

NZHTA TECH BRIEF SERIES

July 2004

Volume 3 Number 3

Review of the effectiveness of infrared
thermal imaging (thermography) for
population screening and diagnostic
testing of breast cancer

Jane Kerr

New Zealand
Health Technology Assessment

Department of Public Health and General Practice
Christchurch School of Medicine
Christchurch, NZ.

Division of Health Sciences, University of Otago

NEW ZEALAND HEALTH TECHNOLOGY ASSESSMENT (NZHTA)

Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
Christchurch, New Zealand

Review of the effectiveness of infrared thermal imaging (thermography) for population screening and diagnostic testing of breast cancer

Jane Kerr

This report should be referenced as follows:

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(3)*

Titles in this Series can be found on the NZHTA website:
http://nzhta.chmeds.ac.nz/thermography_breastcancer.pdf

2004 New Zealand Health Technology Assessment (NZHTA)

ISBN 1-877235-64-4

ISSN 1175-7884

ACKNOWLEDGEMENTS

This Tech Brief was commissioned by the National Screening Unit of the New Zealand Ministry of Health.

The report was prepared by Dr Jane Kerr (Research Fellow) who selected and critically appraised the evidence. The research protocol for this report was developed by Ms Marita Broadstock (Research Fellow). The literature search strategy was developed and undertaken by Mrs Susan Bidwell (Information Specialist Manager). Mrs Ally Reid (Administrative Secretary) provided document formatting. Internal peer review was provided by Dr Robert Weir (Senior Research Fellow), Dr Ray Kirk (Director) and Ms Broadstock.

We are grateful to Dr Simon Baker (Public Health Physician, National Screening Unit, Ministry of Health), Dr Madeleine Wall (Clinical Leader, Breastscreen Aotearoa, National Screening Unit, Ministry of Health) and Dr Richard Tremewan (Medical Physics and Bioengineering Department, Christchurch Hospital) who provided valuable advice and input.

We are grateful to the authors who provided assistance in the retrieval of articles for this review: Professor E. F. Ring (Medical Imaging Research Group, School of Computing, University of Glamorgan), Dr N. A. Diakides (Advanced Concepts Analysis Inc., Falls Church, Virginia) and Professor R. G. Margolese (Jewish General Hospital, McGill University, Montreal). We are also grateful to those from the fields of breast screening, medical physics/bioengineering, radiology and infrared thermal imaging who provided advice regarding available literature.

DISCLAIMER

NZHTA takes great care to ensure the accuracy of the information supplied within the project timeframe, but neither NZHTA nor the University of Otago can accept responsibility for any errors or omissions that may occur. NZHTA and the University of Otago along with their employees accept no liability for any loss of whatever kind, or damage, arising from the reliance in whole or part, by any person, corporate or natural, on the contents of this paper. This document is not intended to be used as personal health advice; people seeking individual medical advice are referred to their physician. The views expressed in this report are those of NZHTA and do not necessarily represent those of the University of Otago, or the New Zealand Ministry of Health.

COPYRIGHT

This work is copyright. Apart from any use as permitted under the *Copyright Act 1994* no part may be reproduced by any process without written permission from New Zealand Health Technology Assessment. Requests and inquiries concerning reproduction and rights should be directed to the Director, New Zealand Health Technology Assessment, Christchurch School of Medicine and Health Sciences, P O Box 4345, Christchurch, New Zealand.

CONTACT DETAILS

New Zealand Health Technology Assessment (NZHTA)
Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
PO Box 4345
Christchurch
New Zealand
Tel: +64 3 364 3696 Fax: +64 3 364 3697

Email: nzhta@chmeds.ac.nz

Website: <http://nzhta.chmeds.ac.nz>

LEVEL OF EVIDENCE CONSIDERED IN TECH BRIEFS

Tech Briefs are rapidly produced assessments of the *best available evidence* for a topic of highly limited scope. They are less rigorous than systematic reviews. Best evidence is indicated by research designs which are least susceptible to bias according to evidence levels adapted from Bandolier (<http://www.jr2.ox.ac.uk/bandolier/band70/b70-5.html>, accessed 22 June 2004) that categorise studies of diagnostic methods according to susceptibility for bias (see **Appendix 1**). Where methodologically acceptable and applicable, appraised evidence is limited to systematic reviews, meta-analyses, evidence-based clinical practice guidelines, health technology assessments and randomised controlled trials (RCTs). Where not available, poorer quality evidence may be considered.

CONFLICT OF INTEREST

None.

EXECUTIVE SUMMARY

Background

Over the last few years, modern digital infrared thermal imaging (thermography) has become available to patients in New Zealand through privately-financed providers. The National Screening Unit (NSU) in New Zealand is receiving inquiries about the effectiveness of infrared thermography in the early detection and diagnosis of breast cancer. The aim of this review is to provide the NSU with independent information from which to respond to these queries and to assist New Zealand patients and clinicians in making informed choices about the role of infrared thermography in the screening and diagnosis of breast cancer.

Aim

To review the evidence for the effectiveness of infrared thermography for population screening and diagnostic testing of breast cancer.

Topic One - What is the international evidence on the effectiveness, benefits, harms and costs of using infrared thermography as a screening tool for breast cancer?

Topic Two - What is the international evidence on the effectiveness, benefits, harms and costs of using infrared thermography as an adjunctive diagnostic tool for breast cancer?

Method

The MeSH headings (Medline subject headings) thermography and breast neoplasms were used. Additional keywords included thermography, thermometry, thermology, infrared, infra-red, breast adj3 cancer, breast adj3 malignan\$, breast adj3 carcinoma, breast adj3 neoplas\$. The NZHTA Core Search was employed and included major bibliographic databases (Medline, Embase etc.) and review databases (EBM reviews, Cochrane, DARE etc). The literature search for this evaluation was not limited by publication date or language.

Results

The literature search identified 1,154 potentially relevant articles in abstract form. After a series of selection criteria were applied to the abstracts of the articles, 85 were retrieved in full text, from which a final group of three papers were identified as eligible for inclusion in the review. One study was formally appraised and included in this review for Topic One. This study was of prospective cohort design. Two studies were formally appraised and included for Topic Two, one of these studies was a case-control design and the other was a clinical trial design.

Conclusions

A significant finding of the review conducted for the purpose of this NZHTA Tech Brief was that there were no studies that evaluated the effectiveness of the infrared technology devices that are available in New Zealand. In addition, there were few that evaluated comparable infrared technologies, or technologies that may become available to New Zealand. The evidence that is currently available does not provide enough support for the role of infrared thermography for either population screening or adjuvant diagnostic testing of breast cancer. The major gaps in knowledge at this time can only be addressed by large-scale, prospective randomised trials. More robust research on the effectiveness and costs of technologically advanced infrared thermography devices for population screening and diagnostic testing of breast cancer is needed, and the conclusions of this review should be revisited in the face of additional reliable evidence.

GLOSSARY

Asymptomatic - asymptomatic people are those who do *not* have a symptom (e.g., breast lump) that may be due to a disease (e.g., breast cancer).

Benign tumour - a benign tumour is an abnormal growth that is neither malignant, nor a cancer. A benign tumour is not capable of spreading, and usually does not recur after being removed.

Bias - deviation of results or inferences from the truth, or processes leading to such deviation. Any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth.

Biopsy - in a breast biopsy, a sample of tissue is removed to be examined under a microscope, as an aid to diagnosis.

Breast cancer - a histologically-proven malignant lesion that is classified as ductal carcinoma in situ or invasive breast cancer.

Breast cancer classification - breast cancers are classified in terms of tumour cell type, grade, size, lymph node involvement and stage.

Breast implant - a round or tear-shaped sack inserted into the chest in order to restore or enhance the shape of the breast. A breast implant may be filled with saline, silicone or a synthetic material.

Blinded study - a study in which observers and/or subjects are kept ignorant of the group to which they are assigned.

Cancer - a general term for a large number of diseases that all display uncontrolled growth and spread of abnormal cells (also called malignant tumours). Cancer cells have the ability to continue to grow, invade and destroy surrounding tissue, and leave the original site and travel via the lymph or blood systems to other parts of the body where they may establish further cancerous tumours.

Carcinoma - a malignant tumour made up of epithelial cells that may infiltrate surrounding tissues and spread to other parts of the body via the blood or lymph nodes systems.

Case-control study - an epidemiological study involving the observation of *cases* (persons with the disease, such as breast cancer) and a suitable *control* (comparison, reference) group of persons without the disease. The relationship of an attribute (e.g., positive diagnostic test result) to the disease is examined by comparing retrospectively the past history of the people in the two groups with regard to how frequently the attribute is present.

Cohort study - an epidemiological study in which subsets of a defined population can be identified who are, have been, or in the future may be exposed or not exposed in different degrees, to a factor or factors (e.g., receiving a screening test for breast cancer) hypothesised to influence the probability of occurrence of a given disease or other outcome (e.g., positive biopsy for breast cancer). Studies usually involve the observation of a large population, for a prolonged period (years), or both.

Confidence interval (CI) - the computed interval with a given probability – e.g., 95%, that the true value of a variable such as a mean, proportion, or rate is contained within the interval. The 95% CI is the range of values in which it is 95% certain that the true value lies for the whole population.

Confounder - a third variable that indirectly distorts the relationship between two other variables.

Core Biopsy - needle is used to remove a core of tissue for histological examination.

Coverage - the number, percent, or proportion of eligible women reached by a programme.

Cross-sectional study - a study that examines the relationship between diseases or other health related characteristics (e.g., patient diagnosed with definite breast cancer) and other variables of interest (e.g., positive diagnostic test result for breast cancer) as they exist in a defined population at one particular time.

Descriptive study - a study concerned with, and designed only, to describe the existing distribution of a variable, without regard to causal or other hypotheses.

Diagnosis - the process of identifying a disease by its characteristic signs, symptoms and findings on investigation.

Diagnostic test efficacy - the impact and usefulness of a diagnostic test expressed in terms of its technical properties.

Ductal carcinoma-in situ (DCIS) - a form of breast cancer, which spreads along the ducts of the breast, but has not invaded the duct wall.

Effectiveness - a measure of the extent to which a specific intervention, procedure, regimen, or service, when deployed in the field in routine circumstances, does what it is intended to do for a specified population.

Epidemiology - the study of the distribution and determinants of health-related states or events in specified populations.

Evidence based - based on valid empirical information.

Evidence table - a summary display of selected characteristics (e.g., of methodological design, results) of studies of a particular intervention or health problem.

False negative result - a negative test result in a person who *does* have the condition being tested for.

False positive result - a positive test result in a person who *does not* have the condition being tested for.

Fine needle aspiration biopsy (FNAB) - a fine needle is used to remove some cells or fluid from a suspicious area for histological examination.

Generalisability - applicability of study results to other populations.

Histology - the microscopic study of the minute structure and composition of tissues.

Lesion - an area of tissue damaged by disease or injury.

Mammogram - a specialised X-ray of the breast. It may be used as a screening test in women with no signs or symptoms of breast cancer, or it may be used to evaluate breast symptoms (e.g., breast lump).

Mammography - the process of taking a mammogram.

Māori - the indigenous people of New Zealand.

Mean - calculated by adding all the individual values in the group and dividing by the number of values in the group.

Median - any value that divides the probability distribution of a random variable in half. For a finite population or sample, the median is the middle value of an odd number of values (arranged in ascending order) or any value between the two middle values of an even number of values.

Meta-analysis - the process of using statistical methods to combine the results of different studies. The systematic and organised evaluation of a problem, using information from a number of independent studies of the problem.

Microcalcification - microcalcification means “very small calcification”. Calcifications (areas where calcium has been deposited) are frequently seen on mammograms of the breast, and are usually due to benign breast processes. Certain appearances and patterns of calcification are associated with malignancy.

Morbidity - illness.

Mortality - the number of deaths from a specified disease that are diagnosed or reported during a defined period of time in a given population.

Negative predictive value (NPV) - the probability that a person having a negative result on a test does not have the condition that the test is designed to detect.

Positive predictive value (PPV) - the probability that a person having a positive result on a test has the condition that the test is designed to detect.

Population-based screening programme - a population-based screening programme is one in which screening is systematically offered by invitation to a defined, identifiable population: this requires a means of identifying and inviting the target population – e.g., through a population register.

Randomised controlled trial (RCT) - an epidemiologic experiment in which subjects in a population are randomly allocated into groups to receive or not receive an experimental preventive or therapeutic procedure, manoeuvre or intervention. The groups are compared prospectively. RCTs are generally regarded as the most scientifically rigorous method of hypothesis testing available in epidemiology.

Reference standard - an independently applied test that is compared to a screening or diagnostic test being evaluated in order to verify the latter’s accuracy. A reference standard therefore provides an accurate or “truth” diagnosis for verification of positive and negative diagnoses. It is sometimes referred to as providing “final truth determination”.

Screening - screening is the examination of asymptomatic people in order to classify them as likely or unlikely to have the disease that is the object of screening. The aim of screening is to detect disease before it is clinically apparent, and for this to improve the outcome for people with the disease.

Selection bias - error due to systematic differences in characteristics between those who are selected for inclusion in a study and those who are not (or between those compared within a study and those who are not).

Sensitivity (Se) - sensitivity is the proportion of truly diseased persons in a tested population who are identified as diseased by a test. Sensitivity is a measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test.

Specificity (Sp) - specificity is the proportion of truly non-diseased persons in a tested population who are so identified by a test. Specificity is a measure of the probability of correctly identifying a non-diseased person with a test.

Surgical (open) biopsy - the surgical removal (performed under a local or general anaesthetic) of all, or part, of a suspicious area of tissue for histological examination.

Symptomatic - symptomatic people are those who do have a symptom (e.g., breast lump) that may be due to a disease (e.g., breast cancer).

Systematic review - literature review reporting a systematic method to search for, identify and appraise a number of independent studies.

True negative - a test correctly identifies a person without the disease.

True positive - a test correctly identifies a person with the disease.

Tumour - an abnormal growth of tissue. A breast tumour may be: localised without potential (benign); malignant and growing inside the milk ducts (DCIS); malignant and invading nearby tissues (invasive), or; malignant and invading distant tissues (metastatic).

Ultrasound - the use of high frequency sound waves to study an organ or tissue. Ultrasound is particularly useful for distinguishing fluid-filled structures from solid lesions.

OBJECTIVE OF THIS REPORT AND RESEARCH QUESTIONS

The aim of this Tech Brief was to review the evidence for the effectiveness of infrared thermography for population screening and diagnostic testing of breast cancer. Because there were two topics under consideration, the research questions and results of this review for each topic have been considered separately.

Topic One – Screening

What is the international evidence on the effectiveness, benefits, harms and costs of using infrared thermography as a screening tool for breast cancer?

Topic Two – Diagnosis

What is the international evidence on the effectiveness, benefits, harms and costs of using infrared thermography as an adjunctive diagnostic tool for breast cancer?

BACKGROUND

This Tech Brief was requested by Simon Baker, Public Health Physician, National Screening Unit, Ministry of Health, New Zealand.

Rationale for this Report

In New Zealand, breast cancer is an important health concern. For non-Māori women, it is the commonest cause of cancer registrations and death, and for Māori women it is the second commonest, after lung cancer (BreastScreen Aotearoa, 2004). Over the last few years, modern digital infrared thermal imaging (thermography) has become available to patients in New Zealand through privately-financed providers. The National Screening Unit (NSU) in New Zealand is receiving inquiries about the effectiveness of infrared thermography in the early detection and diagnosis of breast cancer. The aim of this review is to provide the NSU with independent information from which to respond to these queries and to assist New Zealand patients and clinicians in making informed choices about the role of infrared thermography in the screening and diagnosis of breast cancer.

Definitions of Thermography

Thermography is the recording of temperature, and has various applications in industrial (e.g., for testing of materials) and military (e.g., for surveillance) contexts (Anbar, 1995). Clinical thermography is the recording of temperature to form an image (thermogram) of the temperature distribution on the surface of the body. Clinical thermography has been considered as a diagnostic instrument for a variety of medical conditions since the 1960s, although its application for many of these remains controversial (McLean, 1999). There are several thermography methods (also known as thermometry and thermology). These include “contact thermography” (where a needle inserted into a suspicious area measures temperature directly, or a sheet of heat-sensitive liquid crystal film is applied to the breast), and “microwave thermography” (where temperature distribution is indicated by microwave radiation emitted by the surface of the body).

The most common thermography method is “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. Since the 1970s, it has been possible to process the infrared information using computers, that can then display and store images of the infrared thermal patterns (Ring and Ammer, 2000). Infrared thermal imaging can be either ‘static’ or ‘dynamic’. Static imaging (also sometimes called ‘steady-state’ imaging) is when a patient’s infrared data is detected under controlled conditions (Ohashi and Uchida, 2000). During dynamic infrared imaging, the patient is also subjected to one of several methods of cooling, such as cold air blowing over the breast area being imaged, or cold water submersion of the hands (Ohashi and Uchida, 2000). The manner and rate by which the temperature of the skin on the breasts’ surface recovers its equilibrium is then measured by the infrared detector (Jones, 1998).

Infrared thermography using digital computer processing systems (sometimes also referred to as digital infrared thermal imaging) is currently being marketed in New Zealand for breast cancer detection, and is the focus of this report. Obvious potential benefits of infrared thermography are that it is a non-invasive technique and that it does not involve any patient (or operator) exposure to ionising radiation. However, it is limited in its role as a primary (stand-alone) breast cancer diagnostic modality by its inability to inform clinicians where an abnormality lies to permit even a blind biopsy (Homer, 1985; van Dam et al., 1988; Head et al., 2000; Head and Elliott, 2002; Arena et al., 2003).

