SHIATSU

A review of the evidence

指圧

Professor Nicola Robinson Julie Donaldson Ava Lorenc

October 2006

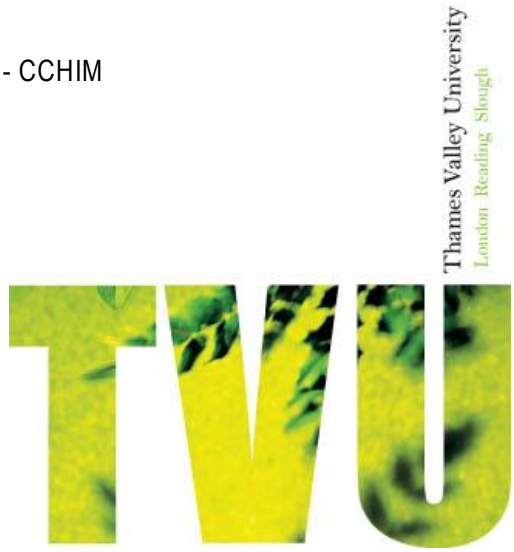
Thames Valley University
London Reading Slough

Commissioned and funded by the Shiatsu Society UK www.shiatsusociety.org

This report was commissioned and funded by:
Shiatsu Society UK
Eastlands Court
St Peters Road
Rugby
CV21 3QP
Tel: 0845 130 4560
Email: info@shiatsusociety.org
Web: www.shiatsusociety.org

Centre for Complementary Healthcare & Integrated Medicine - CCHIM
Faculty of Health and Human Sciences
Thames Valley University
Walpole House
18-22 Bond Street
Ealing
London W5 5AA, UK

Tel. 0208 280 5172
Email: nicky.robinson@tvu.ac.uk
www.cchim.com



Contents

Glossary of terms

Attrition rate: The rate at which participants are 'lost' during the course of a study. (Also called 'loss to follow up'). Participants that are 'lost' during a study are often called dropouts and are usually untraceable.

Bias: When a point of view prevents impartial judgment on issues relating to the subject of that point of view. In clinical studies, bias is controlled by blinding and randomisation.
A systematic distortion of research results due to the lack of objectivity, fairness, or impartiality on the part of the evaluator or assessor. Alternatively, there are disparities in research or test results due to using improper assessment tools or instruments across groups.
A systematic error or deviation in results or inferences from the truth. In studies of the effects of health care, the main types of bias arise from systematic differences in the groups that are compared (**selection bias**), the care that is provided, exposure to other factors apart from the **intervention** of interest (**performance bias**), withdrawals or exclusions of people entered into a study (**attrition bias**) or how outcomes are assessed (**detection bias**). Reviews of studies may also be particularly affected by **reporting bias**, where a biased subset of all the relevant data is available.

Blinded study: A study done in such a way that the patients or subjects do not know (is blinded as to) what treatment they are receiving to ensure that the results are not affected by a placebo effect (the power of suggestion).

Blinding: The process of preventing those involved in a trial from knowing to which comparison group a particular participant belongs. The risk of bias is minimised when as few people as possible know who is receiving the experimental intervention and who the control intervention. Participants, caregivers, outcome assessors and analysts are all candidates for being blinded. Blinding of certain groups is not always possible, for example surgeons in surgical trials.

Bonferroni correction for Type 1 error: This is an example of a multiple comparison techniques. It adjusts the Type 1 error level to compensate for multiple comparisons between three or more groups or two or more response variables.

Carryover effect: The persistence, into a later period of treatment, of some of the effects of a treatment applied in an earlier period.

Control group: The subjects in a controlled study who do not receive the treatment.

Controlled study: A study that uses the method of comparison to evaluate the effect of a treatment by comparing treated subjects with a control group, who do not receive the treatment. (See also uncontrolled study)

Convenience sample: A group of individuals being studied because they are conveniently accessible in some way. This could make them particularly unrepresentative, as they are not a random sample of the whole population. A convenience sample, for example, might be all the people at a certain hospital, or attending a particular support group. They could differ in important ways from the people who haven't been brought together in that way: they could be more or less sick, for example**.**

Double-blind: In a double-blind study, neither the subjects nor the people evaluating the subjects know who is in the treatment group and who is in the control group. This mitigates the placebo effect and guards against conscious and unconscious prejudice for or against the treatment on the part of the evaluators.

Duplicate bias: A study that has been published more than once using the same data but written as a separate study.

Hawthorne effect/bias: This can be summarized as "Individual behaviours may be altered because they know they are being studied." This means that the act of measurement, itself, impacts the results of the measurement. In science, dipping a thermometer into a vial of liquid can affect the temperature of the liquid being measured. In the same way, the act of collecting data, where none was collected before creates a situation that did not exist before, thereby affecting the results**.**

Intention to treat analysis: A strategy for analysing data from a randomised controlled trial. All participants are included in the arm to which they were allocated, whether or not they received (or completed) the intervention given to that arm. Intention-to-treat analysis prevents bias caused by the loss of participants, which may disrupt the baseline equivalence established by randomisation and which may reflect non-adherence to the protocol. The term is often misused in trial publications when some participants were excluded.

Inter-rater reliability: The degree of stability exhibited when a measurement is repeated under identical conditions by different raters. Reliability refers to the degree to which the results obtained by a measurement procedure can be replicated. Lack of inter-rater reliability may arise from divergences between observers or instability of the attribute being measured.

Intervention group: A group of participants in a study receiving a particular health care intervention. Parallel group trials include at least two intervention groups.

Intra-rater reliability: The degree of stability exhibited when a measurement is repeated under identical conditions by the same rater. Reliability refers to the degree to which the results obtained by a measurement procedure can be replicated. Lack of intra-rater reliability may arise from divergences between instruments of measurement, or instability of the attribute being measured.

Language bias: Exclusion, in a meta-analysis of controlled trials, of those published in languages other than English.

Mean: The sum of a list of numbers, divided by the number of numbers. This is also often referred to as the average.

Meta-analysis: A statistical procedure to combine a number of existing studies. Through such a procedure, effects which are hard or impossible to discern in the original studies because of a too small sample size can be made visible, as the meta-analysis is (in the ideal case) equivalent to a single study with the combined size of all original studies. A weakness of the method is that problems with any of the studies will affect the result of the meta-analysis, so a good meta-analysis of bad studies will still result in bad data.

Null hypothesis: A statement concerning one or more parameter(s) that is subjected to a statistical test; a statement that there is no relationship between the two variables of interest; the belief that any apparent relationship between or among variables in one or more research samples has been caused by sampling error; the hypothesis that is tested when seeking to gain statistical support for a research hypothesis.

Order effects: Where the effects of two different interventions (A, B) are both being studied for all participants divided into two groups (1, 2). The order in which these interventions are administered may have an effect on the outcome e.g. group 1 has intervention A followed by B and group 2 has intervention B followed by A.

Placebo: An inactive substance or procedure administered to a participant, usually to compare its effects with those of a real drug or other intervention, but sometimes for the psychological benefit to the participant through a belief that s/he is receiving treatment. Placebos are used in clinical trials to blind people to their treatment allocation. Placebos should be indistinguishable from the active intervention to ensure adequate blinding.

Placebo effect: The belief or knowledge that one is being treated can itself have an effect that confounds with the real effect of the treatment. Subjects given a placebo as a pain-killer report statistically significant reductions in pain in randomised studies that compare them with subjects who receive no treatment at all. This very real psychological effect of a placebo, which has no direct biochemical effect, is called the placebo effect. Administering a placebo to the control group is thus important in studies with human subjects; this is the essence of a blind experiment.

Powered sample size: The sample size calculated for a study will ensure that it is sufficient in order to detect a significant difference.

Practice effect: The effect of receiving an intervention for a second time. This can also be referred to as a learning effect. When you split the subjects, the group that gets the control first has the practice effect added to the intervention, whereas the group that gets the intervention first has the practice effect added to the control treatment. So when you average the difference scores, the practice effect disappears and you are left with the treatment effect, provided the two groups have the same number of subjects.

Pragmatic design: A trial that aims to test a treatment policy in a 'real life' situation, when many people may not receive all of the treatment, and may use other treatments as well. This is as opposed to an explanatory trial, which is done under ideal conditions and is trying to determine whether a therapy has the ability to make a difference at all (i.e. testing its efficacy).

P-value: The probability (ranging from zero to one) that the results observed in a study (or results more extreme) could have occurred by chance if in reality the null hypothesis was true. In a meta-analysis, the P-value for the overall effect assesses the overall statistical significance of the difference between the intervention groups, whilst the P-value for the heterogeneity statistic assesses the statistical significance of differences between the effects observed in each study.

Probability samples: Samples in which each element in the population has a known chance of being selected into the sample.

Purposive sample: A non probability sampling technique wherein investigators use their judgment and prior knowledge to choose people for the sample who would best serve the purposes of the study.

Random Sample: A random sample is a sample whose members are chosen at random from a given population in such a way that the chance of obtaining any particular sample can be computed. The number of units in the sample is called the sample size, often denoted as ‘n’.

Randomised block experiment: This study design splits the experiment into a number of "mini-experiments" or blocks for convenience, or to increase power. Typically, each block has one experimental unit of each treatment.

Randomised control trial (RCT): A clinical trial in which chance is deliberately introduced in assigning subjects to the treatment and control groups. For example, write an identifying number for each subject on a slip of paper, stir up the slips of paper, and draw slips without replacement until half of them have been drawn. The subjects identified on the slips drawn could then be assigned to treatment, and the rest to control. Randomising the assignment tends to decrease confounding of the treatment effect with other factors, by making the treatment and control groups roughly comparable in all respects but the treatment.

