementazione e i tempi di attesa.	Giudizio espresso all'unanimità.

Desirable Effects How substantial are the desirable anticipated effects?		
Judgement	**Research evidence**	**Additional considerations**
○ Trivial ○ Small ● Moderate ○ Large ○ Varies ○ Don't know	Hernström 2025 e Lauritzen 2024 (1,2) indicano che lo screening integrato con IA è associato ad un aumento dei tassi di identificazione dei tumori. I tassi di identificazione dei tumori sono aumentati, rispettivamente, del 18% e del 29% nei due studi inclusi. L'effetto assoluto su 10.000 donne sottoposte a screening va da 4 a 25 casi in più in Hernström 2025 (1) e da 3 a 24 in più in Lauritzen 2024 (2). In Hernström 2025 (1) si osserva un aumento dell'8% non statisticamente significativo del tasso di richiamo con lo screening integrato con IA, con effetto assoluto da 2 richiami in meno a 33 richiami in più ogni 10.000 donne sottoposte a screening. Lo studio di Lauritzen 2024 (2) ha registrato una riduzione del tasso di richiamo del 21% statisticamente significativa, con effetto assoluto da 80 a 46 richiami in meno ogni 10.000 donne sottoposte a screening. In Hernström 2025 (1) il tasso di falsi positivi è simile tra i due bracci dello studio, mentre Lauritzen 2024 (2) ha riportato una riduzione del tasso di falsi positivi statisticamente significativa del 32%, con effetto assoluto da 88 a 62 falsi positivi in meno ogni 10.000 donne sottoposte a screening nel braccio di intervento. (Appendice 3)	Risultati simili sono stati ottenuti in uno studio prospettico svedese (3), in cui l'IA ha sostituito la seconda lettura e si dimostra una riduzione del tasso di richiamo del 4% (relative recall 0.96; 95%CI 0.94-0.97) e un aumento del 4% di detection (relative detection 1.04; 95%CI 1.00-1.09) utilizzando il software Lunit INSIGHT MMG. Uno studio retrospettivo italiano più piccolo (4) ha riportato risultati simili con lo stesso strumento di IA (Lunit INSIGHT MMG). Sharma 2023 (5), utilizzando lo strumento di IA "MIA", ha dimostrato che la doppia lettura con IA, rispetto alla doppia lettura umana storica, garantisce la generalizzazione su quattro fornitori diversi di apparecchiature mammografiche e sedi, nonché a livello geografico, con almeno un tasso di richiamo, CDR, sensibilità, specificità e PPV non inferiori. Interval cancer: uno studio retrospettivo su oltre 1,000,000 di mammografie mostra che l'IA individua circa il 38% dei cancri negativi alla doppia lettura e il 93% degli screen detected (6). Studi simili condotti con diversi sistemi di IA riportano dati simili con un range di cancri screen detected identificati da 89% a 93%, di cancri intervallo fra 20% e 50%, a soglie di specificità >=90% (7)(8)(9) (10) (11)(12)(13)(14). Queste evidenze indirette da studi retrospettivi suggeriscono che sia poco probabile che l'introduzione dell'intervento sia associata ad un aumento dei cancri intervallo. Il panel sottolinea che non sono ancora presenti evidenze sull'andamento dei cancri intervallo da studi prospettici. Giudizio espresso all'unanimità.

Undesirable Effects How substantial are the undesirable anticipated effects?		
Judgement	**Research evidence**	**Additional considerations**
○ Large ○ Moderate ○ Small ● Trivial ○ Varies ○ Don't know	Appendice 3	Parte dei casi di richiamo aggiuntivi con IA nello studio randomizzato sono controbilanciati dall'aumento dei casi rilevati. (1) Lauritzen 2024 (2) osserva i pazienti per un breve periodo di follow-up (massimo 180 giorni). La valutazione dei tassi di cancri intervallo di Hernström 2025 (1) è attesa a breve e offrirà informazioni più solide sull'accuratezza diagnostica. Dati preliminari dallo studio di Hernström riportati all' European Congress of Radiology 2025 hanno segnalato una riduzione dei cancri d'intervallo non-luminal A del 27%. Gli studi inclusi sono stati condotti utilizzando un solo strumento di IA (Transpara), ciò suggerisce un rischio di dinamiche monopolistiche nel mercato per questa specifica applicazione. Tuttavia, come già menzionato, altri strumenti di IA hanno mostrato risultati comparabili in studi retrospettivi e sono attualmente in fase di ulteriore valutazione in studi prospettici (es. (3); trial MIA nel Regno Unito (ISRCTN - https://doi.org/10.1186/ISRCTN95571932); (15)). Nel giudizio espresso sul possibile aumento del tasso di richiamo il panel ha tenuto conto del contestuale aumento di tasso di rilevazione dei tumori e dell'aumento del valore predittivo positivo. Il panel sottolinea che non sono ancora presenti evidenze sull'andamento dei cancri intervallo e dei cancri ai successivi passaggi. Queste evidenze sono necessarie a escludere che l'aumento del tasso di identificazione sia a carico di cancri a lentissima crescita e dunque possa aumentare la sovradiagnosi. Giudizio espresso all'unanimità.

Certainty of evidence

What is the overall certainty of the evidence of effects?

Judgement	Research evidence	Additional considerations
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Le evidenze complessive sugli effetti derivano da due studi, uno dei quali è un trial controllato randomizzato, entrambi monocentrici e condotti all'interno di programmi di screening di popolazione organizzati. Si rileva che lo studio di Lauritzen 2024 (2) sebbene non randomizzato presenta un rischio di bias basso in quanto le caratteristiche generali dello screening e della popolazione, poiché inserito in uno screening organizzato, dovrebbero essere molto simili nei due periodi confrontati. In questo senso, il disegno pre/post di questo studio non dovrebbe porre problemi di bias di selezione. I due studi differiscono per le caratteristiche della popolazione: Lauritzen 2024 (2) si concentra esclusivamente sulle donne tra i 50 e i 69 anni, mentre Hernström 2025 (1) include donne tra i 40 e gli 80 anni, di cui il 19,4% sotto i 45 anni. La riduzione del carico di letture per i radiologi è molto rilevante e con un elevato livello di certezza (come riportato nella sezione Resources) Al contrario, la certezza di un miglioramento dell'accuratezza è moderata.	Sono due le principali criticità metodologiche in Lauritzen 2024 (2), entrambe riconosciute dagli autori. La prima riguarda la diversa durata dell'intervallo di screening nelle due fasi: nella fase pre-IA l'intervallo medio dall'ultima mammografia era di 845 giorni, nella fase IA di 993 giorni. Il maggiore intervallo nella fase IA potrebbe aver contribuito a un tasso di rilevazione più alto. Tuttavia, le analisi stratificate per tempo dall'ultima mammografia suggeriscono che non sia presente un effetto confondente. La seconda criticità riguarda la modifica della soglia per classificare i casi a basso rischio (da 5 a 7, che approssimano il cinquantesimo e settantesimo percentile del rischio di malignità), non prevista nel protocollo. Al contrario, Hernström 2025 (1) ha utilizzato una soglia fissa pari a 9, corrispondente approssimativamente al novantesimo percentile del rischio di malignità. Inoltre, i pazienti a rischio molto elevato (10H) sono stati automaticamente richiamati. I due studi differiscono anche per flusso operativo e modalità di "mascheramento" dell'IA. In Lauritzen 2024 (2), i radiologi non utilizzavano direttamente l'IA ma sapevano se il punteggio del paziente superasse o meno la soglia di triage (mascheramento parziale). Inoltre, tutti i casi a basso rischio, letti da un solo radiologo, erano assegnati ad un radiologo senior. In Hernström 2025 (1), i radiologi vengono assegnati in modo randomizzato ai due bracci dello studio per ciascuna sessione, garantendo un numero equilibrato di letture in entrambi i bracci. Nei casi del braccio IA, tutti i radiologi avevano accesso ai punteggi di rischio generati dall'IA. Per i casi ad alto rischio, un secondo radiologo rivedeva la valutazione del primo, secondo il normale flusso operativo. In caso di disaccordo o su richiesta di uno dei due radiologi, si teneva una riunione di consenso. Non sono emerse differenze significative nel numero di tali riunioni tra i due bracci. Lo studio randomizzato è stato condotto nel contesto dello screening organizzato svedese. Le pazienti sono state reclutate in quattro centri screening nella Svezia sud-occidentale (Malmö, Lund, Landskrona, and Trelleborg). Tuttavia, la lettura è avvenuta in un solo centro. Lo studio non randomizzato (pre-post), è stato condotto nel contesto dello screening organizzato nella regione di Copenhagen. Anche in questo caso le letture sono state effettuate in un solo centro. Giudizio espresso all'unanimità.

Values Is there important uncertainty about or variability in how much people value the main outcomes?		
Judgement	**Research evidence**	**Additional considerations**
○ Important uncertainty or variability ○ Possibly important uncertainty or variability ● Probably no important uncertainty or variability ○ No important uncertainty or variability	La ricerca di evidenze sui valori e le preferenze dei pazienti sono state effettuate dal GDG ECIBC. L'ultima ricerca risale al 2022. (rapporto tecnico JRC PICO 10-11, contratto FWC443094012015). La revisione ha rilevato che le partecipanti attribuiscono scarso valore agli effetti psicosociali e fisici dei risultati falsi positivi e della sovradiagnosi. *Appendice 4*	L'impatto dell'intervento su esiti potenzialmente rilevanti per le donne (tasso di rilevazione dei tumori e tasso di richiamo) risulta modesto e non particolarmente rilevante. Lo studio randomizzato Hernström 2025 (1) è stato disegnato come di non inferiorità, anche se ha poi mostrato un aumento dell'1,4 per 1000 nel tasso di rilevazione dei tumori e un aumento non significativo del tasso di richiamo. Nello studio non randomizzato di Lauritzen 2024 (2), entrambi gli indicatori mostrano un miglioramento statisticamente significativo degli esiti potenzialmente rilevanti dal punto di vista delle donne. Giudizio espresso all'unanimità.

