Thermography as a screening and diagnostic tool: a systematic review

Anita Fitzgerald, Jessica Berentson-Shaw

Abstract

Aims To determine the effectiveness of digital infrared thermography for the detection of breast cancer in a screening population, and as a diagnostic tool in women with suspected breast cancer.

Methods A comprehensive search of electronic databases together with a search of international websites was conducted. Diagnostic studies comparing thermography with mammography for screening in asymptomatic populations; or comparing thermography with histology in women with suspected breast cancer; were eligible for inclusion. Quality of included studies was appraised using the QUADAS criteria.

Results One study reported results for thermography in screening population and five studies reported diagnostic accuracy of thermography in women with suspected breast cancer. Overall, studies were of average quality. Sensitivity for thermography as a screening tool was 25% (specificity 74%) compared to mammography. Sensitivity for thermography as a diagnostic tool ranged from 25% (specificity 85%) to 97% (specificity 12%) compared to histology.

Conclusions Currently there is not sufficient evidence to support the use of thermography in breast cancer screening, nor is there sufficient evidence to show that thermography provides benefit to patients as an adjunctive tool to mammography or to suspicious clinical findings in diagnosing breast cancer.

Clinical thermography has been in use since the 1960s and detects temperature variation on the surface of the skin; in breast cancer, thermography involves using a thermal imaging device to detect and record the heat pattern of the breast surface.[1] There are several methods of thermography; this review will focus on the most common method used by commercial companies in New Zealand (NZ) and—infrared thermography where infrared radiation emitted by the skin surface is detected. Information from an infrared detector is relayed to a processing system, which produces images of temperature distribution.[2]

Thermography does not provide information on the morphological characteristics of the breast, rather it provides functional information on thermal and vascular conditions of the tissue. The role of thermography is considered to be complimentary to other techniques; as it is a test of physiology that alone is not sufficient for medical practitioners to make or confirm a diagnosis[1,3]

The current method of breast cancer screening in both New Zealand and Australia is by mammography. BreastScreen Australia was launched in 1991, followed by BreastScreen Aotearoa (New Zealand) in 1998; both services offer mammography to

women aged 45–69, although in Australia, women from 40 years, and women over 70 years are able to attend for screening.

A New Zealand-conducted Health Technology Assessment (HTA) reported high sensitivity and specificity of screening mammography and showed that test accuracy improves with the age of patients.[4] The current method of breast cancer diagnosis in both New Zealand and Australia, is the 'triple test' including clinical breast examination, diagnostic mammography, and fine-needle aspiration biopsy (FNAB). This combination is considered positive if any of the three components are positive, and negative if all three components are negative.[5]

The use of thermography is controversial; it is promoted as a tool to monitor breast health by private thermography clinics, while in New Zealand it is not part of any national breast cancer health program. One previous systematic review on the effectiveness of thermography for detection of breast cancer was conducted in 2004[2] but since then new studies on thermography have been published; the authors are unaware of any current systematic reviews including these recent articles.

The objective of this review is two-fold: to determine the effectiveness of digital infrared thermography for the detection of breast cancer in a screening (asymptomatic) population, and to determine the effectiveness of digital infrared thermography as a diagnostic tool in women with suspected breast cancer.

Methods

Searching the literature

The literature was systematically searched for English language articles that fitted the inclusion criteria from 1984 to the end of April 2011. Additionally, reference lists of retrieved studies were searched and websites discussing thermography were searched for potential studies.

The following databases were searched: MEDLINE, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Review, Database of Abstracts of Reviews of Effects, EMBASE, Cumulative Index to Nursing and Allied Health Literature, PsychoINFO and Web of Science. The clinical questions were able to be combined in one search and complete search strategies are available from the corresponding author on request.

Several additional sources were searched to minimise the likelihood of missing an important study. The following resources were searched for guidelines on thermography: Guidelines International Network, National Guideline Clearing House, National Library for Health (UK), SIGN, TRiP (Turning Research into Practice).

Several international websites, including all available HTA sites were searched for reports on thermography for breast screening or diagnosis; a full list is available on request. Additionally, a number of New Zealand-specific resources were searched, including: KRIS (Kiwi Research Information Service), Australasian Digital Theses Programme, Index New Zealand, Te Puna and Digital NZ.