Sample with low attrition rate: This indicates that there was a low level of drop outs from the group of study participants (See **Attrition rate**)

Single blind study: A study in which one party, either the investigator or participant, is unaware of what medication the participant is taking; also called single-masked study.

Single-group pretest posttest design: There is no control group in this type of study. The results are therefore measured by:
Pretest - a means to measure existing knowledge or ability prior to the implementation of an instructional activity, innovation or program
Posttest - a means to measure knowledge or ability after an instructional activity, innovation or program is implemented, using one or more research methods . Also sometimes referred to as a "post-assessment."

Three armed RCT: A randomised clinical trials where there are three groups receiving different treatments / interventions for comparison.

Type I error: Error that occurs when the null hypothesis is rejected when a true relationship between variables does not exist; also called alpha (α) error.

t–distribution: A statistical distribution describing the distribution of the means of samples taken from a population with unknown variance.

t-test: The t-test employs the statistic (t) to test a given statistical hypothesis about the mean of a population (or about the means of two populations). The most common t-test is a test for a difference of two means

Uncontrolled study: A study in which there is no control group; i.e., in which the method of comparison is not used: the experimenter decides who gets the treatment, but the outcome of the treated group is not compared with the outcome of a control group that does not receive treatment.

Nomenclature of points

It is necessary in an academic report of this kind to name points strictly as they appear in published papers. This is why the same points appear with different abbreviations across this report; for example Pericardium 6 appears as P6, PC6, Pc6 and HP6.

The various abbreviations as they appear in the report are listed below:

Heart meridian points are abbreviated as: Ht; HT

Small Intestine meridian points are abbreviated as: Si; SI

Pericardium meridian (also known as Heart Protector, Heart Governor and Heart Constrictor) points are abbreviated as: P; Pc; PC; HP

Triple Heater meridian (also known as Triple Warmer) points are abbreviated as: TH; TW

Spleen meridian points are abbreviated as: Sp; SP

Stomach meridian points are abbreviated as: ST; St

Lung meridian points are abbreviated as: L; LU; Lu

Large Intestine meridian points are abbreviated as: LI; Li

Kidney meridian points are abbreviated as: K; KI; Ki

Bladder meridian (also known as the Urinary Bladder) points are abbreviated as: UB; BL; Bl; B

Liver meridian points are abbreviated as: Liv; LIV; LR

Gall Bladder meridian points are abbreviated as: GB

Conception Vessel meridian (also known as Ren Mai) points are abbreviated as: Ren; CV

Governing Vessel meridian (also known as Du Mai) points are abbreviated as: Du; GV

Chinese point names are given in some abstracts and the meridian/number format has been inserted where appropriate.

The “Third Eye Point” between the medial ends of the eyebrows on the bridge of the nose has no meridian/number format associated with it. It is named in the following ways in the text: Yintang; Extra 1

1. Executive summary

The aim of this evidence review was to identify and appraise scientific publications on the practice of Shiatsu in order to determine the direction of future research for the Shiatsu profession.

Comprehensive searches were conducted (Feb 1990-June 2006) of databases; MEDLINE, Cochrane, EMBASE, CINAHL, AMED, PsycINFO, BNI, Blackwell Synergy, Ingenta, Science Direct and Index to Theses. Acupressure and Shiatsu use the same points and are based on the meridian system of Traditional Chinese Medicine, but Shiatsu techniques cover more than just acupressure. On this basis, it was agreed that acupressure studies should be included in the review.

Initial search results identified 602 studies. After applying exclusion criteria and quality assessment, 5 Shiatsu and 41 acupressure publications remained for review and appraisal. The Shiatsu studies comprised three uncontrolled studies and two quasi-experimental studies. For acupressure, three were systematic reviews, 23 randomised controlled trials (RCTs), 14 quasi-experimental studies and one uncontrolled study.

The Shiatsu studies provide very limited evidence on a diverse range of health issues (angina, low back pain, fibromyalgia, chemotherapy side effects/anxiety and inducing labour). The methodological quality of these studies was generally poor.

Studies on acupressure provided fairly strong evidence for its use in the treatment of pain. Evidence for acupressure for nausea and vomiting was inconsistent, with the strongest evidence for post-operative nausea. Weak evidence for renal symptoms and COPD/asthma was found. The remaining acupressure studies provided evidence of variable quality on psycho-social health issues, consciousness/anaesthesia and other disparate health issues.

The methodological quality of studies and the health issues investigated were heterogeneous and therefore study results could not be pooled. The main methodological limitations of the studies identified included: small sample sizes, insufficient details on sampling and follow up, high drop out rates, uncontrolled design, and lack of blinding.

The research base for Shiatsu is very much in its infancy and the profession will need to work closely with its practitioners and researchers in order to build up evidence of effectiveness. Well-designed efficacy studies are needed on Shiatsu as an intervention.

Recommendations following this review include:

- Conduct further research on the effectiveness of Shiatsu as an intervention
- Encourage practitioners to engage in research using well designed studies
- Clarification of the relationship between Shiatsu and acupressure for marketing and public awareness
- Consider the development and piloting of an adverse event reporting system for Shiatsu
- Explore clinical and the cost effectiveness of Shiatsu in an integrated setting
- Identify specific topic areas for initial research investment
- Develop an evaluative framework for integrated Shiatsu practice
- Develop a research resource for the profession
- Investigate the appropriateness of various research methodologies for Shiatsu research

2. Introduction

The word SHIATSU is Japanese and means pressure ("ATSU") with fingers ("SHI"), i.e. "finger pressure". The term has been used over the last 200 years to describe the practice of a massage therapy which incorporates gentle manipulations and stretches combined with pressure techniques exerted through the fingers, thumbs, elbows, knees and feet. Shiatsu is an oriental medicine which has its roots in Chinese medicine and may even have pre-dated acupuncture. It embraces the philosophy of Yin and Yang, the energy meridians, the five elements and the concept of Ki, or energy. Practitioners use points on the meridians to rebalance the body's energy. These pressure points are known as "tsubos" in Japanese and are points that allow the therapist to act on the energy meridians. The concept of affecting the balance of energy through tsubos on the meridians is similar to that of acupuncture where needles are placed at these specific points or where heat is applied to chosen points on the meridians, and in Shiatsu where pressure is applied on both points and meridians.

However, more recently Shiatsu is known as a form of bodywork which primarily developed in Japan. It has been recognized by the Japanese Government as a therapy in its own right during the last 50 years[1]. It is now practiced in many European countries and was one of eight non-conventional, complementary medicine disciplines named in the Collins Report.[2]

Shiatsu has a number of different styles, philosophical approaches and theoretical bases. The Shiatsu Society UK encourages an eclectic outlook so that practitioners and students become familiar with and respect the different forms and styles of Shiatsu, The approaches most commonly found in Britain are Zen Shiatsu (most common), Macrobiotic Shiatsu, Healing Shiatsu, Tao Shiatsu, Seiki, Namikoshi Shiatsu and Hara Shiatsu.

Shiatsu aims is to balance, restore and maintain the body's energy balance and prevent the build-up of stress. It is used to treat a wide range of conditions, from specific injuries to more general symptoms of poor health, and is a deeply relaxing experience. Shiatsu is complementary to mainstream Western medicine, not an alternative. Some of the most common syndromes which may be amenable to treatment by Shiatsu include: headaches; migraine; stiff necks and shoulders; backaches; coughs; colds; menstrual problems; respiratory illnesses including asthma and bronchitis; sinus trouble and catarrh; insomnia; tension; anxiety and depression; fatigue and weakness; digestive disorders and bowel trouble; circulatory problems; rheumatic and arthritic complaints; sciatica and conditions following sprains and injuries.

Acupressure is the treatment through massage of specific pressure points as defined within with meridian system of Traditional Chinese Medicine (TCM). Acupressure is embedded within Shiatsu training, theory and practice. Shiatsu practitioners are trained in the anatomical location, functions and uses of over 150 pressure points on the body. Shiatsu and acupressure are not identical, since Shiatsu training and practice covers more than acupressure and includes diagnosis from the Hara, and the treatment of entire meridians as well as points. However, Shiatsu incorporates acupressure, and there is an argument to be made that evidence for the efficacy of acupressure can validly be used to support claims about the efficacy of Shiatsu for specific conditions. [3]

3. Aim

The aim of this report was to systematically review the current evidence base for Shiatsu by identifying relevant scientific publications and appraising the quality of the research published to date.

---

[1] Lundberg .P (1992) *The New Book of Shiatsu*. New York: Fireside Books.

[2] European Parliament (1997*). The Collins Report, Resolution on the Status of Non-Conventional Medicine*. Strasbourg (disseminated by the European Parliament, May 1999).

[3] Bewley D (2006) Director Shiatsu Society, letter to Committee of Advertising Practice, June 2006.

4. Objective

To inform future research directions for the Shiatsu profession in order to build their evidence base.
To support evidence driven practice, marketing and advertising of Shiatsu.
To support the development of Shiatsu training and education.

5. Methods and search strategy

5.1 Databases

Comprehensive searches were conducted for publications from February 1990 to June 2006. No date restrictions were set therefore initial results included publications prior to this date. The initial searches were conducted in February 2006. The MEDLINE searches were updated in March, April and June 2006 with a final update in August 2006 in preparation for reviewing the evidence. All search results were collated in individual Reference Manager® databases for review.