Balance of effects

Does the balance between desirable and undesirable effects favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ● Probably favors the intervention ○ Favors the intervention ○ Varies ○ Don't know	**Giudizio del panel su effetti desiderati: moderati (moderate)** L'intervento è associato a un aumento dei tassi di identificazione dei tumori, aumentati con range di effetto assoluto di 13-14 casi in più ogni 10.000 donne sottoposte a screening (1)(2). L'intervento è associato ad una riduzione del tasso di falsi positivi in uno studio e ad una sostanziale equivalenza nell'altro, con range di effetto assoluto di 1 falso positivo in più a 76 in meno ogni 10.000 donne sottoposte a screening (1)(2). **Giudizio del panel su effetti indesiderati: trascurabili (trivial)** L'intervento è associato ad una riduzione del tasso di richiamo in uno studio e ad un aumento nell'altro, con range di effetto assoluto da 16 richiami in più a 65 in meno ogni 10.000 donne sottoposte a screening (1)(2).	Dal punto di vista dell'accuratezza diagnostica, l'intervento mostra esiti favorevoli o, quantomeno, non inferiori rispetto allo standard della doppia lettura da parte dei radiologi. Il panel sottolinea come il tasso di richiamo sia contesto specifico e che la riduzione del tasso di richiamo si è osservata nello studio in cui il tasso di richiamo nel controllo era più alto (2). È verosimile che l'impatto sul tasso di richiamo dipenda anche dal tasso di richiamo del programma, prima dell'introduzione dell'IA. Il panel ritiene rilevante riconsiderare questo giudizio quando saranno disponibili evidenze sul tasso di cancri intervallo ed eventuale distribuzione più favorevole dello stadio alla diagnosi nel round successivo. Giudizio espresso all'unanimità.

Resources required

How large are the resource requirements (costs)?

Judgement	Research evidence	Additional considerations
○ Large costs ○ Moderate costs ○ Negligible costs and savings ● Moderate savings ○ Large savings ○ Varies ○ Don't know	Hernström 2025 e Lauritzen 2024 (1)(2) indicano che lo screening integrato con IA è associato ad una riduzione del carico di lavoro dei radiologi in termini di numero letture: (1): -44%; (2): -33,5%. Assumendo che il sistema di IA sia stato validato adeguatamente, l'intervento richiede sia l'acquisizione del software di IA, sia l'implementazione di una serie di attività di supporto per garantirne un uso efficace. Queste includono l'integrazione con i sistemi esistenti, test di interoperabilità e attività di validazione. Devono inoltre essere previste sessioni di formazione ed esercitazioni di simulazione per i radiologi. Inoltre, è necessario un monitoraggio continuo, soprattutto nel caso di applicazioni dinamiche auto-apprendenti. Un recente elaborato PoliS-Lombardia (Università di Milano) su diversi fornitori ha stimato un costo "all inclusive" di €2,5 per lettura con IA (https://www.myecole.it/ooc/wp-content/uploads/2024/11/PW_Santicchia-Sonia.pdf). Tuttavia, questa stima si riferisce ad un prezzo per un singolo centro e verosimilmente è sovrastimata rispetto ad un contesto di gare d'appalto competitive con volumi di attività più elevati. Si considera quindi €2 per lettura un valore di riferimento più realistico ma ancora conservativo. Questa stima è coerente con le stime di uno studio condotto in UK dal NICE nel 2021 che riporta un range di costo per lettura tra **£0.60 a £3.00**, al netto dell'IVA (16). Questi sono costi emergenti, che richiedono quindi risorse aggiuntive. Essi sono compensati dalla significativa riduzione del tempo di lettura da parte dei radiologi. Secondo una simulazione recente effettuata al Policlinico di Milano, il costo per ora di lettura (inclusi tempi preparatori e tempi morti fisiologici) è pari a € 20. In base ad una media ponderata della retribuzione lorda (RAL) di radiologi senior e junior, questo corrisponde a € 4,2 per lettura. Assumendo la stessa riduzione del numero di letture osservata in Hernström 2025 (1) (44,2%), l'intervento permetterebbe di risparmiare 913 letture ogni 1.000 donne sottoposte a screening. Questo comporterebbe un risparmio di € 1.835 per 1.000 donne. *Appendice 5* Per un centro che esegue 15.000 mammografie all'anno, ciò si tradurrebbe in un risparmio di €27.520. Dall'analisi di sensibilità, l'intervento risulta cost-saving sino ad un costo per esame dell'IA pari a **€ 3,83.**	I prezzi futuri delle soluzioni IA sono difficili da prevedere, ma è probabile che in una fase iniziale le aziende mantengano prezzi contenuti per acquisire quote di mercato. Più critiche per il SSN potrebbero risultare le strategie di pricing nel medio periodo, anche in relazione al rischio di situazioni di monopolio o lock-in da parte del SSN o dei mercati globali. La riduzione delle seconde letture può inoltre diminuire il tempo di refertazione delle mammografie e liberare tempo ai radiologi per i secondi livelli, riducendo così i tempi d'attesa per le risposte negative e per il prosieguo del percorso diagnostico/terapeutico. Sharma 2023 (5), utilizzando lo strumento di IA "MIA", ha dimostrato che la doppia lettura con IA, rispetto alla doppia lettura umana storica, garantisce generalizzazione su quattro fornitori diversi di apparecchiature mammografiche e sedi, nonché a livello geografico. In particolare, la riduzione del carico di lavoro (dal 30% al 44,8%) potrebbe significativamente alleviare la pressione sui servizi di screening in cui è limitata la disponibilità di lettori qualificati. Il panel ritiene che il ritorno di informazione prodotto dai programmi di screening nell'uso dell'IA ai produttori di software abbia un valore che potrebbe essere monetizzato in fase di contrattualizzazione. Le risorse richieste variano significativamente tra le diverse organizzazioni del SSN. Alcune dispongono già delle infrastrutture tecnologiche necessarie per una rapida implementazione (es. PACS e RICS), mentre altre potrebbero necessitare di investimenti più consistenti per acquisire e integrare tali sistemi. Nell'esprimere il giudizio, il panel non ha considerato i costi di formazione del personale e i costi di potenziamento hardware e integrazione software necessari all'introduzione dell'IA nei programmi di screening. Questi costi impattano sui costi iniziali. L'entità dei costi legati all'information technology dipende dal grado di sviluppo del sistema gestionale e organizzativo già presenti nel programma: in sistemi che sono già ben funzionanti e molto informatizzati questi possono essere contenuti, in sistemi poco informatizzati e non organizzati possono essere più impattanti. Giudizio espresso all'unanimità.

Certainty of evidence of required resources What is the certainty of the evidence of resource requirements (costs)?		
Judgement	**Research evidence**	**Additional considerations**
○ Very low ○ Low ● Moderate ○ High ○ No included studies	Non sono stati inclusi studi che riportano valutazioni economiche.	La stima dei costi richiesti è complessa a causa dell'elevata variabilità tra prodotti disponibili, modelli di pricing, dinamicità del mercato dell'IA per lo screening mammografico e incertezza sui volumi di mammografie che saranno effettivamente processati con IA. Tuttavia, le evidenze attuali forniscono una base sufficientemente solida per ritenere che l'intervento sia costo-efficace se non addirittura cost-saving. I due fattori critici legati ai costi sono: (a) la capacità delle organizzazioni del SSN di effettuare gare d'appalto efficaci; (b) il volume di servizi che si prevede di acquistare. Il panel evidenzia comunque l'incertezza e la variabilità relative ai costi di implementazione non considerati nel giudizio delle risorse richieste. Giudizio espresso all'unanimità.

Cost effectiveness

Does the cost-effectiveness of the intervention favor the intervention or the comparison?

Judgement	Research evidence	Additional considerations
○ Favors the comparison ○ Probably favors the comparison ○ Does not favor either the intervention or the comparison ○ Probably favors the intervention ● Favors the intervention ○ Varies ○ No included studies	Non sono stati inclusi studi di costo-efficacia. I dati di efficacia dei due studi suggeriscono un effetto positivo o almeno neutro dal punto di vista dell'efficacia complessiva. Allo stesso tempo, entrambi gli studi mostrano una riduzione importante dei costi relativi alle letture dei radiologi che dovrebbero almeno compensare i costi dell'IA. Sebbene manchino ancora evidenze solide, il panel ritiene che l'intervento sia probabilmente costo-efficace, se non cost-saving.	Un'analisi completa di costo-efficacia relativa al PICO è fuori dallo scopo di questo adolopment, poiché richiederebbe dati reali e stime affidabili di parametri chiave attualmente non disponibili o molto difficili da determinare. Tuttavia, due simulazioni di base possono offrire spunti utili. Nello scenario descritto nel criterio "Resources required", la stima più probabile è che l'intervento sia cost-saving, ovvero consenta un risparmio di risorse per il SSN producendo esiti positivi o al peggio neutrali dal punto di vista dei pazienti. In termini qualitativi, se i dati sull'accuratezza fossero utilizzati per stimare i benefici dal punto di vista del paziente, l'intervento rileva 1,4 casi di tumore in più (ogni 1.000 pazienti), ma aumenta di 1 i falsi positivi. Qualsiasi ipotesi credibile sul valore (utilità/disutilità) degli esiti potenziali (sopravvivenza, disutilità dei falsi positivi o del sovradiagnosi/sovratrattamento) porta molto probabilmente a concludere che l'intervento sia costo efficace se non cost-saving. In altre parole, questo intervento molto probabilmente domina la doppia lettura da parte dei radiologi in quanto preferibile sia in termini di salute, sia di costi. Giudizio espresso a seguito di votazione: 15 panelisti hanno votato per "Favors the intervention" 9 panelisti hanno votato per "Probably favors the intervention" Nessun astenuto.

Equity What would be the impact on health equity?		
Judgement	**Research evidence**	**Additional considerations**
○ Reduced ○ Probably reduced ○ Probably no impact ● Probably increased ○ Increased ○ Varies ○ Don't know	Entrambi gli studi non hanno raccolto dati sull'etnia e le condizioni socio-economiche delle pazienti. Non emergono implicazioni dirette e significative in termini di equità riconducibili all'implementazione dell'intervento. Non vi è evidenza che l'uso dell'IA nello screening mammografico crei barriere di accesso al programma per i gruppi svantaggiati. Anzi, la possibilità di liberare risorse interne tramite l'uso dell'IA potrebbe favorire azioni per aumentare la partecipazione al programma di screening. Tuttavia, è possibile che segmenti della popolazione (non necessariamente svantaggiati o meno istruiti) esprimano forme di opposizione all'IA in generale.	Un tema critico è l'equità territoriale. Alcuni centri o regioni potrebbero non disporre di infrastrutture tecnologiche adeguate o delle competenze professionali necessarie, con il rischio di un'implementazione disomogenea dell'intervento. I programmi di implementazione dovrebbero prestare particolare attenzione a questo aspetto. Le differenze nella capacità dei programmi di screening nell'integrare l'IA nel flusso della lettura e gestione delle donne può generare iniquità geografica nella diffusione sul territorio nazionale. Un'implementazione uniforme porterebbe a un aumento dell'equità per maggiore sostenibilità dei programmi e minori differenze di qualità di lettura. La diffusione dell'IA potrebbe avere un impatto positivo sull'accuratezza delle letture effettuate nello screening spontaneo, in cui è attualmente prevista una singola lettura. In questo contesto, l'impatto potrebbe essere maggiore rispetto a quello dello screening organizzato. Il panel ha espresso preoccupazione sul fatto che l'uso della IA possa essere promosso in maniera strumentale nell'offerta opportunistica rispetto ai modelli organizzati. Il panel ha considerato che l'impatto sull'equità potrebbe variare nel tempo. In un primo tempo, le differenze nella velocità di implementazione possono introdurre una iniquità geografica. A lungo termine, l'intervento può ridurre la variabilità nella qualità della lettura dei programmi e aumentare la sostenibilità dello screening, aumentando l'equità nell'accesso a programmi con elevati standard di qualità. Giudizio espresso all'unanimità.