Selection of studies for inclusion

Study design—This review included diagnostic accuracy studies of which there are two basic types, defined by the Centre for Reviews and Dissemination[6]; single-gate design and two-gate design. Full details of the designs of these studies is reported elsewhere.[6] Single and two-gate studies were eligible for inclusion if they compared digital infrared thermography with mammography in screening asymptomatic women, or if they compared with digital infrared thermography with histology in women with suspected breast cancer.

Studies were required to have sufficient data to construct a 2×2 contingency table which displays numbers of true-positive cases, false-positive cases, false-negative cases, and true-negative cases of breast cancer.

Participants—For studies investigating thermography for screening, asymptomatic women with unknown disease status were eligible for inclusion. For studies investigating thermography for diagnosis, women with suspicious symptoms (e.g. presenting with a breast lump or nipple discharge), women with suspicious findings on clinical examination or women with an abnormal mammogram were eligible for inclusion.

Studies of patients younger than 16 years, animal studies, and studies with fewer than ten participants were excluded.

Index test—Digital infrared thermography was the index test considered in this review. Other methods of thermography and outdated methods no longer available were excluded.

Studies which sought to develop interpretive software or models to assess the accuracy of different imaging parameters and that were not primarily designed to assess accuracy of thermography in testing for breast cancer in a normal patient population (diagnostic or screening), were excluded.

Reference standard—For studies investigating thermography as a screening tool, a reference standard of histology was not considered appropriate. In this case, mammogram or clinical diagnosis was accepted as the reference standard. For studies investigating thermography as a diagnostic tool, the reference standard was histology.

Data collection and analysis

For each included study, we used standard evidence tables to extract characteristics of participants, data about the index tests and reference standard, and aspects of study methods. We extracted indices of diagnostic performance from data presented in each primary study by constructing 2×2 contingency tables of true-positive cases, false-positive cases, false-negative cases, and true-negative cases. If these were not reported, we reconstructed the contingency table using the available information on relevant parameters (sensitivity, specificity or predictive values).

Study quality was assessed using the QUADAS checklist,[7] with each item scored as "yes", "no", or "unclear". Results of the quality assessment are presented in the text, in graphs and in a table using the Cochrane Collaborations Review Manager 5 software.[8] The authors did not calculate a summary score estimating the overall quality of an article since the interpretation of such summary scores is problematic and potentially misleading.[9,10]
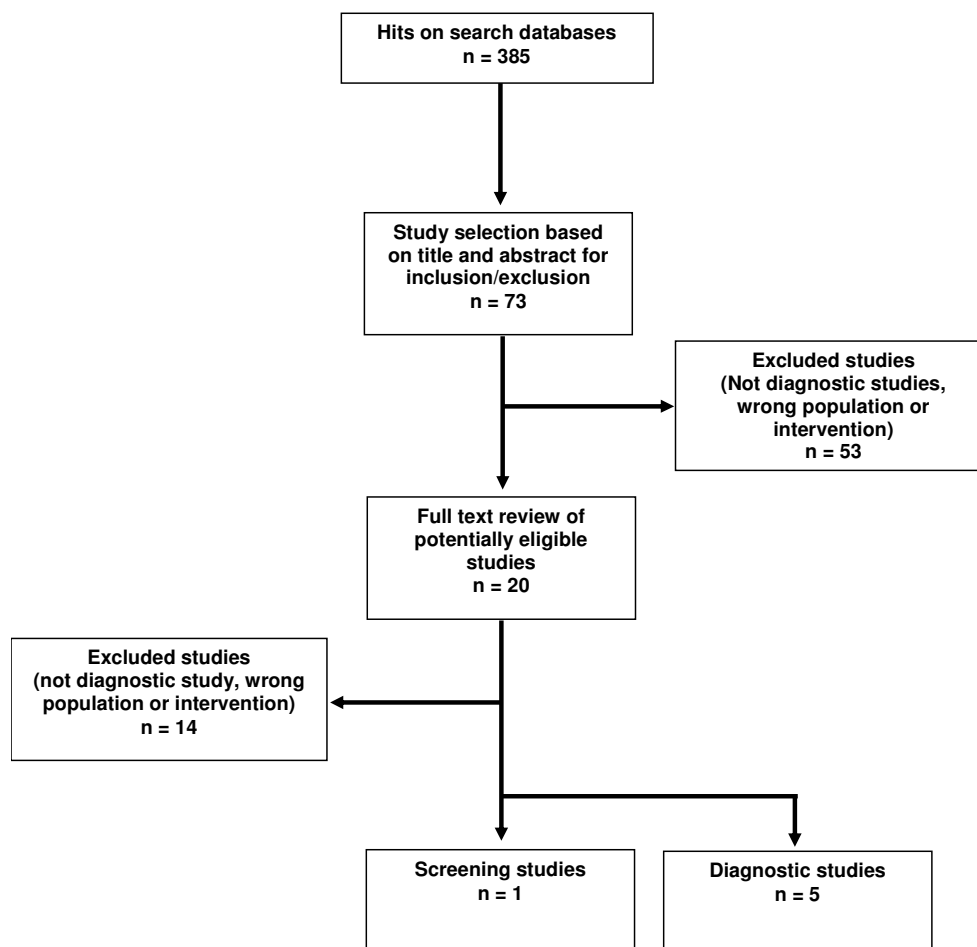
Sensitivity, specificity, negative predictive values, positive predictive values, and likelihood ratios (with 95% confidence intervals) were calculated for each test in each study using the methods described by the Centre for Reviews and Dissemination[6] and results tabulated and presented in ROC space. Area under the ROC curve gives a graphical representation of sensitivity and specificity of a test.

