Influence of Study Design on Outcomes Following Reflexology Massage: An Integrative and Critical Review of Interventional Studies

Andrew McVicar, PhD,1 Christina Greenwood, PhD,2 Carol Ellis, MSc,2 and Chantelle LeForis, BSc3

Abstract

Background: Interpretation of the efficacy of reflexology is hindered by inconsistent research designs and complicated by professional views that criteria of randomized controlled trials (RCTs) are not ideal to research holistic complementary and alternative medicine practice. The influence of research designs on study outcomes is not known. This integrative review sought to evaluate this possibility.

Materials and methods: Thirty-seven interventional studies (2000–2014) were identified; they had RCT or non-RCT design and compared reflexology outcomes against a control/comparison group. Viability of integrating RCT and non-RCT studies into a single database was first evaluated by appraisal of 16 reporting fields related to study setting and objectives, sample demographics, methodologic design, and treatment fidelity and assessment against Jadad score quality criteria for RCTs. For appraisal, the database was stratified into RCT/non-RCT or Jadad score of 3 or greater or less than 3. Deficits in reporting were identified for blind assignment of participants, dropout/completion rate, and School of Reflexology. For comparison purposes, these fields were excluded from subsequent analysis for evidence of association between design fields and of fields with study outcomes.

Results: Thirty-one studies applied psychometric tools and 20 applied biometric tools (14 applied both). A total of 116 measures were used. Type of measure was associated with study objectives \((p < 0.001; \chi^2)\), in particular of psychometric measures with a collated “behavioral/cognitive” objective. Significant outcomes were more likely \((p < 0.001; \chi^2)\) for psychometric than for biometric measures. Neither type of outcome was associated with choice of RCT or non-RCT method, but psychometric responses were associated \((p = 0.007)\) with a nonmassage control strategy.

Conclusions: The review supports psychometric responses to reflexology when study design uses a nonmassage control strategy. Findings suggest that an evaluation of outcomes against sham reflexology massage and other forms of massage, as well as a narrower focus of study objective, may clarify whether there is a relationship between study design and efficacy of reflexology.

Introduction

Reflexology is a complementary and alternative medicine (CAM) therapy in which pressure is applied to specific points, usually but not exclusively located on the foot, to stimulate the internal organs with which they are claimed to associate. Systematic reviews of randomized controlled trials (RCTs) do not support efficacy of reflexology for clinically relevant outcomes,1–5 but the value of RCT criteria to CAM research has been questioned on the basis that the holistic approach typical of CAM therapies is contrary to the reductionist approach of RCT designs.6 Research designs appropriate for the study of CAM therefore present an unresolved issue depending on professional viewpoint, yet the extent of influence an interventional design might have on reflexology outcomes is unknown. This paper reports an integrative review of RCT and non-RCT interventional designs to re-evaluate evidence for reflexology outcomes in the context of study protocols. RCTs provide the “gold standard” for evidence-based healthcare practice, but for CAM this biomedical cause-and-effect model is claimed to be impractical because it fails to
acknowledge the nonlinearity of CAM practice and how the practitioner–client interaction may contribute an incidental effect on trial outcomes. The importance of the therapeutic relationship also complicates the identification of an adequate placebo and makes blinding of the practitioner and/or participant difficult, perhaps even impossible, within a research design. Systematic reviews of RCTs involving reflexology identify such design issues as hindrances to data interpretation.

Despite methodologic debate, the regulatory requirements for the adoption of research findings in mainstream healthcare seem likely to continue to lean heavily toward evidence produced by RCTs. For wider acceptance of integrated care approaches, it therefore falls to CAM researchers to find a way to reasonably meet criteria for study rigour within an acceptable interventional framework but also requires both CAM and biomedical researchers to acknowledge that some compromises have to be made. For example, RCTs for CAM are increasingly anticipated to report on intervention fidelity, such as the number and length of reflexology treatments. An emphasis on a standardized method could go some way toward meeting at least some criticisms from biomedicine, but it would also represent a compromise on viewpoints of the holistic aspects of CAM.

What remains unclear, however, is the extent to which research design affects study outcomes after a CAM intervention. In this respect, the published reflexology literature provides an opportunity to review the influence of methodologic heterogeneity on study outcomes because in addition to RCTs it also includes peer-reviewed non-RCT studies with designs that also involve a comparator group. This was the approach taken by Steenkamp and colleagues in their integrative review of RCTs, experimental designs, case studies, case reports, conference literature, and grey literature to evaluate outcomes of reflexology in people with chronic disorders. In common with systematic reviews of RCT studies, noted above, they also identified the variability of designs and typologies of the disorders investigated as limitations. However, they did not seek to evaluate the potential influences on study outcomes that design issues per se might have introduced.