Current understandings of the underlying pathological mechanisms for increased temperature in breast cancer are that breast cancer cells produce nitric oxide (NO). This NO interferes with the normal neuronal (nervous system) control of breast tissue blood vessel flow by causing regional vasodilation in the early stages of cancerous cell growth, and enhancing angiogenesis (new blood vessel formation) in later stages (Anbar, 1998; Anbar et al., 2001). The subsequent increased blood flow in the area causes a temperature increase relative to the normal breast temperature, and even deep breast lesions seem to

have the ability to induce changes in skin temperature (Snyder et al., 2000). Breast cancer metabolic processes may also contribute to the detectable increase in heat (Foster, 1998). These changes relate to physiological breast processes. It is believed that in healthy individuals, temperature is generally symmetrical across the midline of the body (McLean, 1999; Ng et al., 2002). Subjective interpretation of many diagnostic imaging modalities, including infrared thermographic images relies on the underlying philosophy that normal contralateral images are relatively symmetrical, and that small asymmetries may indicate a suspicious lesion (Qi and Head, 2001). Therefore, in breast cancer, infrared thermography detects disease by identifying areas of asymmetric temperature distribution on the breasts' surface (Qi and Head, 2001).

Medical History of Thermography

The use of thermography as a means of detecting breast cancer has a substantial history. Earlier research is still referred to in promotional material – for example, in references to “800 peer reviewed breast thermography studies” (International Academy of Clinical Thermology [IACT], website accessed 20 March 2004-IACT is a non-profit, non-regulatory agency with interests in research, education, and the establishment of standards and guidelines in clinical thermography).

Before the passage of the Medical Device Amendment in 1976 requiring FDA approval for devices, thermography was actively studied and used as a tool for the early detection of breast cancer in the US (Nass et al., 2001). Up to 3,000 thermography clinics operated in the US during this time (Foster, 1998). Data generated by researchers from between the 1960s and the early 1970s suggested that thermography may be a viable screening tool for breast cancer. A subsequent reanalysis of this data has identified major problems with its relevance to screening, including that much of it was based in research generated from symptomatic patients, rather than asymptomatic populations (see Moskowitz, 1995, for an overview). However, several important studies in this period that were undertaken to evaluate breast screening methods did include thermography, and shall be briefly summarised here.

Twenty-seven Breast Cancer Detection Demonstration Projects (BCDDP) were set up in 1972 in the United States, with the aim of screening 10,000 asymptomatic women per project over a five year period, with an additional five year follow-up (Letton and Mason, 1986). The plan was to compare thermography, mammography, and clinical examination as screening tools, using state of the art equipment for the time. However, thermography was dropped early in these studies due to high false positive rates and low sensitivity (Moskowitz, 1985; Moskowitz, 1995; Foster, 1998). Some controversy followed after thermography was dropped from the BCDDPs. The controversy mainly concerned how well the investigators had been trained in thermography methods and interpretation, but the equipment used was state of the art for that time (Moskowitz, 1995). Furthermore, Threatts et al. (1980) subsequently reported the results of an investigative review in which ten experienced thermographers conducted a blinded interpretation of 576 thermograms from the BCDDP. They concluded that the overall detection rate of thermography for all breast cancers (including in situ and minimal cancers) for the population studied was similar to that that would result from chance (Threatts et al., 1980, in Margoese, 1998).

Feig et al. (1977) compared the use of thermography, an early form of mammography (xeromammography), and clinical examination as a screening tool in a separate clinical trial including 16,000 women. Overall, thermography's sensitivity and specificity were 39% and 82%, respectively, compared with xeromammography's 78% sensitivity and 98% specificity.

Following these trials, medical interest in thermography has steadily decreased (Foster, 1998). In 1984, the American College of Radiology issued a policy statement that thermography was still an experimental procedure with no established clinical indications for the detection of breast cancer (Dankiw, 1990).

Since 1990, two Health Technology Assessment (HTA) reviews of thermography have been published. These reviews included evaluations of the evidence for its application in the detection and management of breast cancer. The first of these reports, in the form of a narrative review (which included an overview of existing institutional reviews of the technology), found no conclusive evidence supporting the use of thermography as either a screening or diagnostic tool for breast cancer (Dankiw, 1990). The second HTA report considered both contact and infrared thermography methods, and reviewed the

literature from 1983 to 1999, as well as pertinent earlier articles identified by this process. This report also found insufficient evidence to support the use of thermography in any aspect of the management of breast cancer, and in addition concluded that it had "...been shown convincingly that thermography has no place in mass primary or secondary screening for breast cancer..." (McLean, 1999, p7).

Also, in Canada, Margolese et al. (1998) developed a set of clinical guidelines for the management of women presenting with a palpable breast lump, after conducting a systematic review of the literature published to January 1996 (with non-systematic review to January 1997) to gather information for these guidelines. The systematic review itself was not published, but in the published report of the final guidelines, the authors clearly stated that thermography was not a recommended diagnostic procedure in the work-up for women presenting with a palpable breast lump. The authors of the guidelines further concluded that there was no role for thermographic techniques at that time, outside the testing of improved technology in structured clinical trials (Margolese et al., 1998).

Technological History of Infrared Thermography

A brief outline of the technological history of infrared thermography is presented to enable understanding of the subsequent rationale for the focus of this current report. Although infrared thermal imaging has been used in medicine since the early 1960s, infrared camera technologies that were available from the 1970s up until the mid-1980s suffered from several limitations. These limitations included poor spatial resolution (uncertainty up to 1cm²), poor thermal resolution (only able to diagnose abnormalities > 1°C difference), slow systems (taking up to 4 seconds per image) that therefore also substantially reduced both spatial and temperature resolutions, poor calibration systems, and optical aberrations on the camera lenses that resulted in unreliability at image peripheries (Anbar, 1998). In addition, older infrared cameras required the addition of liquid nitrogen to cool the detector. This required a period of equilibrium before image acquisition, limited the angle of camera orientation, and necessitated regular camera maintenance. Newer systems contain efficient electronic cooling systems, which have overcome these problems (Jones, 1998; Ring and Ammer, 2000). By the mid-1980s, some thermography practitioners stated that, at that time, new and more accurate thermography systems had supplanted "older technology" (Gautherie et al., 1987).

The advances in infrared camera technology over the last decade have been accompanied by progress in computerised image processing systems (Diakides, 1998; Foster, 1998; Jones, 1998). It was not until the 1970s that data acquired by infrared cameras was processed by computers into digital images for viewing (Ring and Ammer, 2000). Now, some sophisticated modelling programmes can enhance the spatial resolution of images already acquired (Snyder et al., 2000). In addition, although infrared images are often analysed subjectively, which is therefore dependent on interpreter experience, recently some researchers have been trialling new computer software statistical programmes that analyse the data from infrared breast images quantitatively. Examples of the uses that such software programmes allow include the development of algorithms to calculate quantitative scores of risk for breast lesions (Arena et al., 2003), and the objective detection of small asymmetries on infrared images (Head and Elliott, 1997; Qi and Head, 2001). Technological advances in recent years have resulted primarily from military developments in infrared surveillance, and have led to renewed medical interest in the technology (Diakides, 1998; Nass et al., 2001). Given the advances in both infrared cameras and digital processing of the acquired images, several authors have suggested that the systems currently available are not comparable to those that were previously used (Anbar, 1998; Ring and Ammer, 2000; Nass et al., 2001; Arena et al., 2003).

The digital infrared thermal imaging (DITI) device "Med 2000®" was developed around 1990 (Hobson, 2002). MedithermNZ is promoting this technology from a base in Auckland. In addition, an "Infrared Thermal Imaging Medical Clinic" has been established in Tauranga, offering infrared digital imaging with a system from Micro Health Systems, Inc. (MHS 5000 Thermal Imager®). The detectors in these systems are lightweight, do not require liquid nitrogen cooling, and have thermal sensitivities ranging from 0.01 to 0.1°C (http://www.meditherm.com/mms_products.htm, accessed 22 March 2004; <http://www.mhs5000.com/specs.htm>, accessed 23 March 2004). Computer software specific for each system enables the images acquired by the infrared detector to be viewed, manipulated, and stored on computers. The providers of infrared thermal imaging in New Zealand therefore use infrared imaging systems that are technologically more advanced than those available up until the mid-1980s.

Screening for Breast Cancer

The aim of screening for any disease is to reduce mortality (and morbidity) from this disease by performing a test on an asymptomatic population, to detect disease at an earlier stage than it would be encountered in an unscreened population (Feig, 1999; Sunshine and Applegate, 2004). It aims to reduce the risk of development of, or dying from, a disease for a screened population. A reduction in mortality as a result of a screening programme depends upon high levels of coverage of the population (in the case of breast cancer, this population is defined as women asymptomatic of breast cancer) as well as quality screening processes and follow-up services (BreastScreen Aotearoa, 2004).

In New Zealand, a nationwide breast screening programme, called BreastScreen Aotearoa, was launched in December 1998. This programme offers free screening mammography every two years to women who: are aged 45 to 69 years; have not had mammography within the previous 24 months; are free from breast cancer (or at least five years since diagnosis if previously diagnosed); are asymptomatic; are New Zealand citizens (or whose usual abode is in the Cook Islands, Niue or Tokelau); hold an immigration permit that allows a stay in New Zealand of two or more years (BreastScreen Aotearoa, 2004).

A recent systematic review conducted in New Zealand for the Royal New Zealand College of General Practitioners considered methods for the early detection of breast cancer. The report concluded that screening mammography has a high sensitivity (between 80-95%) and specificity (93-95%), and that these measures of test accuracy generally increase with patient age (Hider, 1999). International evidence has shown that delivering mammographic breast screening through a properly organised programme is effective in reducing mortality from breast cancer by 30 percent for women aged 50 to 69 years if coverage is 70 percent (International Agency for Research on Cancer, 2002, cited in BreastScreen Aotearoa, 2004). The international evidence for screening mammography has been acquired from a number of prospective randomised controlled trials, which have demonstrated statistically significant reductions in mortality from breast cancer (see Mahoney, 2002, for overview). Prior to the establishment of the nationwide programme in New Zealand, it had been demonstrated in pilot programmes that mammographic screening could be done effectively and efficiently in New Zealand (BreastScreen Aotearoa, 2004).

Thermography used for screening for breast cancer

Criteria for assessing the validity of screening programmes (adapted from Wilson and Jungner, 1968) include that the screening test be reliable, valid, and repeatable; acceptable, safe and easy to perform; have a high positive predictive value (PPV); and be sensitive and specific. Furthermore, the screening programme's cost should be commensurate with the benefits of early detection, and there should be effective treatment available for managing the abnormality identified by the screening test. These criteria highlight necessary outcomes for studies of the effectiveness of infrared thermography in the domain of breast cancer screening. The outcomes of particular interest in this review included measures of clinical effectiveness including PPV, test sensitivity (Se) and test specificity (Sp)-particularly relating to small cancers, which are most treatable; and measures of cost effectiveness including health care and other costs.

Infrared thermography practitioners have suggested that it should not be used alone as a method of screening, or as a replacement for mammography, but rather it should be used and evaluated as a complementary screening and detection procedure adjunctive to physical examination and mammography (Head et al., 2000; IACT, website accessed 20 March 2004). Previously, clinicians have also used thermography as a stand-alone screening procedure, possibly "encouraged in doing so by the commercial vendors of the various devices" (Martin, 1983).

For breast cancer screening, this Tech Brief therefore considered the role of infrared thermography for breast cancer screening from both these perspectives. Firstly, the role of infrared thermography was considered as an adjunctive one (e.g., with screening mammography or other tests), to detect possible incremental improvements in breast cancer detection, or any other any other potential benefits that this technique may offer. In addition to the health care costs, the psychological costs (i.e., anxiety and confusion) that may be caused by receiving conflicting screening test results were also considered as relevant factors for adjunctive testing in breast cancer screening. Secondly, the role of infrared

thermography was considered as a comparative one (e.g., thermography versus mammography as a primary screening tool).

The IACT website (accessed 20 March 2004) proposes that thermography can act as an “early warning” system by identifying signs of possible cancer or pre-cancerous cell growth that would not be detectable by other screening methods for up to ten years. As detailed in Wilson and Jungner’s (1968) screening programme criteria above, for this to be an advantage, studies would need to demonstrate how thermography leads to the earlier detection and treatment of cancer, and therefore reduced mortality from the disease, compared with conventional screening approaches. Evaluating the use of thermography in risk assessment (i) without cancer detection and mortality outcomes, or (ii) without being part of a multimodal screening programme, was therefore not within the scope of this review.

Diagnosis of Breast Cancer

The role of a diagnostic test for any disease is to further evaluate abnormalities that have been pre-detected by either clinical symptoms, or screening tests. To be clinically efficacious, a diagnostic technique must permit a confident characterisation of the nature of a lesion that can be shown to alter patient management (Orel and Troupin, 1993).

For breast cancer diagnosis in New Zealand, current recommendations are that the “triple test” is the method of choice for diagnosing any palpable abnormality of the breast (Royal New Zealand College of General Practitioners, 1999). The components of the “triple test” are: 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. Most studies have found that the combination of tests in the “triple test” is more reliable than each alone for evaluating a breast mass (Hider, 1999; Donegan, 2002). Furthermore, when the results are concordant, the sensitivity is around 99%, and when all three tests are negative, breast cancer is highly unlikely (Hider, 1999; Donegan, 2002). FNAB appears to be the most accurate part of the triple test, but this partly results from its incremental value after clinical examination and diagnostic mammography have occurred (Hider, 1999). Ultrasound is unlikely to add any diagnostic information beyond the triple test, but is recommended in certain clinical situations – e.g., if clinical findings are suggestive of a cystic lesion (Hider, 1999). For younger women (under 35 years) with symptomatic breast lesions, ultrasound is more sensitive than mammography, and is recommended as the first imaging test (Royal New Zealand College of General Practitioners, 1999). Other imaging tests that are sometimes used in the diagnosis of breast cancer in New Zealand include scintomammography, colour doppler and Magnetic Resonance Mammography, but these are not currently recommended as routine diagnostic tests (Royal New Zealand College of General Practitioners, 1999).

If an asymptomatic woman has a possible breast abnormality detected by a screening mammogram from BreastScreen Aotearoa (the nationwide screening programme in New Zealand), the current recommendations are for that women to have further mammographic assessment, and then ultrasound or other work-up as considered appropriate by radiologists experienced in breast imaging. If an abnormality cannot be confirmed as benign by these tests, the next recommendation is for a clinical examination to be performed by an experienced clinician, and then FNAB or breast tissue biopsy as appropriate (BreastScreen Aotearoa, 2004).

Thermography used for diagnosis of breast cancer

As noted previously, infrared thermography is limited in its role as a primary (stand-alone) breast cancer diagnostic modality by its inability to inform clinicians where an abnormality lies to permit even a blind biopsy (Homer, 1985; van Dam et al., 1988; Head et al., 2000, Head and Elliott, 2002; Arena et al., 2003). For breast cancer diagnosis, the role of infrared thermography was therefore considered by this Tech Brief as an adjunctive one (e.g., with clinical breast examination, mammogram, and/or ultrasound) in studies of patients with suspected breast cancer, to detect possible incremental improvements in breast cancer detection, or any other potential benefits that this technique may allow. The use of thermography in predicting prognosis for patients with breast cancer, or in monitoring treatments of breast cancer was not considered.

TECH BRIEF METHODS – SELECTION CRITERIA

Explanations of selection criteria are detailed in **Appendix 2**.

Topic 1: Screening

Study inclusion criteria

Publication type

Studies published between 1985 and 20 May 2004 (inclusive) in the English language, including primary (original) research (published as full original reports) and secondary research (systematic reviews and meta-analyses).

Setting

Studies evaluating the use of infrared thermography as an adjunctive or stand-alone tool for the population screening of breast cancer.

Population

Asymptomatic women at (any) risk for breast cancer. Notably, thermography has been promoted in New Zealand as particularly useful in the detection of abnormalities in women aged 30 to 50 years, women with small breasts, and women with breast implants (Wharton, 2003), so these factors were not considered to be exclusion criteria.

Sample size

Studies with samples of at least 20 participants.

Index test

- studies considering “technologically advanced” infrared imaging devices that are (or are comparable with those) currently available, or which may become available in the future to patients in New Zealand. To identify devices as “technologically advanced”, publications from 1985 were considered for appraisal.

Comparators

- studies evaluating multimodal screening approaches (e.g., screening mammography, clinical breast examination) including infrared thermography compared with the same approaches without infrared thermography
- studies evaluating infrared thermography compared with mammography as primary screening tools.

Reference standard

Histological confirmation – e.g., core breast biopsy, open surgical breast biopsy.

Outcomes

Measures including any of the following presented in the results:

- significant differences in estimates of sensitivity (Se), specificity (Sp), positive predictive value (PPV) or negative predictive value (NPV), detection of disease at an earlier stage between comparators in the detection of cancer
- quality of life (including psychological costs)

- health care costs
- safety outcomes
- reduction in breast cancer mortality.

Study exclusion criteria

Research papers were excluded if they:

- were not published in English
- concerned studies of infrared thermography devices that are no longer available (or that are not technologically comparable to those available in New Zealand)
- concerned studies of methods of thermography that were not infrared thermography, or did not specify the method of thermography under study
- reported on the use of thermography in aspects of breast cancer management other than screening
- had only an abstract available or were conference proceedings
- were an expert opinion commentary, a narrative review or a book chapter
- were discussion articles – e.g., presented as a letter or editorial
- reported a sample size lower than 20 participants or a case presentation
- reported animal studies
- did not clearly describe their methods and results, or had discrepancies in data within the article.