Results

The searches identified 385 citations of which 73 appeared to be relevant. Of these, 20 were considered relevant to the purpose of our review the fulltexts were retrieved (Figure 1). Fourteen articles were subsequently excluded.

The most common reason for exclusion was that the study was either not a primary diagnostic study of test accuracy or it did not involve appropriate comparisons. One study, with a total of 306 participants fulfilled the inclusion criteria for screening, and five studies, with a total of 1224 participants fulfilled inclusion criteria for diagnosis in women with suspected breast cancer.

Figure 1. Selection of studies for review

```
┌─────────────────────────┐
│  Hits on search databases │
│        n = 385          │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Study selection based │
│ on title and abstract for│        ┌──────────────────────────┐
│   inclusion/exclusion   │───────▶│     Excluded studies     │
│        n = 73           │        │ (Not diagnostic studies, │
└─────────────────────────┘        │  wrong population or      │
            │                      │      intervention)        │
            │                      │         n = 53           │
            ▼                      └──────────────────────────┘
┌─────────────────────────┐
│    Full text review of  │
│  potentially eligible   │
│        studies          │
│        n = 20           │
└─────────────────────────┘
            │
┌──────────────────────────┐
│     Excluded studies     │
│ (not diagnostic study,   │◀──────┐
│  wrong population or      │       │
│     intervention)        │       │
│        n = 14           │       │
└──────────────────────────┘       │
            │                      │
     ┌──────┴──────────┐
     ▼                 ▼
┌──────────────┐  ┌──────────────┐
│  Screening   │  │  Diagnostic  │
│   studies    │  │   studies    │
│    n = 1     │  │    n = 5     │
└──────────────┘  └──────────────┘
```

Breast thermography for screening

One study was identified investigating the accuracy of thermography to determine the diagnostic accuracy of digital infrared thermography for the detection of breast cancer in a screening (asymptomatic) population by Williams and colleagues in 1990 (Table 1).[11]

Quality of included study—The quality of the included study was poor. The QUADAS tool reports that 'reported estimates of diagnostic accuracy may have limited clinical applicability (generalisability) if the spectrum of tested patients is not similar to the patients in whom the test will be used in practice'; this study may have been subject to spectrum bias, since volunteers may have had greater risk of developing breast cancer than those not screened.

Spectrum bias occurs when the participants included in a study are not similar to those in whom the test would be used in practice and can limit the generalisability. There was confusion regarding the role of clinical examination as it seemed to occur in both

the index test and the reference standard groups; this meant that blinding, and accuracy of both the index test and reference standard were not clear.

Verification bias occurs when not all of the study group receive confirmation of the diagnosis by the reference standard (partial verification bias) or when some of the index test results are verified by a different reference standard (differential verification bias).

Verification bias is likely in this study because not all participants received the same reference standard, only those with positive findings on thermogram had a mammography; this could cause biased estimates of the performance of thermography as the negative results were not confirmed as being accurate. The effect of those lost to follow-up without explanation is unclear.