Search engines and journal databases accessed are listed below (Table 1)

Table 1 - Search engines and journal databases accessed:

Via PubMed:
MEDLINE
Via OVID:
EBM reviews (includes all Cochrane Library resources)
Allied and Complementary Medicine (AMED)
British Nursing Index (BNI)
Cumulative Index to Nursing & Allied Health Literature (CINAHL)
EMBASE
MEDLINE in process & non indexed
OVIDMEDLINE
PsycINFO
Journal databases:
Science Direct
Blackwell Synergy
Ingenta Select
Wiley Interscience

The following databases were also searched:

1. Index to Theses
http://www.theses.com/

2. ZETOC (British Library electronic table of contents)

The Shiatsu Society UK provided a copy of a commissioned report[4]

[4] Mackay H & Long A (2003) The Experience and Effects of Shiatsu: Findings from a Two Country Exploratory Study. University of Salford, UK

In addition, information and unpublished data was collected from the Shiatsu Society UK. The references of retrieved information were checked to identify any further studies. Any duplicates identified by systematically searching the database were removed.

5.2 Definition of search terms

Shiatsu was used as the main search term for most searches as it is included in the MeSH term 'acupressure' in MEDLINE. MeSH is the National Library of Medicine's (NLM) "controlled vocabulary used for indexing publications for MEDLINE/PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts" See Appendix 1
More information on MeSH – Medical Subject Headings can be found at:

http://www.nlm.nih.gov/mesh/meshhome.html

The full details of searches carried out are given in Appendix 2.

5.3 Assessment of the evidence

The stages used to assess the evidence are given in Appendix 3 and are shown graphically in the Flow Chart Figure 1. Abstracts were retrieved and reviewed against the inclusion criteria (Appendix 3) and if accepted into the review they were retrieved for classification and appraisal. Studies could be classified into one of the following: a systematic review, randomized controlled trial, and uncontrolled studies. Two reviewers independently categorized the evidence and an independent adjudicator used if there was any disagreement about inclusion.

As part of the review process, the references of any systematic or literature reviews and meta-analyses were checked against the search results to ensure accuracy of searches. It was during this process, that it was found that a small number of these references, relating to acupressure studies that had not been captured in any of the above searches. By obtaining MEDLINE abstracts in citation format, it became clear that acupressure was also included in a second MeSH tree and therefore not all of the acupressure citations in MEDLINE had been included in the initial searches in February 2006. See Appendix 4 for the second MeSH tree description and subsequent search details.

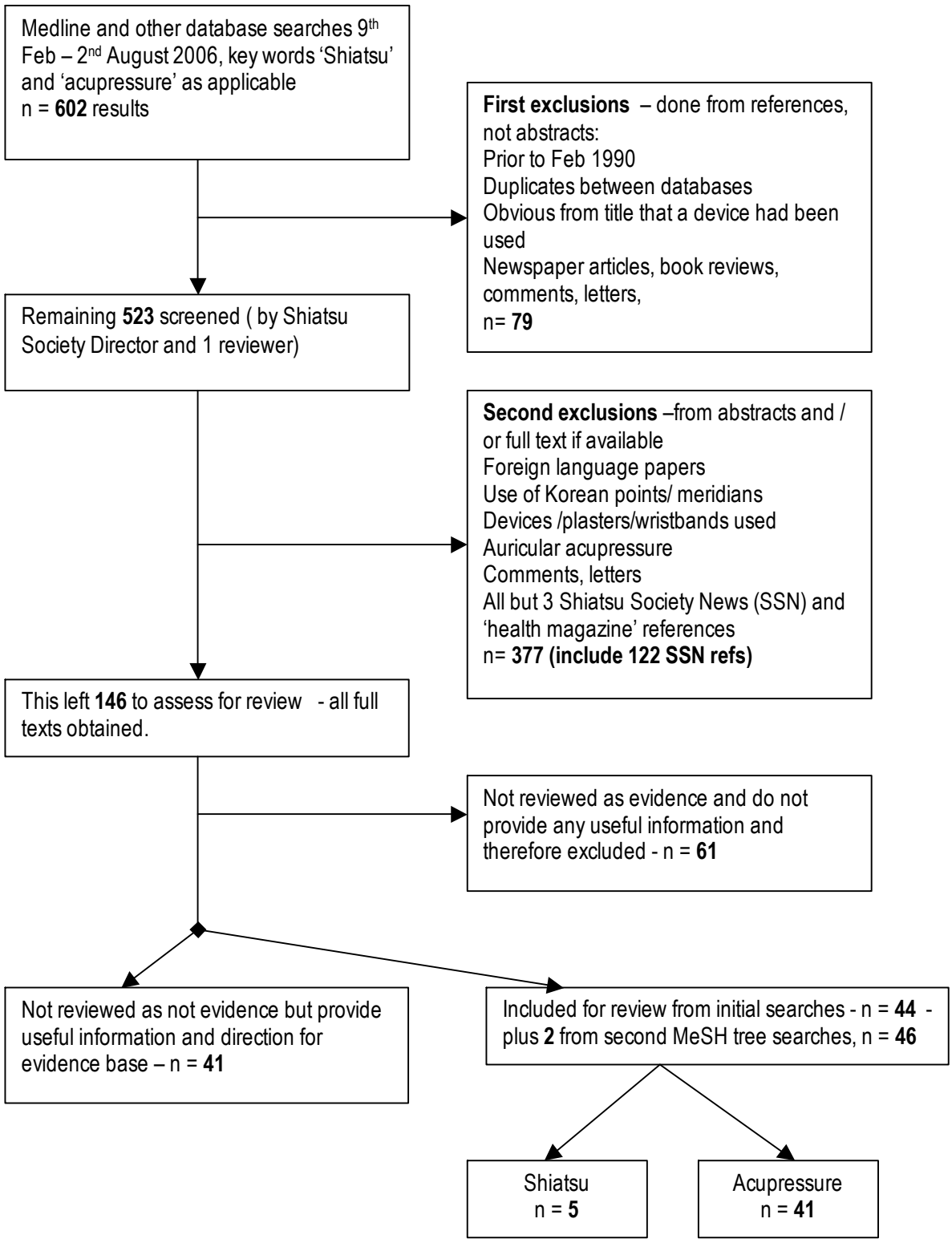
All relevant studies were appraised and their methodological quality assessed. The categorisation of the quality, weight and direction of evidence for each study was graded using criteria developed and adapted from Waddell[5].

Category 1: Generally consistent finding in a range of evidence from well-designed experimental studies
Category 2: Either based on a single acceptable study, or a weak or inconsistent finding in some multiple acceptable studies.
Category 3: Limited scientific evidence, which does not meet all the criteria of acceptable studies, or an absence of directly applicable studies of good quality. This includes published and unpublished expert opinion.

Relevant information was extracted independently by 2 reviewers using a standardised extraction form. See Appendix 5 and Figure 1.

[5] Waddell G, Feder G, McIntosh A, Lewis M, Hutchinson A. (1996) Clinical Guidelines for Management of Acute Low Back Pain (Low Back Pain Evidence Review). Royal College of General Practitioners. London.

5.4 Figure 1: Flowchart of evidence review process



6. Results and analysis

After carrying out the initial database searches, a total of 602 publications were identified which had a keyword of 'Shiatsu' and/or 'acupressure' (Figure 1) After duplicates between databases, comments, newspaper articles and letters were excluded (79), 523 publications were screened online using published abstracts and full articles where available. A further second screening using agreed exclusion criteria (Appendix 3) resulted in 146 publications for review. Full texts of the 146 publications were further screened by two reviewers.

After applying exclusion criteria and quality assessment, 5 Shiatsu and 39 acupressure publications remained for review and appraisal. Two acupressure studies were further added after the second MeSH term searches, leaving a total of 46 studies to review, 5 Shiatsu and 41 acupressure (Figure 1 and Appendices, 4, 6, 7.). The 46 included studies were critically appraised by two reviewers using the checklist in Appendix 8. Evidence tables of these publications were constructed (Appendix 9). Data collected on each study included: study design, setting, sample, health issue, analysis of results, conclusions and comments on quality.

A total of 13 Shiatsu and 48 acupressure publications were excluded as a result of screening the 146 publications. Details of the excluded publications can be found in Appendix 10. Note: two acupressure publications were excluded after the second MeSH tree searches.

Of the remaining 41 publications which were considered useful for background information, 22 referred to Shiatsu and 19 to acupressure. Details of these can be found in Appendix 11.

Reference Manager® databases for included, excluded publications and those providing background information were also constructed (Appendices 12, 13, 14). No further publications were included for review from the Index to Theses and ZETOC searches. Details of these searches can be found in Appendix 15.

The Shiatsu publications comprised three uncontrolled studies and two quasi-experimental studies (without a randomised control group). For acupressure, three were systematic reviews, 23 randomised controlled trials (RCTs), 14 quasi-experimental studies (without a randomised control group) and one uncontrolled study. The majority of studies used a standardised acupressure/Shiatsu procedure, only five studies were pragmatic.

It was felt inappropriate to combine the Shiatsu and acupressure studies in case there were differences in the techniques. Pooling of data or a meta-analysis of all included studies could not be carried out due to heterogeneity in study methodology, the range of health conditions studied, variety of interventions and outcome measures employed.

Tables 2 and 3 give the number of included Shiatsu and acupressure articles by health issue.