Topic 2: Diagnosis

Study inclusion criteria

Publication type

Studies published between 1985 to 20 May 2004 (inclusive) in the English language, including primary (original) research (published as full original reports) and secondary research (systematic reviews and meta-analyses).

Setting

Studies evaluating the use of infrared thermography as an adjunctive tool for the diagnosis of breast cancer.

Population

Patients symptomatic for breast cancer (e.g., presenting with a breast lump, thickening, asymmetrical glandular prominence, pain, or nipple discharge) or who have had an abnormal mammogram.

Notably, thermography has been promoted in New Zealand as particularly useful in the detection of abnormalities in women aged 30 to 50 years, women with small breasts, and women with breast implants (Wharton, 2003), so these factors were not considered to be exclusion criteria.

Sample size

Studies with samples of at least 20 participants.

Index test

- studies considering “technologically advanced” infrared imaging devices that are (or are comparable with those) currently available, or which may become available in the future to patients in New Zealand. To identify devices as “technologically advanced”, only publications from 1985 were considered for appraisal.

Comparators

- studies evaluating multimodal diagnostic approaches for patients with suspected breast cancer (e.g., clinical breast examination, mammogram, and/or ultrasound) including infrared thermography compared with the same approaches without infrared thermography.

Reference standard

Histological confirmation – e.g., core breast biopsy, open surgical breast biopsy).

Outcomes

Measures including any of the following presented in the results:

- significant differences in estimates of sensitivity (Se), specificity (Sp), positive predictive value (PPV) or negative predictive value (NPV), detection of disease at an earlier stage between comparators in the detection of cancer
- quality of life (including psychological costs)
- health care costs
- safety outcomes
- reduction in breast cancer mortality.

Study exclusion criteria

Research papers were excluded if they:

- were not published in English
- concerned studies of infrared thermography devices that are no longer available (or that are not technologically comparable to those available in New Zealand)
- concerned studies of methods of thermography that were not infrared thermography, or did not specify the method of thermography under study
- reported on the use of thermography in aspects of breast cancer management other than diagnosis
- had only an abstract available or were conference proceedings
- were an expert opinion commentary, a narrative review or a book chapter
- were discussion articles – e.g., presented as a letter or editorial
- reported a sample size lower than 20 participants or a case presentation
- reported animal studies
- did not clearly describe their methods and results, or had discrepancies in data within the article.

MAIN SEARCH TERMS

MeSH headings (Medline subject headings): thermography/ breast neoplasms/

Additional keywords: thermography, thermometry, thermology, infrared, infra-red, breast adj3 cancer, breast adj3 malignan\$, breast adj3 carcinoma, breast adj3 neoplas\$,

Full details of the search strategies for the main sources are presented in **Appendix 3**.

SEARCH SOURCES

The NZHTA CORE Search was employed. Characteristics of the Core search include: essential sources only, major databases and secondary sources, and mostly published and indexed literature. For more detail about the search sources, refer to the NZHTA Search Protocol at <http://nzhta.chmeds.ac.nz/nzhtainfo/protocol.htm> Steps 1-9 (Core sections).

Bibliographic databases

- Medline
- Embase
- Current Contents
- Cinahl
- WebofScience (Science Citation Index)
- Cochrane Library Controlled Trials Register

Review databases

- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness (DARE)
- NHS Economic Evaluation Database (EED)
- Health Technology Assessment (HTA) Database
- ACP Journal Club

In addition to the sources above, the bibliographies of retrieved papers were scanned for references that had not been identified in the search. Experts in the field of breast screening, medical physics/bioengineering, radiology and infrared thermal imaging provided advice regarding available literature. Relevant websites, including those of manufacturers, were searched where possible. Authors were contacted to attain copies of publications identified by the search strategy that could not be retrieved through our usual sources.

An initial scoping search of Medline, Current Contents, and the review databases was done in November 2003. This search was updated on 4 to 5 March 2004, and a full search of the other sources undertaken on the same days.

Animal studies and correspondence were excluded in the search where this was possible within the limitations of the individual database. Because earlier research is still referred to in promotional material, to provide an overview for the client of the quality and relevance of all published research, the literature search for this evaluation was not limited by publication date or language. The overview is presented on **page 12**.

An update of the search was done on 19 and 20 May 2004. Databases searched were Current Contents, Pubmed (last 90 days), WebofScience, Medline and Premedline (Ovid version), and Embase. The CRD databases (DARE, NHS EED and HTA) and clinical trials websites were checked on 20 May 2004 for any additional information.

APPRAISAL METHODOLOGY

Summaries of appraisal results are shown in tabular form (known as *Evidence Tables*) which detail study aim and design, study setting, study comparators and reference standard, sample, methods, results, reported conclusions and NZHTA reviewer conclusions/comments based on the limitations and validity of the study.

The evidence presented in the selected studies were assessed and classified according to evidence levels adapted from Bandolier (<http://www.jr2.ox.ac.uk/bandolier/band70/b70-5.html>, accessed 22 June 2004) that categorise studies of diagnostic methods according to susceptibility for bias (**Appendix 1**).

ESTIMATION OF INDEX TEST RESULTS

The diagnostic test performance includes consideration of validity and reliability of the test (infrared thermography). Specifically, sensitivity, specificity, and positive and negative predictive values were calculated when possible to assess the validity of infrared thermography testing. These measures were calculated based on presentation of results as shown in **Table 1** below.

Table 1. Assessment of validity of a diagnostic test

		Reference test	
		Positive	Negative
Diagnostic test	Positive	a	b
	Negative	c	d
Total sample size		n ₁	n ₂

Based on **Table 1** above, measures of validity and 95 percent confidence intervals were calculated using the following formulae:

$$\begin{aligned} \text{Sensitivity} &= a/(a+c) \\ &= a/n_1 \end{aligned}$$

Confidence interval for sensitivity: $p \pm 1.96(pq/n_1)^{1/2}$

$$\text{Where } p = a/(a+c)$$

$$q = c/(a+c)$$

$$\begin{aligned} \text{Specificity} &= d/(b+d) \\ &= d/n_2 \end{aligned}$$

Confidence interval for specificity: $p \pm 1.96(pq/n_2)^{1/2}$

$$\text{Where } p = d/(b+d)$$

$$q = b/(b+d)$$

$$\text{Positive predictive value (PPV)} = a/(a+b)$$

Confidence interval for specificity: $p \pm 1.96(pq/n_2)^{1/2}$

$$\text{Where } p = a/(a+b)$$

$$q = b/(b+d)$$

$$\text{Negative predictive value (NPV)} = d/(c+d)$$

Confidence interval for specificity: $p \pm 1.96(pq/n_2)^{1/2}$

$$\begin{aligned} \text{Where } p &= d/(c+d) \\ q &= c/(b+d) \end{aligned}$$

If either $n \cdot p$ or $n(1-p)$ were less than five for sensitivity or specificity, confidence intervals based on the normal approximation to the binomial distribution using the formulae above were considered unreliable and exact methods based on the binomial distribution were used to calculate the confidence interval. Stata version 7.0 was used for these calculations (StataCorp, 2001).

RESULTS OF SEARCH STRATEGY

From the above search strategy, 1,154 potentially relevant articles/abstracts were identified of which 85 were retrieved.

An overview of the quality and relevance of all published research is provided by the following list of broad reasons for non-retrieval at the abstract stage (it should be noted that papers were often excluded for several reasons, and the following list is only indicative of the key reason for exclusion). These reasons were:

- non-English (485)
- non-thermographic technology (198)
- non-relevant thermography devices (this category included non-infrared thermography devices as well as technologically outdated infrared thermography devices) (215)
- articles relating aspects of breast cancer management other than screening or diagnosis (42)
- technological articles about physical aspects of devices or computational processes (51)
- conference proceedings (8)
- expert opinion commentary, a narrative review or a book chapter (12)
- discussion articles (including letters and editorials) (48)
- case presentations (3)
- animal studies (7).

Of the retrieved articles, 82 were excluded. These were excluded for the following reasons:

- non-thermographic technology (2)
- concerned infrared thermography devices that are no longer available (or that are not technologically comparable to those available in New Zealand) (3)
- concerned methods of thermography that were not infrared thermography, or did not specify the method of thermography under study (12)
- articles relating to aspects of breast cancer management other than screening or diagnosis (4)
- technological articles about physical aspects of devices or computational processes (12)
- conference proceedings (3)
- expert opinion commentary, a narrative review or a book chapter (33)
- discussion articles (including letters and editorials) (11)
- did not clearly describe methods and results (1)
- not retrieved within the review timeframe (1).

These papers, annotated with the reason for exclusion and a brief description of content, are listed in **Appendix 4**.

TOPIC 1: RESULTS RELEVANT FOR SCREENING

One retrieved article was eligible for review and was appraised for the screening topic (listed in **Appendix 5**). The included paper is presented in the evidence table below (see **Table 2, page 14**). There were no papers providing level one, two or three evidence. The study that was included and critically appraised was level four.

Table 2. Evidence table of appraised article relating to the use of infrared thermography as a screening tool

Authors Country Evidence Grading	Study Design, Index Test, Comparators and Reference Standard	Study Aim, Methods and Selection Criteria	Sample Characteristics	Results	Limitations and Conclusions
Williams et al. (1990) Bath, United Kingdom Grade: Level Four	<p>Study setting Breast screening clinic, Royal United Hospital, Bath.</p> <p>Study design Prospective cohort.</p> <p>Index test Thermography (TH) using either a system by Barr and Stroud® (digitally recorded on magnetic tape) or a system by Rank Precision Industries® (recorded on black and white film).</p> <p>TH result assessed by examining doctor and classified by them as positive or negative according to predetermined criteria. If a pre-menopausal woman had an abnormal TH result, she was invited back for a second TH test in second week of menstrual cycle.</p> <p>Other diagnostic tests Physical examination (PE) was performed on all participants, but PE findings were not specified. If either TH or PE was classified as positive at initial visit, then woman referred straight for mammography and other tests as deemed clinically required.</p> <p>Reference standard Not specified.</p>	<p>Study aim To determine the sensitivity and specificity of TH as a screening test for breast cancer, and to show whether or not it could be used to identify women at high-risk for developing the disease over five years.</p> <p>Methods Participants identified by two methods. (1) Women between 40-65 years invited to participate from the registers of six general practices (GPs). (2) Women volunteered to participate as a result of study publicity.</p> <p>Once enrolled, each woman gave medical history, underwent TH, and then underwent physical examination.</p> <p>Documentary follow-up conducted of each woman five years later through GP records to identify those who developed breast cancer.</p>	<p>Participants (1) From 8235 women invited from GP records, 52% accepted, n=4284 (2) Volunteers from publicity, n=5954. (Of these, n=229 were symptomatic).</p> <p>Total n = 10,238 (female)</p> <p>Age range 40-65 years. No other demographic details specified.</p> <p>Inclusion criteria None specified.</p> <p>Exclusion criteria None specified.</p>	<p>After first tests, missing data n=9. Results left n=10,229.</p> <p>59 women found to have breast cancer at time of tests. 19/59 were symptomatic (16 positive TH, 3 negative TH). Not specified how many symptomatic women without breast cancer had positive and negative TH.</p> <p>TH results: Positive, n=2681 (Breast cancer n=36, no breast cancer n=2645). Negative n=7548 (Breast cancer n=23, no breast cancer n=7525).</p> <p>(NB: From positive TH n=2681, n=2444 were negative on PE and M, but data not provided of actual cancer status for this n=2444).</p> <p>From this for TH: Se=61% (95% CI=49,73), Sp=74% (73, 75), PPV=1% (0, 2), NPV=99.7% (99.6, 99.8).</p> <p>Remove symptomatic women with breast cancer, approximate calculations for asymptomatic women for TH: Se=50% (35, 66), Sp=74% (73, 75), PPV=1% (0,1), NPV=99.7%, (99.6, 99.9).</p> <p>After 5 years, from n=10170, data from n=9819 (97%) available. Positive TH results found in 72% who developed breast cancer and in 73% of those who did not develop breast cancer.</p>	<p>Limitations Insufficient information on type of TH device used. Reviewer unable to source any further information about these devices. Unclear if all participants were tested by both TH devices or only one.</p> <p>Study population included both symptomatic and asymptomatic women, and not all data for these subgroups presented separately, so cannot extrapolate results to an asymptomatic screening situation.</p> <p>Large number of refusals to participate from GP approach introduces potential for recruitment bias. Potential for selection bias-selection criteria are not specified. Potential information bias due to lack of blinding.</p> <p>Possible verification bias as possibility that TH positive patients were followed more closely than those negative-this would favour the index test (TH).</p> <p>Authors' conclusions TH is not sufficiently sensitive to be used as a screening tool for breast cancer (nor as an indicator of risk).</p> <p>Reviewer conclusions Sensitivity and specificity estimates do not suggest the use of TH would be appropriate as a replacement test. It was not possible to evaluate the use of TH as an adjunctive test in this study. Methodological flaws in the design of this study are not clearly discounted in the report.</p>

Summary of Results Relevant to Screening

The study by Williams et al. (1990) was appraised for this topic and used infrared thermography devices for which technological details were not specified. It was also unable to allow for any comparison to be made between infrared thermography and any other screening modality, as either a stand-alone or adjunctive modality. However, it is the only study conducted since 1985 found as a result of the search strategy employed here that has examined the role of infrared thermography in a screening situation, and therefore is presented as the best available evidence.

Williams et al.'s (1990) study was graded as level four evidence because there was no masked reference standard applied. Several other methodological and reporting flaws detailed in **Table 2 (page 14)**, contributed to why this study did not provide adequate evidence relating to the use of infrared thermography in breast cancer screening. Of these, perhaps the most significant was that the population studied included both symptomatic and asymptomatic women, without presenting a breakdown of the data that allowed for these sub-groups to be considered separately, which limited the generalisability of these results to a population screening situation.

The study by Williams et al. (1990) employed a prospective cohort design. Leitch (2002) recently discussed the possible options for improvements in breast cancer screening. She concluded that scrutiny in prospective randomised controlled trials was essential for the benefit of screening mammography to be accepted, and that no other imaging technique has been sufficiently evaluated to replace mammography as the best available tool (Leitch, 2002).

The outcomes of particular interest in this review included measures of clinical effectiveness including PPV, test sensitivity (Se) and test specificity (Sp) – particularly relating to early (small) cancers, which are most treatable; and measures of cost effectiveness including health care and other costs. There was no evidence from the paper appraised to support the clinical effectiveness of infrared thermography as a screening tool for breast cancer. An abnormal result in a screening situation may lead to indefinite monitoring through mammography and breast palpation (Homer, 1985). Avoidable investigation of benign lesions (i.e., investigation of false positives) will drive up the cost of screening, cause unnecessary morbidity and anxiety for those women involved, and may also be used as a criticism of screening programmes (Feig, 1999). This review found no papers examining the costs of using currently available infrared thermography as a screening tool for breast cancer, and this is highlighted as a major gap in knowledge.

TOPIC 2: RESULTS RELEVANT FOR DIAGNOSIS

Two retrieved articles were appraised for the diagnosis topic (listed in **Appendix 5**). Included papers are presented in **Table 3 (pages 16-17)** in alphabetical order of authors' surname. There were no papers with level one, two, or three evidence. The studies that were included and critically appraised were both level four.