Diagnostic accuracy—A prospective single-gated (diagnostic cohort) study aimed to determine whether thermography could be used to identify women with breast cancer during screening, or identify women at risk of developing breast cancer within 5 years.[11]

10,229 women aged 40–65 were invited and attended a breast screening clinic. At the time of screening, infrared imaging reported a sensitivity of 61%, specificity of 74%, a positive predictive value of 0.01% and a negative predictive value of 1.00%. Five years following initial screening, infrared imaging reported a sensitivity of 28%, specificity of 74%, positive predictive value of 0.01% and a negative predictive value of 0.99%.

Thermography is not sufficiently sensitive to be used as a screening test for breast cancer, nor is it useful as an indicator of risk developing within 5 years. Currently there is not sufficient evidence to support the use of thermography in breast cancer screening.

Table 1. Included studies investigating thermography in a screening population (Williams 1990)

Participants	Index test	Reference standard	Method of analysis	Sens	Spec	PPV	NPV	LR+	LR-
n=10229	Infrared imaging**	Mammography	At screening	61%	74%	0.01%*	1.00%	2.35 (1.91–2.88)*	0.53 (0.38–0.73)*
			At 5-year follow-up	28%	74%	0.01%*	0.99%	1.09 (0.73–1.63)*	0.97 (0.83–1.14)*

Abbreviations: n – number of participants; PPV – positive predictive value; NPV – negative predictive value; LR+ - positive likelihood ratio; LR- - negative likelihood ratio.

* indicates NZGG calculated values.

** Device used – two devices were used in this study, one by AWRE (Aldermaston, in conjunction with Barr and Stroud) and one by Rank Precision Industries. No further details were reported.

Breast thermography for diagnosis

Five studies were identified assessing the use of thermography as a diagnostic tool in women with suspicious symptoms (Table 2).[12–16]

Quality of included studies—Overall the included studies were of average quality. All studies reported a high risk of bias for at least one item on the QUADAS checklist. Overall the most common sources of bias were insufficient descriptions of the reference standard and index tests; this is important because variations in diagnostic accuracy can often be traced back to differences in the execution of the index test or reference standard. It is also important because a clear and detailed description is needed to implement the test in another setting.

Another source of bias in the included studies was the spectrum of patients within the studies not being representative of the population in whom the test would be used in practice; this can limit the generalisability. Poor reporting of the delay between index tests and reference standards was evident in all included studies, and blinding of reference or index test results to the other was also poorly reported.

Diagnostic accuracy—A limited number of studies were identified comparing digital infrared thermography to histology in women with symptoms, suspicious clinical findings, or abnormal mammogram. Four studies used a single-gate (diagnostic cohort) design, while one study used a two-gate (diagnostic case-control) design. Two were conducted in the UK,[14,16] two in the USA,[12,15] and one in Canada.[13]

While most studies were able to show sensitivity over 70% for at least one mode of digital infrared thermography, the specificity of thermography for diagnosting breast cancer was generally low, between 12% and 85% for most studies (Table 2). One study reported results that conflicted with other studies, showing low sensitivity (25%) and a high specificity (85%)[14] and another study showed high (83%) sensitivity and high 81% specificity (81%)[13].

In the studies presented in this review, low specificities are due to a high number of false-positive results. For example, the study by Parisky[15] reported a false-positive rate of 1544 and a false-negative rate of 13 out of the 2299 patients tested. This means that for 68% of the patients in this study thermography provided an incorrect diagnosis. Another study by Arora[12] that showed a higher specificity reported a false-positive rate of 19 and a false-negative rate of 6 in a study of 92 participants. This means that for 27% of the patients in the study, thermography provided an incorrect diagnosis.

The study by Keyserlingk[13] provided figures for combined modality approaches to breast cancer diagnosis, however there was not enough data presented in that particular study to confirm the accuracy of the different combinations of tools.

When plotted in ROC space, overall the included studies show poor performance for accurately diagnosing breast cancer (Figure 2). Currently there is not sufficient evidence to show that thermography provides benefit to patients as an adjunctive tool to mammography or to suspicious clinical findings in diagnosing breast cancer.