Table 2: Shiatsu studies by health issue

Health Issue	Number of studies
Angina	1
Low back pain	1
Fibromyalgia	1
Chemotherapy side effects/anxiety	1
Inducing labour	1
Total	**5**

Table 3: Acupressure studies by health issue			
Health Issue – Initial search results	**Sub sections**		**Number of studies**
Pain	Dysmenorrhoea	3	11
	Labour pain	3	
	Lower back pain	3	
	Minor trauma	1	
	Neck pain	1	
Nausea & vomiting	Post-operative	4	9
	Chemotherapy	3	
	Pregnancy	2	
Renal Symptoms			5
Chronic obstructive pulmonary disease/asthma			4
Anxiety/stress/sleep problems			3
Measures of anaesthesia/ Consciousness			3
Angina			1
Gastrointestinal motility			1
Gagging			1
Nocturnal enuresis			1
		Subtotal	**39**
Health Issue – Subsequent search			
Angina			1
Nausea & vomiting (pregnancy)			1
		Total	**41**

Full details of each included study, comprising methodology, design, sample, intervention, results, conclusion and critical appraisal comments, can be found in Appendix 9. A narrative summary of the studies is given below, which discusses the evidence found for Shiatsu and acupressure grouped according to the health condition investigated.

6.1 Shiatsu

The Shiatsu studies identified investigated quite separate health issues and did not use comparable methodology and therefore could not be pooled. These studies appear in alphabetical order in Appendix 9 and are discussed and summarised individually below.

Ballegaard et al (1996) investigated the effect of Shiatsu, with acupuncture and lifestyle adjustment, on patients with angina pectoris. The focus of this study was on cost benefit rather than efficacy. 69 consecutive patients were treated and compared with those from a separate trial of two invasive treatments for angina[6]. Incidence of death/myocardial infarction (MI) was 7% in this sample, compared to 21% and 15% in the comparison group (undergoing coronary-artery bypass grafting and percutaneous transluminal coronary angioplasty respectively). There was no significant difference in pain relief between groups. Additionally a cost-saving of $12000 per patient was estimated. This was a convenience sample and was not powered. The main flaws were the absence of an equivalent control group and lack of blinding. The comparison group were from the USA and the study was done in Denmark, additionally 56% of the participants would have been excluded from the one of the comparison groups. Also, due to the pragmatic design, it is difficult to isolate the effects of acupressure from co-interventions of acupuncture and lifestyle adjustments.

[6] King SB, Lembo N J, Weintraub W S et al (1994) A randomised trial comparing coronary angioplasty with coronary bypass surgery. *New England Journal of Medicine,* 331(16):1044-50

Brady et al (2001) administered Shiatsu massage to a convenience sample of 66 volunteers complaining of lower back pain. This was a single-group pretest-posttest design. Pain and anxiety significantly decreased after treatment ($p<0.001$), which did not change when demographic variables were controlled for. The absence of a control group and use of a volunteer sample who paid for treatment limits the validity of these results. 13 patients had previously received Shiatsu, further limiting the generalisability of findings.

Faull (2005) conducted a pilot study to compare the effectiveness of Watsu (water Shiatsu) to Aix massage for fibromyalgia syndrome (FMS). 17 female participants were randomly assigned to receive either Watsu then Aix or vice versa, with a 3 week break between treatment blocks. A significant improvement was seen after treatment with Watsu ($p=0.01$) for SF-36 subscales of physical function, bodily pain, vitality and social function, but not for Aix. This was only a pilot study and used a very small volunteer sample, only 13 of whom completed the study. No control group was used, although the counterbalanced design should reduce carryover effects of using repeated measures design. However, order effects may have occurred due to high dropout rate from Watsu first group (4 out of 8).

Iida et al (2000) investigated the relaxation effects of Shiatsu on anxiety and other side effects in patients receiving cancer chemotherapy. Nine patients were divided into strong anxiety or weak anxiety groups and all were given Shiatsu massage on the hands and feet. The strong anxiety group showed a significant decrease in anxiety after intervention ($p=0.09$). The weak anxiety group showed a significant increase in the relaxation score ($p=0.01$). There was a slight relief of physical symptoms in both groups but significance is not stated. This is a very small study, limiting the validity of results, aggravated by the further division of the sample into two groups, reasons for which are not clear. The use of the t-test on such a small sample will only detect differences that are huge, and may be the reason why few effects were seen. No control group was used.

Ingram et al (2005) investigated the effects of Shiatsu on post-term pregnancy in 142 women attending a consultant clinic appointment at 40 weeks gestation. Two groups were used, a Shiatsu group who received thumb pressure on points GB21, Li4 and Sp6 and who were taught breathing techniques and exercises. The control group received no intervention. The Shiatsu group was significantly more likely to labour spontaneously than the control ($p=0.038$) and had a longer labour ($p=0.03$). The main flaw was that groups were selected according to which midwife was on duty (only one midwife was trained in Shiatsu), although groups were homogenous for maternal age, parity and delivery details. The frequency of use of self-administered Shiatsu was not monitored. As a preliminary audit this study gives some useful results, although Shiatsu was not compared to a sham treatment.

There was insufficient evidence both in quantity and quality on Shiatsu in order to provide consensus.

6.2 Acupressure

The studies described as giving acupressure as an intervention form the second part of Appendix 9 and are in alphabetical order and summarised below.

6.2.1 Pain

Pain was the most common issue addressed by acupressure studies. These included studies on dysmenorrhoea (3 studies), labour pain (3 studies), lower back pain (3 studies), one study on minor trauma and one on neck pain. Seven of the 11 studies were RCTs, with control groups and random assignment; the remainder did not have a control and/or random assignment.

Dysmenorrhoea

Chen and Chen (2004) randomised 69 students with primary dysmenorrhoea into an intervention group who received acupressure at Sp6 and a control group who rested. Acupressure significantly reduced menstrual pain ($p<0.05$). The sample size was powered and 72% completed the study. The placebo effect was not controlled for as a sham treatment arm was not included. The generalisability of the findings is limited as the participants were volunteers and aged 17-19.

Jun et al (2006) carried out a controlled trial of acupressure compared to light touch at Sp6 for primary dysmenorrhoea, also on a sample of students. Sample size (58) was powered. The severity of dysmenorrhoea was significantly reduced in the acupressure group compared to control ($p=0.000$) and this effect lasted for up to 2 hours after treatment ($p=0.032$). Allocation to study groups was performed sequentially not randomly, although groups were homogenous in their baseline demographics and the factors affecting dysmenorrhoea. Students and data collectors were blinded. In both this and the study above, Hawthorne bias may be present as it is possible (although not stated) that the participants were students of the researchers.

Pouresmail and Ibrahimzadeh (2002) carried out a three-armed RCT of 216 high school students (aged 14-18 years), to compare the effects of acupressure, acupressure at sham points and Ibuprofen on primary dysmenorrhoea. Results indicated that all three techniques significantly reduced pain ($p<0.01$). Both acupressure and Ibuprofen were better than placebo. This is a high quality study with random group assignment and a large sample with a low attrition rate. However, the validity of the outcome measures was not disclosed and it is not clear if blinding was used.

Labour pain

Chung et al (2003) randomly assigned 127 parturient women to an intervention group who received acupressure at Li4 and BL67, placebo group who received light skin stroking at these points and a control group (conversation only). All groups showed a significant decrease in labour pain during the active first phase of labour ($p=0.041$) and acupressure was significantly more effective than control ($p=0.017$) but not compared to light stroking. This indicated that effects of acupressure may be due to tactile stimulation rather than meridian effects. Additionally a third of women receiving acupressure qualitatively reported that it had reduced their pain. This is a high quality three-armed RCT, with homogenous groups, although sample size in each group was only 42/43 and response rate was very low at the transitional phase of labour (31 out of 127). Both the outcome measure (VASs) and the acupressure procedure were shown to be reliable and valid by the researchers. The three steps to ensure validity and reliability of the acupressure were 1) Protocol was established by experienced Chinese physicians and using a pilot study; 2) Intrarater reliability test was used to control pressure force, measured for each practitioner and three experts evaluated the accuracy of the acupoint location for each practitioner; 3) Practitioners underwent a 2 hour training session and monthly meetings.

Lee et al (2004) conducted a double-blind RCT of acupressure compared to touch on Sp6 acupoint for labour pain. A volunteer sample of seventy-five women in labour were matched for five characteristics of labour and randomly assigned. There were significant differences between the groups in subjective labour pain scores immediately after the intervention ($p=0.012$), 30 mins after ($p=0.021$) and 60mins after ($p=0.012$). Anxiety was also significantly lower in the acupressure group compared to the control ($p=0.03$). Groups were homogenous. Bias may be introduced by using a volunteer sample. Blinding was used where possible (patients and data collectors) and the use of a placebo treatment controlled for the emotional supportive effects of human touch.

Waters and Raisler (2003) used ice massage on acupoint Li4 during labour contractions in a one-group pretest-posttest study. As measured by the visual analogue scale (VAS), pain was reduced after the intervention. This study had a number of methodological limitations, the main flaw being the absence of a control group. In addition, no sample size is given, convenience sampling was used, and only early labour pain was investigated due to the limitations of the outcome measure.

Lower back pain

Hsieh et al (2004) conducted an RCT of acupressure compared to physical therapy for chronic low back pain. 146 participants were randomly assigned to receive four weeks of either acupressure or physical therapy (thermotherapy, infrared, electrical stimulation, exercise and traction). Mean post-treatment pain scores were significantly lower in the acupressure group ($p=0.0002$) and also after 6 months ($p=0.0004$). This is a high quality trial with a powered sample (although convenience), homogenous groups, valid outcome measures and using intention to treat analysis to protect against attrition bias. Blinding was used where possible; practitioners and patients were blinded to pretest scores and follow-up staff were blind to treatment allocation. Although no placebo treatment was used, it can be assumed that physical therapy is usual care in Taiwan. This study was pragmatic as acupressure treatment was individualised rather than using a standardised protocol.

Hsieh et al (2006) conducted another RCT of acupressure compared to physical therapy for chronic low back pain, on 129 orthopaedic outpatients. The methodology was very similar to Hsieh et al 2004, comparing acupressure to physical therapy in randomised groups. This study also showed significantly lower pain and disability scores in the acupressure group compared to physical therapy ($p<0.05$). Again, no placebo treatment was used and the treatment was pragmatic rather than standardised. As it used the same methodology, this study is of a similarly high quality to Hsieh et al (2004).