Table 3. Evidence table of appraised articles relating to the use of infrared thermography as an adjunctive diagnostic tool

Authors Country Evidence Grading	Study Design, Index Test, Comparators and Reference Standard	Study Aim, Methods and Selection Criteria	Sample characteristics	Results	Limitations and Conclusions
Keyserlingk et al. (1998) Montreal, Canada Grade: Level Four	<p>Study setting Ville Marie Breast and Oncology Center.</p> <p>Study design Case Control.</p> <p>Index test Infrared imaging (IR) (Bales Scientific®). High-resolution, scanning, electronically cooled system. Four images. Computer reading graded by examining physician. Final results categorised as 'normal' or 'abnormal'.</p> <p>Comparators Mammography (M). At least four standard-view images. Interpreted by examining physician and radiologist. Graded as 'suspicious', 'equivocal' or 'non-specific'.</p> <p>Physical examination (PE) Details not specified. Graded as 'suspicious', 'equivocal' or 'non-specific'.</p> <p>Reference standard Histological diagnosis (surgical biopsy).</p> <p>* This company has now been taken over by Computerized Thermal Imaging®.</p>	<p>Study aim To assess the potential contribution of currently available high-resolution digital IR as an adjuvant imaging technique in the early detection of breast cancer</p> <p>Methods Retrospective chart review from August 1995 onwards (end date not specified) to identify consecutive cases (post-operative patients having initial diagnosis of breast cancer, with final staging as either DCIS, stage I, or stage II) and controls (post-operative patients with benign breast histology following open surgical breast biopsy).</p> <p>Inclusion criteria Patient pre-operative evaluation included clinical exam, mammography and IR imaging.</p> <p>Definitive surgical management as first therapeutic modality.</p> <p>Exclusion criteria None specified.</p>	<p>Participants n = 200 (gender not clearly specified).</p> <p>Clinical reason for referral to breast centre not specified.</p> <p>Cases n=100 (from 128 charts reviewed) Age range 31-84 years (mean=53). Final staging of breast cancer DCIS (n=4), stage I (n=42) stage II (n=54). Mean tumour size 2.5cm.</p> <p>Controls n = 100 (from unknown number of charts reviewed) No further demographics provided for this group. No pathological diagnostic details supplied for this patient group.</p> <p>No statistical analyses presented comparing demographics of patient groups at baseline.</p>	<p>Cases PE (suspicious)=61% M (susp)=66% PE or M (suspicious)=83% IR (abnormal)=83% M (suspicious or equivocal)=85% IR (abn) or M (susp)=93% IR (abn) or M (susp or equiv)=95% IR (abn) or M (susp) or PE (susp)=98% Controls IR (abn)=19% M* ('abnormal')= 30%</p> <p>Independent calculations (by JK) of diagnostic accuracy: IR alone Se 83% (95%CI=76, 90) Sp=81% (73, 89) PPV=81% (74, 89) NPV=83% (75, 90).</p> <p>M* alone Se=85% (78,92) Sp=70% (61,78) PPV=74% (66, 82) NPV=82% (74, 90).</p> <p>Cases Incremental difference between M (66%) and IR +M (93%) Exact McNemars χ^2 value=27 (1d.f.), p<0.0001.</p> <p>Controls Comparing M* (30%) with IR (19%), Yates corrected χ^2 value=2.70 (1d.f.), p= 0.10.</p> <p>M*=Definition of 'abnormal' mammography assumed as suspicious + equivocal.</p>	<p>Limitations Unable to fully assess adjuvant diagnostic value of IR as insufficient information provided about results of IR, M and PE for control group. Specifically, there were insufficient data to compare IR + M* versus M* in the control group. Statistical analysis for IR alone versus M* for control group cannot exclude the role of chance.</p> <p>Potential selection bias. Method of selection for control group unclear. Potential interpretation bias, as interpretation of both the index test and the comparators were not blinded, and it is unclear what information was available to those interpreting the reference standard. Unclear whether there was a significant gap in time between IR and surgical biopsy for any patients.</p> <p>Authors' conclusions Although study results suggest that IR can provide safe, objective and practical additional information to both PE and M, the results should not be extrapolated outside controlled diagnostic environments without further, prospective, evaluation.</p> <p>Reviewer conclusions Although the sensitivity results present highly significant data suggesting an additive benefit of IR to M, there is inadequate data reported, specifically the proportion of false positive results, to fully confirm a beneficial role of IR as an adjuvant diagnostic modality for breast cancer.</p>

Table 3. Evidence table of appraised articles relating to the use of infrared thermography as an adjunctive diagnostic tool (continued)

Authors Country Evidence Grading	Study Design, Index Test, Comparators and Reference Standard	Study Aim, Methods and Selection Criteria	Sample characteristics	Results	Limitations and Conclusions
Parisky et al. (2003) USA Grade: Four	<p>Study setting Multicentre. Clinical sites in Los Angeles, Baltimore, Washington DC, Boston and Miami.</p> <p>Design Non-controlled clinical trial (analysed like a case-control).</p> <p>Index test Infrared imaging (IR). BSC 2100® system, temperature specificity < 0.05 °C. Dynamic imaging process (breast cooling). Mammogram used by physician evaluators to locate region(s) of interest on IR image. Each region of interest then analysed by computer algorithm to produce numeric score. Score assigned as positive or negative result from pre-determined scale.</p> <p>Other diagnostic tests All subjects had mammography (M). No technical details, nor result categories specified.</p> <p>All subjects had physical examination (PE). No details specified.</p> <p>45% subjects had ultrasound (US). No technical details, nor result categories given.</p> <p>Reference standard Histological diagnosis (core or surgical biopsy).</p>	<p>Study aim To determine the efficacy of a dynamic computerised IR imaging system for distinguishing between benign and malignant lesions in patients undergoing biopsy on the basis of mammographic findings.</p> <p>Methods Recruited patients for whom breast biopsy was recommended on basis of abnormal M, abnormal PE, or both. Patients then had IR imaging, then surgery. Each subject's IR images analysed by 3 evaluators (so 875 lesions produced 2625 IRI results) Evaluators were blinded to biopsy result, but knew certain PE and M details.</p> <p>Inclusion criteria None specified.</p> <p>Exclusion criteria Breast surgery in last year Breast implants Breast reduction surgery Radiation or histologically proven cancer in breast of interest Pregnancy Weight more than 135kg.</p>	<p>Participants Patients recruited n = 1293 (Exclusions n = 524) Reasons for exclusions Unacceptable IR (229), unacceptable/unavailable M (209), unable to assess a region of interest on IR because did not concur with M or PE findings (71), other (15).</p> <p>Patients whose data remained in study for evaluation n = 769</p> <p>Gender female (766), male (3). Age <40 years (68), 40-60 years (433), > 60 years (268). No age range, median or mean reported (although median age must have been between 40-60 years).</p> <p>Ethnicity white (463), black (207), hispanic (81), asian (13), other (5).</p> <p>From this group of 769 patients, 875 lesions were analysed (77 patients had two lesions biopsied, 13 had more than two biopsied).</p> <p>A subset analysis excluded lesions defined as microcalcifications on M. This subset therefore left for analysis 479 lesions from 448 patients.</p>	<p>Of the total 875 lesions analysed, 187 were malignant and 688 were benign. So, for these lesions, the PPV for standard work-up (M +/- PE, +/- US) is 21% (95%CI= 19,24).</p> <p>From 875 lesions 2625 IR results were reported. 326 IR results were then excluded (because evaluators could not concur IR area with M or PE area of suspicion), leaving 2299 IR results for analysis: Se 97% (95%CI =96, 99) Sp 14% (13, 16) PPV 24% (22, 26) NPV 95% (93, 98). From these 2299 results, the PPV for standard work-up can be calculated and is 22% (20, 23)*.</p> <p>The subset analysis of 479 lesions produced 1437 IR results, (206 excluded because of same reasons as above) leaving 1231 IR results: Se 100% (95%CI =99, 100), Sp 18% (16, 21), PPV 27% (25, 30), NPV 99% (98, 100). From these 1231 results, the PPV for standard work-up can be calculated, and is 24% (21, 26)*.</p> <p>Note: * Insufficient data was presented in this report to enable independent calculations of Se and Sp for standard work-up.</p>	<p>Limitations Study sponsored by manufacturers of BSC2100® (Computerized Thermal Imaging®).</p> <p>The possibility that random chance accounts for the additional value that IR adds to PPV cannot be excluded.</p> <p>Excluding participants with indeterminate results and the non-matching of M and IR leads to an overestimation of sensitivity and NPV.</p> <p>Inadequate description of patient M, PE and US methods presented. Cannot rule out inter-site variation in interpretation of these modalities as a confounding factor in pre-IR work-up.</p> <p>Unclear whether histological assessment was blinded.</p> <p>Authors' conclusions IR imaging has a high NPV, is noninvasive and safe, and therefore has an adjunctive role in determining whether immediate biopsy is warranted.</p> <p>Reviewer conclusions The results of this study suggest that unnecessary biopsies may be avoided by the addition of this IR method to standard work-up in lesions that are not microcalcifications on M. When all breast lesions detected by standard work-up are considered, there exists the unacceptable clinical situation that IR may classify 2.6% of malignancies as benign.</p>

Summary of Results Relevant to Diagnosis

Of the two studies appraised, one (Keyserlingk et al., 1998) considered the potential adjuvant benefit that could be gained from using infrared imaging in a multi-modality diagnostic setting, and used a retrospective case-control design. This study was classed as level four evidence because interpretation of both the index test and the comparators were not blinded, and it is unclear what information was available to those interpreting the reference standard. The other study (Parisky et al., 2003) considered the efficacy of infrared imaging when used specifically for the purpose of further clarifying, from routine non-invasive diagnostic modalities, whether to proceed to a more invasive test. This study used a clinical trial design. This study was also classed as level four, because it was unclear whether histological assessment was blinded. The other major flaws identified are detailed in **Table 3 (pages 16-17)** for this topic. The quality of evidence from many studies of diagnostic techniques has been questioned by authors such as Knottnerus et al. (2002), who commented that methodological flaws are common in diagnostic studies.

The retrospective case-control study by Keyserlingk et al. (1998) provided efficacy data that was highly suggestive that infrared thermography provided an adjuvant benefit. However, as the authors themselves cautioned this needs to be further investigated in a controlled, prospective manner. The non-controlled clinical trial by Parisky et al. (2003) suggested that for a subset of patients (those without microcalcifications on mammography), the addition of infrared imaging would indeed provide a valuable tool if used under the same conditions in clinical practice as it was in their trial. However, this finding cannot be extrapolated to all patients who undergo breast investigations. When the whole study population was considered, a major factor of potential clinical concern was that infrared imaging misdiagnosed 2.6% of breast malignancies as benign.

Thermograms need to be captured under controlled conditions, as they are sensitive to environmental changes in temperature, humidity, and air circulation. Such parameters must therefore be considered when choosing a venue for patient examination (Jones, 1998). The patient is usually required to rest for 10 to 20 minutes with clothing removed from the relevant area to achieve thermal equilibrium (Jones, 1998). In addition, on the day the test is performed, patients are ideally required to avoid other controllable factors that could potentially influence skin temperature such as alcohol consumption, physical exertion and application of cosmetic preparations (Ring and Ammer, 2000). Some interest has also focused on physiological factors such as hormonal fluctuations, and temperature fluctuations during the normal female menstrual cycle, as potential affecting the stability of thermographic measurement (Ng et al., 2002). Neither of the studies appraised for the diagnostic topic have reported that menstrual factors were considered, nor report the menstrual status of their participants, despite having in their participant populations women who may have been at differing stages in their reproductive lives.

A 1988 study by van Dam et al. illustrates the necessity for large sample sizes in studies that aim to detect incremental differences in diagnostic accuracy in a multi-modality situation (see **Table 4, page 48** in **Appendix 4** for a detailed overview of this study). Although their study used an infrared device that is no longer available, van Dam et al. (1988) aimed to evaluate the diagnostic characteristics of independently conducted physical examination, mammography, ultrasound scanning, infrared thermography, and various combinations of these tests in the evaluation of solid palpable breast masses. Their study included 201 patients, but was unable to show any statistically significant differences in effectiveness between any of the clinically relevant combinations of the diagnostic modalities examined.

To assess validity and reliability of diagnostic tests, the ideal study design is a cross-sectional survey in which both the new test and the reference standard are performed (Greenhalgh, 1997). However, to assess the usefulness of a diagnostic test in a situation of clinical decision-making ideally requires a prospective comparison between using and not using the test, for which the randomised controlled trial is the most robust method (Knottnerus et al., 2002). Therefore, one major flaw identified by this review is that there was no evidence from randomised controlled trials to support or refute the clinical effectiveness, benefits or harms that may occur from the adjuvant diagnostic use of infrared thermography in clinical decision-making. An equally important finding of this review is that there were no papers identified that examined economic or other costs of relevant infrared thermography devices in an adjuvant diagnostic setting.

CONCLUSIONS OF THIS REVIEW

A significant finding of the review conducted for the purpose of this Tech Brief was that there were no studies that evaluated the effectiveness of the infrared technology devices that are available in New Zealand. In addition, there were few that evaluated comparable infrared technologies, or technologies that may become available to New Zealand. This was despite using a systematic approach, which included a comprehensive search strategy that identified 1,154 abstracts. Research in this area is hindered by lack of consistency between manufacturers' device descriptions and specifications. The review found that much of the recent literature on infrared thermography is in the form of narrative review, discussion or opinion articles. Most of the published study reports on infrared thermography refer to studies of infrared devices that are outdated or no longer available, or non-infrared methods of thermography. No studies of this technology have been conducted in New Zealand. The evidence that is currently available does not provide enough support for the role of infrared thermography for either the population screening or adjuvant diagnostic testing of breast cancer. The major gaps in knowledge at this time can only be addressed by large-scale, prospective randomised trials.

Our conclusions are consistent with recommendations by key professional clinician groups. The Royal Australian and New Zealand College of Radiologists Breast Imaging Reference Group do not recommend the use of thermography for the early detection of breast cancer in their policy (2001). Similarly, the American Medical Association do not recommend it as a technique for diagnostic purposes, although they note that research protocols using thermography are continuing and data derived from these studies will require careful evaluation (American Medical Association, website accessed 20 May 2004).

More robust research on the effectiveness and costs of technologically advanced infrared thermography devices for population screening and diagnostic testing of breast cancer is needed, and the conclusions of this review should be revisited in the face of additional reliable evidence.

REFERENCES

- Anbar, M. (1995). Quantitative and dynamic telethermometry - A fresh look at clinical thermology. *IEEE Engineering in Medicine & Biology Magazine*, 14, 15-16.
- Anbar, M. (1998). Clinical thermal imaging today. *IEEE Engineering in Medicine and Biology Magazine*, 17, 25-33.
- Anbar, M., Milescu, L., Naumov, A., Brown, C., Button, T., Carty, C., & Al Dulaimi, K. (2001). Detection of cancerous breasts by dynamic area telethermometry. *IEEE Engineering in Medicine & Biology Magazine*, 20, 80-91.
- American Medical Association. (undated). Thermography update H-175.988: AMA Policy Finder. Available from: <http://www.ama-assn.org/> Accessed on 20.5.04.
- Arena, F., Barone, C., & DiCicco, T. (2003). Use of digital infrared imaging in enhanced breast cancer detection and monitoring of the clinical response to treatment. *IEEE EMBC*, 1129-1132.
- BreastScreen Aotearoa (2004). *National policy and quality standards for BreastScreen Aotearoa*. Wellington: Ministry of Health National Screening Unit.
- Dankiw, W. (1990). *Medical thermography. Health care technology series No. 4*. Canberra: Australian Institute of Health.
- Diakides, N.A. (1998). Infrared imaging: an emerging technology in medicine. *IEEE Engineering in Medicine & Biology Magazine*, 17, 17-18.

- Donegan, W. L. (2002). Diagnosis of breast cancer. In W. L. Donegan & J. S. Spratt (Eds.), *Cancer of the Breast* (pp. 311-345). Philadelphia: W B Saunders.
- Feig, S. A., Shaber, G. S., Schwartz, G. F., Patchefsky, A., Libshitz, H. I., Edeiken, J., Nerlinger, R., et al. (1977). Thermography, mammography, and clinical examination in breast cancer screening. Review of 16,000 studies. *Radiology*, 122, 123-127.
- Feig, S. A. (1999). Role and evaluation of mammography and other imaging methods for breast cancer detection, diagnosis, and staging. *Seminars in Nuclear Medicine*, 29, 3-15.
- Foster, K. R. (1998). Thermographic detection of breast cancer. *IEEE Engineering in Medicine & Biology Magazine*, 17, 10-14.
- Greenhalgh, T. (1997). How to read a paper: getting your bearings (deciding what the paper is about). *BMJ*, 315, 243-246.
- Gautherie, M., Haehnel, P., Walter, J. P., & Keith, L. G. (1987). Thermovascular changes associated with in situ and minimal breast cancers. Results of an ongoing prospective study after four years. *Journal of Reproductive Medicine*, 32, 833-842.
- Head, J. F., Wang, F., Lipari, C. A., & Elliott, R. L. (2000). The important role of infrared imaging in breast cancer - New technology improves applications in risk assessment, detection, diagnosis, and prognosis. *IEEE Engineering in Medicine and Biology Magazine*, 19, 52-57.
- Head, J.F., & Elliott, R.L. (2002). Infrared imaging: making progress in fulfilling its medical promise. *IEEE Engineering in Medicine & Biology Magazine*, 21, 80-85.
- Hider, P., & Nicholas, B. (1999). The early detection and diagnosis of breast cancer: a literature review - an update. *NZHTA Report*, 2, 1-150.
- Hobson, J. (2002). New techniques in medicine: thermal imaging. *Choice*, September, 60-61.
- Homer, M. J. (1985). Breast imaging: pitfalls, controversies, and some practical thoughts. *Radiologic Clinics of North America*, 23, 459-472.
- International Academy of Clinical Thermology. (undated). Mammography vs thermography. Available from: <http://www.iact-org.org/patients/breastthermography/mammography-vs-therm.html> Accessed on 20.3.04
- Jones, B. F. (1998). A reappraisal of the use of infrared thermal image analysis in medicine. *IEEE Transactions on Medical Imaging*, 17, 1019-1027.
- Keyserlingk, J. R., Ahlgren, P. D., Yu, E., & Belliveau, N. (1998). Infrared imaging of the breast: Initial reappraisal using high-resolution digital technology in 100 successive cases of stage I and II breast cancer. *Breast Journal*, 4, 245-251.
- Knottnerus, J. A., van Weel, C., & Muris, J. W. (2002). Evaluation of diagnostic procedures. *BMJ*, 324, 477-480.
- Leitch, A. M. (2002). Screening for breast cancer. In W. L. Donegan & J. S. Spratt (Eds.), *Cancer of the Breast* (pp. 269-286). Philadelphia: W B Saunders.
- Mahoney, M. C. (2002). Breast imaging: mammography, sonography, and emerging technology. In W. L. Donegan & J. S. Spratt (Eds.), *Cancer of the Breast* (pp. 287-310). Philadelphia: W B Saunders.
- Margolese, R. G., Cantin, J., Bouchard, F., Caines, J., Beaulieu, M. D., Little, C. D., Levine, M. N., et al. (1998). 1. The palpable breast lump: information and recommendations to assist decision-making when a breast lump is detected. *Canadian Medical Association Journal*, 158, S3-S8.