Table 2. Included studies comparing thermography with biopsy/histology

Reference (study design)	Participants	Index test	Reference standard	Unit of analysis	Method of analysis	Sens	Spec	PPV	NPV	LR+ (95%CI)	LR- (95%CI)
Arora 2008 Single-gate	n=92	Infrared imaging	Histology	Patients	Screening mode (overall score determined by software from 0–7 where 0=normal and 1–7=abnormal)**	97%	12%	66%*	67%	1.10 (0.96–1.25)*	0.28 (0.05–1.47)*
					*The location of the lesion under question based on prior imaging was assessed to generate a positive or negative clinical assessment **	90%	44%	74%*	71%	1.61 (1.18–2.20)*	0.23 (0.10–0.53)*
					Score of positive or negative generated by the artificial neural network	97%	26%	70%*	82%	1.31 (1.07–1.62)*	0.13 (0.03–0.55)*
Kontos 2011 Single-gate	n=63 (126 breasts)	Infrared imaging	Histology	Breasts	Manual review of images	25%	85%	24%	86%	1.67 (0.68–4.09)	0.89 (0.69–1.14)
Wishart 2010 Single-gate	n=100 (106 biopsies)	Infrared imaging	Histology	Biopsies	Screening mode (overall score determined by software from 0–5 where 0=normal and 1–5=abnormal)	53%	41%	59%	36%	0.92 (0.65–1.30)*	1.11 (0.71–1.74)*
					Score of positive or negative generated by the artificial neural network	48%	74%	73%	48%	1.79 (1.01–3.17)*	0.72 (0.54–0.96)*
					Manual review of images	78%	48%	69%	59%	1.49 (1.09–2.05)*	0.46 (0.26–0.81)*

					NoTouch BreastScan (artificial intelligence program, algorithm-based)	70%	48%	67%	51%	1.34 (0.97–1.87)*	0.62 (0.38–1.02)*
Keyserlingk 1998 Two-gate (included in HTA)	n=200 (100 cases, 100 controls)	Infrared imaging	Histology	Patients	Manual review of images	83%	81%	83%	81%	4.37 (2.89–6.61)	0.21 (0.13–0.33)
Parisky 2003 Single-gate (included in HTA)	2625 evaluations reported for 875 lesions	Infrared imaging	Histology	Number of evaluations	Manual review of images	97%	14%	24%	95%	1.14 (1.11–1.17)	0.18 (0.11–0.32)

Abbreviations: n: number of participants; PPV: positive predictive value; NPV: negative predictive value; LR+ : positive likelihood ratio; LR-: negative likelihood ratio.

* Indicates NZGG calculated values.

** This is a direct quote from the study describing the 'clinical mode' used. It is unclear whether the scans were manually read or utilised the device software or were read by some other method.

† Device used: Meditherm med2000 thermal imaging system (Meditherm Beaufort, NC, USA).

‡ Device used: Sentinel BreastScan (Infrared Sciences Corp., Bohemia, NY, USA).

Figure 2. Include studies plotted in ROC space
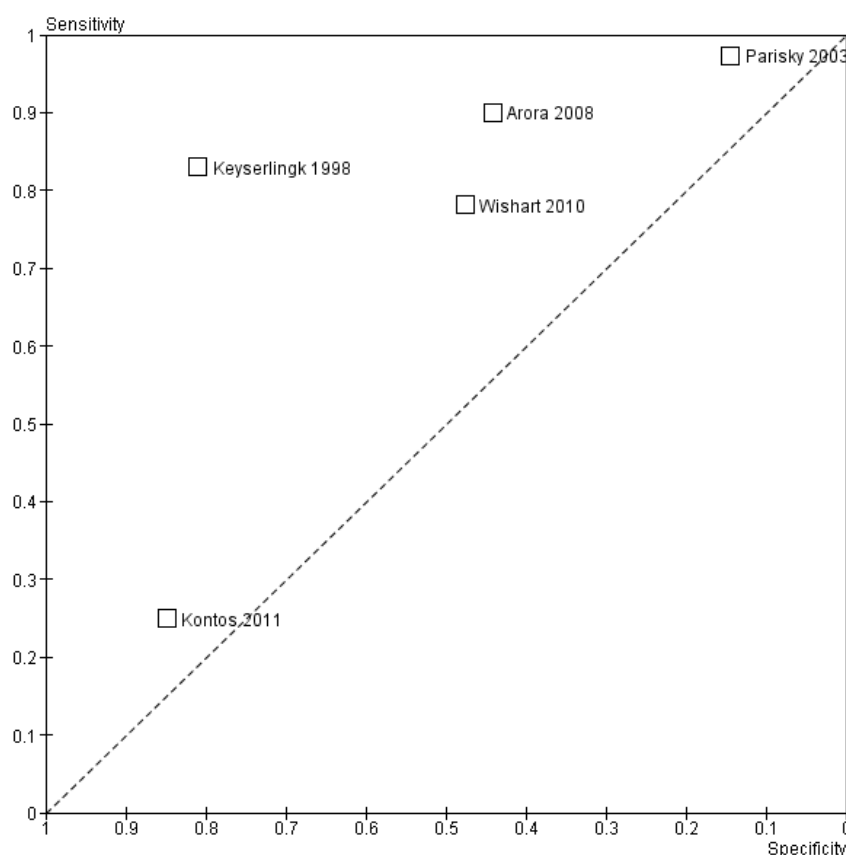
Note: This graph represents single sensitivity and specificity measures for the manually reviewed thermograms (see Table 2). Studies by Arora and Wishart included other measures of accuracy (neural network interpretations of thermograms) but the thermogram interpretation by manual expert review was common to all studies and has been used here.