The aim of this review was to evaluate whether research design influences study outcomes after reflexology treatment. To achieve this, the objective was to undertake an integrative review of reflexology research to enable critical appraisal of interventional study designs in which reflexology outcomes were compared against data from control or comparator groups, but without constraining the analysis by selecting studies that only refer to pathologies or only apply to RCT designs.

Materials and Methods

This paper reports the results of an integrative review of RCT and non-RCT research designs used to investigate the effectiveness of reflexology. Integrative reviews typically select studies that have applied different methods and designs, an approach that is suggested to strengthen data interpretation. In this respect the resultant introduction of heterogeneity contributed to the review objective because it increased the scope for analysis of study designs and their impact. Nevertheless, a structured approach to source selection, data extraction, and data synthesis is still required.

Research literature published between January 2000 and December 2014 was chosen because it spans much of the period since RCTs have become emphasized as sources of evidence for healthcare practice. In searching for published research, an initial network approach with systematic reviews of RCTs from that period was used because these provided a useful resource of earlier RCTs. Those studies were supplemented with subsequent hand searches of studies that they cited. The database was extended and updated by searching electronic databases (Academic One, British Nursing Index, CINAHL, MEDLINE, ScienceDirect, Wiley Online, Cochrane Library) considered the most likely source of publication in this field, using the terms “Reflexology AND” in combination with “RCT,” “trial,” “quasi,” “experimental,” “pilot,” and “asthma” (conditions for which reflexology is a very popular form of CAM) in the title and abstract fields. All studies that appeared to be interventional and published in English in peer-reviewed journals within the selected timeframe were considered, regardless of their being RCTs or non-RCTs.

The initial search identified 232 papers for further consideration. Duplicates and citations that were not actually in the date range were removed. In the relatively few instances

FIG. 1. Literature retrieval flow chart.
where an abstract was not available but the title looked relevant, the paper was sought via the British Library. Descriptive, discussion, or editorial papers or dissertations/theses were then removed from the database. Remaining publications included both quantitative and qualitative studies and were secondarily filtered to ensure that the database included only studies in which outcome data were supported by statistical comparison against control or comparator group data. Thirty-seven quantitative studies were finally selected for the review. Rigor in the filtering and subsequent data extraction was assured by inter-rater agreement between at least two of the reviewers. The literature flow-through is summarized in Figure 1.

Twenty-three of the selected studies were RCTs, 7 used crossover designs, and 7 were identified as using pre–post or experimental designs. Median publication year was 2009. Most studies were located in the United Kingdom, but there was an international profile overall: Eighteen were in the United Kingdom; 7 in the United States; 4 in Iran; and 1 each in Australia, China, Denmark, Israel, Italy, Japan, Taiwan, and Turkey.

**Data extraction and appraisal of reporting**

Selected studies were examined for information pertaining to reporting fields. Study setting (clinical or other), focus (objectives), demographic data (sample, age, and sex), methods (e.g., RCT), control or comparator group, identification of randomization and the method of randomization, identification of blinding, and dropout/noncompletions were noted for each study. Additionally, guidance on design for the reporting of nonpharmacologic trials (Consolidated Standards of Reporting Trials statement), and for CAM in particular, was adopted to identify aspects of intervention fidelity by extracting the location of delivery of reflexology, who delivered the treatment, the school of reflexology whose guidelines for practice were followed, the number of treatments each participant received, and the duration of each treatment. Sixteen reporting fields in total comprised study setting and focus, three demographic fields, six methodologic fields, and five fidelity fields.

Before integration of studies into a single database, any limitations within the data were appraised first by evaluating the breadth of reporting across these fields for RCTs and non-RCTs (Table 1) and second according to Jadad quality criteria (Table 1), in each case using the framework of Brown and colleagues. In this framework, a design field is assigned a score of 1 if information is supplied in the report and 0 if not or if it was nonspecific. The average total score therefore reflects the breadth of reporting coverage by studies in the dataset, and comparison of average scores for individual reporting fields enables a more detailed appraisal.