- Martin, J. E. (1983). Breast imaging techniques. Mammography, ultrasonography, computed tomography, thermography, and transillumination. *Radiologic Clinics of North America*, 21, 149-153.
- McLean, R. (1999). Thermography: technology brief. Montreal: Conseil d'Evaluation des Technologies de la Santé.
- Moskowitz, M. (1985). Thermography as a risk indicator of breast cancer. Results of a study and a review of the recent literature. *Journal of Reproductive Medicine*, 30, 451-459.
- Moskowitz, M. (1995). Breast imaging. In W. L. Donegan & J. S. Spratt (Eds.), *Cancer of the Breast* (pp. 206-239). Philadelphia, PA.: W B Saunders.
- Nass, S. J., Henderson, C., & Lashof, J. C. (2001). *Mammography and beyond: developing technologies for the early detection of breast cancer*. Washington, DC: National Academy Press.
- Ng, E. Y., Fok, S. C., Peh, Y. C., Ng, F. C., & Sim, L. S. (2002). Computerized detection of breast cancer with artificial intelligence and thermograms. *Journal of Medical Engineering & Technology*, 26, 152-157.
- Ohashi, Y., & Uchida, I. (2000). Applying dynamic thermography in the diagnosis of breast cancer - Techniques for improving sensitivity of breast thermography. *IEEE Engineering in Medicine and Biology Magazine*, 19, 42-51.
- Orel, S. G., & Troupin, R. H. (1993). Nonmammographic imaging of the breast: current issues and future prospects. *Seminars in Roentgenology*, 28, 231-241.
- Parisky, Y. R., Sardi, A., Hamm, R., Hughes, K., Esserman, L., Rust, S., & Callahan, K. (2003). Efficacy of computerized infrared imaging analysis to evaluate mammographically suspicious lesions. *AJR. American Journal of Roentgenology*, 180, 263-269.
- Qi, H., & Head, J. F. (2001). Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms. *Proceedings of the 23rd Annual International Conference IEEE Engineering in Medicine and Biology Society*, 2866-2869.
- Ring, E. F., & Ammer, K. (2000). The technique of infra red imaging in medicine. *Thermology International*, 10, 7-14.
- Royal Australian and New Zealand College of Radiologists Breast Imaging Reference Group. (2001). Policy on the use of thermography to detect breast cancer. Available from: http://www.ranzcr.edu.au/open/policies/diagnostic_imaging/pol7_3.htm Accessed on 25.6.04
- Royal New Zealand College of General Practitioners (1999). *Early detection of breast cancer: guidelines for primary care providers*. Wellington: The College.
- Snyder, W., Qi, H., Elliott, R. L., Head, J. F., & Wang, C. X. (2000). Increasing the effective resolution of thermal infrared images: an algorithm based on mean-field annealing that also removes noise and preserves image edges. *IEEE Engineering in Medicine & Biology*, 19, 63-70.
- StataCorp (2001). *Stata statistical software: release 7.0*. College Station, TX.: Stata Corporation.
- Sunshine, J. H., & Applegate, K. E. (2004). Technology assessment for radiologists. *Radiology*, 230, 309-314.
- van Dam, P. A., Van Goethem, M. L., Kersschot, E., Vervliet, J., Van den Veyver, I. B., De Schepper, A., & Buytaert, P. (1988). Palpable solid breast masses: retrospective single- and multimodality evaluation of 201 lesions. *Radiology*, 166, 435-439.
- Wharton, L. (2003). Thermography offers alternative in breast cancer detection. *National Business Review*, October 17, 56.

Williams, K. L., Phillips, B. H., Jones, P. A., Beaman, S. A., & Fleming, P. J. (1990). Thermography in screening for breast cancer. *Journal of Epidemiology & Community Health*, 44, 112-113.

Wilson, J. M., & Jungner, G. (1968). *Principles and practice of screening for disease*. Geneva: World Health Organisation.

APPENDIX 1: EVIDENCE LEVELS FOR DIAGNOSTIC TESTS

Susceptibility for bias in studies of diagnostic tests

*[Adapted from **Diagnostic Testing emerging from the gloom?***

Bandolier <http://www.jr2.ox.ac.uk/bandolier/band70/b70-5.html>]

- Level One An independent, masked comparison with reference standard among an appropriate population of consecutive patients.
- Level Two An independent, masked comparison with reference standard among non-consecutive patients or confined to a narrow population of study patients.
- Level Three An independent, masked comparison with an appropriate population of patients, but reference standard not applied to all study patients.
- Level Four Reference standard not applied independently or masked.
- Level Five Expert opinion with no explicit critical appraisal, based on physiology, bench research, or first principles.

APPENDIX 2: EXPLANATION OF SELECTION CRITERIA

Inclusion criteria

Publication Type

Studies published between 1985 and 20 May 2004 (inclusive) - These dates were chosen to identify studies that concerned the index test of interest to this review.

English language - Tech Briefs are rapidly produced assessments of the best available evidence. Because of time restrictions and the resultant limitation on resources for translation, it is usual to exclude non-English articles. Nevertheless, it should be noted that during the review of abstracts for this review, there were no significant non-English language studies identified that fulfilled all the other selection criteria.

Setting

Topic One - The rationale for this inclusion criteria for screening is explained in the Background section of this report, under the sub-heading “Thermography used for screening for breast cancer”.

Topic Two - The rationale for this inclusion criteria is explained in the Background section of this report, under the sub-heading “Thermography Used for Diagnosis of Breast Cancer”.

Population

Topic One - The aim of screening for any disease is to reduce mortality (and morbidity) from this disease by performing a test on an asymptomatic population, which is why asymptomatic women were the population of interest for this topic.

Topic Two - The role of a diagnostic test for any disease is to further evaluate abnormalities that have been pre-detected by either clinical symptoms, or screening tests, which is why patients symptomatic for breast cancer or who have had an abnormal mammogram were the population of interest for this topic.

Sample Size

Studies with small sample size are prone to sampling error. A lower limit of 20 is usual for critical reviews.

Index Test

The focus of this Tech Brief was to assess evidence relevant to infrared imaging devices that are (or are comparable with those) currently available, or which may become available in the future to patients in New Zealand. As outlined in the Background section under the heading “Technological History of Infrared Imaging”, infrared cameras that were available up until the mid-1980s suffered from several limitations that have been overcome since that time, and there have been accompanying improvements in computerised processing systems. This provided the rationale for selecting for appraisal only studies that have been published since 1985.

Comparators

Topic One - The rationale for this inclusion criteria for screening is explained in the Background section of this report, under the sub-heading “Thermography used for screening for breast cancer”.

Topic Two - The rationale for this inclusion criteria is explained in the Background section of this report, under the sub-heading “Thermography Used for Diagnosis of Breast Cancer”.

Index Test

Histological diagnosis is generally considered to be the “gold-standard” measure for detection of cancer.

Outcomes

The outcome measures chosen for this Tech Brief are standard for critical reviews of diagnostic modalities.

Exclusion Criteria

English language - As explained above for inclusion criteria

Concerned studies of infrared devices that are no longer available (or that are not technologically comparable to those available to those available in New Zealand) - As explained above for index test.

Concerned studies of methods of thermography that were not infrared thermography, or did not specify the method of thermography under study - The focus of this review was on infrared thermography devices, because that is the type of thermography that is available in New Zealand, therefore studies of other types of thermography were not relevant to this review. A realistic assessment of a study could not be made in the context of this review if the method of thermography used for the study was not specified.

Reported on the use of thermography in aspects of breast cancer management other than screening (Topic One) or diagnosis (Topic Two) - Screening and diagnosis were the aspects of breast cancer management under consideration in this review. Therefore, articles concerning thermography use in other aspects of breast cancer management, including technological articles about physical aspects of devices or computational processes, were not relevant to the aims of this review.

Had only an abstract available, or were conference proceedings - These types of articles contain limited data on the methods or results of studies.

Were an expert opinion commentary, a narrative review or a book chapter - These types of article are opinion articles that usually collect and report information in a non-systematic manner. They are therefore prone to selection bias and/or reporting bias, and furthermore frequently do not present data that can be used to assess effectiveness.

Were discussion articles (e.g., presented as a letter or editorial) - These types of article are opinion articles that usually collect and report information in a non-systematic manner. They are therefore prone to selection bias and/or reporting bias, and furthermore frequently do not present data that can be used to assess effectiveness.

Reported studies with fewer than 20 participants or were a case presentation - Studies with small sample size are prone to sampling error. A lower limit of 20 is usual for critical reviews.

Reported animal studies - Humans were the focus of this review, therefore animal studies were not considered relevant.

Did not clearly describe their methods or results, or had discrepancies in data within the article - Such articles do not provide reliable data that can be used to assess effectiveness using rigorous critical appraisal techniques.

APPENDIX 3: SEARCH STRATEGY

Medline & Cochrane Controlled Trials Register

- 1 thermography/
- 2 thermal imaging.mp.
- 3 (infrared imag\$ or infra-red imag\$).mp.
- 4 or/1-3
- 5 breast neoplasms/
- 6 breast cancer.ti,ab.
- 7 5 or 6
- 8 4 and 7
- 9 letter.pt.
- 10 8 not 9
- 11 animal/
- 12 animal/ and human/
- 13 11 not 12
- 14 10 not 13
- 15 (thermology or thermometry or thermography).tw.
- 16 (breast adj3 carcinoma).mp.
- 17 (breast adj3 malignan\$).tw.
- 18 (breast adj3 neoplas\$).tw.
- 19 (breast adj3 cancer\$).tw.
- 20 or/16-19
- 21 15 and 7
- 22 (4 or 15) and 20
- 23 14 or 22

Embase

- 1 thermography/ or thermometry/
- 2 (thermal imaging or thermography or thermometry or thermology).tw.
- 3 (infra-red imag\$ or infrared imag\$).mp.
- 4 or/1-3
- 5 Breast Cancer/
- 6 (breast adj3 cancer).tw.
- 7 (breast adj3 malignan\$).tw.
- 8 (breast adj3 carcino\$).mp.
- 9 (breast adj3 adenocarcino\$).mp.
- 10 or/5-9
- 11 cancer screening/
- 12 4 and 10
- 13 4 and 11
- 14 13 not 12
- 15 12 or 13
- 16 letter.pt.
- 17 15 not 16

Cinahl & Amed

- 1 thermography/
- 2 thermal imaging.mp.
- 3 (infrared imag\$ or infra-red imag\$).mp.
- 4 (thermology or thermometry or themography).tw.
- 5 (breast adj3 carcinoma).mp.
- 6 (breast adj3 malignan\$).tw.
- 7 (breast adj3 neoplas\$).tw.
- 8 (breast adj3 cancer\$).tw

- 9 (breast adj3 carcino\$).mp.
- 10 (breast adj3 adenocarcino\$).mp.
- 11 Breast Neoplasms/
- 12 or/1-4
- 13 or/5-11
- 14 12 and 13

Current Contents/Science Citation Index

- 1 thermometry OR thermal imaging OR thermology OR thermography
- 2 (infrared OR infra-red) SAME imag*
- 3 #1 OR #2
- 4 breast SAME cancer
- 5 Breast SAME carcinoma
- 6 breast SAME malignanc*
- 7 breast SAME adenocarcin*
- 8 breast SAME neoplas*
- 9 #4 OR #5 OR #6 OR #7 OR #8
- 10 #3 AND #9

The vocabulary used in the strategies above was adapted for use in remaining sources.

APPENDIX 4: EXCLUDED RETRIEVED PAPERS

Note: The summaries of the excluded retrieved papers in this appendix are not the result of a critical appraisal. Any conclusions presented in these summaries are those of the author(s) of the individual paper only, and not the author of this report.

Anbar, M. (1995). Quantitative and dynamic telethermometry - A fresh look at clinical thermology. *IEEE Engineering in Medicine & Biology Magazine*, 14, 15-16.

This article was excluded because it was an expert opinion article. This short article briefly outlines the academic history of clinical thermal imaging, explaining that a scarcity of academic research in this field has (understandably) meant that the merits of clinical thermology have not been firmly established by systematic academic studies. The author opines that it is essential to understand the mechanisms that underlie thermal changes in breast tissue, and that this understanding will enable the efficacy of diagnostic testing by this modality to be optimised. The author states that reliable, modern, and fast, computerised equipment must be used and studied to achieve this aim.

Anbar, M. (1998). Clinical thermal imaging today. *IEEE Engineering in Medicine and Biology Magazine*, 17, 25-33.

This article was excluded because it was a narrative review. It outlines recent technological advances in infrared equipment and also discusses current understandings of the physiological processes that underlie thermal changes in the breast and other tissues.

Anbar, M. (2002). Assessment of physiologic and pathologic radiative heat dissipation using dynamic infrared imaging. *Annals of the New York Academy of Sciences*, 972, 111-118.

This report was excluded due a lack of clear description of methods and results. Most of this article is a narrative review that outlines in detail the physiological processes behind skin temperature regulation. The author explains that because the skin's temperature modulation is so complex, it can only be accurately measured using highly sensitive dynamic infrared detectors. These allow for measurement across multiple frequencies, and additionally for computer-generated procedures to objectively assess localised attenuations of temperature modulation. The local attenuations signal possible underlying pathology. The author then presents some graphical results from an objective computer-generated assessment of different breast lesions. The methods used to attain the graphical results were not clearly described, including study design, sample selection and characteristics, there was no comparator diagnostic modality used, and no numerical results presented in this report. The author concludes that this type of computer-assisted interpretation has potential usefulness in breast cancer diagnosis.

Anbar, M., Brown, C.A., & Milescu, L. (1999). Objective identification of cancerous breasts by dynamic area telethermometry (DAT). *Thermology International*, 9, 137-143.

This paper was requested by electronic mail from the first author, after the failure to attain the journal through all our usual library sources, but it was not retrieved within the timeframe (several months). The paper is likely to have presented a study report concerning the development of an objective diagnostic parameter using infrared imaging techniques and quantitative computational processes.

Anbar, M., Brown, C., Milescu, L., Babalola, J., & Gentner, L. (2000). The potential of dynamic area telethermometry in assessing breast cancer. *IEEE Engineering in Medicine & Biology Magazine*, 19, 58-62.

An article describing the mechanisms of using 'dynamic' IR imaging techniques, where highly sensitive and fast responding IR detectors are coupled with quantitative computer analysis to measure over time areas of abnormal modulation of skin temperature. This article was excluded because it mainly described technological aspects of this process, and although it presented some clinical examples of use, these were in case report form. The authors describe how the use of this technique is based on the hypothesis that abnormal secretion of nitrous oxide (NO) by breast cancer cells affects the usual vascular control of blood vessel flow, and therefore identifying the areas of abnormal blood flow locates abnormal or malignant breast tissue.

Anbar, M., Milescu, L., Naumov, A., Brown, C., Button, T., Carty, C., & Al Dulaimi, K. (2001). Detection of cancerous breasts by dynamic area telethermometry. *IEEE Engineering in Medicine & Biology Magazine*, 20, 80-91.

Report from a study that was undertaken using a high-speed infrared detector to develop a single diagnostic parameter that distinguishes benign from malignant breast lesions. This paper was excluded because it focuses on reporting in detail the various computational and statistical processes that were undertaken to develop this single objective diagnostic parameter. Furthermore, there was insufficient data presented to estimate any of the pre-specified outcomes of interest for this review. The study was based on the assumption that breast cancer cells secrete nitrous oxide, that the nitrous oxide attenuates the normal vascular control of the breast tissue, and that this produces minute changes in skin temperature over time. To develop this diagnostic parameter, a total of 100 women (with presumed benign breast lesions, or in-situ breast cancer, or invasive breast cancer on excisional biopsy) were imaged. The authors conclude that by allowing a computer to assess the infrared image, this type of dynamic imaging can yield meaningful diagnostic information, and that their method warrants testing in clinical trial situations to determine its scope and limitations.

Anonymous (1985). The use of diagnostic tests for screening and evaluating breast lesions. Health and Public Policy Committee, American College of Physicians. *Annals of Internal Medicine*, 103, 143-146.

This article was excluded because it was a narrative review article. It presents an author's assessment of breast self-examination, mammography, thermography and other techniques for the screening of asymptomatic women and an assessment of mammography, thermography, ultrasound and fine-needle aspiration for the diagnostic evaluation of breast masses. For each modality, a description is presented followed by the author's assessment of efficacy, safety (where applicable) and cost. For screening, thermography is assessed as safe, but of insufficient accuracy, with cost difficult to assess. For the diagnosis of breast lumps, this author concluded that thermography was still generally not helpful for the evaluation of breast masses. The author's final conclusions are that mammography is a necessary part of breast screening, although ages for screening are still controversial, and that for the assessment of breast masses mammography and/or ultrasound plus fine-needle aspiration provide sufficient information in most cases.

Anonymous (1986). Diagnosis and first-line treatment of breast cancer in Italian general hospitals. G.I.V.I.O. (Interdisciplinary Group for Cancer Care Evaluation) - Italy. *Tumori*, 72, 273-283.

This report was excluded because it was a descriptive report that outlines and discusses the quality of diagnostic procedures and treatment processes used in the management of women with breast cancer at 63 Italian hospitals between March 1983 and April 1984. It is mainly concerned with breast cancer treatment issues.

Arena, F., Barone, C., & DiCicco, T. (2003). Use of digital infrared imaging in enhanced breast cancer detection and monitoring of the clinical response to treatment. *IEEE EMBC*, 1129-1132.

Report from a study that had the aim of demonstrating that an automated computer generated algorithm can be developed to predict likely risk of breast cancer when advanced computer hardware and software is used in conjunction with advanced high-resolution digital infrared technology. This study was excluded from appraisal because it was a report from conference proceedings. It lacked a clear description of methods and results. The researchers in the study pre-selected three patient populations on basis of: assumed normal; known biopsy results of wide variety of cancer; or previous cancer treatment. Although the report details the mechanism and nature of the IR examinations undergone by the study population, there is little information given concerning the results of any other investigations that the study population may have undergone.

Barraclough, B. M., Picker, R. H., & Barraclough, B. H. (1985). Toward breast screening. *Australasian Radiology*, 29, 181-184.

Two-part journal article. The first part discusses potential breast screening methods, including thermography. The second part presents an overview of the results from a small study in which 100 female patients underwent ultrasound scanning as part of their breast examinations in a hospital in Sydney, Australia, and was excluded from appraisal because it only provided information about ultrasound scanning in these patients. The authors surmised that ultrasound scanning should be considered for inclusion in a breast-screening programme.