Yip and Tse (2004) randomly assigned 61 adults with sub-acute or chronic low back pain into an intervention or control group (usual care only). The intervention consisted of acupoint stimulation using an electronic device on acupoints Li10, Li11, Si10, TW15 and BL10 and acupressure with lavender oil on UB 22, 23, 25 and 40. The intervention group showed a significant reduction in pain intensity compared to the control ($p=0.0001$) but not for duration of pain. The sample size was powered, however participants were volunteers, 16% dropped out and these were older which may cause bias. It is difficult to isolate the effect of acupressure in this study, due to co-interventions of electrical stimulation and lavender oil, and due to the lack of a control treatment.

Neck pain

Yip and Tse (2006) used the same protocol as above to treat 28 adults with sub-acute non-specific neck pain. The acupressure group showed a significantly greater reduction in pain than control ($p=0.001$). Although group assignment was random, this trial used a very small sample and no blinding or placebo. Again, it is difficult to isolate the acupressure effect.

Minor trauma

Kober et al (2002) conducted a double-blind RCT with 60 minor trauma patients who were randomly allocated to acupressure, sham acupressure or control groups. All were treated for 3 minutes during transportation in ambulances. At the end of transport they found significantly less pain, anxiety and heart rate in the acupressure group but not in either sham or control groups. Sampling bias may be present as eligible patients were purposively selected by paramedics. All groups were homogenous, the trial was truly double blinded (paramedic giving treatment and patient) and intention to treat analysis was used, although there were no dropouts.

Overall, the evidence for the efficacy of acupressure for pain is fairly strong and can be graded as category 1 evidence (*generally consistent findings in a range of evidence from well-designed experimental studies*) (see p. 3). Although some studies have methodological flaws, a number of RCTs consistently show that acupressure is more effective than control for reducing pain, namely dysmenorrhoea (Chen & Chen 2004; Jun et al 2006; Pouresmail & Ibrahimzadeh 2002), lower back pain (Hsieh et al 2004; Hsieh et al 2006; Yip & Tse 2004) and labour pain (Chung et al 2003; Lee et al 2004).

6.2.2 Nausea & vomiting

Nausea and vomiting was the second most common health issue to be studied. We found 10 studies, which investigated nausea and vomiting in three main situations; post-operative including caesarean (four studies), as a side effect of chemotherapy (three studies) and during pregnancy (three studies). Nearly all studies used the P6 acupoint.

Post-operative

Chen et al (2005) investigated the use of acupressure at P6 on reducing nausea, vomiting, anxiety and pain in 104 post-caesarean women. They found that acupressure significantly reduced nausea, vomiting and retching up to 10 hours post-caesarean compared to a control group who received standard care. Anxiety and pain were also reduced. Although this study had a fairly large sample and a control group, a convenience sample was used and group assignment was not random (first 52 recruited were in intervention group). Although this may introduce seasonal/time-related bias, this was in order to prevent participants discussing the study and groups were shown to be homogenous for demographic and physiological variables and pre-test scores.

Two reviews for postoperative nausea and vomiting were found, one systematic review (Lee & Done 2004) and one meta-analysis (Shiao & Dune 2006). Lee & Done found 26 trials specifically using P6. Although studies were heterogeneous, they concluded that acupressure reduced the risk of both nausea and vomiting compared to sham treatment, and reduced the risk of nausea but not vomiting compared to antiemetic medication. As a Cochrane review this is a high quality systematic review, which used comprehensive search terms and combined data from the trials. It was limited to acupoint P6. Shiao and Dune pooled the data on 33 trials using some form of acupoint stimulation versus placebo or control, 30 of which used the P6 acupoint. Two further trials compared acupoint stimulation to medication. Their results showed that all modalities of acupoint stimulation were effective in reducing postoperative nausea and vomiting compared to controls, and as effective as medication. This is a well conducted meta-analysis using comparable studies and a good selection process. 18 of the trials were for acupressure, providing a large body of evidence in this area, although most of these used bands to apply pressure. The pooled data from these studies showed that acupressure reduced nausea ($p<0.0001$) and there was no evidence of bias.

Ming et al (2002) conducted a randomised block experiment comparing finger-pressing, wrist-band and control (conversation only) in a sample of 150 patients undergoing endoscopic sinus surgery. They found that post-operative nausea and vomiting were significantly different between the three groups ($p=0.001$ and $p<0.001$ respectively). This study has a good sample size and very low attrition (98.7% follow up) but was not blinded. Patients were matched for motion-sickness before being randomly assigned (it is unclear why this was variable was used) and groups were homogenous. Although internal validity was high, the study was not blinded which may have introduced placebo/observer bias.

Chemotherapy

Acupressure for nausea as a side-effect of chemotherapy was investigated by Dibble et al (2000), Ezzo et al (2006) and Shin et al (2004).

Dibble et al (2000) conducted a pilot RCT of 17 women undergoing chemotherapy for breast cancer in oncology outpatient clinics. Patients were randomised (stratified based on setting and treatment regimen) to receive usual care or usual care plus acupressure at P6 and ST36. Nausea experience and intensity were significantly reduced in the acupressure group ($p<0.01$ and $p<0.04$ respectively). Results of this study are inconclusive due to the very small sample size and the lack of a placebo treatment (discussed as unethical), although groups were homogenous. Also, the Hawthorne effect may have been present due to the extra attention given to the treatment group.

Ezzo et al (2006) conducted a Cochrane Systematic review on 11 trials of acupoint stimulation for chemotherapy-induced nausea and vomiting. Pooled data showed that all methods combined reduced the incidence of acute vomiting ($p=0.04$), but not severity of nausea compared to control. Acupressure reduced mean acute nausea severity ($p=0.04$) but not acute vomiting or delayed symptoms, although studies did not use placebo controls. This is a well conducted review which reports all methodological details. Data was pooled using intention to treat analysis and using original data where possible. Additionally duplicate bias and language bias were controlled for. Evidence for acupressure is however limited as the review included all acupoint stimulation (including acupuncture) only three of which were acupressure trials and which include those which used bands.

Shin et al (2004) compared the effects of self-acupressure on P6 with antiemesis medication to medication alone, in a sample of 40 postoperative gastric cancer patients receiving the first cycle of chemotherapy. A significant reduction was found between intervention and control groups in the severity of nausea and vomiting, duration of nausea and frequency of vomiting (all $p<0.01$). Although these results are highly significant, a number of methodological issues are present. The sample is small and convenience sampling was used, group allocation was also not random (allocation used, first 20 patients in control group), although groups are homogenous for demographic, disease and treatment variables. Again, the intervention group had additional attention, which may have introduced the Hawthorne effect.

Pregnancy

Three studies investigated nausea and vomiting in pregnancy; Habek et al (2004) looked at hyperemesis gravidarum (HG), which is a more severe and rare form of the nausea and vomiting investigated by Markose et al (2004) and Belluomini et al (1994).

Habek et al (2004) randomised 36 pregnant women with HG to four groups; acupuncture, placebo acupuncture, acupressure and placebo acupressure. Results showed that acupressure significantly reduced the occurrence of HG ($p<0.01$). This study was double-blinded which is unusual in these studies. Sampling was not given, but group allocation was random. Statistical analysis of group composition was not performed. The main flaw with this study is the small sample, which is then divided into four groups, so the power in each group is very low. Also, the outcome measure appears to be simply the disappearance of nausea and vomiting as assessed by the patient and gynaecologist, which is subject to bias. The acupressure protocol was not controlled and was self-administered.

Markose et al (2004) conducted a one group uncontrolled study of acupressure on P6 for nausea, vomiting and dry retches in 35 women pregnant under 12 weeks. After treatment (day 7) there was a significant reduction in the frequency of nausea from day 3 (before treatment) ($p=0.008$), vomiting ($p=0.000$) and retching ($p=0.016$). This study was of poor quality as it was uncontrolled and used a very small sample. In addition, only 17 of the 35 women completed the study. Sampling procedure is not given.

Belluomini et al (1994) randomised 90 pregnant women (12 weeks gestation or less) to receive either acupressure at Pc6 or sham acupressure at a non acupoint. Both groups showed significant

reduction in nausea and emesis over time, but this improvement was significantly greater in acupressure group (p=0.0021) than control. There were no differences in severity or frequency of emesis between groups. The sample was selected from referred patients, details of this are not clear. Only 60 out of 90 completed the study and intention to treat analysis was not used. Drop out was however similar between study groups. A randomised block design was used which can give more powerful treatment effects, but criteria for blocking were not given (may be gestational age). Groups homogenous for pregnancy characteristics and pre-test scores. Maternal age was associated with nausea and vomiting score. Gestational age was controlled for. This study was single blind and used a sham treatment arm. Acupressure was self administered acupressure and reliability of the procedure was not checked.

In summary, the evidence for acupressure for nausea and vomiting is somewhat inconsistent and varies with type of nausea investigated. Studies investigating post-operative nausea provide the strongest evidence, which can be as graded as Category 1 evidence (see protocol 3b) as the studies are generally well designed (Chen 2005; Ming 2002), and include a systematic review (Lee & Done 2004) and a meta analysis (Shiao & Dune 2006). The two trials reviewed for chemotherapy-induced nausea and vomiting (Dibble 2000; Shin 2004) give little reliable evidence, mainly due to small sample size, and although the Cochrane review (Ezzo 2006) gives quality evidence, little of it is on true acupressure. The three studies of acupressure for nausea in pregnancy are of poor quality with small samples and/or uncontrolled study design (Belluomini et al 1994; Habek 2004; Markose 2004)

6.2.3 Renal Symptoms

Five studies were identified which investigated the use of acupressure for renal symptoms. All of these studies have a number of similarities as Tsay SL was lead researcher in four and co-researcher in the fifth study.