**RCTs vs non-RCTs**

The average total reporting score for the entire database (n = 37 studies) was 14.0 of 16 fields (0.88 per field), thus indicating some reporting deficits. Stratifying the database according to RCT and non-RCT subsets produced average total scores of 14.3 of 16 (0.89 per field) and 13.6 of 16 (0.85 per field), respectively, indicating that the deficits were present in both subsets. These predominantly related to methodologic and fidelity fields (Table 1). For RCTs, the six
methodologic fields collectively averaged 0.87 per field, and the five fidelity fields averaged 0.82, whereas for non-RCTs the averages were 0.71 and 0.87 per field, respectively; this indicated greater deficit of methodologic reporting but slightly better fidelity reporting in the non-RCT studies. Close examination identified that blinding was poorly reported in both subsets, but especially so in the non-RCT subset (Table 1), where reporting of dropout/completion rate was also very weak. The reporting of intervention fidelity fields by RCTs and non-RCTs also varied, but the main weakness was related to school of reflexology: Just 24 of 37 studies overall identified the therapeutic tradition, with the greatest deficit being in the RCT studies.

**Jadad scores**

The rigor of RCT designs in published reviews of reflexology is often evidenced by application of the scale put forward by Jadad et al.,20 which focuses on key methodologic criteria. The scale refers to sources of potential bias within an experimental design and so also has relevance within interventional designs other than RCTs. A decision therefore was taken to additionally appraise the reporting quality of all selected studies according to those criteria.

The Jadad scale is based on five questions: (1) Was the study described as randomized? (2) Was the study described as double-blind? (3) Was there a description of withdrawals and dropouts? (4) Was the method of randomization described in the paper? (5) Was the method of blinding described? Affirmative answers are each scored 1 point, but points are deducted if the method of randomization lacked rigor and/or the method of blinding was inappropriate. Blinding is a particular issue for reflexology because sham treatments are very difficult to deliver without being obvious. For example, 66.7% of control participants in one study in the authors’ database successfully identified which study arm they had been in. Few studies (n = 13 of 37) in this review attempted blinding at any point of delivery, and most that did report the process described it poorly. A liberal approach to scoring this criterion therefore was taken to acknowledge reference to at least single-blinding, with or without evaluation of its quality or success.

As a validity check, 16 studies that had been scored by Ernst and colleagues5 in their 2011 review were identified and compared to the Jadad evaluations undertaken without prior reference to those published scores. Studies were scored by all authors and, if necessary, discussed among the team for consensus view. The mean allocated score of 3.1 compared favorably with the published average of 3.0.

For the whole database, the allocated Jadad scores ranged from 0 to 5 but averaged only 2.4, indicating a weighting toward lower values. Reporting deficiencies occurred in most key criteria. Although 36 of 37 studies (97%) claimed random allocation to groups, just 25 (68%) identified the process of randomization. Reporting of blinding was even more incomplete (13 of 37 studies [35%]), and only 21 of 37 studies (57%) identified noncompleters among the participants. For further appraisal, the database was dichotomized according to Jadad scores based on a quality threshold of 3 or above, in line with its adoption by Ernst and colleagues.2 A high-score subgroup (n = 18) averaged 3.6 (median, 3), and a low-score subgroup (n = 19) averaged 2.2 (median, 1).

Fifteen of the high-score studies were identified as RCTs, the remaining as crossover (n = 1) or experimental (n = 2) designs. Low-score studies were also a combination of RCT (n = 9), crossover (n = 5), and experimental (n = 5) designs. The distribution of RCTs and non-RCT designs was significantly different between the subgroups (chi square; p < 0.02). However, for the purposes of this review the appearance of both in each subgroup supported the secondary appraisal of reporting fields according to Jadad criteria additional to a simple RCT/non-RCT division.

Overall (Table 1), the reporting of the 16 design fields in the low Jadad score subgroup averaged 13.5 of 16 fields (0.84 per field) and in the high Jadad score subgroup averaged 14.6 of 16 fields (0.91 per field), indicating a differential between the two. Much of this resulted from the reporting of the five components of the Jadad score itself. Notwithstanding these, further stand-out differentials arose from the reporting of location of delivery of reflexology, which was higher in the low-score subgroup, and of the school of reflexology and number of treatments, which were higher in the high-score subgroup.

**Integration**

For the purpose of this review it was necessary to consider the viability of integrating the data within a single database in light of identification of variable reporting within the studies. Although there was variability across and between a number of fields, Table 1 highlights that those most poorly reported were blinding, dropout/completion rate, and school of reflexology, with some scores being lower than 0.6 per field in at least one subgroup. To improve parity within the database without significant detriment to the aims of this review these three fields were excluded from further analysis, leaving 13 fields to be taken forward for the review. Table 2 summarizes the overall reporting appraisal of these remaining fields. Following exclusions, the overall average score of all fields (n = 37 studies) increased from 0.88 to 0.97 per field. Collective average scores for just the methodologic and fidelity fields increased from 0.79 to 0.92 per field and from 0.84 to 0.91 per field, respectively.