Acceptability Is the intervention acceptable to key stakeholders?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Non è stata effettuata una revisione sistematica della letteratura per studi su accettabilità. **DONNE:** Lo studio di Hernström 2025 (1) non ha rilevato una resistenza significativa da parte delle pazienti: solo lo 0,2% ha rifiutato di partecipare allo studio. Pesapane 2024 (17) riporta una chiara preferenza da parte delle pazienti italiane per un utilizzo dell'IA a supporto, piuttosto che in totale sostituzione del radiologo nel processo diagnostico, sottolineando l'importanza di combinare l'expertise dei radiologi con l'IA per migliorare l'accuratezza diagnostica. Un orientamento analogo è emerso anche in altri contesti internazionali, quali Norvegia, Svezia, Australia e Stati Uniti. (18, 19, 20, 21, 22). **PROFESSIONISTI SANITARI**: Inoltre, secondo quanto riportato da Fattore et al. nella presentazione "Screening, Digitalizzazione e Innovazione" (Bari, 14–15 aprile 2025), relativa a un sondaggio condotto tra 305 professionisti operanti nei centri di screening italiani sull'atteggiamento verso l'IA, il 75% ha espresso interesse verso il potenziale dell'IA in ambito screening. Inoltre, il 67% si è dichiarato favorevole a investire in applicazioni basate sull'IA. Il 59% ritiene che la resistenza dei pazienti non costituisca un ostacolo all'implementazione di tali sistemi nella pratica clinica. Tuttavia, il 29% del campione dichiara di non avere un'opinione in proposito. Concentrandosi sullo screening mammografico, il 76% dei rispondenti (che hanno dichiarato di avere esperienza in questo ambito) si è detto disposto a sperimentare l'intervento oggetto del PICO (triage da parte di IA, doppia lettura umana + IA per pazienti ad alto rischio e singola lettura + IA per i pazienti a basso rischio) nel proprio centro.	**DONNE:** Le evidenze considerate non fanno emergere problemi relativi all'accettabilità dell'introduzione dell'IA nello screening da parte delle donne. È emersa una preoccupazione nei confronti di processi in cui l'IA sostituisca completamente il giudizio del radiologo. Il panel ritiene importante ribadire che è fondamentale investire su una corretta e trasparente comunicazione alle donne. **RADIOLOGI:** Tra i radiologi sono emersi interesse e disponibilità verso l'introduzione dell'intervento proposto nei centri in cui operano. **COORDINATORI DEI PROGRAMMI DI SCREENING E POLICYMAKER:** Tra i coordinatori di programmi di screening e i policymakers è emerso interesse verso il potenziale dell'intelligenza artificiale in ambito screening e non sono emerse particolari resistenze all'introduzione dell'IA nei protocolli di screening. Giudizio espresso all'unanimità.

Feasibility Is the intervention feasible to implement?		
Judgement	**Research evidence**	**Additional considerations**
○ No ○ Probably no ● Probably yes ○ Yes ○ Varies ○ Don't know	Il sondaggio condotto tra 305 professionisti dei centri di screening italiani indica un generale supporto all'intervento, pur sottolineando la necessità di un'adeguata infrastruttura tecnologica e di risorse finanziarie (Fattore et al., 2025). Diversi aspetti critici devono essere affrontati per garantire una corretta implementazione: (a) sviluppare strategie efficaci per la trasformazione digitale richiesta e gestire le relazioni con i fornitori durante tutte le fasi: selezione del prodotto, validazione locale, integrazione con i sistemi informativi esistenti; (b) definire chiaramente il flusso di lavoro, compresi i cut-off di triage, l'uso dei risultati IA da parte dei radiologi, e una possibile riorganizzazione dei servizi (es. lettura immediata dei casi a basso rischio post-mammografia); (c) assegnare i casi in modo appropriato tra radiologi con diversi livelli di esperienza; (d) implementare un sistema di monitoraggio continuo delle performance dell'intervento; (e) gestire pazienti e professionisti che potrebbero rifiutare l'utilizzo dell'IA (probabilmente una minoranza); (f) identificare e affrontare adeguatamente le questioni legali e regolatorie, come protezione dei dati, responsabilità civile e responsabilità professionale, anche in relazione alla legislazione europea in corso di finalizzazione (EU AI act).	Le applicazioni di IA possiedono caratteristiche economiche peculiari. Non sono rivali nel consumo, ovvero l'uso della tecnologia in un contesto non ne limita la disponibilità in altri. In altre parole, il costo del software non dipende dal volume. Anzi, se lo strumento IA migliora grazie all'addestramento con ulteriori dati, l'incremento dei volumi potrebbe contribuire a migliorare la performance. Questa caratteristica apre scenari incerti sul pricing e sulla struttura di mercato. Per questo, è particolarmente necessario che il SSN monitori e gestisca attivamente le dinamiche del mercato dell'IA per questa applicazione. Il panel evidenzia che sarebbe necessario condurre un'analisi dell'impatto sull'ambiente e in particolare sul consumo di energia dell'applicazione dell'IA alla lettura delle mammografie e confrontarlo con quello della doppia lettura. Il panel ribadisce l'importanza di valutare la fattibilità a livello di singolo programma di screening, considerando i costi iniziali e di implementazione, inclusi i costi di formazione del personale, i costi di potenziamento hardware e di integrazione dei software nei flussi di lavoro; tali costi possono variare considerevolmente in base alle caratteristiche del sistema e dell'organizzazione esistenti e alle risorse umane disponibili. Giudizio espresso all'unanimità.

SUMMARY OF JUDGEMENTS

	JUDGEMENT						
PROBLEM	No	Probably no	Probably yes	**Yes**		Varies	Don't know
DESIRABLE EFFECTS	Trivial	Small	**Moderate**	Large		Varies	Don't know
UNDESIRABLE EFFECTS	Large	Moderate	Small	**Trivial**		Varies	Don't know
CERTAINTY OF EVIDENCE	Very low	Low	**Moderate**	High			No included studies
VALUES	Important uncertainty or variability	Possibly important uncertainty or variability	**Probably no important uncertainty or variability**	No important uncertainty or variability			
BALANCE OF EFFECTS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	**Probably favors the intervention**	Favors the intervention	Varies	Don't know
RESOURCES REQUIRED	Large costs	Moderate costs	Negligible costs and savings	**Moderate savings**	Large savings	Varies	Don't know
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	Very low	Low	**Moderate**	High			No included studies
COST EFFECTIVENESS	Favors the comparison	Probably favors the comparison	Does not favor either the intervention or the comparison	Probably favors the intervention	**Favors the intervention**	Varies	No included studies
EQUITY	Reduced	Probably reduced	Probably no impact	**Probably increased**	Increased	Varies	Don't know
ACCEPTABILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know
FEASIBILITY	No	Probably no	**Probably yes**	Yes		Varies	Don't know

TYPE OF RECOMMENDATION

Strong recommendation against the intervention	Conditional recommendation against the intervention	Conditional recommendation for either the intervention or the comparison	**Conditional recommendation for the intervention**	Strong recommendation for the intervention
○	○	○	●	○

CONCLUSIONS

Recommendation

Per donne asintomatiche con un rischio medio di cancro della mammella, il panel italiano **suggerisce** di **utilizzare** una strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana ciascuna mammografia **invece che** la doppia lettura umana per tutte le mammografie nello screening mammografico basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico. (Raccomandazione **su condizione**, **a favore dell'intervento**; grado di certezza delle evidenze **moderato**)

Justification

Overall justification

Il panel ha espresso la raccomandazione tenendo conto di un bilancio degli effetti desiderabili e non desiderabili probabilmente favorevole all'intervento, con un livello di certezza delle evidenze moderato. La raccomandazione si è basata principalmente sui risultati di due studi population based multicentrici con elevata numerosità del campione incluso (in totale, oltre 200.000 donne). La valutazione delle risorse richieste sembra suggerire possibili risparmi moderati derivanti dall'introduzione dell'intervento con un grado di certezza delle evidenze a supporto classificato come moderato, con un profilo costo-beneficio giudicato a favore dell'intervento. Il panel giudica inoltre come probabilmente positivo l'impatto sull'equità dell'introduzione dell'intervento che viene considerato probabilmente accettabile e probabilmente fattibile.

Alla votazione sulla direzione e la forza della raccomandazione, hanno partecipato 21 componenti del panel. Il panel ha concordato all'unanimità sulla direzione a favore dell'intervento e ha votato per la forza della raccomandazione: 19 si sono espressi per una raccomandazione su condizione a favore dell'intervento e 2 per una raccomandazione forte a favore dell'intervento. Non ci sono stati astenuti.

Sebbene la raccomandazione sia basata su due studi che hanno utilizzato lo stesso software, il panel ritiene che possa essere applicata ad altri sistemi che abbiamo dimostrato una simile accuratezza nella classificazione della mammografia sulla base del rischio di presenza di lesioni maligne.

Detailed justification:

Desirable Effects:

Gli effetti desiderabili sono stati considerati dal Panel come moderati. Fra questi sono stati considerati l'aumento del tasso di identificazione di tumori, la possibile riduzione del tasso di falsi positivi e l'aumento del valore predittivo positivo dei richiami. Non sono attualmente disponibili risultati sui cancri intervallo per cui mancano evidenze dirette su quanto l'aumento del tasso di identificazione si traduca in benefici di salute. Evidenze indirette da studi retrospettivi suggeriscono che sia poco probabile un impatto negativo sui cancri intervallo dell'introduzione dell'intervento. Il panel ritiene importante riconsiderare l'entità degli effetti desiderabili appena disponibili evidenze dirette sui cancri intervallo.

Undesirable Effects:

Gli effetti indesiderabili sono stati considerati dal Panel come trascurabili (trivial) considerando l'impatto sul tasso di richiami, le cui stime sono risultate discordanti tra i due studi inclusi. Uno studio ha mostrato una riduzione del tasso di richiami mentre l'altro ha riportato un possibile aumento del tasso di richiami. Nella valutazione complessiva il Panel ritiene importante valutare il tasso di richiami alla luce dell'aumento del tasso di identificazione e del valore predittivo positivo. Il Panel ribadisce l'importanza di riconsiderare l'entità degli effetti indesiderabili appena disponibili evidenze dirette sui cancri intervallo, per escludere un aumento della sovradiagnosi.