Discussion

Extensive systematic literature searches were conducted, study quality was carefully assessed using a validated tool,[7] and the authors attempted to maximise available data by deriving accuracy data from those studies where not all diagnostic measures were reported. In terms of its use as a screening tool, this review found that digital infrared thermography is not sufficiently sensitive to be used as a screening test for breast cancer, nor is it useful as an indicator of the risk of developing breast cancer within five years. In terms of its use as a diagnostic tool, this review found that there is not sufficient evidence to show that thermography provides benefit to patients as an adjunctive tool to mammography or to suspicious clinical findings in diagnosing breast cancer.

One of the limitations of reviewing the accuracy of diagnostic studies is poor reporting in the included; where authors of studies have not reported elements necessary to answer criteria included in a QUADAS appraisal, the authors cannot be

certain whether this indicates poor methodology with its subsequent consequence for bias, or simply poor reporting of a methodologically sound study.

The introduction and implementation of the STAndards for the Reporting of Diagnostic accuracy studies (STARD) guidelines may improve reporting of diagnostic studies in the future.[17,18] The objective of the STARD initiative is to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study (internal validity) and to evaluate its generalisability to populations of interest (external validity).

Industry sponsoring appears to have played a role in the conclusions of some of the included studies investigating thermography as a diagnostic tool. Three industry sponsored studies[12,15,16] concluded that thermography was a valuable adjunctive test to mammography and/or clinical examination, despite the low specificity reported.

Two studies did not state the source of funding; of these, one study[14] reported that due to the low specificity, thermography should not be used as an adjunctive tool to diagnose breast cancer; the other[13] reported more favourable results for thermography, but indicated that thermography trials are conducted in highly controlled environments and stated that "*Our initial data should not be extrapolated to either formal screening or non-controlled diagnostic environments without appropriate evaluation, preferably in prospective controlled multicentre trials*."[13] It is concerning that results differ between those industry sponsored studies reviewed and those conducted independent of industry. High quality, large scale diagnostic studies, with particular attention to sources of funding are needed.

This systematic review of thermography as a screening and diagnostic tool has some limitations. Overall, our findings are limited by the small number of studies available in the literature; incomplete reporting of studies' characteristics and results; limited methodological quality of those reviewed studies; and relatively small sample sizes. Only one study was identified investigating thermography as a screening tool.

For the studies investigating thermography as a diagnostic tool, pooling studies in a diagnostic meta-analysis was not possible because of limited data, and the heterogeneity between studies. Similarly, due to the limited number of identified studies, sensitivity analyses were not possible to assess which methodological aspects may have contributed to clinical heterogeneity (for example the timing of imaging, the different characteristics of the patient population) or heterogeneity related to study design (for example prospective versus retrospective studies, presence of incorporation bias).

Studies were heterogeneous in a number of areas:

- The units of analysis differed across studies; two studies reported results by number of patients, one by number of breasts, one by number of biopsies and one by number of evaluations

- The method of thermogram analysis differed between the included studies. Three studies used expert physicians to manually review the images, while two studies used modern artificial neural networks to review images.

- Study design differed, one study used a two-gated approach, four used single-gated approaches.