Cho and Tsay (2004) randomly assigned 62 haemodialysis patients to acupressure and control groups to test the effect of acupressure on fatigue and depression on End-Stage Renal disease (ESRD). Acupressure group received acupoint massage on zusanli (St36), sanyinjiao (Sp6), taixi (Ki3) and yung chuan (Ki1) while the control group received routine care. Results showed a significantly greater reduction in fatigue ($p<0.004$) and depression ($p=0.045$) in the acupressure group than the control. Sample size was powered, group assignment random, and treatment groups were homogenous except for age. Differences in pre-test scores and age were also controlled for. The extra attention the treatment group received may have had an effect as a sham treatment arm was not included.

Two articles by Tsay & Chen (2003) and Tsay et al (2003) appear to be based on the same RCT of acupressure for quality of sleep in ESRD patients, but are included as separate studies as they were published as individual papers. However, we have only described the methodology/quality once below, as details are identical. 98 ESRD patients from four hospitals were randomly assigned into three groups, acupressure (on points H17 and K11), sham (massage not on acupoints) and control (standard care). Results indicate that improvement in quality of sleep was significantly greater in acupressure compared to control ($p<0.01$). However there were no differences between the acupressure and the sham group, or the sham and control group, except that subjective sleep quality was improved in the sham group compared to the control ($p=0.003$). Blinding was used, for interviewer/data collector, usual care provider and participant, but not acupressure nurse. The outcome measures and acupressure procedure were reliable. Bonferroni correction controlled for type 1 error. Group assignment was random, and groups were homogenous for demographics, sleep affecting behaviour and ESRD related factors. Attrition was low (98 from 105).

Tsay (2004) conducted an RCT of 106 ESRD patients, investigating acupressure for fatigue. Again, patients were randomised to three groups, acupressure, sham and control. Acupoints Ki1, St36, GB34 and Sp6 were used. Results, adjusted for differences in baseline fatigue, showed that patients in the acupressure group ($p=0.01$) and sham group ($p=0.003$) both had significantly lower

fatigue scores than control. Although reduction of fatigue was greater in acupressure than sham groups, this difference was not significant, indicating that non acupoints massage also had an effect on reducing fatigue. Participants were not blinded; the researchers stated that participants knew which group they were in. Control and intervention groups were demographically and clinically homogenous and co-variables of depression and quality of sleep were controlled for in analyses. The reliability and validity of the procedure was evaluated by expert validation, and the internal consistency of the outcome measures was good.

Tsay et al (2004) tested the effects of acupuncture or transcutaneous electrical acupoint stimulation (TEAS) on fatigue, sleep quality and depression in a prospective RCT. They randomly assigned 106 haemodialysis patients to three groups to receive either acupressure or TEAS on points Ki1, St36, GB34 and Sp6 or control who received routine care only. Acupressure and TEAS patients had significantly lower fatigue (p=0.05 and p=0.016 respectively) and less depressed moods (p=0.009 and p=0.008 respectively) than control, adjusted for baseline differences. There were no differences between acupressure and TEAS groups. This study used random group assignment and three arms, with homogenous groups. It also had a powered sample size and low attrition rate (2 out of 108). The reliability and validity of the procedure was evaluated by expert validation, and the internal consistency of the outcome measures was good. However, no details of blinding are given and results are limited to haemodialysis patients.

These five studies provide category 2 evidence for the use of acupressure for renal symptoms (*evidence based on a single acceptable study, or a weak or inconsistent finding in multiple acceptable studies*) (see p. 3). This categorisation has been chosen mainly because they do not represent a range of studies, as all are fairly similar in design and setting and mainly led by one researcher. The individual studies provide some evidence for the efficacy of acupressure for ESRD/haemodialysis patients, but this is limited in generalisability. Although they did control for a number of factors and biases, most studies were not blinded which reduces the quality of the evidence.

6.2.4 Chronic obstructive pulmonary disease/asthma

Three studies on chronic obstructive pulmonary disease (COPD) and one on chronic obstructive asthma were identified (Maa et al 1997; Maa et al 2003; Tsay et al 2005; Wu et al 2004).

Maa et al (1997) investigated the effects of self-administered acupressure on reducing dyspnoea and other associated symptoms in 31 patients with COPD. Patients were those beginning a pulmonary rehabilitation program at two private hospitals and acupressure was used as an adjunct to standard care. The study was a pretest-postest crossover design; group 1 had 6 weeks of acupressure followed by sham acupressure and group 2 vice versa. Real acupressure was more effective than sham for reducing dyspnoea (p=0.009) and minimally effective for reducing decathexis (p=0.044) but had no effect on any other symptoms. This study had a small sample, although sensitivity analysis was performed and did not identify any idiosyncratic individuals. Also, dropout was high (20 of 51), and mostly due to medical reasons, which may have biased the results. Also the study was only single-blind and many patients could identify sham from real acupressure. The crossover design should reduce effects of retesting, carryover or time-related effects of patients acting as their own controls, which controls for heterogeneity. Outcomes were valid and reliable.

Maa et al (2003) conducted a pilot randomised trial of acupuncture and acupressure for improving the quality of life of patients with chronic obstructive asthma. 41 outpatients were randomly assigned to receive acupuncture, acupressure or control, all groups received standard care. Acupressure patients had a significantly greater reduction in health related quality of life (p=0.05) and in irritability (p=0.06) but not in any other scores. Again, this study had a small and purposive sample, although again, sensitivity analysis was performed and did not identify any idiosyncratic individuals. There was also a very high attrition rate (41%), which was again mostly due to medical reasons and dropout was also greater from the acupuncture group. Intention to treat analysis was not used. The study was not blinded.

Tsay et al (2005) used a two group experimental blocking design to investigate acupressure (at points Li4, PC6 and HT7) for dyspnoea, anxiety, heart rate and respiratory rate in patients with COPD. 52 patients, all on mechanical ventilation support, were matched for sex, age and length of ventilation use then randomly assigned to acupressure or control (massage and handholding) groups. Dyspnoea ($p=0.009$), anxiety ($p=0.011$), heart rate ($p=0.005$) and respiratory rate ($p<0.0001$) improved significantly in the acupressure group compared to control. This study had a powered sample although there was no information about dropout. The groups were homogenous for demographic and clinical factors. Clinical outcome measures were used and the procedure was reliable and validated by experts. Patients, data collectors and usual care givers were blinded, but not acupressure nurses or researchers.

Wu et al (2004) matched 44 outpatients with COPD for age, sex, pulmonary function, smoking and steroid use then randomly assigned them to receive acupressure (points GV14, CV22 , B13, B23, L10) or sham acupressure (Sp5, Sp3, Li1). Scores from the Pulmonary Status and Dyspnoea Questionnaire modified scale showed that the true acupressure group improved significantly more than the sham group for all three subscales; dyspnoea ($p<0.05$), fatigue ($p<0.01$) and activity ($p<0.001$). Tolerance for activity also significantly improved ($p<0.001$) as did anxiety ($p<0.001$). Although this study used a small sample, the randomised block design should give more powerful treatment effects. The sham points were on different meridians and ganglionic sections, indicating the efficacy of the specific acupoints chosen. The acupressure protocol was highly reliable and validated by the researchers, using three tests: 1) independent rating for validity to achieve 100% agreement, 2) observation of accuracy of points by TCM practitioner and 3) true and sham points compared on video for homogeneity in timing. The outcome measures were also reliable and valid.

Overall, the evidence for acupressure for COPD/asthma is category 2 evidence (see p. 3) as there are only a small number of studies and these have a number of methodological flaws. All studies had fairly small samples, and two of the identified studies had a high dropout rate which is likely to have biased their results (Maa et al 1997; Maa et al 2003).

6.2.5 Anxiety/stress/sleep problems

Three studies were identified which investigated psycho-social aspects of health, specifically pre-operative anxiety, quality of sleep and alertness.

Agarwal et al (2005) conducted an RCT with 76 adults undergoing elective surgery. Patients were randomised to receive acupressure at Extra 1 point or sham acupressure at an inappropriate site. Anxiety decreased in both groups, but both returned to baseline after 30 minutes. The decrease in anxiety was greater in the Extra 1 group ($p<0.05$). Bispectral index values were also lower during treatment in both groups, and were lower for Extra 1 group ($p<0.05$). The sample size was powered, group allocation was random and groups were homogenous. A sham treatment arm was included, although there was no other control group. The study was single-blinded (patient).

Chen et al (1999) performed a three armed RCT testing the effectiveness of acupressure in improving the sleep quality of institutionalised residents. 246 elderly residents with sleep disturbances were matched for hypertension, hypnosis, naps and exercise then randomly assigned to acupressure (points baihui (GV20), fengchi (GB20), anmian (BL18) and shenmen (Ht7) x2), sham (1cm-3cun from real points) or control (conversation) groups. Quality of sleep improved in all three groups and improvements were significantly greater in acupressure group (scheffes post hoc comparison). This was a high quality trial, with a large sample size, systematic random sampling and random group assignment, matched to give more powerful treatment effects. The control and intervention groups were homogenous for a huge range of factors (demographics age, gender, living conditions, drug use, chronic disease, time at facility, naps, exercise, time in bed, milk tea and coffee consumption, smoking, sleep indices).