Blume, S. S. (1993). Social process and the assessment of a new imaging technique. *International Journal of Technology Assessment in Health Care*, 9, 335-345.

This article was excluded because it was a narrative review. It presents an historical review of the sociological processes that have occurred since the initial discovery of infrared thermography, tracing the various changes in medical, industrial, and thermography proponents' opinions and attitudes to the technology since 1955. Concludes that the history of thermography shows how the process of continuous assessment of medical technologies is dominated by certain values, priorities and assumptions. Not directly relevant to the topic, focussing on sociological issues not scientific evidence.

Cady, B., Steele, G. D., Jr., Morrow, M., Gardner, B., Smith, B. L., Lee, N. C., Lawson, H. W., et al., (1998). Evaluation of common breast problems: guidance for primary care providers. *Ca: a Cancer Journal for Clinicians*, 48, 49-63.

This report was excluded because it was a narrative review/expert opinion article that describes an organised approach for the management of women presenting to primary care providers with breast symptoms or a positive screening test. It mentions thermography only to state that it has no current place in breast cancer screening or diagnostic evaluation.

Ciatto, S., Palli, D., Rosselli del Turco, M., & Catarzi, S. (1987a). Diagnostic and prognostic role of infrared thermography. *Radiologia Medical*, 74, 312-315.

Report from an Italian study that used an older infrared device that is no longer available (AGA Thermovision 680). This study was excluded from appraisal because it used older equipment that is no longer obtainable, and for which device specifications were not easily available. The study had two aims, firstly to evaluate the diagnostic accuracy of infrared thermography (TH) compared with physical examination (PE) and mammography (M), and secondly to determine whether thermographic patterns are related to prognostic outcome. Between 1976 and 1983, 6,832 women attending a specialist oncology centre were examined by PE and M, and were selected to have TH on a "clinical

basis without well defined criteria" (p 312). From this population, 469 women with breast cancer were identified, with their diagnoses confirmed by either cytology or histology. The authors found that the sensitivity of TH as a diagnostic method in their study was 47 percent overall, for M was 89 percent and for PE was 86 percent, and furthermore that combining TH with PE or M did not appear to increase the combined sensitivity. These authors concluded that their study results did not encourage the use of infrared thermography in either diagnostic or prognostic processes.

Ciatto, S., Rosselli Del Turco, M., Cecchini, S., Grazzini, G., & Iossa, A. (1989). Telethermography and breast cancer risk prediction. *Tumori*, 75, 110-112.

Report from a study that aimed to assess the association between different infrared thermographic patterns and likelihood of developing breast cancer. This paper was excluded from appraisal because it was concerned with risk factor identification, did not provide evidence that earlier detection of breast cancer occurred as a result of utilising thermography in this manner, and the type of infrared thermography device was not reported. The study used a retrospective chart review method to identify patients who had had certain thermographic patterns. Patient histories were then reviewed to ascertain their subsequent breast health. As a result of this study, the authors concluded that they had found no association between abnormal thermographic patterns and risk of developing breast cancer.

Ciatto, S., Smith, A. H., Di Maggio, C., Pescarini, L., Lattanzio, E., Ancona, A., Punzo, C., et al. (1987b). Breast cancer diagnosis under the age of forty years. *Tumori*, 73, 457-461.

Report from an Italian study that aimed to assess the sensitivity of single tests, and combinations of tests, for the diagnosis of breast cancer in women under the age of 40 years. This study was excluded because it did not specify the method of thermography used. The study was conducted in 1984, and the method employed was a retrospective review of case records of women aged 39 years and under with breast cancer from six Italian institutions. The time period varied between centres, but went as far back as 15 years prior to 1984 in some centres. The authors report the rates of histologically-positive cases that were correctly diagnosed by each modality, and were only able to determine from their results that physical examination was the test of choice for women in this study (although it was only positive in 77% of all the cancers). The authors conclude that physical examination cannot be used alone, that their study was unable to define the best test(s) to be combined with physical examination, but that thermography could probably be ruled out from the poor results that they had found in their preliminary diagnostic calculations.

Cotton, P. (1992). AMA's Council on Scientific Affairs takes a fresh look at thermography. *JAMA-Journal of the American Medical Association*, 267, 1885-1887.

This paper was excluded because it was a discussion article, which presents a brief review of the applications of thermography in the USA, including issues of health insurance reimbursement.

Cuttino, J. T., Jr. (1987). Breast imaging techniques. Current status. *North Carolina Medical Journal*, 48, 303-304.

This paper was excluded because it was a narrative review article. It is a short article that describes the various diagnostic modalities available to image the breast. The author concludes that, at that point in time, mammography is the best modality for breast screening, that ultrasound is useful as an adjunctive diagnostic modality but not for screening, that CT and MRI are still at developmental stages, and that thermography has no role at all in either screening or diagnostic situations.

de Thibault de Boesinghe, L. (1990). The value of thermography for the diagnosis, prognosis and surveillance of non-palpable breast cancer. *Journal Belge de Radiologie*, 73, 375-378.

This paper was excluded because it was a narrative review article. It briefly outlines the history and technique of thermography, and the interpretation of infrared thermograms. The author concludes with favourable comments concerning the role of infrared thermography as a technique to be combined with other available imaging modalities for the screening and diagnosis of breast cancer.

Diakides, N. A. (1997). New developments in low cost IR imaging systems. *European Journal of Thermology*, 7, 213-215.

This paper was excluded because it was a narrative review article. It outlines developments in infrared technologies, their applications and their future potentials. The author states that there is potential for clinical research to determine the benefits of advanced thermal imaging in comparison to other imaging modalities.

Diakides, N.A. (1998). Infrared imaging: an emerging technology in medicine. *IEEE Engineering in Medicine & Biology Magazine*, 17, 17-18.

This paper was excluded because it was an editorial article. It introduces a journal issue that is dedicated to the subject of the application of new technologies in infrared imaging that have emerged since the early 1990s.

Donnelly, L. F., Freundlich, I. M., Wallace, J. D., & Dodd, G. D. (2000). Thermography and the venous diameter ratio in the detection of the non-palpable breast carcinoma. 1968. *AJR. American Journal of Roentgenology*, 174, 1092.

This paper was excluded because it was an historical pictorial article that shows a reproduction of a 1968 journal illustration.

Feig, S. A. (1999). Role and evaluation of mammography and other imaging methods for breast cancer detection, diagnosis, and staging. *Seminars in Nuclear Medicine*, 29, 3-15.

This paper was excluded because it was a narrative review/expert opinion article. It outlines the requirements that the author thinks that any new supplementary imaging modality would have to meet to provide additional benefits over mammography alone in breast cancer screening, or over mammography plus the standard additional radiological techniques in breast cancer diagnosis.

Foster, K. R. (1998). Thermographic detection of breast cancer. *IEEE Engineering in Medicine & Biology Magazine*, 17, 10-14.

This paper was excluded because it was a narrative review article. It outlines and discusses the history and major studies of thermography and mammography as techniques for breast cancer screening and diagnosis. The author stresses that it is important that sufficient evidence is gained from extensive clinical studies before a test is adopted as a tool for the widespread screening of asymptomatic populations.

Gautherie, M., Haehnel, P., Walter, J. P., & Keith, L. G. (1987). Thermovascular changes associated with in situ and minimal breast cancers. Results of an ongoing prospective study after four years. *Journal of Reproductive Medicine*, 32, 833-842.

A report from a study that aimed to assess the thermal and vascular disorders associated with the early stages of breast malignancy, with the aim of identifying those women who may be a higher risk for breast cancer. This study was excluded because it used both contact thermography and infrared thermography methods, and the data presented in this report did not allow for a distinction to be made between these thermographic methods. Furthermore, the study focussed on thermography patterns as being a risk factor for breast cancer, and the authors state that their study was not designed to assess the sensitivity and specificity of thermography versus other diagnostic methods. A cohort of 25,782 women, which included both asymptomatic women (some of whom had a positive family history of breast cancer) and women who had presented with a variety of breast symptoms, was followed for four years. At first visit, all women were examined by thermography (76% by both contact and infrared methods), then had mammography, and physical examination. The authors' report that 1.2 percent of the asymptomatic women in their study were found to have breast cancer, which was higher than they would have expected in an asymptomatic population. The latter part of this paper discusses hypotheses, which were current at that time, to explain why breast tumours generate heat.

Ghys, R. (1987). Cancer diagnosis in a "breast clinic": a 15 years retrospection. *Journal of Nuclear Medicine & Allied Sciences*, 31, 315-320.

A report from a study that aimed to assess the independent diagnostic accuracies of the techniques of clinical examination, mammography, thermography, anamnesis and diaphanoscopy, as well as the adjunctive value of the imaging modalities when added to clinical diagnosis for breast cancer. The techniques of anamnesis and diaphanoscopy were not explained in the report. This study was excluded because it used old thermographic equipment that is no longer easily available, and for which technical specifications are not clearly reported. The authors conclude from their study that the combined use of diaphanoscopy, thermography and mammography provided the best detection of breast cancers that were not clinically obvious.

Gold, R. H., Bassett, L. W., & Kimme-Smith, C. (1986). Breast imaging. State-of-the-art. *Investigative Radiology*, 21, 298-304.

This paper was excluded because it was a narrative review article of various breast imaging procedures. It includes a short paragraph about thermography, in which the authors state that thermography cannot reliably differentiate benign from malignant breast disease.

Grenn, M. W. (1996). Recent advances in portable infrared imaging systems Amsterdam. Proc 18th International Conference of the IEEE Engineering in Medicine and Biology Society. Paper 1091, 2083-2084.

This paper was excluded because it was the results of a conference proceedings. It is a short descriptive article that details technical specifications of portable infrared imaging systems that developed over the five years prior to that time.

Harrison, B., & Mabuchi, K. (1998). Biomedical infrared imaging in Japan. *IEEE Engineering in Medicine & Biology Magazine*, 17, 66-70.

This paper was excluded because it was a discussion article about the status and use of infrared imaging in Japan. It notes particularly the cultural and language barriers that are present for Japanese researchers in this field.

Head, J. F., & Elliott, R. L. (1997). Thermography. Its relation to pathologic characteristics, vascularity, proliferation rate, and survival of patients with invasive ductal carcinoma of the breast. *Cancer*, 79, 186-188.

This paper was excluded because it was a journal correspondence article. It discusses the results from a study in which the authors had used infrared thermography to assess disease characteristics and prognosis for patients with breast cancer.

Head, J.F., & Elliott, R.L. (2002). Infrared imaging: making progress in fulfilling its medical promise. *IEEE Engineering in Medicine & Biology Magazine*, 21, 80-85.

This paper was excluded because it was a narrative review/expert opinion article. It discusses the background to, and current and potential applications of, infrared imaging in medicine. The authors' opinion is that thermography has not yet received the attention it merits as a potential part of multi-modal screening for breast cancer. The authors support thermography as a risk assessment and prognostic tool for breast cancer. The article states that current issues in medical applications of IR technologies revolve around poor quality IR images, which require visual (rather than quantitative) analysis, and therefore require training and agreed diagnostic standards. The paper outlines differences between different generations of IR technologies, and upcoming developments. The authors' opinion is that early instruments were not sensitive enough to detect the subtle changes in temperature required to accurately detect and monitor disease. The authors state that Dynamic Area telethermometry (DAT) (developed by Anbar et al., 2000) may be the most promising new development in infrared imaging.

Head, J. F., & Lipari, C. A. (1999). Comparison of mammography and breast infrared imaging: sensitivity, specificity, false negatives, false positives, positive predictive value and negative predictive value. *Proceeding of the 21st Annual International Conference IEEE Engineering in Medicine and Biology Society*, 1116.

This paper was excluded because it was proceedings from a conference. The authors report having undertaken a literature search to identify papers for a literature review, but do not clearly specify the methods and selection criteria that they employed to search and identify papers for their review. The literature review was followed by an averaging of the performance characteristics of mammography from seven papers (reporting from various different diagnostic settings), and a comparison with the averaged results of the performance characteristics of infrared imaging derived from one study of infrared imaging in a screening setting. The authors conclude from their review that infrared imaging has performance characteristics similar to mammography for the detection of breast cancer.

Head, J. F., Wang, F., & Elliott, R. L. (1993). Breast thermography is a noninvasive prognostic procedure that predicts tumor growth rate in breast cancer patients. *Annals of the New York Academy of Sciences*, 698, 153-158.

A report from a two-part study that used data from a specialist breast clinic to investigate the prognostic accuracy of breast thermography. This study was excluded because it examined the role of thermography in predicting prognosis in breast cancer patients. The first part of the study compared previously acquired thermographic results from 126 deceased breast cancer patients, 100 living breast cancer patients and 100 non-cancer patients, finding that a significantly higher percentage of the deceased group had 'abnormal' thermograms. No indication was given of whether the results of the thermographic interpretation were taken from the original examination, or whether the original thermograms were re-reviewed for the purposes of this study. No information is given as to the results of any other clinical or radiological investigations that these patients may have undergone. The authors conclude that this study demonstrates the prognostic significance of thermography. The second part of this study compared the thermographic results from an unspecified number of breast cancer patients to other known/suspected prognostic factors, finding that only factors relating to tumour size and growth rate were significantly related to thermographic results. The authors conclude that this suggests that breast cancer patients with abnormal thermograms have faster growing, more aggressive tumours.

Head, J. F., Wang, F., Lipari, C. A., & Elliott, R. L. (2000). The important role of infrared imaging in breast cancer - New technology improves applications in risk assessment, detection, diagnosis, and prognosis. *IEEE Engineering in Medicine and Biology Magazine*, 19, 52-57.

A paper reporting from authors who undertook two separate studies to assess the use of infrared (IR) imaging in breast cancer management. The first study aimed to assess risk assessment, detection and prognosis, and used a retrospective case-control design to review IR results of patients who had attended a specialist breast centre since 1973. This study was excluded because the type of IR device(s) used was unspecified, but in addition it focuses mainly on the use of IR in predicting prognosis in breast cancer patients. The selection criteria for this study were not specified, but three groups of patients were studied: 126 who had died of breast cancer since 1973 (who had had IR imaging within one-year of breast cancer diagnosis); 100 randomly selected living 'breast cancer' patients (who also had had IR imaging within one-year of breast cancer diagnosis); 100 patients with a variety of breast conditions who had not been diagnosed as having breast cancer (who then underwent IR imaging and were classified as high risk if certain IR patterns were seen). No other clinical information, other imaging modality results, nor pathological confirmation for breast disease status is specified in this report. When the IR results between the patient groups are presented, the deceased breast cancer group had 88 percent abnormal IR results, the living breast cancer group had 65 percent abnormal IR results, and the other breast disease group had 28 percent abnormal results ($p < 0.0001$). In addition, when the size of 74 breast cancer tumours was considered in relation to IR results, there was a statistically significant association between increased tumour size and likelihood of an abnormal IR result (it is noted that, of these tumours, 47 percent that were <2cm in size had a normal IR result). The authors surmised from this that an increased incidence of abnormal IR results is related to the likelihood of progression of disease. Given that few details of study methods and patients' clinical status are specified, the validity of the authors' conclusion is difficult to confirm.

The second study reported by these authors used a cross-sectional method to compare the quality of IR images produced by two modern IR detectors. This study was excluded because it concerned mainly technical aspects of infrared detector systems. One device had a scanning IR detector from which analysed IR images were categorised as normal, slightly abnormal, or abnormal. The other device had a focal plane array detector from which IR images were categorised quantitatively into one of eight numerical categories. The authors reported that the focal plane array detector produced better quality images than the scanning detector. Images were taken using both detectors from a group of 220 patients, whose clinical status was not specified in the report other than that the group consisted of women who were both 'normal' and 'high-risk'. A physician had categorised the patients as 'high-risk' or 'normal' by subjectively assessing their past medical history. The report does not specify any further clinical details for these patients, nor results from any other diagnostic modality that these

patients may have undergone. The authors compared the IR results that were classified as possibly abnormal to the risk categorisations from the medical histories, but found no correlation. They surmised from this that the IR categories that represented abnormal thermal patterns were therefore an independent risk factor for breast cancer. It is difficult to accept this conclusion, given the lack of clinical data provided and the fact that these patients were not followed prospectively to determine if they then developed breast cancer.

At the end of this paper, the authors conclude that the role of IR imaging in breast cancer risk assessment, detection, diagnosis and prognosis has not yet been determined, and also that IR imaging can only be used to complement mammography and physical exam in the detection and diagnosis of breast cancer.

Hearn, K. (1991). Choosing the right tools. *Radiologic Technology*, 62, 240-242.

This paper was excluded because it was a narrative article. It is a brief paper, written by a radiography student, which describes the different techniques of clinical thermography. The author suggests that thermography may be a useful breast imaging modality in combination with other screening methods.

Heywang-Kobrunner, S. H. (1992). Nonmammographic breast imaging techniques. *Current Opinion in Radiology*, 4, 146-154.

This paper was excluded because it was a narrative review article. It discusses and compares the potential of available non-mammographic imaging techniques to improve diagnostic capabilities in the early detection of breast malignancy. The author considers that there is no role for thermography in breast screening, diagnosis, prognostic prediction, or detection of recurrence. This paper has an annotated bibliography.

Homer, M. J. (1985). Breast imaging: pitfalls, controversies, and some practical thoughts. *Radiologic Clinics of North America*, 23, 459-472.

This paper was excluded because it was a narrative review/expert opinion article. It covers the topic of breast imaging. The first section discusses the benefits and controversies of mammography, and the second section discusses other imaging modalities that may be of use for breast imaging. A subsection on thermography states that the role of thermography in the evaluation of breast disease is unclear. The author concludes that it seems reasonable to await further scientific investigations to establish the correct place of thermography, before it can be used with confidence for either screening or diagnostic breast imaging.