**Data analysis**

All extracted data were stored as an electronic database (SPSS software, version 17; IBM, Armonk, NY). Statistical analysis was descriptive, but cross-tabulation by chi-square analysis using Fisher exact correction for small numbers was used to identify any significant associations between fields. The variability of designs posed some issues for data analysis because field data in some instances had to be merged to strengthen the association tests, meaning that assumptions had to be made as to cognate links. The collation of individual reporting fields is noted in the Results section where appropriate. Significance of association between fields and between them and study outcomes was set at p < 0.05 (two-tailed).

**Results**

This section presents the descriptive data for the integrated database, followed by analysis of relationships with study outcomes. Regarding application of Jadad criteria, the
poor reporting and subsequent exclusion of both blinding and dropout/completion rate as reporting fields meant that the criteria were now only supported by randomization. Because the strategy applied to randomize samples was one of the methodologic fields recorded and entered into the review analysis, Jadad scoring was not included as a variable for the between-field analysis. However, Jadad criteria provide a standard for evaluating the rigor of an interventional design and so a post hoc analysis was also undertaken to identify whether or how findings from the analyses might also be related to Jadad scores. That evaluation appears later.

Study designs

Study setting and objectives. The setting of 25 of 37 studies related to patients who had a specified diagnosis (Table 3). Most of these (n = 10) referred to cancer, whereas others referred to neurologic or neuromuscular disorders (n = 8), cardiovascular disorders (n = 2), respiratory problems (n = 2), renal/urogenital problems (n = 2), or gastrointestinal disorder (n = 1). It was not the intent of this review to select only studies that involved patients diagnosed with an illness, and 12 of the selected studies involved evaluation of reflexology treatment in “healthy” individuals. Seven of the latter related to the effect of a specific physiologic challenge (e.g., labor, menopause, or cold immersion), and the remaining 5 studies involved unchallenged participants. For further analytical purposes, the objectives (study foci) of the intervention were collated as primarily relating to pathologic/physiologic indicators (n = 12) or to behavioral/cognitive indicators (n = 25). Setting and objective were not significantly associated.

Demographic data. Sample sizes across the database ranged from 10 to 243 participants (n = 37 studies). Mean sample size (±standard deviation) was 59.0 ± 58.6, indicating a skewed distribution (median, 36). Mean participant age across all studies (n = 31) was 49.4 ± 16.4 years (range, 19.8–90 years; median, 49.1 years). A total of 74.1% ± 24.0% of participants were female (range, 25%–100%; median, 74.5%; n = 35 studies). None of the demographic fields showed any statistical association with the study setting or objective, or with any of the reporting fields described below.