The internal validity of the procedure was extensively controlled by inter-rater reliability and expert validation. However, the study was only single-blind and the principal investigator, who knew the participants, both administered treatment and collected data, which may introduce Hawthorne effect and researcher bias. Generalisability is limited as setting was a very specific home for elderly people with low income and without a son.

Harris et al (2005) used a crossover design to test acupressure to modify alertness in the classroom. They randomly assigned 39 students to two acupressure treatment sequences: stimulation-relaxation-relaxation or relaxation-stimulation-stimulation. Compared to relaxation, stimulation acupressure gave a greater alertness score ($p=0.019$). Day of study and hours of overnight sleep also significantly affected the score. The study was single-blind (subjects), although the majority of students could correctly discern the treatment. This did not significantly affect the results, although it came close, raising p to 0.0484. There is a chance that participants were students of the researchers, it which case it would appear that Hawthorne effect may be present. Small sample size (39) and low generalisability as all medical students (well educated, scientific researchers who were highly motivated to comply) were also issues. Group allocation was random and control and intervention groups were homogenous. Crossover design should reduce effects of retesting, carryover or time-related effects, although participants acting as their own controls can cause practise effect (especially with self-report). Validity of the outcome measure was not given. Nine students provided missing data retrospectively which may cause recall bias. Statistical analysis was very comprehensive, accounting for effects of sequence, period, treatment and 'other covariates', masking, and co-variates, including caffeine, sleep, medication, anxiety and compliance.

The quality of studies for acupressure for psycho-social health issues is variable; reliable conclusions cannot be drawn from the existing evidence base, although evidence for improving sleep quality in institutionalised elderly is strong (Chen et al 1999).

6.2.6 Anaesthesia/consciousness

Three studies have investigated the effects of acupressure on levels of anaesthesia or consciousness. These levels include the acoustic evoked potential (AEP), changes in which reflect the depth of anaesthesia and transition from awake to anaesthetised (Dullenkopf et al 2004); bispectral index (BIS) and spectral edge frequency (SEF) which are measures of the level of consciousness during anaesthesia/sedation (Fassoulaki et al 2003; Litscher 2004).

Dullenkopf et al (2004) used a repeated measures, counterbalanced design to investigate the influence of acupressure at Extra 1 point on the AEP of unsedated adult volunteers. 15 volunteers received acupressure at Extra 1 point followed by acupressure on a control point the following day or vice versa, the order was chosen randomly. Subjects acted as their own controls and results showed that AEP reduced significantly after 10 minutes of pressure on Extra 1 point ($p=0.0044$), but this effect only lasted for 5 minutes. Stress levels were also reduced ($p=0.0066$). This study had a very small sampling and no details of sampling were given. Patients acted as their own controls, which can cause danger of attrition, and practice effect. It can also cause carryover effects but these should be addressed by counterbalancing, as there were no differences in changes in AEP between participants who had Extra 1 or sham acupressure first.

Fassoulaki et al (2003) used a similar repeated measures design to give 25 volunteers acupressure on extra 1 point or a control on alternate days in a randomised manner, with the aim of reducing self-reported stress levels. BIS was significantly reduced during pressure on extra 1 point ($p<0.001$) but returned to baseline after pressure release. Pressure on the control point also reduced BIS but reductions from extra 1 were greater ($p<0.001$). Sample size is small, although it was powered. Again, sampling/follow-up details are not given. Participants were excluded if they believed in Traditional Chinese Medicine theory, which may well bias results. Acupressure was given for 10 minutes at Extra 1 and for only 5 minutes in the control group, which is a major flaw. Again, patients acting as their own controls can cause danger of attrition, practice and carryover effects.

Litscher (2004) conducted a crossover trial of acupressure on yintang, acupuncture, laserneedle acupuncture and sham acupressure on the BIS and SEF in 25 healthy volunteers. Participants each received all four interventions, the order of which was randomised for each patient. Results showed a significant reduction of BIS and SEF during acupressure ($p=0.001$). Stress was also reduced by acupressure ($p<0.001$) but also by sham acupressure ($p<0.012$). This was a volunteer sample and quite small, also participants were paid to take part. Subjects and data collectors were blinded. Again, the use of subjects as their own controls raises issues of bias, especially as subjects only had 20 minutes between treatments so treatment effects may overlap.

Overall, the evidence for the effects of acupressure on consciousness/anaesthesia is weak, rated as category 3 (see protocol 2b) as only three studies have been identified, all of which use a repeated measures design rather than RCT and small sample size (Dullenkopf et al 2004; Fassoulaki et al 2003; Litscher 2004).

6.2.7 Other conditions

The remaining five articles on acupressure investigated distinct health conditions which were not grouped but are considered separately below.

Ballegaard et al conducted two studies of acupressure for angina (1999 and 2004). The 1999 study was mainly a cost-benefit analysis of using acupressure as part of a self-care program for outpatients with angina pectoris. 105 patients were given acupressure at CV17, UB14 and UB15, along with acupuncture and a range of other lifestyle modifications based on self-care. Three groups were used for comparison of risk; published data on invasive treatments[78], a random sample of the Danish population and the group used for this study. The intervention group had a 90% reduction in hospitalisation and a 70% reduction in needed surgery. Medication intake and degree of disease were significantly reduced and quality of life improved after treatment (all $p<0.0001$). The risk of cardiac death or myocardial infarction was lower in the treatment group than general population (significance not given). As this study was designed as a cost benefit analysis rather than an efficacy study, a different study design may have given different results, the main problem being the use of non equivalent control groups. Also the sample was volunteer and convenience. It is difficult to isolate the effects of acupressure from co-interventions of acupuncture and the self-care program.

Ballegaard et al (2004) investigated the long-term effects of acupressure, as part of integrated rehabilitation (IR), to reduce the risk of angina pectoris sufferers dying from myocardial infarction (MI). 168 patients (103 candidates for surgery, 69 inoperable) with angina in a private clinic in Denmark received 12 sessions of IR which included acupressure at CV17, UB14 and UB15. 3 historical controls were used; 1) General Danish population, 2) New York clinical database[9] and 3) Patients who underwent surgery from a study in New York[10]. The 3 year accumulated risk of death was 2% (confidence limits 0 - 4.7%) for the 103 surgery candidates, compared to 6.4% (confidence limits 4.7 – 6.1%) for the Danish population and 8.4% (confidence limits 7.7 – 9.1%) for New York surgery patients. Risk of death was 7.7% (3.9-11.5%) for the 69 inoperable patients, compared to 16% (10-34%) and 25% (18-36%) for American patients treated with laser surgery or medication respectively.

[7] King et al (1994) A randomised controlled trial comparing coronary angioplasty with coronary bypass surgery. *N Eng J Med,* 331:1044-1050

[8] Yusef et al (1994) Effect of coronary artery bypass graft surgery on survival: Overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet,* 344:563-570

[9] Hannan, Racz M, McCallister B et al (1999) A Comparison of 3 year survival after coronary artery bypass graft surgery and percutaneous transluminal coronary angioplasty. *Jour Am Coll Cardiol*, 33(1):63-72

[10] Schofield P M, Sharples L D, Caine N et al (1999) Transmyocardial laser revasculisation in patients with refractory angina: A randomised controlled trial. *Lancet,* 353:519-524

The accumulated risk of operation/MI/death was reduced in groups who had undergone treatment for longer ($p<0.05$ for trend). In addition, the IR program resulted in cost savings of $36000 and $22000 for surgical/inoperable patients respectively, although these costs were based on an American study. This study had a good sample size although sampling was not random and a very long follow-up period. The main flaws are the absence of an equivalent control group and lack of blinding. Also, due to the pragmatic design, it is difficult to isolate the effects of acupressure from co-interventions of acupuncture, self-care program including Chinese health philosophy, stress management and lifestyle adjustments. The sample was not significantly different in baseline variables to Scandinavian heart patients. The researchers described this study as a quality control review, which is subject to selection bias, expectation bias and social bias as patients have chosen and are paying for the treatment. However, evidence is cited which claims that no bias is introduced by patients choice of a particular treatment or paying for treatment[11].

Chen et al (2003) conducted an RCT using acupressure to improve gastrointestinal (GI) motility in women after trans-abdominal hysterectomy. 41 patients were randomly assigned to intervention (acupressure on Pc6, St36, Sp6) and control (acupressure on sham points) groups. The acupressure group had significantly improved GI motility ($p<0.05$), higher self-awareness of GI motility ($p<0.05$) and satisfaction ($p<0.001$) compared to control. The sample was small and not powered and the study was only single-blind. However, groups were homogenous for a wide range of factors (identified from previous research) (demographics, bowel movements, GI history, surgery history, duration of surgery, blood loss, analgesics, pain, post-surgical activities, leaving the bed and food intake patterns). The reliability of the procedure was verified by training, expert verification of point location and GI motility.

Lu et al (2000) investigated the anti-gagging effects of acupressure in 109 dental patients. Patients were randomly assigned to three groups; acupuncture (P6 or sham point), acupressure (P6 or sham) and pharmacological sedation with either acupressure or acupuncture. Acupressure was additionally performed with either thumb, device or Sea-band. There was a significant reduction in gagging with acupuncture (team evaluation $p=0.047$, patient $p=0.009$) and with device acupressure (team $p=0.002$, patient $p=0.001$) at P6 versus sham point, but no other significant differences for acupressure (using thumb or Sea-band). The study was described as double-blind although blinding procedures are not evident. The use of so many comparison groups results in very small group size (between 9 and 18). The outcome measure (subjective rating by the dental team and patient) was not validated and may not be reliable. Groups were not compared for homogeneity in baseline characteristics.