Balance of effects:

Alla luce delle valutazioni su effetti desiderabili e non desiderabili, il Panel ritiene che il bilancio degli effetti sia probabilmente a favore dell'utilizzo di una strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana ciascuna mammografia.

Resources required

Il panel valuta che l'introduzione dell'intervento potrebbe portare a risparmi moderati sulla necessità di risorse. In particolare, a fronte di un possibile incremento di costi diretti legati all'introduzione dei software di IA, entrambi gli studi riportano una riduzione del numero di letture umane da 670 a 913 ogni 1.000 donne sottoposte a screening con conseguente risparmio di tempo radiologo. Il panel ritiene inoltre che, affinché la riduzione del numero di letture si traduca in un effettivo risparmio, è necessario che la strategia di triage sia introdotta in un contesto con flussi di lavoro efficienti e che il tempo radiologo risparmiato venga reimpiegato in maniera efficiente dall'organizzazione sanitaria. Non sono stati considerati i costi di formazione del personale e i costi di potenziamento hardware e integrazione software necessari all'introduzione dell'IA nei programmi di screening.

Cost effectiveness:

Il panel giudica il profilo di costo-efficacia a favore dell'intervento. Non sono stati inclusi studi di costo-efficacia. Il giudizio si è basato sul giudizio dato riguardo al bilancio degli effetti (probabilmente a favore dell'intervento) e sull'analisi dei costi effettuata ad hoc (risparmi moderati).

Equity

Il panel giudica che l'introduzione dell'intervento porterebbe ad un probabile aumento dell'equità. Il giudizio si è basato considerando principalmente la maggior sostenibilità dei programmi di screening e le minori differenze di qualità di lettura tra programmi, in uno scenario di introduzione uniforme a livello nazionale. Il panel ritiene importante porre l'attenzione sulla possibilità che si determinino iniquità geografiche per diffusione eterogenea a livello nazionale, per limitare le quali risulta necessario adeguare infrastrutture tecnologiche, competenze professionali e di governo dei flussi di lavoro.

Acceptability

Il panel giudica che l'introduzione dell'intervento sia probabilmente accettabile. Il giudizio è basato principalmente sui risultati di una survey condotta a livello nazionale. Per le donne è emersa una preferenza per l'utilizzo dell'intelligenza artificiale che non elimini la valutazione da parte del radiologo nel processo diagnostico. Tra i radiologi sono emersi interesse e disponibilità verso l'introduzione dell'intervento proposto nei centri in cui operano. Tra i decisori è emerso interesse verso il potenziale dell'intelligenza artificiale in ambito screening.

Feasibility

Il panel giudica probabilmente fattibile l'introduzione dell'intervento. Il giudizio è basato principalmente sui risultati di una survey condotta a livello nazionale. Il panel sottolinea la necessità di un'adeguata infrastruttura tecnologica, di risorse finanziarie nella fase di introduzione e di una governance dei flussi di lavoro.

Subgroup considerations

Un'analisi per classi di età nello studio MASAI suggerisce che l'aumento del tasso di identificazione dei tumori sia osservato solo nelle donne di età superiore a 50 anni. Il tasso di identificazione nel braccio d'intervento nelle donne di età inferiore a 50 anni rimane comunque uguale a quello della doppia lettura (1). Non si suggeriscono di conseguenza implementazioni diversificate per sottogruppi.

Implementation considerations

La strategia di triage basato sull'intelligenza artificiale consiste nell'identificazione di una soglia di score di malignità attribuito da software basati su IA, utilizzata per l'assegnazione di ciascuna mammografia a singola lettura umana (mammografie con score di malignità inferiore alla soglia) o a doppia lettura umana (mammografie con score di malignità superiore alla soglia). Tale strategia di triage prevede arbitrato in caso la singola lettura del radiologo proponga un richiamo per una mammografia classificata a basso rischio dall'IA e arbitrato e/o consenso in caso i due radiologi diano pareri discordanti sul richiamo per una mammografia classificata ad alto rischio dall'AI (Appendice 2).

La soglia di score per mandare a singola o doppia lettura può variare da contesto a contesto e in base al sistema di IA in uso. I criteri di discordanza per attivare l'arbitrato o il richiamo possono essere differenti. Alcuni sistemi di IA possono identificare un piccolo gruppo di mammografie ad altissimo rischio per le quali in alcuni casi (1) i radiologi hanno convenuto che è opportuno mandare a richiamo a prescindere dall'esito delle letture umane.

La lettura del radiologo può essere in cieco rispetto alle informazioni elaborate dall'IA o informata. Nella modalità informata, il radiologo può ricevere solo l'informazione sulle lesioni individuate o anche sullo score di rischio, per lesione e complessivo. Si può anche definire la soglia di rischio per le lesioni che l'IA segnala al lettore umano come di possibile interesse.

In caso di introduzione di una strategia di triage basata sull'intelligenza artificiale, è necessario prevedere una fase di ottimizzazione del flusso di lavoro a livello di singolo programma di screening, che includa una valutazione delle modalità e tempistiche ottimali per fornire al radiologo l'informazione prodotta dall'algoritmo di IA e che tenga conto delle performance specifiche del singolo programma per la definizione della soglia ottimale di score di rischio. In particolare, in questa fase, si potrà decidere: la soglia di score per il triage, le modalità di arbitrato, l'informazione a cui il radiologo accede in singola lettura o in doppia lettura, la soglia di rischio per le lesioni da segnalare, eventuali eccezioni alle modalità di richiamo. Il panel auspica che vengano definite delle modalità di implementazione comuni all'interno dei programmi di screening italiani per quanto riguarda le soglie accettabili e le modalità di richiamo.

In caso di introduzione di una strategia di triage basata sull'intelligenza artificiale è utile prevedere una fase di introduzione controllata e di monitoraggio prospettico, anche attraverso progetti pilota, per valutarne l'accuratezza diagnostica del software applicato ai sistemi di imaging locali, l'impatto su performance del programma, costi e accettabilità in contesti specifici.

Per garantire pluralità di operatori sul mercato è opportuno che la comunità scientifica e le autorità regolatorie condividano dei criteri di validazione dei sistemi di IA che siano basati sulla valutazione retrospettiva di determinate casistiche dove possano validare la loro sensibilità e specificità. È opportuna una valutazione periodica dei costi e delle performance dei sistemi di intelligenza artificiale sul mercato, al fine di garantire trasparenza e ridurre il rischio di lock-in tecnologico verso un singolo fornitore.

Nell'implementazione è fondamentale garantire il diritto delle persone a ricevere un'informazione completa, chiara e trasparente, così da favorire decisioni consapevoli, rafforzare fiducia e accettabilità, nel rispetto dell'autonomia del soggetto. L'informazione va veicolata attraverso materiali dedicati ed elaborati. in condivisione con i portatori di interesse. Particolare attenzione è posta alla tutela dei dati personali, considerato che l'uso dell'IA implica il trattamento di informazioni cliniche sensibili. Ciò richiede conformità al Regolamento (UE) 2016/679 (GDPR) e al Regolamento (UE) 2024/1689 (AI Act), con specifici obblighi di trasparenza e garanzia per sistemi definiti ad alto rischio. L'utilizzo di IA deve inoltre avvenire nel rispetto del Regolamento (UE) 2017/745 sui dispositivi medici, ma la responsabilità complessiva resta in capo al professionista e alla struttura sanitaria, chiamati a definire ambiti di utilizzo, procedure alternative e modalità di monitoraggio. È infine auspicabile che tali scelte siano condivise e coordinate a livello regionale (idealmente, a livello nazionale), al fine di assicurare unitarietà di esercizio, maggiore tutela per le persone e garanzia per le professioni e le strutture coinvolte.

Nell'implementazione è necessario considerare che la performance dei sistemi di IA al momento è probabilmente differente in caso di uso della tomosintesi, dato che i sistemi sono stati per lo più sviluppati e addestrati con mammografie 2D.

In ogni caso la compatibilità tecnica e le performance dei diversi software di IA con le diverse immagini mammografiche (2D, tomosintesi a diversa ampiezza di angolo) fornite da apparecchiature diverse per azienda produttrice, modello e versione, andranno sempre preliminarmente verificate da ogni centro mediante un congruo periodo di uso in prova. Si segnala che la compatibilità tecnica (ovvero le immagini sono inviate al sistema, che fornisce il feedback in tempi ragionevoli), non corrisponde al fatto che si avranno localmente per quel sistema di IA le performance attese sulla base della letteratura.

Nella valutazione delle risorse richieste non sono stati considerati i costi relativi alla formazione del personale e al potenziamento della componente hardware e all'integrazione dei software nei flussi di lavoro. Tali costi dipendono dalle caratteristiche del sistema e dell'organizzazione esistenti a livello di ciascun programma di screening e vanno considerate nella valutazione di fattibilità dell'intervento.
Nell'implementazione dei sistemi di IA è importante considerare le opportunità di supporto alla formazione continua dei radiologi offerte dal confronto con le valutazioni date dall'IA alle mammografie precedenti di cancri intervallo o di cancri identificati ai round successivi.

Monitoring and evaluation

Al momento del calcolo degli indicatori per programma, dovrebbe essere fornita l'informazione sulla presenza di sistemi di IA e sul protocollo adottato. In caso di introduzione parziale, si dovrebbe valutare l'opportunità di raccogliere indicatori separati nelle schede di monitoraggio dell'Osservatorio Nazionale Screening (ONS) e del Gruppo Italiano per lo Screening Mammografico (GISMa).

A livello locale deve essere monitorata la performance del sistema di lettura e dei singoli lettori quando affiancati da IA in termini di tasso di richiamo e valore predittivo positivo (e tasso di identificazione).

Research priorities

Condurre validazioni multicentriche con differenti apparecchiature mammografiche per produttore e modelli, differenti RIS-PACS, differenti algoritmi e in contesti organizzativi diversi per informare la trasferibilità delle evidenze a supporto della raccomandazione.

Sviluppare evidenze sull'effetto sui cancri intervallo dell'utilizzo di algoritmi basati su IA per la lettura delle mammografie di screening.

Sviluppare evidenze sull'effetto della densità mammografica sulle performance degli algoritmi basati su IA.

Sviluppare evidenze specifiche relative alla performance dei sistemi e delle strategie di lettura con IA nelle fasce d'età 40-44 anni e 45-49 anni.

Sviluppare evidenze a supporto dell'ottimizzazione delle modalità di messa a disposizione dell'informazione prodotta dagli algoritmi di IA per il radiologo e sui possibili bias cognitivi introdotti dalla conoscenza o meno delle elaborazioni dell'AI.