The high false-positive and false-negative rates noted in thermography are problematic in the context of commercially driven fee-for-service screening tests that are not part of an organised screening programme because of the ongoing ability to generate repeat business. Those with a negative or equivocal test result are often encouraged to 'monitor' their breast health by those organisations providing the service in order to identify future abnormalities; the consequences of this may be twofold, on one hand those with a positive result (abnormality on thermogram) are likely to seek unnecessary mammography at additional cost, even though there is a high chance that a positive result derived from a thermogram is false.

On the other hand, the very idea that their breast health is being monitored is likely to lead some consumers to the conclusion that they have been adequately screened and that mammography is unnecessary. The psychological cost of having a positive thermogram cannot be ignored, particularly when the rate of false-positives is likely to be high. Screening and diagnostic tools offered to at risk individuals in the context of, or as an adjunct to, tests within a comprehensive and organised screening programme must be sufficiently accurate and cost-effective to keep these issues to a minimum and to provide the best possible care for the patient.

To date, no studies of infrared thermography have been conducted in New Zealand or in Australia, although thermography is offered in both countries to members of the public on a fee for service basis.

In conclusion, currently there is insufficient evidence to support the use of thermography in breast cancer screening, nor is there sufficient evidence to show that thermography provides benefit to patients as an adjunctive tool to mammography or to suspicious clinical findings in diagnosing breast cancer.

Competing interests: None declared.

Author information: Anita Fitzgerald, Assistant Research Manager; Jessica Berentson-Shaw, Manager – Research Services; New Zealand Guidelines Group, Wellington

Correspondence: New Zealand Guidelines Group, Level 10, 15 Murphy Street, PO Box 10 665, The Terrace, Wellington, New Zealand. Email: afitzgerald@nzgg.org.nz

References:

1. Ng EYK, Kee EC. Advanced integrated technique in breast cancer thermography. J. Med. Eng. Technol. 2008;32(2):103-14.
2. Kerr J. Review of the effectiveness of infrared thermal imaging (thermography) for population screening and diagnostic testing of breast cancer. NZHTA Tech Brief Series, 2004.
3. Kennedy DA, Lee T, Seely D. A comparative review of thermography as a breast cancer screening technique. Integr. Cancer Ther. 2009;8(1):9-16.
4. Hider P, Nicholas B. The early detection and diagnosis of breast cancer: a literature review - an update. In: Report N, editor, 1999.
5. Royal New Zealand College of General Practitioners. Early detection of breast cancer: guidelines for primary care providers. . Wellington, 1999.
6. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. University of York: CRD, University of York, January 2009.

7. Whiting P, Rutjes A, Reitsma J, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med. Res. Methodol. 2003;3(25).

8. Review Manager (RevMan) Version 5.1 [program]. Copenhagen: The Nordic Cochrane Centre, 2011.

9. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054-60.

10. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med. Res. Methodol. 2005;5(19).

11. Williams K, Phillips B, Jones P, et al. Thermography in screening for breast cancer. J. Epidemiol. Community Health 1990;44:112-13.

12. Arora N, Martins D, Ruggerio D, et al. Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. Am. J. Surg. 2008;196(4):523-6.

13. Keyserlingk J, Ahlgren P, Yu E, Belliveau N. Infrared imaging of the breast: initial reappraisal using high-resolution digital technology in 100 successive cases of stage I and II breast cancer. Breast J. 1998;4:245-51.

14. Kontos Mea. Digital infrared thermal imaging (DITI) of breast lesions: sensitivity and specificity of detection of primary breast cancers. Clin. Radiol. 2011;6(6):536-39.

15. Parisky Y, Sardi A, Hamm R, et al. Efficacy of computerized infrared imaging analysis to evaluate mammographically suspicious lesions. Am. J. Roentgenol. 2003;180:263-69.

16. Wishart GC, Campisi M, Boswell M, et al. The accuracy of digital infrared imaging for breast cancer detection in women undergoing breast biopsy. Eur. J. Surg. Oncol. 2010;36(6):535-40.

17. Bossuyt P, Reitsma J, Bruns D, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med 2003;138:40-4.

18. Bossuyt P, Reitsma J, Bruns D, et al. Standards for Reporting of Diagnostic Accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann. Intern. Med. 2003;138:W1-12.