Yuksek et al (2003) randomised 24 patients to receive either acupressure or oxybutinin for nocturnal enuresis. Acupressure was applied to points Gv4, Gv15, Gv20, B23, B28, B32, H7, H9, St36, Sp4, Sp6, Sp12, Ren2, Ren3, Ren6, K3 and K5. There were no significant differences in incidence of bed-wetting between groups after treatment. No bed-wetting was seen in 83.3% of children who received acupressure and 58.3% who received oxybutinin. The main flaw was the very small sample size, with no details of sampling, comparison of groups or randomisation. Additionally selection bias was introduced from moving 3 patients who had previously unsuccessful pharmacological treatment from oxybutinin to acupressure group. Acupressure was not compared to a placebo/sham group.

Studies retrieved and collated as background information on Shiatsu (Appendix 11), identified four single case reports of adverse events occurring following Shiatsu massage (Herskovitz et al 1992; Mumm et al 1993; Tsubo 2001; Wada et al 2005). This is an important area for the profession regarding safety issues and possible causal links between Shiatsu and adverse events. Note, these may not have been the only case reports.

[11] Morrison D A, Sethi G, Sacks J et al (2002) The VA AWESOME (angina with extremely serious operative mortality evaluation) Multicenter Registry. Percutaneous coronary intervention versus coronary bypass graft surgery for patients with medically refractory myocardial ischemia and risk factors for adverse outcomes with bypass: the VA AWESOME multicenter registry: comparison with the randomised clinical trial. *J Am Coll Cardiol,* 39:266-273

A limited number studies assessed qualitative aspects of Shiatsu as a therapy (Cheesman et al 2001; Long & Mackay 2003) but the data was either not presented scientifically or was not carried out in controlled circumstances. Other studies mentioned acupuncture massage techniques and it is unclear if this was about acupressure (Furlan et al 2002). There were also general articles mentioning Shiatsu as an intervention (Galantino et al 2003) and some mentioned Shiatsu as part of a service provision (Peace & Manasse 2002; Sommers et al 2002; Yates 2005). A survey was funded by the Research Council for Complementary Medicine (Harris & Pooley 1998) to investigate what conditions practitioners currently treated and to ascertain the direction of future research into the efficacy of Shiatsu. The survey found that the most common conditions presenting for treatment were musculo-skeletal and psychological problems and concluded that future efficacy research should focus on these areas, in particular neck/shoulder, lower back problems, arthritis, depression, stress and anxiety. Two studies referred to Watsu but one was a personal account (Davis 2003) and the other a series of case reports (Vogtle et al 1998).

Background information on acupressure mainly included acupressure and its effects on nausea and vomiting and literature reviews (Collins & Thomas 2004; Harris 1997). The articles on nausea and vomiting included some reviews (Oates & Whitehead 2003; Oates & Whitehead 2004; Jewell 2003; Aikins 1998) but tended to be very broad and included either various types of intervention or were non systematic or related to devices or combinations of interventions (Anderson & Aikins 1998; Johnson 2005). Non-pharmacological management was shown to be effective for post operative and chemotherapy induced nausea and vomiting but Shiatsu was not specifically mentioned (Lee & Done 1999; King 1997; Pan et al 2000).

A thesis on delivering Shiatsu in general practice (Pirie 2003) looked at the impact of Shiatsu on GP consultations and whether frequency of prescriptions for medications were reduced (Appendix 15). This was a qualitative study and the researcher was also the practitioner. The research concluded that complementary medicine could be effectively delivered in general practice and that further research in clinical and cost effectiveness was warranted.

These findings provide an important addition to the existing knowledge base on Shiatsu but are very limited.

7. Discussion

The research base for Shiatsu is very much in its infancy and the profession will need to work closely with practitioners and researchers in order to build up a larger body of evidence. This evidence review has considered and included research studies which have a conventional RCT design. The methodological limitations of the studies reported in this systematic literature review included small sample sizes, non reporting of follow up, insufficient details on sampling, high drop out rates, uncontrolled design, lack of blinding etc. Many studies were also underpowered i.e. their sample size at the outset was insufficient to detect a significant difference.

Complementary medicine is under pressure to provide scientific evidence of efficacy if they are to be accepted and integrated within the prevailing framework of conventional medicine. The relevance of evidence based medicine depends on the quality of the research carried out. While much of the research carried out with Shiatsu or acupressure as an intervention is of insufficient quality to provide consensus on its use, some high quality (Category 1) clinical research (particularly around pain) does exist. This provides a model for future research.

Contention is slowly emerging about how complementary medicine should be evaluated[12,13,14,15,16,17]. The complexity of interventions and their potential synergistic effect requires innovative evaluative approaches using whole systems research that include qualitative and quantitative methods.[18,19] Analysis of the literature indicates that important refinements are being generated in complementary medicine research and clinical trial design in response to the challenges posed by the forced encounter of the paradigms of holistic and conventional medical practice.

Shiatsu is no different to other complementary therapies in that a pragmatic RCT approach (reflecting normal practice) should be an inherent part of study design. In addition, qualitative data provides additional information on patients' and/or practitioners' views on the effectiveness of treatment. Many studies are including such qualitative data as part of their design to provide a broader picture of patient outcomes.

If Shiatsu is to be evaluated at least in part through RCTs, these must be appropriately designed. A recent paper concluded that RCTs would be more effective in studying acupuncture if participants were randomised to groups based on acupuncture diagnosis, not solely on conventional western criteria[11]. This may be true for other complementary therapies including Shiatsu which uses elements of Traditional Chinese Medicine diagnosis as well as Hara diagnosis. The authors felt that although treatments must be standardised to ensure replicability of the study, blinding was not absolutely necessary for a good quality RCT, however, if used, control groups need to be standardised. They felt that homogeneity of groups based on specific acupuncture diagnostic criteria (which takes into account the different philosophy and Chinese medicine system) could be used as evidence of efficacy of the intervention and satisfy both acupuncture and conventional medicine critics.

Future clinical research on Shiatsu and acupressure also needs to take into account practitioner variability in terms of point selection (based on differences in their education and training). The issue of blinding, which involves "sham" or placebo treatments, is also difficult to resolve. Shiatsu (as distinct from acupressure) presents futher complexities as treatments are based on Hara diagnosis and rarely if ever "standardised".

In the case of acupuncture there has been controversy about the use of sham acupuncture and it is now generally felt that it cannot be used as an inert control though it could be one arm of an RCT[12] The strength of any conclusions from research will depend on the quality of the evidence included and even with gold standard RCT there can still be bias[20]. As the included studies have shown, "sham" acupressure including light touch at acupoints does have an effect. This means

[12] Walji R, Boon H (2006). Redefining the randomised controlled trial in the context of acupuncture research. *Complement Ther Clin Pract.*

[13] Shea J. (2006) Applying evidence-based medicine to Traditional Chinese Medicine: Debate and strategy. *J Altern Complement Med* 12(3): 255-263.

[14] Walker LG, Anderson J. (1999)Testing Complementary and alternative therapies within a research protocol. *European J Cancer* 35(11) 1614-1617

[15] Herman PM, D'Huyvetterk . Mohler MJ ((2006) Are health services research methods a match for CAM. *Altern Ther* 12(3): 78-83.

[16] Broom A. (2005) Using qualitative interviews in CAM research: A guide to study design, data collection and data analysis. *Complement Ther Med* 13: 65-73.

[17] Walach H, Falkenberg T, Fonnebo V, Lewioth G, Jonas WB. (2006) Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. 6: 29. <http://www.biomedcentral.com/1471-2288/6/29. doi 10.1186/1471-2288-6-29>.

[18] Verhoef MJ, Lewith G, Ritenbaugh C, Boon H, Fleishman S & Leis A. (2005) Complementary and alternative medicine whole systems research: Beyond identification of inadequacies of the RCT. *Complement Ther Med* 13(3): 206-212

[19]Giordano J, Garcia MK, Strickland G.(2004) Integrating Chinese Traditional medicine into a US Public Health Paradigm. *J Altern Complement Med* 10(4): 706-710.

[20] Leibovici L. (1999) Alternative (complementary) medicine: a cuckoo in the nest of empiricist warblers. *BMJ (Clinical Research ED.);* 319: 1629 -32.

that “sham” acupressure may not be an appropriate control unless a study is very carefully designed.

Safety is best established with prospective studies and the four reports identified in this review highlighted the importance of having good evidence on safe practice. The type and frequency of adverse events and any transient reactions after Shiatsu therapy needs exploration as although this was not the focus of this review, several incidences of adverse reactions to Shiatsu were found.

8. Conclusions

The summaries of the best quality evidence to date suggest that, due to the small number of studies specifically relating to Shiatsu, well designed research in any area would be a welcome addition to the current evidence base. For acupressure and pain, the evidence is generally consistent and has demonstrated that acupressure can control pain. Acupressure studies for nausea and vomiting have been somewhat inconsistent and may merit further research. Similarly the studies on COPD and asthma, psycho-social aspects of health, anaesthesia and other health conditions are generally weak due to study design. From these studies reviewed only pain, nausea and vomiting have provided some evidence of benefit but are too heterogeneous and therefore cannot be amalgamated.

9. Recommendations

- Significant research needs to be carried out if Shiatsu is to develop an evidence base
- Further research is needed to investigate the effectiveness of Shiatsu as an intervention
- Encourage practitioners to engage in research using well designed studies
- The relationship between Shiatsu and acupressure needs clarification for marketing and public awareness
- Consider the development and piloting of an adverse event reporting system for Shiatsu
- Explore clinical and the cost effectiveness of Shiatsu in an integrated setting
- Identify specific topic areas for initial research investment
- Develop an evaluative framework for integrated Shiatsu practice
- Develop a research resource for the profession
- Investigate the appropriateness of various research methodologies for Shiatsu research