Isard, H. J. (1985). Thermography for breast-cancer diagnosis. *JAMA-Journal of the American Medical Association*, 253, 41-42.

This paper was excluded because it was a journal correspondence article (letter to the editor). The author expresses an opinion that the American College of Radiology has dismissed too lightly the potential benefit of thermography as an adjunct to other breast cancer screening and diagnostic modalities.

Isard, H.J. (1992). Thermography in breast-cancer. *JAMA-Journal of the American Medical Association*, 268, 3074-3074.

This paper was excluded because it was a journal correspondence article (letter to the editor). The author expresses an opinion that the discussion article by Cotton (1992) published in a previous edition of JAMA presented an unbalanced view of the usefulness of thermography. The author points out that a 1972 study that followed 10,000 women for four years found that from 5,662 symptomatic patients, 270 had proven breast cancer, and that for these symptomatic women thermography had a true-positive rate of 72 percent. The author further states that breast thermography may have some use as a prognostic indicator for cancer survival. This letter from Isard is followed by a reply from Cotton, who justifies his previous view of thermography on the basis that the findings from the 1972 study had not been replicated in several subsequent attempts. Cotton then agrees that thermography may well be useful as a prognostic indicator for breast cancer survival, but points out that there was as yet insufficient evidence for use in this role outside controlled clinical trials.

Jackson, V. P., Hendrick, R. E., Feig, S. A., & Kopans, D. B. (1993). Imaging of the radiographically dense breast. *Radiology*, 188, 297-301.

This paper was excluded because it was a narrative review/expert opinion article. It discusses the issues and alternative options for imaging mammographically dense breasts. The authors do not consider thermography as providing potential assistance in this clinical situation.

Jones, B. F. (1998). A reappraisal of the use of infrared thermal image analysis in medicine. *IEEE Transactions on Medical Imaging*, 17, 1019-1027.

This paper was excluded because it was a narrative review/expert opinion article. It assesses the value of IR thermography to medicine in the light of new developments in camera technology and software techniques for dynamic analysis of temperature variation. It states that the potential of infrared thermography has been revolutionised by the development of lightweight digital cameras that are electrically cooled or uncooled, with high spatial and temperature resolutions.

Joseph, L. P. (1999). Breast cancer detection. *Science*, 284, 743.

This paper was excluded because it was a correspondence article. It is a letter from the Devices and Radiological Health Department of the FDA responding to a recent public statement made by an actor in the USA who had recommended thermography because "mammograms are now known to be causing cancer". The letter states that thermography has not been approved for screening by the FDA, that the American Medical Association declined to recommend it for medical applications, and that studies have proven that mammograms are the most effective way to detect early breast cancer.

Kaczmarek, M., & Nowakowski, A. (2003). Analysis of transient thermal processes for improved visualization of breast cancer using IR imaging. *IEEE EMBC*, 1113-1116.

This paper was excluded because it was a descriptive report of the technical development of a mechanism for performing 'active dynamic thermography', in which the target object (the breast) is thermally excited and then multiple images are taken. The aim of this process is to improve resolution, and improve recognition of internal structures of the breast. The authors concluded that this is a promising new technique that needs to be studied further.

Kastberger, G., & Stachl, R. (2003). Infrared imaging technology and biological applications. *Behavior Research Methods Instruments & Computers*, 35, 429-439.

This paper was excluded because it was a discussion article about the biological applications of IR technology. The paper describes some of the technological details of infrared thermography detectors and then focuses on the applications of infrared thermography in the study of animal behaviours.

Keith, L. G., Oleszczuk, J. J., & Laguens, M. (2001). Circadian rhythm chaos: a new breast cancer marker. *International Journal of Fertility & Womens Medicine*, 46, 238-247.

This paper was excluded because it was a narrative review article. It focuses on evidence for a connection between disturbances of normal circadian thermal variations and breast cancer. The paper provides a brief overview of some of the historical aspects of thermography up until the mid-1980s, and uses results adapted from a 1982 study of liquid crystal thermography to support a proposed theory that an increased disturbance of thermal circadian rhythm may indicate a greater degree of malignancy of breast cancer.

Keith, L. G., Oleszczuk, J. J., & Laguens, M. (2002). Are mammography and palpation sufficient for breast cancer screening? A dissenting opinion. *Journal of Womens Health & Gender Based Medicine*, 11, 17-25.

This paper was excluded because it was a discussion/opinion article that outlines various drawbacks of mammography as a screening modality. The authors propose that the detection of abnormal circadian fluctuations in breast temperature may be a sign for increased risk of breast cancer.

Keller, C. E., & Watt, A. C. (1985). Mammography: still the imaging standard. *Henry Ford Hospital Medical Journal*, 33, 82-87.

This paper was excluded because it was a narrative review article. It examines and discusses the use of a variety of imaging methods for the detection of breast cancer. The authors state that no data are available to establish thermography as a useful screening method in detecting early occult breast cancer and conclude that thermography remains an experimental method.

Keyserlingk, J. R., Ahlgren, P. D., Yu, E., Belliveau, N., & Yassa, M. (2000). Functional infrared imaging of the breast. *IEEE Engineering in Medicine & Biology Magazine*, 19, 30-41.

This paper was excluded because it was a narrative review article. It overviews the history of infrared imaging of the breast, discusses the major studies since the 1960s, their designs, and postulates reasons for the poor acceptance of thermographic imaging by the medical community. This narrative review includes a summary of the retrospective study of infrared imaging as an adjunctive technique in breast cancer diagnosis reported by Keyserlingk et al. (1998). The authors conclude that newer infrared imaging technologies are vastly superior to those previously available, although they need stringent regulations and quality control standards.

Laine, H., Harjula, A., & Tallroth, K. (1985). Sonography in palpable breast tumours. Comparison with palpation, thermography and X-ray mammography. *Journal Belge de Radiologie*, 68, 5-8.

This paper reports from a study that aimed to evaluate the reliability of ultrasound scanning in the diagnosis of palpable breast lumps. The study was excluded because it does not identify the method of thermography used. Twenty-eight women with breast lumps were examined by mammography, ultrasound and thermography. Despite the small sample size, the authors concluded that ultrasound was the imaging modality of choice, and that thermography was useful but tended towards 'overdiagnosis'.

Lally, J. (1993). Thermography: a prototype of failed technology. *Delaware Medical Journal*, 65, 789-790.

This paper was excluded because it was an editorial article. It discusses the difficulties involved in accurately assessing the effectiveness of new health technologies. The author argues that societal resources should not be used on technologies that do not have solid scientific evidence for their worth, and uses thermography as an example of such a technology.

Letton, A. H., & Mason, E. M. (1986). Routine breast screening. Survival after 10.5 years follow-up. *Annals of Surgery*, 203, 470-473.

This paper was excluded because it was a narrative review. It reports from one of 27 centres (Georgia) in the U.S., which undertook a Breast Cancer Detection Demonstration Project in the 1970s, stating that thermography was abandoned at that centre after only a few months.

Letton, A. H., Mason, E. M., & Ramshaw, B. (1996). Twenty-year follow-up breast screening project. *Journal of the Medical Association of Georgia*, 85, 31-32.

Report from a study of long-term 20 year follow-up of a cohort of 9,043 women who underwent breast cancer screening over a five year duration at one of the BCDDP centres from 1973. This study was excluded because it focused on non-thermographic imaging modalities. Thermography was dropped at an early stage at this centre after the finding that only 46 percent of the women with breast cancer had positive thermograms, and therefore screening consisted of mammography plus physical examination. Compared the long-term outcome for 128 women who had cancers detected during the five-year screening period with a matched group of 1,069 who had negative examination during the same period, and were followed-up as well. It found that the tumours detected by the routine screening process were on average 0.8cm smaller than those subsequently detected by clinical symptoms or voluntary mammograms.

Liszka, G., Herics, I., Bodo, M., & Ringwald, G. (1987). Diagnostic value of infrared thermography in breast cancer, and in proliferative and non proliferative mastopathies. *Archiv fur Geschwulstforschung*, 57, 487-491.

A report from a descriptive study that presents the thermographic results from a group of 201 patients who each had a palpable breast mass. This study was excluded because it used an older infrared detector, which is no longer manufactured. All patients had infrared thermography, as well as physical examination, mammography, and needle aspiration cytology. However, only the results from thermography are presented in this report, so no comparative or adjunctive comparisons of performance can be made with the other diagnostic modalities. Thermography results were classified as 'suspicious' or 'malignant' for 30 out of 37 breast carcinomas diagnosed in this group, but were also 'suspicious' or malignant for 24 out of 50 'proliferative mastopathies', and for 20 out of 114 'non-

proliferative mastopathies'. The larger malignant tumours were more likely to have a 'malignant' thermographic result. The authors concluded that infrared thermography by itself is not a reliable method to establish or exclude the diagnosis of breast cancer, but that they consider it as a complementary examination. This is somewhat surprising, as the information reported by these authors does not allow for an adequate assessment of the performances of thermography as an adjunctive diagnostic modality. However, these authors did compare their results with older studies that had used the same classification of thermography results as they had, pointed out that there was a large variation in findings between different studies, but did not attempt to explain the differences.

Maisey, M. N. (1988). Imaging techniques in breast cancer. What is new? What is useful? - A review. *European Journal of Cancer & Clinical Oncology*, 24.

This paper was excluded because it was a narrative review/ expert opinion article. It examines various 'new' imaging techniques and discusses their applications for the management of breast cancer under the broad headings of: diagnosis and screening; imaging axillary nodes; staging procedures, and; choice and assessment of therapy. The author considers that, at that time, thermography is suitable for neither screening nor diagnosis of breast cancer.

Margolese, R. G., Cantin, J., Bouchard, F., Caines, J., Beaulieu, M. D., Little, C. D., Levine, M. N., et al. (1998). 1. The palpable breast lump: information and recommendations to assist decision-making when a breast lump is detected. *Canadian Medical Association Journal*, 158, S3-S8.

This paper was excluded because it was a narrative report presenting a set of evidence-based clinical guidelines to assist in the clinical decision-making process for women with palpable breast lumps, developed for the Canadian Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. A systematic review of published evidence to January 1996 (which included a non-systematic review of literature until January 1997, as well as expert opinion) was conducted to identify the evidence upon which the guidelines were based, and the guidelines were developed as a result of a peer-reviewed, consensual research process. Thermography was recommended as having no current role in the management of women with breast lumps, except in the context of structured clinical trials. The first author of this report was contacted to request a report from the original systematic review conducted for the guideline development, but the data was no longer available.

May, T. S. (2002). Military surveillance system for breast cancer detection. *Drug Discovery Today*, 7, 1111-1112.

This paper was excluded because it was a discussion article. It explains the various technologies that were previously used and that are now being developed in the field of infrared imaging, providing opinions from various experts in the field of thermal imaging. The author concludes that results from long-term, large clinical trials of modern infrared detectors are required before it can be determined whether infrared thermal imaging has sufficient sensitivity and specificity to be used for the diagnosis of breast cancer.

McKinna, J. A., Davey, J. B., Walsh, G. A., A'Hern, R. P., Curling, G., Frankland, H., & Viggers, J. (1992). The early diagnosis of breast cancer--a twenty-year experience at the Royal Marsden Hospital. *European Journal of Cancer*, 28A, 911-916.

A report of a observational study that assessed the rates of cancer detection in women attending a specialist breast diagnostic unit in London from 1968 to 1987. This study was excluded because the thermography method and type of device are not specified. Two discrete patient populations were identified, as the clinic policy changed between 1981 and 1982. From 1968 to 1981, the population

under study included both those with breast symptoms and also asymptomatic women attending for screening (often because of familial history risk factors). From 1982 to 1987, only those women who had been assessed at the clinic and then categorised as 'high-risk' for breast cancer were followed up by ongoing breast screening, entering a prospective 'screening interval' study. This clinic initially undertook thermography on all women. Following low sensitivities in both time periods, the use of thermography was discontinued at the breast unit.

Moskowitz, M. (1985). Thermography as a risk indicator of breast cancer. Results of a study and a review of the recent literature. *Journal of Reproductive Medicine*, 30, 451-459.

This two-part paper firstly presents the results from a retrospective investigation into thermography as a risk factor for breast cancer that was undertaken by the author using data collected from 1973 to 1983 in the Cincinnati Breast Detection Demonstration Project (which had included thermography as part of screening procedures between 1973 and 1976). This paper was excluded because it did not state the method of thermography, nor type of device used and focused on risk prediction. The author concluded as a result of this study that there was no association between a 'positive' thermogram and increased risk of developing breast cancer. The latter part of this paper presents a narrative review/expert opinion discussion on the topic of thermography as a risk indicator for breast cancer. The author concludes that, at that point in time, thermography had not been proven useful in the evaluation of breast disease, and suggests that well-designed prospective trials needed to be undertaken to assess its true usefulness.

Mushlin, A. I. (1985). Diagnostic tests in breast cancer. Clinical strategies based on diagnostic probabilities. *Annals of Internal Medicine*, 103, 79-85.

This paper was excluded because it was a narrative review/expert opinion article. It provides an overview of the screening and diagnostic modalities then available for the detection of breast cancer. The author outlines the history of infrared thermography and other thermography methods, and reviews the major evidence for the technique of thermography to that date. As a result of his review the author summarises that, although there is no radiation exposure involved, thermography is not suited to screening due to low sensitivity and high false-positive results, and that mammography has higher accuracy in the diagnostic situation and would not be assisted by thermography. For risk assessment, the author concludes that thermography's role is limited because up to 20 percent of women screened have abnormal thermographic results and only a small percentage of them go onto develop malignancy.

Negri, S., Bonetti, F., Capitanio, A., & Bonzanini, M. (1994). Preoperative diagnostic accuracy of fine-needle aspiration in the management of breast lesions: comparison of specificity and sensitivity with clinical examination, mammography, echography, and thermography in 249 patients. *Diagnostic Cytopathology*, 11, 4-8.

A report from a study that aimed to verify the pre-operative diagnostic accuracy of physical examination (PE), mammography (M), ultrasound (US), thermography (TH) and fine needle aspiration (FNA). This study was excluded because it did not specify the method of thermography utilised. The researchers retrospectively reviewed the case notes of 2,178 women who had attended a breast unit with "mammary problems" (p. 4) between April 1988 and February 1989, and identified 249 women who had undergone all five tests. There is no report of how or in what order the five tests were done, what equipment was used, nor how they were interpreted. From 249 women, 94 patients underwent excision and histological examination on the basis of a result indicating malignancy from one or more of the five tests. From this, 57 cases of histologically-proven breast cancer were identified. The 155 women who did not have excision biopsy were re-examined after another 12 to 18 months to look for any false-negative results, but no further cases of breast cancer were identified after that time. The authors report only the data from the 94 patients who had the histological diagnosis. From this group, TH alone had a sensitivity (Se) = 32% and specificity (Sp) = 92%, compared to: PE alone Se=51%,

Sp=89%; M alone Se=88%, Sp=57%; US alone Sp=56%; FNA alone Se=95%, Sp=92%. In combinations: PE +TH Se=58%, Sp=81%; PE + M Se=89%, Sp=54%; PE + M + TH Se=89%, Sp=49%; PE + M + FNA Se=100%, Sp=49%; PE + M + FNA + TH Se=100%, Sp=40%, PE + US + M Se=89%, Sp=46%; PE + US + M + TH Se =89%, Sp=38%; PE + US + M + FNA Se=100%, Sp=40%, PE + US + M + FNA + TH Se= 100%, Sp=35%. The authors' conclusion from this study is that a single procedure is not sufficient to detect all breast cancers, that the best results are seen when both TH and US were excluded from the combination, and that the best diagnostic combination was PE + M + FNA because each modality detected cases that the others did not (and that others have found that combination to be the most cost-effective). The authors further conclude that the low values of sensitivity of TH make it unsuitable for the diagnosis and treatment of breast lesions.

Ng, E. Y., Fok, S. C., Peh, Y. C., Ng, F. C., & Sim, L. S. (2002). Computerized detection of breast cancer with artificial intelligence and thermograms. *Journal of Medical Engineering & Technology*, 26, 152-157.

Report from a study in which the authors developed a statistical computer tool called an 'artificial neural network' (ANN) to be used in conjunction with infrared thermography. This paper was excluded because it reported on computational processes, and did not provide data that could be used to estimate the outcomes of interest of this review. The purpose of the ANN was to objectively control for factors that may confound the temperature measurement of female breast tissue, such as the hormonal and temperature fluctuations during the normal female menstrual cycle. The study methods are not well described, but the authors appear to have used data from between 200 to 207 female patients of unspecified age with normal breasts, benign breast disease and malignant breast disease to develop their tool. The authors appear to have collected infrared data at different times during the participants' menstrual cycles. The authors report that one ANN tool that was developed as a result of their research correctly predicted the health status of 124 of their participants. They conclude that the relatively small population of participants studied limited the diagnostic accuracy of this ANN, and that a larger population group would need to be studied to develop a more accurate tool.

Ng, E. Y., & Sudharsan, N. M. (2004). Computer simulation in conjunction with medical thermography as an adjunct tool for early detection of breast cancer. *BMC Cancer*, 4, 17(unpaginated electronic journal).

Report from a study that aimed to improve computerised infrared images of the female breast by using numerical modelling processes. This paper was excluded because it is concerned with the technical and mathematical aspects of developing this computational model. The authors concluded that their model may assist in the assessment of breast lumps.

Ng, E. Y., Ung, L. N., Ng, F. C., & Sim, L. S. (2001). Statistical analysis of healthy and malignant breast thermography. *Journal of Medical Engineering & Technology*, 25, 253-263.