Sviluppare evidenze a supporto della definizione delle soglie di score di rischio per l'ottimizzazione delle performance dei protocolli di screening e del flusso di lavoro.

Sviluppare evidenze sulla trasferibilità delle stime di performance della strategia di triage basata su IA a donne appartenenti a gruppi etnici non adeguatamente rappresentati nei dataset di training degli algoritmi di IA utilizzati.

Sviluppare evidenze sulle performance degli algoritmi in donne con protesi mammarie.

Sviluppare evidenze in relazione all'attuazione di protocolli di screening basati sul rischio, incluso le possibilità di predire il rischio di cancro intervallo o di cancro a medio temine dei sistemi di AI.

Valutare, in termini di accuratezza e di implementation research, l'evoluzione tecnologica riguardo l'uso delle precedenti mammografie e valutare l'applicazione dell'IA alla tomosintesi.

Mettere a punto protocolli di validazione e standard di accuratezza minimi che debbano essere soddisfatti dai nuovi sistemi di IA per essere efficienti e sicuri, e che possano essere condotti anche su database di immagini da analizzare in modo retrospettivo.

Prevedere valutazioni di Health Technology Assessment per l'impatto organizzativo ed economico contesto specifico.

Condurre valutazioni sull'impatto sull'ambiente dell'applicazione dell'IA alla lettura delle mammografie in programmi di screening di popolazione.

Sviluppare evidenze mediante trial randomizzati sull'impiego della IA come sostituzione della doppia lettura da parte dei radiologi. Nelle ultime versioni, i sistemi di IA più noti hanno sviluppato la capacità di confrontare l'esame corrente con i precedenti (fino a 6). Ciò potrà essere in grado di migliorare la specificità di questi sistemi e, potenzialmente, evitare la necessità dell'intervento umano nel confrontare i casi richiamati dalla IA con i precedenti. Non vi è al momento una chiara evidenza di quale modalità di utilizzo dell'IA sia superiore. In una prospettiva di sempre maggiore accuratezza dei sistemi di IA, il modello di "triage", che prevede la costante valutazione umana di tutti i casi, potrebbe risultare progressivamente meno accurato rispetto al modello di sostituzione di uno dei due lettori.

Bibliografia

1. Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025;7(3):e175-e183. doi:10.1016/S2589-7500(24)00267-X

2. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. Radiology. 2024;311(3):e232479. doi:10.1148/radiol.232479

3. Dembrower K, Crippa A, Colón E, Eklund M, Strand F; ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. Lancet Digit Health. 2023;5(10):e703-e711. doi:10.1016/S2589-7500(23)00153-X

4. Brancato B, Magni V, Saieva C, et al. AI-supported approaches for mammography single and double reading: A controlled multireader study. Eur J Radiol. 2025;187:112101. doi:10.1016/j.ejrad.2025.112101

5. Sharma N, Ng AY, James JJ, et al. Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms. BMC Cancer. 2023;23(1):460. Published 2023 May 19. doi:10.1186/s12885-023-10890-7

6. Hovda T, Larsen M, Bergan MB, Gjesvik J, Akslen LA, Hofvind S. Retrospective evaluation of a CE-marked AI system, including 1,017,208 mammography screening examinations. Eur Radiol. 2025;35(9):5685-5694. doi:10.1007/s00330-025-11521-4

7. Bergan MB, Larsen M, Moshina N, et al. AI performance by mammographic density in a retrospective cohort study of 99,489 participants in BreastScreen Norway. Eur Radiol. 2024;34(10):6298-6308. doi:10.1007/s00330-024-10681-z

8. Oberije CJG, Currie R, Leaver A, et al. Assessing artificial intelligence in breast screening with stratified results on 306 839 mammograms across geographic regions, age, breast density and ethnicity: A Retrospective Investigation Evaluating Screening (ARIES) study. BMJ Health Care Inform. 2025;32(1):e101318. Published 2025 May 14. doi:10.1136/bmjhci-2024-101318

9. Larsen M, Aglen CF, Lee CI, et al. Artificial Intelligence Evaluation of 122 969 Mammography Examinations from a Population-based Screening Program. Radiology. 2022;303(3):502-511. doi:10.1148/radiol.212381

10. Byng D, Strauch B, Gnas L, et al. AI-based prevention of interval cancers in a national mammography screening program. Eur J Radiol. 2022;152:110321. doi:10.1016/j.ejrad.2022.110321

11. Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I. Can artificial intelligence reduce the interval cancer rate in mammography screening?. Eur Radiol. 2021;31(8):5940-5947. doi:10.1007/s00330-021-07686-3

12. Graewingholt A, Rossi PG. Retrospective analysis of the effect on interval cancer rate of adding an artificial intelligence algorithm to the reading process for two-dimensional full-field digital mammography. J Med Screen. 2021;28(3):369-371. doi:10.1177/0969141320988049

13. Wanders AJT, Mees W, Bun PAM, et al. Interval Cancer Detection Using a Neural Network and Breast Density in Women with Negative Screening Mammograms. Radiology. 2022;303(2):269-275. doi:10.1148/radiol.210832

14. Yu TT, Hoyt AC, Joines MM, et al. Mammographic classification of interval breast cancers and artificial intelligence performance. J Natl Cancer Inst. 2025;117(8):1627-1638. doi:10.1093/jnci/djaf103

15. Chang YW, Ryu JK, An JK, et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study. Nat Commun. 2025;16(1):2248. Published 2025 Mar 6. doi:10.1038/s41467-025-57469-3

16. NICE 2021. https://www.nice.org.uk/advice/mib242/resources/artificial-intelligence-in-mammography-pdf-2285965629587653.2021

17. Pesapane F, Giambersio E, Capetti B, et al. Patients' Perceptions and Attitudes to the Use of Artificial Intelligence in Breast Cancer Diagnosis: A Narrative Review. Life (Basel). 2024;14(4):454. Published 2024 Mar 29. doi:10.3390/life14040454

18. Ozcan BB, Dogan BE, Xi Y, Knippa EE. Patient Perception of Artificial Intelligence Use in Interpretation of Screening Mammograms: A Survey Study. Radiol Imaging Cancer. 2025;7(3):e240290. doi:10.1148/rycan.240290.

19. Holen ÅS, Martiniussen MA, Bergan MB, Moshina N, Hovda T, Hofvind S. Women's attitudes and perspectives on the use of artificial intelligence in the assessment of screening mammograms. Eur J Radiol. 2024;175:111431. doi:10.1016/j.ejrad.2024.111431

20. Pearce A, Carter S, Frazer HM, et al. Implementing artificial intelligence in breast cancer screening: Women's preferences. Cancer. 2025;131(9):e35859. doi:10.1002/cncr.35859

21. de Vries CF, Morrissey BE, Duggan D, Staff RT, Lip G. Screening participants' attitudes to the introduction of artificial intelligence in breast screening. J Med Screen. 2021;28(3):221-222. doi:10.1177/09691413211001405.

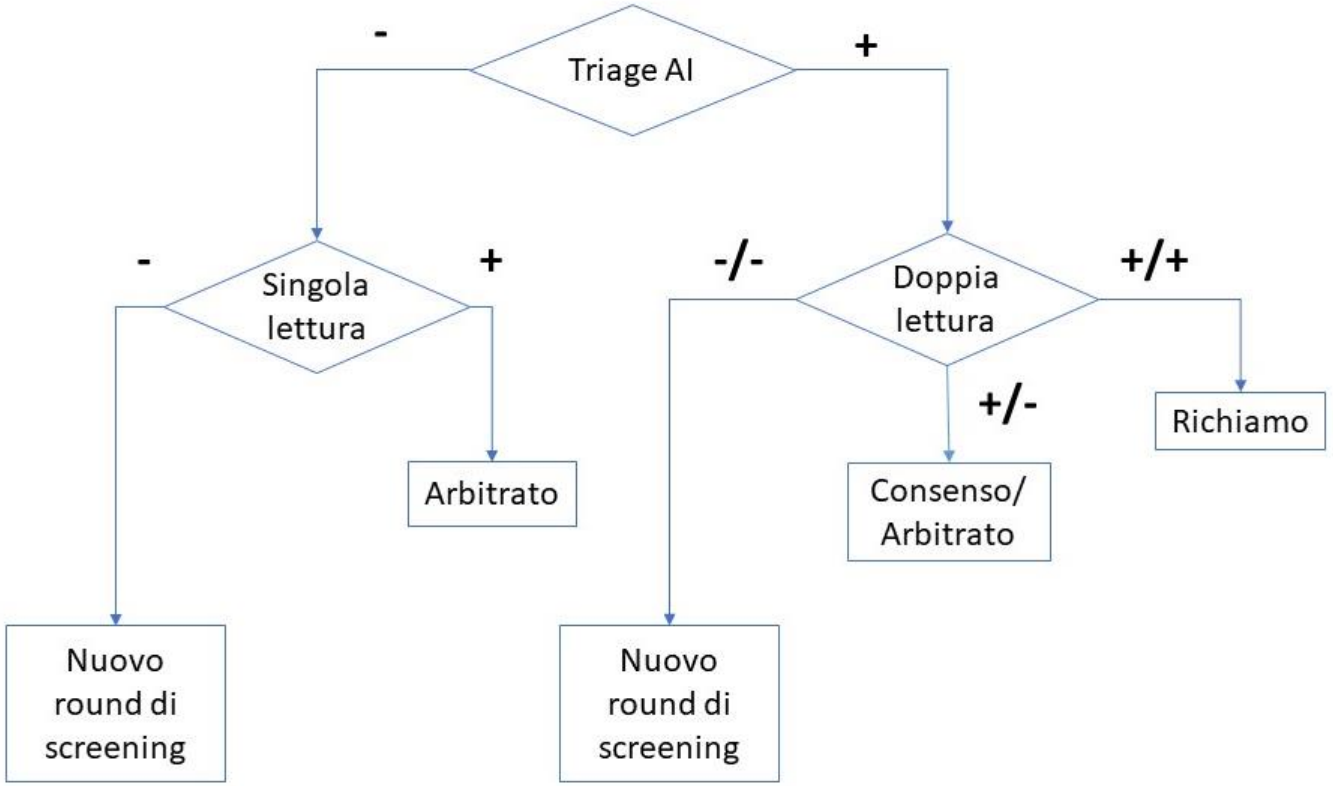
22. Tang JS, Frazer HM, Kunicki K, Basnayake P, Omori M, Lippey J. Australian healthcare workers' views on artificial intelligence in BreastScreen: Results of a mixed method survey study. Prev Med Rep. 2024;48:102917. Published 2024 Oct 28. doi:10.1016/j.pmedr.2024.102917

APPENDICI

Appendice 1. Parole chiave utilizzate per la ricerca bibliografica in PubMed, Scopus e Ovid (Journals@Ovid Full Text):

PubMed:
(breast cancer[Title/Abstract])
AND (screening[Title/Abstract]
OR mammograph[Title/Abstract]*
OR tomosynthesis[Title/Abstract])
AND (artificial intelligence[Title/Abstract]
OR AI[Title/Abstract]
OR machine learning[Title/Abstract]
OR deep learning[Title/Abstract])
Scopus:
(TITLE-ABS-KEY("breast cancer"))
AND (TITLE-ABS-KEY("screening")
OR TITLE-ABS-KEY("mammograph")*
OR TITLE-ABS-KEY("tomosynthesis"))
AND (TITLE-ABS-KEY("artificial intelligence")
OR TITLE-ABS-KEY("AI")
OR TITLE-ABS-KEY("machine learning")
OR TITLE-ABS-KEY("deep learning"))
Ovid (*Journals@Ovid Full Text):*
(breast cancer.ti,ab.)
AND (screening.ti,ab.
OR mammograph.ti,ab.*
OR tomosynthesis.ti,ab.)
AND (artificial intelligence.ti,ab.
OR AI.ti,ab.
OR machine learning.ti,ab.
OR deep learning.ti,ab.)

Appendice 2 - Figura 1. Caratteristiche generali della strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana considerata come Intervento nel PICO della raccomandazione votata dal panel italiano.



Appendice 3 – Summmary of findings table

Outcomes	№ of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with double human reading	Risk difference with triage using artificial intelligence to assign single or double human reading
Interval Cancer Rate - not measured	-	-	-	-	-
Cancer detection rate	105915 (1 RCT)[1]	⨁⨁⨁◯ Moderate[a]	**RR 1.29** (1.09 to 1.51)	Study population	
				50 per 10.000	**14 more per 10.000** (4 more to 25 more)
Cancer detection rate	118997 (1 non-randomised study)[2]	⨁⨁◯◯ Low[a,b]	**RR 1.18** (1.04 to 1.35)	Study population	
				70 per 10.000	**13 more per 10.000** (3 more to 24 more)
Recall rate	105915 (1 RCT)[1]	⨁⨁⨁◯ Moderate[c]	**RR 1.08** (0.99 to 1.17)	Study population	
				194 per 10.000	**16 more per 10.000** (2 fewer to 33 more)
Recall rate	118997 (1 non-randomised study)[2]	⨁⨁⨁◯ Moderate[c]	**RR 0.79** (0.74 to 0.85)	Study population	
				309 per 10.000	**65 fewer per 10.000** (80 fewer to 46 fewer)
False positive rate	105915 (1 RCT)[1]	⨁⨁⨁◯ Moderate[c]	**RR 1.01** (0.91 to 1.11)	Study population	
				145 per 10.000	**1 more per 10.000** (13 fewer to 16 more)
False positive rate	118997 (1 non-randomised study)[2]	⨁⨁⨁◯ Moderate[c]	**RR 0.68** (0.63 to 0.74)	Study population	
				239 per 10.000	**76 fewer per 10.000** (88 fewer to 62 fewer)
Positive predictive value	2137 (1 RCT)[1]	⨁⨁⨁⨁ High	**RR 1.19** (1.04 to 1.37)	Study population	
				2.551 per 10.000	**485 more per 10.000** (102 more to 944 more)
Positive predictive value	3305 (1 non-randomised study)[2]	⨁⨁⨁◯ Moderate[b]	**RR 1.49** (1.33 to 1.66)	Study population	
				2.256 per 10.000	**1.105 more per 10.000** (744 more to 1.489 more)

Outcomes	№ of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects* (95% CI)	
				Risk with double human reading	Risk difference with triage using artificial intelligence to assign single or double human reading
Ductal carcinoma in situ	105915 (1 RCT)[1]	⨁⨁⨁◯ Moderate[a]	**RR 1.48** (1.01 to 2.18)	Study population	
				8 per 10.000	**4 more per 10.000** (0 fewer to 10 more)
Ductal carcinoma in situ	118997 (1 non-randomised study)[2]	⨁⨁◯◯ Low[a,b]	**RR 1.56** (1.17 to 2.19)	Study population	
				11 per 10.000	**6 more per 10.000** (2 more to 13 more)

1. Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025;7(3):e175-e183. doi:10.1016/S2589-7500(24)00267-X
2. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. Radiology. 2024;311(3):e232479. doi:10.1148/radiol.232479.

a. Imprecisione valutata come seria perché l'intervallo di confidenza della stima attraversa una delle soglie di entità dell'effetto. L'entità del beneficio dell'intervento è stato classificato come Moderato dal panel mentre il margine inferiore dell'intervallo di confidenza delle stime è stato valutato da ERT e confermato dal panel come effetto piccolo o nullo e pertanto l'intervallo di confidenza delle stime attraversa almeno una soglia decisionale.
b. Rischio di bias valutato come serio per possibile confondimento dovuto a differenze tra il periodo pre-e post-introduzione dell'AI in termini di tempo dall'ultimo round di screening e in termini di storia di screening delle donne arruolate.
c. Inconsistenza valutata come seria sulla base del confronto tra la stima di effetto (e relativo intervallo di confidenza) riportata dallo studio randomizzato e la stima di effetto (e relativo intervallo di confidenza) riportata nello studio non-randomizzato inclusi.

Appendice 4 – Sintesi della ricerca di evidenze sui valori e le preferenze dei pazienti effettuata dal Guidelines Development Group ECIBC. L'ultima ricerca risale al 2022. (rapporto tecnico JRC PICO 10-11, contratto FWC443094012015).

Outcome	Finding	Certainty
False-positive related adverse effects (psychological distress/biopsies and surgeries)	Women associate false-positive results with significant psychosocial distress. At baseline, the psychosocial impact is similar to being diagnosed with breast cancer. In most women, it tends to diminish at six months, while for some, it results in longer-term anxiety. Most women are willing to accept false-positive results to some extent*, as it is recognized as an expected consequence of breast cancer screening. False positives do not deter women from future screening (even when it implies an invasive procedure). ** "Women would be willing to accept a mean of 47.8 false-positive biopsies (median 27.2) (95% CI 24.9 to 70.8) to avoid one breast cancer death. 95% of women would accept between 6.7 and 127.3 false mammograms to save one life. 29% of women would be willing to accept 50 false-positive mammographies to avoid one breast cancer death (Sicsic 2018)".* *On average, women were willing-to-accept nearly 1.72 (CI 1.47–1.97) additional unnecessary follow-ups for one additional cancer detected per 100 women screened (Vass 2018)".*	⊕⊕⊕⊕ **HIGH**
Overdiagnosis	The majority of women are not aware of overdiagnosis. After being exposed to the concept, some women have difficulties understanding its meaning and implications. The term overdiagnosis generates confusion and, to some extent, causes them anxiety. Women who understand overdiagnosis are willing to accept its occurrence to some degree* (implying surgery or treatment of malignancies without impacting life expectancy). Demographic factors influence willingness. ** "Women would be willing to accept, on average, 14.1 cases of overdiagnosis (median 9.6) to avoid one breast cancer-related death (95% CI 12.9 to 15.2). 95% of women would be willing to accept between 2.3 and 49.2 overdiagnosis to save one life. 49% of women would be willing to accept 10 cases of overdiagnosis to avoid one death due to breast cancer (Sicsic 2018). "The median acceptable maximum number of over detected and overtreated cases in breast cancer screening ranged from 150/1,000 women to 313/1,000 women (for a 10% or 50% breast cancer mortality reduction, respectively) (Van den Bruel 2015)".*	⊕⊕⊕⊕ **HIGH** [1]
Burden and inconvenience	***When needing additional diagnostic procedures:*** women prefer precision over comfort. Most women report feeling more relief and confidence if they choose the intervention with a better chance of finding cancer, disregarding its invasiveness (e.g., preferring biopsy to image-based follow up).	⊕◯◯◯ **VERY LOW** due to risk of bias, imprecision, and indirectness

1 Certainty of the quantitative evidence for this outcome was judged as low. The overall confidence rating presented here represents the certainty of the qualitative evidence

Appendice 5 – Summary of findings table

Outcomes	№ of participants (studies) Follow-up	Certainty of the evidence (GRADE)	Relative effect (95% CI)	**Anticipated absolute effects[*] (95% CI)**	
				Risk with double human reading	**Risk difference with triage using artificial intelligence to assign single or double human reading**
Workload reduction	105915 (1 RCT)[1]	⊕⊕⊕⊕ High	**RR 0.56** (-- to --)	Study population	
				20.747 per 10.000	**9.129 fewer per 10.000** (20.747 fewer to 20.747 fewer)
Workload reduction	118997 (1 non-randomised study)[2]	⊕⊕⊕⊕ High	**RR 0.665** (-- to --)	Study population	
				20.000 per 10.000	**6.700 fewer per 10.000** (20.000 fewer to 20.000 fewer)

1. Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI): a randomised, controlled, parallel-group, non-inferiority, single-blinded, screening accuracy study. Lancet Digit Health. 2025;7(3):e175-e183. doi:10.1016/S2589-7500(24)00267-X
2. Lauritzen AD, Lillholm M, Lynge E, Nielsen M, Karssemeijer N, Vejborg I. Early Indicators of the Impact of Using AI in Mammography Screening for Breast Cancer. Radiology. 2024;311(3):e232479. doi:10.1148/radiol.232479.

Appendice 6 – Revisione esterna

REVISORE	PAGINA	RIGHE	TIPO COMMENTO (1, Maggiore; 2, Minore; 3, editing)	COMMENTO	RISPOSTA A COMMENTO
Simona Vecchi	**na**		**3**	Rivedere il formato delle referenze bibliografiche nel testo. In alcuni casi è stato riportato il cognome, l'iniziale del nome e l'anno.	Il formato delle referenze è stato armonizzato.
Simona Vecchi	**8**		**3**	Controllare il sommario. Come titoli principali vengono considerati le varie sezioni dell'EtD. Non so se è stata una scelta e se è coerente con il resto della linea guida.	Il sommario riprende solo alcune sezioni dell'EtD, in coerenza con il resto della linea guida.
Simona Vecchi	**6,15,34**	**20**	**1**	Mix di italiano/inglese che non facilita la lettura. Raccomanderei una revisione editoriale del testo per modificare/eliminare alcune frasi probabilmente derivanti da una traduzione dell'inglese non lineare. **Esempio.** Quesito clinico: dovrebbe una strategia di triage basato sull'intelligenza artificiale per assegnare a singola o doppia lettura umana ciascuna mammografia invece che la doppia lettura umana per tutte le mammografie essere utilizzato nello screening mammografico basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT) per la diagnosi precoce del cancro al seno nei programmi di screening mammografico? Proposta di riformulazione "Nei programmi di screening mammografico, per una diagnosi di cancro al seno basato su mammografia digitale (2DFFDM) o tomosintesi digitale del seno (DBT), dovrebbe essere utilizzata una strategia di triage basato sull'intelligenza artificiale per l'assegnazione a singola o doppia lettura umana rispetto alla doppia lettura umana?"	È stata condotta una revisione editoriale del testo per migliorare la formulazione di alcune frasi tradotte. Il quesito clinico non è stato modificato perché aderente alla formulazione standard prevista da GRADE.