Report from a study that aimed to determine the feasibility of quantifying the analysis of breast thermograms. This paper was excluded because it was not clear what type of infrared thermography was used. Ninety patients were chosen at random, but it is not clear from what population they were chosen (8 were excluded because of previous breast surgery or anatomical distortion). Three patient groups:

Asymptomatic patients [n=30] (used mammography results as gold standard for this group).

Benign breast disease [n=48] (used mammography results as gold standard for this group, plus histology for 11 with suspicious mammograms- so all patients in this group did not receive the same).

Cancer patients [n=4] (used mammography as gold standard and histology for all, as mammography results were suspicious enough to warrant invasive test).

The authors detail much of their bio-statistical development method of quantification of the thermal parameters. The authors conclude that they were able to quantify the temperature measurements of both breasts with cancer, and healthy breasts, although large amounts of thermal data were necessary to do this. They finally state that further studies are needed to evaluate the role of thermography for breast cancer imaging.

Ohashi, Y., & Uchida, I. (2000). Applying dynamic thermography in the diagnosis of breast cancer - Techniques for improving sensitivity of breast thermography. *IEEE Engineering in Medicine and Biology Magazine*, 19, 42-51.

Report from a study that compared the diagnostic accuracy of 'steady state' thermography with the diagnostic accuracy of 'dynamic' (sequential and subtraction) thermography methods. This study was excluded because the equipment used for the study was not specified, and it was not clear what method of thermography was used. Examined approximately 728 patients with breast cancer for this study, and 100 with benign breast disease. It was not clear how these diagnoses were confirmed. It was not clear how the thermographic image analysis was done, whether by an operator or a computer. It was not clear whether the disease status of subjects was known to those interpreting the thermograms. No comparison with any other diagnostic modality was made. Major methodological flaws were present in the design, conduct and reporting of this study.

Ohsumi, S., Takashima, S., Aogi, K., & Usuki, H. (2002). Prognostic value of thermographical findings in patients with primary breast cancer. *Breast Cancer Research & Treatment*, 74, 213-220.

Report from a study that evaluated the independent prognostic significance of a thermal difference (measured by infrared telethermometry) between the affected area of the diseased breast and the corresponding area of the non-diseased breast in 340 women diagnosed with primary unilateral breast cancer. This study was excluded because it considered the role of infrared thermography in breast cancer prognosis. The patients in this study received various forms of treatment for their breast cancer, and were followed for a mean time length of 96 months. From the results of their study, the authors conclude that, although the measurement of thermal differences between breast cancer patients' tumours using the infrared imaging device was straightforward, non-invasive and objective, it was less important for disease prognosis than the conventional prognostic factors of lymph node status, tumor size and histological status, and oestrogen receptor status.

Orel, S. G., & Troupin, R. H. (1993). Nonmammographic imaging of the breast: current issues and future prospects. *Seminars in Roentgenology*, 28, 231-241.

This paper was excluded because it was a narrative review/expert opinion article. It discusses the current value and potential applications of available radiological modalities in the screening and diagnosis of breast cancer, comparing them to mammography. On the topic of thermography, the authors present a brief overview of the history and literature, concluding that there was currently no data to support the use of this technique in screening for, or diagnosis of, breast cancer.

Osman, M. M., & Afify, E. M. (1988). Thermal modeling of the malignant woman's breast. *Journal of Biomechanical Engineering*, 110, 269-276.

This paper was excluded because it was a report from a technical study. It aimed to increase knowledge on how malignant tumours' size, location, and depth affect the temperature distribution in the breast. Much of the report discusses detailed mathematical calculations. The authors conclude that hot spots on breast thermal contour plots may not be directly related to tumour below the surface.

Paul, J. L., & Lupo, J. C. (2002). From tanks to tumors. *IEEE Engineering in Medicine & Biology Magazine*, 21, 34-35.

This paper was excluded because it was a descriptive journal article that summarises the conference proceedings of a technological workshop that looked at means of utilising developments in data management, image processing and IR sensor technology in the management of breast cancer.

Paulus, D. D. (1987). Imaging in breast cancer. *Ca: a Cancer Journal for Clinicians*, 37, 133-150.

This paper was excluded because it was a narrative review/expert opinion article. It discusses mammography and the technologies that compete with it for imaging in breast cancer management. Thermography is critiqued by the author as being an ineffective modality for breast cancer diagnosis, unhelpful for prognosis, questionable for risk-prediction, and considered as almost completely worthless for screening asymptomatic women.

Pera, A., & Freimanis, A. K. (1987). The choice of radiologic procedures in the diagnosis of breast disease. *Obstetrics & Gynecology Clinics of North America*, 14, 635-650.

This paper was excluded because it was a narrative review article. It discusses available radiologic procedures for breast cancer detection in asymptomatic women, in women with indeterminate clinical findings, and in women with an obvious palpable breast mass. A short section on thermography describes it as a non-invasive modality that images the heat radiated by the breast, which has not proven to be useful for either screening or diagnosis.

Ramlau, C., & Sledzikowski, J. (1993). Combination of thermographic and ultrasound methods for the diagnosis of female breast cancer. *European Journal of Gynaecological Oncology*, 14, 152-154.

This paper was excluded because it is a technical study report. It describes an experimental project where infrared thermography and ultrasound were combined together to develop a method of assessing breast cancer risk.

Ring, E. F. (1998). Progress in the measurement of human body temperature. *IEEE Engineering in Medicine & Biology Magazine*, 17, 19-24.

This paper was excluded because it was a narrative article. It outlines the historical development of temperature measurement as an indicator of disease from biblical times to modern infrared imaging techniques.

Ring, E. F., & Ammer, K. (2000). The technique of infra red imaging in medicine. *Thermology International*, 10, 7-14.

This paper was excluded because it was a narrative review/expert opinion article. It describes in non-expert language the specific equipment, location and patient requirements for the acquisition of medical thermological images. The authors' opinion is that good technique and standardisation are essential, and achievable, for the clinical applications of infrared thermography.

Runowicz, C. D., Goldberg, G. L., & Smith, H. O. (1993). Cancer screening for women older than 40 years of age. *Obstetrics & Gynecology Clinics of North America*, 20, 391-408.

This paper was excluded because it was a narrative review. It outlines screening methods for cervical, ovarian and breast cancers. For breast screening, the authors describe thermography as having the drawbacks of high false positives and a 25 percent false negative rate. The authors conclude that, for those reasons, thermography has not achieved widespread clinical use for breast cancer screening.

Seymour, E. Q., & Stanley, J. H. (1985). The current status of breast imaging. *American Surgeon*, 51, 591-595.

This paper was excluded because it was a narrative review of the status of breast imaging. It offers guidance for the use of the different imaging modalities. The authors briefly surmise that thermography is not reliable for the detection of early non-palpable breast lesions, and that there is insufficient evidence for the role of thermography in detecting women at high-risk for breast cancer. The authors conclude that low-dose mammography used with the (then) current guidelines can result in a considerable reduction in mortality from breast cancer.

Simpson, H. W. (1996). Sir James Young Simpson Memorial Lecture 1995. Breast cancer prevention: a pathologist's approach. *Journal of the Royal College of Surgeons of Edinburgh*, 41, 359-370.

A report from a study of the female 'breast menstrual cycle' using a contact thermography method to compare monthly breast temperatures in a group of female breast cancer patients with a control group of non-cancer patients. This study was excluded because it used a non-infrared thermography method. The study found significant differences in the monthly breast temperature cycles between the two groups. The author proposes that this may enable a calculation of lifetime breast cancer risk for particular patients.

Stark, A. M. (1985). The value of risk factors in screening for breast cancer. *European Journal of Surgical Oncology*, 11, 147-150.

Report from a study that aimed to identify whether a positive thermogram was a factor that may classify a woman as being at greater risk of developing breast cancer. This study was excluded from appraisal because it looked at the role of thermography in identifying women at higher risk for breast cancer, but did not investigate whether or not identifying these risk factors led to earlier detection and treatment of breast cancer than would have normally occurred. There was no comparison made with a similar group of women who had not undergone thermographic examination. In addition, information regarding the type of thermography, or model of thermographic device utilised during this study was not easily available (although the study started in 1969, so is likely to have used equipment that is no longer available).

Stevens, R. G., & Beaman, S. A. (1988). The use of difference of Gaussian image filtering to assess objectively the correlation between breast vascularity and breast cancer. *Physics in Medicine & Biology*, 33, 1417-1431.

Study report describing the development of a method to quantify the vascularity and associated temperature gradients of thermograms of the breast. This paper was excluded the report is concerned with describing the detailed physics calculations and statistical analysis undertaken to quantify the parameters of interest. The resultant processing algorithm developed by this method was reported to correlate well with actual tissue vascularity, but the authors conclude that no part of it was of value as a breast screening technique.

Ulmer, H. U., & Frischbier, H. J. (1987). The role of thermography in the identification of high-risk groups in breast cancer. *Breast Disease*, 1, 19-27.

This paper reports from a study that aimed to investigate whether infrared thermography could identify asymptomatic women who were at increased risk of developing breast cancer. This study was excluded because the type of infrared thermography device that was used in the study was not reported, and it looked at the role of thermography in risk factor identification. From an initial cohort of 11,546 asymptomatic women examined between 1973 and 1980 and followed up for a minimum of five years, thermographic abnormalities were found in 17.5 percent. The authors found a rate of breast cancer incidence of approximately 2% in women with at least one abnormal thermogram considered as suspicious or requiring further investigation. The authors conclude, as a result of their study, and taking into consideration previous literature reports, that thermography was inappropriate for selecting women at high-risk of breast cancer for ongoing screening.

van Dam, P. A., Van Goethem, M. L., Kersschot, E., Vervliet, J., Van den Veyver, I. B., De Schepper, A., & Buytaert, P. (1988). Palpable solid breast masses: retrospective single- and multimodality evaluation of 201 lesions. *Radiology*, 166, 435-439.

Report from a study that aimed to evaluate the diagnostic characteristics of independently conducted physical examination, mammography, ultrasound, infrared thermography and various combinations of these tests in the evaluation of solid palpable breast lumps. This study was excluded because it utilised an infrared device that is no longer available, and for which device specifications were not reported.

White, T., Butler, N., & Murphy, R. (1998). An uncooled IR sensor with digital focal plane array. *IEEE Engineering in Medicine & Biology Magazine*, 17, 60-65.

This paper was excluded because it was a detailed technical article. It reports on the development of a microbolometer focal plane array infrared camera (bolometers are thermal detectors heated by incoming radiation, resulting in a temperature rise that is sensed as a change in the element resistance). The paper uses individual case examples of clinical experiments to suggest potential applications for the particular camera system that the authors have been involved in developing.

Wright, T., & McGechan, A. (2003). Breast cancer: new technologies for risk assessment and diagnosis. *Molecular Diagnosis*, 7, 49-55.

This paper was excluded because it was a narrative review article. It discusses the issues relating to the identification of women at high risk of breast cancer, as well as issues relating to screening for breast cancer. The authors describe various technologies (that are in various stages of development and testing) for potential use in both these clinical applications. The authors state that while mammography is not perfect, it is currently the best tool for screening for breast cancer. The authors report that there are two types of new infrared thermography devices being studied to see if they are suitable for the early detection of breast cancer, although the results of the studies will not be available for several years. One type of device, being developed by the US Office of Naval Research using thermal data collected by two cameras, is a modification of a device that is used in military surveillance. The other type of device is being developed by OmniCorder Technologies, Inc., measures changes in skin temperature over time, and is called dynamic area telethermometry.

Zhou, X. H., & Gordon, R. (1989). Detection of early breast cancer: an overview and future prospects. *Critical Reviews in Biomedical Engineering*, 17, 203-255.

This paper was excluded because it was a narrative review/expert opinion article. It provides a historical overview on issues surrounding the early detection of breast cancer, focusing particularly on screening. It discusses the imaging modalities in use at that time, and those being developed for the early detection of breast cancer. After a discussion of the technique of thermography, the authors conclude that, at that time, thermography has severe limitations in the early detection of breast cancer.

Table 4. Evidence table of the appraised excluded article relating to the use of infrared thermography as an adjunctive diagnostic tool

Authors Country Evidence Grading	Study Design, Index Test, Comparators and Reference Standard	Study Aim, Methods and Selection Criteria	Sample Characteristics	Results	Limitations and Conclusions
van Dam et al. (1988) Antwerp, Belgium.	<p>Study setting Department of Gynaecology, Antwerp University Hospital.</p> <p>Study design Cross sectional survey (with retrospective blinded analysis).</p> <p>Index test Telethermography (TH)*. Unit from Philips, Eindhoven. Details of equipment not specified (nor available).</p> <p>Comparators Physical examination (PE). All done by same clinician to assess the breast, and specific local structures and characteristics of the mass.</p> <p>Mammography (M)*. Conventional screen-film contact unit.</p> <p>Ultrasound (US)*. 5MHz sector scanner for first 183 patients, 7.5MHz linear transducer scanner for last 18 patients.</p> <p>Reference standard Histological biopsy (tissue resection). Reviewed by experienced breast pathologist.</p>	<p>Study aim To evaluate the diagnostic characteristics of independently conducted PE, M, US, TH and various combinations of these tests in the evaluation of solid palpable breast masses.</p> <p>Methods Identified consecutive patients who had a solid palpable breast mass that had been sampled by biopsy, who had undergone all four diagnostic modalities, and for whom all necessary data was readily available for retrospective review.</p> <p>Results from all three imaging modalities* were separately read and interpreted by two senior radiologists (blind to all patient data, and to all other results). Discordant results were re-reviewed for consensus opinion. For PE, M and US 'malignant' and 'suspect' results were counted as positive (require same clinical management). TH 'suspect' results were counted as positive. For all four modalities 'benign' was counted as negative.</p>	<p>Participants n = 201</p> <p>All female.</p> <p>Mean age 46.7 years, age range 14-83 years.</p> <p>Pre-menopausal n=94 Peri-menopausal n=18 Post-menopausal n=89.</p> <p>Benign breast mass n=106 (fibrocystic n=70, fibroadenoma n=30, other n=6).</p> <p>Malignant breast mass n=95 (Invasive ductal carcinoma n=79, intraductal carcinoma n=5, other carcinoma n=11).</p> <p>Breast mass size <2cm n=97 (benign n=56, malignant n=41) >2cm n=104 (benign n=50, malignant n=54).</p> <p>Inclusion criteria None specified.</p> <p>Exclusion criteria None specified.</p>	<p>TH alone: Se=49% (95% CI 39, 59) Sp=86% (79, 92), PPV=76% (65, 86) NPV=65% (58, 73)</p> <p>PE alone: Se=88% (82, 95), Sp =71% (62, 79), PPV=73% (65, 81) NPV=87% (80, 94)</p> <p>PE +TH: Se=92% (86, 97), Sp=65% (56, 74), PPV=70% (62, 78) NPV=90% (83, 96)</p> <p>PE + M: Se =95% (90, 99), Sp =51% (41, 60), PPV=63% (55, 71) NPV=92% (84, 99)</p> <p>PE + M +TH: Se=95% (90, 99), Sp =46% (37, 56), PPV=61% (53, 69) NPV=91% (83, 98)</p> <p>PE +US: Se=94% (89, 99), Sp=66% (57, 75), PPV=71 (63, 79), NPV=92 (86, 98)</p> <p>PE +US +TH: Se=94% (89, 99), Sp= 62% (53, 71), PPV=69% (61, 77) NPV=92% (85, 98)</p> <p>PE+ M +US: Se=97% (91, 99), Sp =49% (39, 59), PPV=63% (55, 71) NPV=95% (85, 99)</p> <p>PE+ M +US +TH: Se=97% (91, 99), Sp=43% (34, 53), PPV=61% (52, 68) NPV=94% (83, 99)</p> <p>Three breast cancers were incorrectly diagnosed as being benign masses by all four diagnostic modalities.</p>	<p>Limitations During the period of this study 1,340 women were assessed at this centre by all four diagnostic modalities, but the reasons for only including 201 in this study are not clearly explained.</p> <p>The number of patients in this study was insufficient to permit discrimination of different multimodality protocols. Chance cannot be excluded as an explanation for any differences in effectiveness between diagnostic modalities.</p> <p>Potential selection bias due to low participation.</p> <p>No explanation as to why a different type of US was used for 18 patients, nor whether this may have affected the diagnostic accuracy of this modality. Inadequate details of TH imaging are reported to exclude environmental and patient factors as contributing to measurement error.</p> <p>Authors' conclusions Thermography did not provide additional information in their 201 patients.</p> <p>Reviewer conclusions Despite the evident potential for selection bias, this study does demonstrate that to show any potential adjunctive benefit of TH (if one truly exists) in this diagnostic situation, a far larger number of patients would have to be included.</p>

APPENDIX 5: APPRAISED RETRIEVED PAPERS

Keyserlingk, J. R., Ahlgren, P. D., Yu, E., & Belliveau, N. (1998). Infrared imaging of the breast: Initial reappraisal using high- resolution digital technology in 100 successive cases of stage I and II breast cancer. *Breast Journal*, 4, 245-251.

Parisky, Y. R., Sardi, A., Hamm, R., Hughes, K., Esserman, L., Rust, S., & Callahan, K. (2003). Efficacy of computerized infrared imaging analysis to evaluate mammographically suspicious lesions. *AJR. American Journal of Roentgenology*, 180, 263-269.

Williams, K. L., Phillips, B. H., Jones, P. A., Beaman, S. A., & Fleming, P. J. (1990). Thermography in screening for breast cancer. *Journal of Epidemiology & Community Health*, 44, 112-113.