Simona Vecchi	**17-18**	**na**	**2**	Non è chiara l'organizzazione delle diverse sezioni di questa LG. Nella sezione "metodologia" è stato inserito il paragrafo sui risultati della revisione sistematica del quesito 3 per gli esiti di efficacia, fattibilità e costi. Seguono poi i paragrafi sulle caratteristiche della LG (vedi paragrafo dal 3.9 in poi). Forse per una maggiore comprensione del testo distinguerei queste due parti. Sarebbe più utile una descrizione generale dei metodi per la revisione della letteratura e poi una sezione dedicata ai risultati della revisione sistematica in capitoli dedicati a ogni quesito. Ma non so se questa proposta è coerente con il resto della LG.	Nell'introduzione della metodologia riportata a pagina 7 è stato riportato un riferimento più preciso ai paragrafi relativi ai metodi specifici per la costruzione dell'EtD di questo capitolo. La struttura complessiva del documento è rimasta invariata per coerenza con gli altri capitoli della linea guida.
Simona Vecchi	**24**	**24**	**2**	Per una maggiore comprensione del testo sposterei la bibliografia dopo la sezione 9. "Regole per il panel" inserendo anche i due riferimenti inclusi nel 9.14. Non so se questa proposta è coerente con il resto della LG.	La bibliografia è stata riformattata ma è stata mantenuta la struttura del documento per coerenza con gli altri capitoli della linea guida.
Simona Vecchi	**16**	**6**	**2**	Nella sezione "*Strategia di ricerca e database*" viene citato l'evidence review team, mai citato prima (ERT). Inserire anche nella lista degli acronimi.	Inserito ERT nella lista degli acronimi.
Simona Vecchi	**16**	**12**	**2**	Nella sezione "*Strategia di ricerca e database*" tra le banche dati ricercate è stata riportata la piattaforma OVID. Sarebbe corretto specificare quale banca dati è stata ricercata tramite OVID. Per facilitare la lettura, non si potrebbe inserire la strategia di ricerca in un file allegato o materiale supplementare? è un dettaglio metodologico di interesse secondario rispetto alla LG.	La ricerca effettuata tramite Ovid è stata condotta all'interno della collezione Journals@Ovid Full Text, che offre accesso agli articoli in testo completo delle riviste disponibili sulla piattaforma Ovid. È stata inserita la banca dati specifica nel paragrafo. È stata inoltre spostata la strategia di ricerca in appendice.

Simona Vecchi	**19**		**1**	Nella descrizione del PICO vengono elencati gli esiti di interesse: cancri intervallo, tasso di identificazione dei tumori, tasso di richiamo, tasso di falsi positivi, valore predittivo positivo, carcinoma duttali in situ, riduzione del carico di lavoro. Descrivendo i risultati degli studi non sono invece riportati gli esiti effettivamente valutati nei 2 studi. Forse si potrebbe fare una rapida descrittiva degli esiti all'inizio dei risultati oppure riportare gli esiti nella tabella sulle caratteristiche degli studi inclusi.	Gli esiti di interesse sono riportati nella evidence-to-decision table (Appendice 2). Nei metodi, gli esiti di interesse valutati negli studi inclusi sono stati richiamati nel testo in relazione alla tabella 1 relativa alle caratteristiche degli studi inclusi.
Simona Vecchi	**18**	**4**	**2**	Nella descrizione degli studi, rivedere la coerenza tra diagramma di flusso e numeri di record identificati dalla ricerca riportati nel testo. "…ha comportato l'esclusione di 2.578 pubblicazioni, lasciando 106 studi 4 per la valutazione dei full-text".	Nel riportare il processo di selezione degli studi non è stato esplicitato un passaggio, ora integrato, relativo a 5 articoli per i quali non è stato possibile recuperare il full-text.
Simona Vecchi	**18**		**1**	Verificare il numero di record ottenuto dalla banca dati OVID (68 record?) Sembra un numero un po' basso.	È stata ripetuta la ricerca e si conferma il numero inserito.
Simona Vecchi	**42**		**2**	Questa dimensione del Grade prevede una valutazione della qualità delle prove relative all'uso delle risorse economiche. Non mi sembra sia stata fatta questa valutazione utilizzando strumenti validati e le considerazioni aggiuntive che sono state riportate non sono relative alla qualità degli studi considerati per questa dimensione. È un aspetto che solitamente viene trascurato in molte lg italiane. Solo per trasparenza metterei che non sono stati identificati studi che hanno valutato questo aspetto	Nella dimensione "Certezza delle evidenze sulle risorse richieste" è stato inserito che non sono stati inclusi studi che riportano valutazioni economiche.
Simona Vecchi	**43**		**2**	Per la dimensione Cost effectiveness, dichiarerei per trasparenza che non sono stati identificati studi di costo efficacia, mantenendo le considerazioni aggiuntive	Nella parte relativa alle evidenze è stato esplicitato che non sono stati inclusi studi di costo-efficacia.

Simona Vecchi	**55**	**95**	**1**	Per l'esito "cancer detection rate" e "ductal carcinoma in situ", la nota associata a questa stima (a) non mi sembra coerente con la stima riportata. come mai avete abbassato per imprecisione? Chiarire meglio la nota se possibile	La valutazione del grado di certezza delle evidenze è stata ricontrollata. Sebbene non siano state definite soglie decisionali a priori, la valutazione dell'imprecisione come seria deriva dal fatto che il beneficio dell'intervento è stato classificato come Moderato dal panel mentre il margine inferiore dell'intervallo di confidenza (IC) delle stime è stato valutato da ERT e confermato dal panel come piccolo o nullo e pertanto l'IC delle stime attraversa almeno una soglia decisionale. È stata modificata la nota per esplicitare meglio il ragionamento. Inoltre, durante la fase di revisione è emerso che per un errore fattuale non è stata applicata la valutazione di imprecisione alla stima della Cancer Detection Rate (CRD) da studio non randomizzato. Nella versione revisionata è stata applicata con relativa modifica del grado di certezza delle evidenze, che non modifica però il grado di certezza complessivo per la CDR in quanto si conferma il grado moderato per lo stesso esito da studio randomizzato.
Simona Vecchi	**55**	**95**	**1**	Per gli esiti Recall rate e false positive rate, chiarire meglio la nota c: "Inconsistenza valutata come seria sulla base del confronto tra le stime dei due studi inclusi, randomizzato e non randomizzato"	La nota è stata riformulata per chiarire che l'inconsistenza è stata valutata confrontando le stime dello stesso esito riportate dai due studi inclusi.

Simona Vecchi	**39**		**2**	Per facilitare la lettura, nella dimensione EtD "Values" si fa riferimento alla LG ECIBC e si rimanda all'Appendice 3. Eliminerei l'appendice e inserirei la tabella nell'EtD per una migliore comprensione.	È stata valutata la possibilità di spostare l'appendice nell'EtD ma per motivi di formattazione del documento il contenuto della tabella risulterebbe illeggibile per cui si ritiene più funzionale tenere il contenuto della tabella in appendice.
Simona Vecchi	**48**		**1**	Nella sezione "overall justification" dell'EtD evidenzierei ancora che la raccomandazione si basa solo su due studi e che, nonostante questo, il panel raccomanda l'intervento proposto. Questo spiegherebbe anche le numerose priorità della ricerca che, giustamente, avete identificato	È stata inserita una frase nella overall justification relativa al fatto che la raccomandazione si è basata principalmente sui risultati di due studi, mettendo però in luce che sono studi population based multicentrici con elevata numerosità del campione incluso (in totale oltre 200.000 donne).
Daniela Bernardi	17–18	Tutto il paragrafo	1 (maggiore)	Il periodo temporale di ricerca (gennaio 2022 – marzo 2025) è appropriato, ma sarebbe utile specificare se siano state considerate preprint o pubblicazioni in corso di stampa, poiché nel campo dell'IA mammografica l'aggiornamento delle evidenze è estremamente rapido.	È stato specificato nei metodi che preprint e/o pubblicazioni in corso di stampa non sono stati considerati.
Daniela Bernardi	17–18	"Sono stati inclusi articoli…"	2 (minore)	Potrebbe essere utile esplicitare se gli studi di validazione tecnica (reader studies su dataset selezionati) siano stati esclusi dalla revisione sistematica e, in caso affermativo, indicare se alcune di tali evidenze siano invece state considerate nella sezione 3.6 di integrazione delle evidenze. Una distinzione chiara tra studi di accuratezza tecnica e studi di impatto clinico favorirebbe la coerenza metodologica	Gli studi di validazione tecnica non sono stati inclusi nella revisione sistematica della letteratura. Per l'identificazione di studi utilizzati per considerazioni aggiuntive sui cancri intervallo nei domini GRADE "effetti desiderabili" è stato condotto un aggiornamento rapido della revisione dei 2 PICO adottati ECIBC, che avendo un approccio diagnostico includevano studi di confronto dell'accuratezza diagnostica. È stata integrata una frase esplicativa nella sezione 3.6.

Daniela Bernardi	19–20	"È stata condotta una sintesi narrativa…"	1 (maggiore)	Si potrebbe giustificare meglio la scelta di non condurre una meta-analisi. Pur riconoscendo la differenza di disegno tra i due studi inclusi (RCT e pre/post), l'omogeneità del software e degli outcome principali (CDR, recall rate) avrebbe potuto consentire una sintesi quantitativa esplorativa o, in alternativa, una spiegazione più esplicita delle ragioni metodologiche per cui non è stata effettuata	La scelta di non condurre una meta-analisi è stata spiegata integrando una frase nei metodi per esplicitare meglio le ragioni della scelta.
Daniela Bernardi	20	"Valutazione del grado di certezza…"	2 (minore)	Chiarire se la valutazione GRADE sia stata eseguita indipendentemente da due revisori e come siano state risolte eventuali divergenze, per trasparenza metodologica.	La valutazione è stata condotta in doppio e condivisa tra tutti i membri dell'ERT. È stata inserita una frase per esplicitare meglio questo passaggio nei metodi.
Daniela Bernardi	20–21	Tutto il paragrafo	1 (maggiore)	La sezione elenca numerose evidenze integrate al di fuori dei criteri di eleggibilità. Sarebbe utile distinguere chiaramente le fonti usate *a supporto di EtD* da quelle citate solo come contesto descrittivo, per evitare il rischio di sovrapposizione di livelli di evidenza.	La revisione sistematica della letteratura ha incluso tre report relativi a due studi che costituiscono la base di evidenza diretta per la raccomandazione formulata. Le evidenze identificate al di fuori della revisione sistematica della letteratura sono state utilizzate per informare le considerazioni aggiuntive dei domini dell'EtD. Il testo aggiunto in risposta al commento successivo include il razionale di inclusione delle evidenze aggiuntive.
Daniela Bernardi	21	"Ulteriori pubblicazioni… incluse nelle considerazioni aggiuntive"	2 (minore)	Suggerirei di aggiungere una tabella riassuntiva con i riferimenti principali (autore, anno, tipo di studio, motivo dell'inclusione) per maggiore trasparenza.	È stato integrato il testo dettagliando i tipi di studio ed il razionale per la scelta delle pubblicazioni incluse nelle considerazioni aggiuntive dei domini dell'EtD.

Daniela Bernardi	21–22	"L'inclusione di sistemi di IA…"	1 (maggiore)	L'analisi economica è utile ma presenta un approccio semplificato: si suggerisce di indicare esplicitamente se l'analisi è stata condotta dal punto di vista del SSN o del singolo centro, e se sono stati considerati costi di manutenzione, aggiornamento del software e licenze annuali.	L'analisi è condotta secondo la prospettiva del SSN e ipotizza che il prezzo per lettura sia omnicomprensivo di tutti i costi collegati all'utilizzo di applicazioni di IA nella lettura delle mammografie intesti come costi addizionali a regime. È stato inoltre specificato che non sono inclusi i costi di manutenzione, aggiornamento del software e licenze annuali. È stata inserita una frase nei metodi per chiarire questo punto.
Daniela Bernardi	21–22	"Il costo per lettura con l'ausilio di un software…"	2 (minore)	l riferimento al documento Santicchia 2024 è pertinente come esperienza gestionale locale, ma trattandosi di un project work accademico non pubblicato su rivista scientifica, sarebbe opportuno specificarne la natura (elaborato didattico PoliS-Lombardia, Università di Milano) e la fonte. Ciò migliorerebbe la tracciabilità e la distinzione tra evidenze scientifiche e casi di buona pratica	È stato esplicitato nei metodi il tipo di documento (Elaborato didattico PoliS-Lombardia, Università di Milano).
Daniela Bernardi	22	"L'intervento risulta cost-saving sino ad un costo di €3,83 per esame"	1 (maggiore)	Il paragrafo fornisce una stima economica articolata e coerente, ma sarebbe utile chiarire se l'analisi includa anche i costi indiretti (formazione iniziale, supporto IT, tempi di validazione e aggiornamento del software). Inoltre, pur essendo stata effettuata una valutazione di soglia (€3,83 per esame), si potrebbe esplicitare se siano stati considerati scenari alternativi o una sensitivity analysis per testare la robustezza del risultato in contesti con diversa struttura dei costi	Il paragrafo è stato revisionato, specificando meglio che la stima economica non include costi indiretti ma include i costi addizionali a regime. È stato specificato inoltre che il calcolo del valore soglia è un tipo di analisi di sensibilità utilizzata per valutare il prezzo massimo di una lettura con IA per il quale l'introduzione della lettura con AI rimane cost-saving.
Daniela Bernardi	22	"I prezzi futuri delle soluzioni di IA …. o	2 (minore)	Bene il richiamo al rischio di *lock-in* tecnologico. Potrebbe essere utile raccomandare esplicitamente un sistema di valutazione	Il concetto era in parte già espresso da una considerazione per l'implementazione espressa dal panel. Tale considerazione è

		lock-in da parte delle imprese”		periodica dei costi e delle performance per evitare dipendenza da un singolo vendor.	stata integrata con la proposta di attuazione di un sistema di valutazione periodica dei costi e delle performance.
Daniela Bernardi	22–23	“Survey sull’utilizzo dell’IA nei programmi di screening italiani”	1 (maggiore)	La survey è un contributo originale, ma non è chiaro se sia stata sottoposta a valutazione etica e se i risultati siano pubblicati o solo riportati internamente. Si suggerisce di aggiungere un riferimento al documento tecnico o al protocollo, al fine di garantirne la tracciabilità e la riproducibilità metodologica, come previsto dal processo GRADE	Nei metodi è stato esplicitato che la survey è stata condotta nel rispetto della normativa vigente. Non è stata richiesta l'approvazione etica poiché i partecipanti appartengono a due reti di professionisti che hanno accettato di collaborare alla missione istituzionale delle reti, compresa la partecipazione a sondaggi relativi al loro lavoro. La survey è stata inviata a tutti i referenti regionali screening, appartenenti alla rete dell’Osservatorio Nazionale Screening (ONS), che hanno coinvolto i referenti locali. La survey è stata inoltre inviata attraverso la rete FASO che rappresenta le società scientifiche professionali degli screening oncologici in Italia. I risultati sono stati presentati ad un congresso nazionale.
Daniela Bernardi	23	“Le principali barriere…”	2 (minore)	Si potrebbe commentare che le barriere individuate (risorse, IT, mancanza di LG) confermano la necessità di un approccio graduale all’implementazione, sottolineando il ruolo formativo dei TSRM e dei radiologi	Tale concetto è già stato espresso in una delle considerazioni per l’implementazione.
Daniela Bernardi	23	“Solo il 18% ha dichiarato di utilizzare tali soluzioni”	2 (minore)	Suggerire di distinguere i centri che utilizzano IA in ambito di ricerca da quelli in routine clinica, poiché il significato in termini di fattibilità è diverso	È stata inserito il dato che evidenzia la distinzione tra uso nei programmi di screening (organizzato e opportunistico) e finalità di ricerca.

Daniela Bernardi	34–38	Intero EtD	1 (maggiore)	l livello di dettaglio del framework è eccellente, ma l'adozione di due soli studi primari (entrambi con Transpara) limita la generalizzabilità delle evidenze. Si suggerisce di esplicitare nel paragrafo 'Limitazioni' la necessità di validazioni multicentriche con differenti algoritmi e in contesti organizzativi diversi, per rafforzare la trasferibilità della raccomandazione.	È stata aggiunta una frase relativa a questo punto nelle priorità per la ricerca.
Daniela Bernardi	36	"Interval cancer… non sono ancora presenti evidenze"	2 (minore)	Corretto sottolineare la mancanza di dati sugli *interval cancers*, ma si potrebbe aggiungere che questo aspetto rappresenta una priorità di ricerca per l'aggiornamento futuro della LG	È stata aggiunta una frase relativa a questo punto nelle priorità per la ricerca sebbene il panel abbia discusso l'imminente pubblicazione di dei dati del trial MASAI sui cancri intervallo.
Daniela Bernardi	37	"Gli studi inclusi sono stati condotti utilizzando un solo strumento di IA"	1 (maggiore)	Sebbene il testo precisi che entrambi gli studi siano stati condotti utilizzando un unico strumento di IA (Transpara), non è chiarito se la raccomandazione sia da intendersi vendor-independent. Si suggerisce di esplicitare che, pur basandosi su evidenze relative a un solo sistema, i principi del modello di triage IA possono essere potenzialmente applicati ad altri sistemi CE-marked con prestazioni equivalenti, previa validazione locale. Ciò consentirebbe di garantire neutralità tecnologica e maggiore trasferibilità.	È stata inserita una frase nelle overall justification che riprende questo concetto in maniera più esplicita rispetto a quanto già presente nel documento e in coerenza con quanto discusso dal panel in plenaria.

Daniela Bernardi	38	"La certezza delle evidenze è moderata"	2 (minore)	Si potrebbe rendere più esplicita la distinzione fra certezza *clinica* (esiti di accuratezza) e certezza *organizzativa* (impatto su workflow e costi).	Il grado di certezza delle evidenze è riportato per singolo outcome clinico nella Summary of findings table (Appendice 3) e valutato complessivamente dal panel nel dominio "Certezza delle evidenze" mentre il grado di certezza delle evidenze dell'impatto sul workflow è riportato nella Summary of findings (Appendice 5) mentre il grado di certezza delle evidenze sulle risorse richieste è valutato complessivamente nello specifico dominio dell'EtD.
Daniela Bernardi	40	"Balance of effects – probably favors the intervention"	1 (maggiore)	Il giudizio 'probabilmente favorevole' è coerente con le evidenze disponibili, ma sarebbe utile specificare che il bilancio tra benefici e rischi potrebbe variare in contesti con tassi di richiamo o di sovradiagnosi differenti da quelli osservati nei programmi nordici. Tale precisazione rafforzerebbe l'adattabilità della raccomandazione al contesto italiano	È stata aggiunta una frase relativa a questo punto nelle Additional considerations del dominio Balance of effects.
Daniela Bernardi	41	"Resources required – moderate savings"	2 (minore)	Il giudizio appare prudente; potrebbe essere utile indicare l'intervallo di costo per lettura considerato e chiarire se le stime derivano da simulazioni o dati empirici.	Le stime derivano da dati osservati nella letteratura e nella pratica. Si è scelto di riportare una stima puntuale del costo di lettura e di metterla in relazione ad intervalli riportati nelle fonti informative citate. Il valore soglia di costo della singola lettura con IA perché l'introduzione dell'IA risulti cost saving è stato calcolato tramite analisi di sensibilità.

Daniela Bernardi	47–48	“Conclusions”	1 (maggiore)	Si suggerisce di esplicitare che, data la certezza moderata delle evidenze e la natura condizionale della raccomandazione, l’adozione del modello di triage IA dovrebbe essere preceduta da una fase di implementazione controllata e di monitoraggio prospettico — ad esempio attraverso progetti pilota nazionali o regionali — per valutarne l’impatto reale su performance, costi e accettabilità	È stata inserita una frase nelle considerazioni per l’implementazione che riporta in maniera più esplicita questo concetto, già riportato in maniera meno esplicita in diversi commenti.
Daniela Bernardi			2 (minore)	L’intero documento è redatto con notevole rigore metodologico e rispecchia fedelmente il processo GRADE e di adolopment. Si suggerisce, per la versione finale SNLG, di integrare una breve sezione conclusiva che colleghi il capitolo sull’intelligenza artificiale agli altri capitoli delle linee guida (es. densità mammaria, personalizzazione del rischio, formazione del personale), al fine di rafforzare la coerenza complessiva del percorso di screening e diagnosi del tumore della mammella.	Si ringrazia per il commento. Le raccomandazioni incluse nei vari capitoli del processo italiano di Adoloopment delle Linee Guida ECIBC sono pubblicate tutte insieme sulla stessa pagina del SNLG.