Methodologic issues. The appraisal of study reporting described in the Methods section had already dichotomized the methods as either RCT (n = 23 studies) or non-RCT (14 studies). Thirty-five of 37 studies reported that samples had been randomized to intervention (reflexology) and control groups, but the process of randomization of participants was provided by just 25 of 37 studies (Table 3). The process of randomization varied from approaches viewed as having low rigor (n = 8; e.g., alternation, coin flipping, attendance on clinic day) to those using computerized methods with higher rigor (n = 17; computer-generated random numbers, block randomization, minimization). A computerized randomization strategy was significantly associated with methods, primarily RCT design (p = 0.025; n = 32). Otherwise, methods and randomization were not associated with any other reporting fields.
<table>
<thead>
<tr>
<th>Reference no.</th>
<th>Setting and focus</th>
<th>Setting</th>
<th>Focus</th>
<th>Sample</th>
<th>Mean age or range</th>
<th>Female (%)</th>
<th>Methods</th>
<th>Randomization method</th>
<th>Jadad score</th>
<th>Control strategy</th>
<th>Where delivered?</th>
<th>Delivered by whom?</th>
<th>No. of treatments</th>
<th>Length of treatment (min)</th>
<th>Significant psychometric outcomes</th>
<th>Significant biometric outcomes</th>
<th>No significant effect on psychometric measure</th>
<th>No significant effect on biometric measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Cancer</td>
<td>Pain</td>
<td>Anxiety</td>
<td>86</td>
<td>58.3</td>
<td>51</td>
<td>Pre-post</td>
<td>Partner dyads</td>
<td>0</td>
<td>Reading with partner</td>
<td>Clinic/home</td>
<td>Partner</td>
<td>Not given</td>
<td>30</td>
<td>Pain Anxiety</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>24</td>
<td>Cancer</td>
<td>Pain</td>
<td>36</td>
<td>Not given</td>
<td>50</td>
<td>Stratified</td>
<td>Unit</td>
<td>Reflexologist</td>
<td>Not given</td>
<td>30</td>
<td>Pain</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Postpartum</td>
<td>Sleep quality</td>
<td>65</td>
<td>31.6</td>
<td>Not given</td>
<td>Crossover</td>
<td>Not given</td>
<td>University</td>
<td>Relaxation</td>
<td>1</td>
<td>Reflexologist</td>
<td>Not given</td>
<td>60</td>
<td>Anxiety</td>
<td>None</td>
<td>SBP</td>
<td>Pulse rate</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>None</td>
<td>Stress</td>
<td>30</td>
<td>16–59</td>
<td>Not given</td>
<td>Crossover</td>
<td>Not given</td>
<td>Clinic</td>
<td>Reflectologist</td>
<td>Not given</td>
<td>30</td>
<td>Pain</td>
<td>None</td>
<td>SBP</td>
<td>Pulse rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Chronic obstructive pulmonary disease</td>
<td>Multiple sclerosis</td>
<td>14</td>
<td>76.0</td>
<td>50</td>
<td>RCT</td>
<td>Randomized</td>
<td>1</td>
<td>No treatment</td>
<td>Hospital room</td>
<td>Not given</td>
<td>4</td>
<td>50</td>
<td>Quality of life</td>
<td>None</td>
<td>Pulse rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Menopause</td>
<td>Indicators</td>
<td>50</td>
<td>53.1</td>
<td>76</td>
<td>Crossover</td>
<td>Not given</td>
<td>3</td>
<td>Progressive muscle relaxation</td>
<td>Clinic</td>
<td>Nurse</td>
<td>6</td>
<td>40</td>
<td>General health</td>
<td>None</td>
<td>Cortisol secretion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Sleep quality</td>
<td>Menopausal flushing</td>
<td>100</td>
<td>45-60</td>
<td>100</td>
<td>RCT</td>
<td>Not given</td>
<td>2</td>
<td>Simulated foot massage</td>
<td>Center</td>
<td>Reflexologist</td>
<td>Not given</td>
<td>21</td>
<td>15</td>
<td>Reduced flushing</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>30</td>
<td>Cancer</td>
<td>Quality of Life</td>
<td>60</td>
<td>59.4</td>
<td>100</td>
<td>RCT</td>
<td>Computerized block randomization</td>
<td>3</td>
<td>Scalp massage or Self-initiated support</td>
<td>Center</td>
<td>Reflexologist</td>
<td>8</td>
<td>60</td>
<td>Quality of life</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Parkinson disease</td>
<td>Irritable bowel syndrome</td>
<td>14</td>
<td>54-75</td>
<td>57</td>
<td>Crossover</td>
<td>Not randomized</td>
<td>1</td>
<td>No treatment</td>
<td>Own home</td>
<td>Therapist</td>
<td>8</td>
<td>60</td>
<td>Well-being</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Low back pain</td>
<td>Symptoms</td>
<td>34</td>
<td>48.0</td>
<td>82</td>
<td>Crossover</td>
<td>Alternation</td>
<td>5</td>
<td>Simulated foot massage</td>
<td>Center</td>
<td>Reflexologist</td>
<td>6</td>
<td>30</td>
<td>Pain</td>
<td>None</td>
<td>Dialysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Dementia</td>
<td>Quality of life</td>
<td>243</td>
<td>47.4</td>
<td>Not given</td>
<td>Crossover</td>
<td>Computerized random number table</td>
<td>3</td>
<td>No treatment</td>
<td>Center/GP surgery</td>
<td>Reflexologist</td>
<td>6</td>
<td>60</td>
<td>Mental health</td>
<td>None</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Cancer</td>
<td>Pain</td>
<td>Anxiety</td>
<td>21</td>
<td>80-96</td>
<td>100</td>
<td>Crossover</td>
<td>Computerized random number table</td>
<td>2</td>
<td>Friendly visits</td>
<td>Residential care home</td>
<td>Reflexologist</td>
<td>4</td>
<td>30</td>
<td>Pain</td>
<td>None</td>
<td>Serum amylase</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Asthma</td>
<td>Lung function</td>
<td>23</td>
<td>33.0</td>
<td>65</td>
<td>Crossover</td>
<td>Coin toss</td>
<td>2</td>
<td>No treatment</td>
<td>Unit</td>
<td>Reflexologist</td>
<td>Not given</td>
<td>30</td>
<td>Anxiety</td>
<td>None</td>
<td>Peak flow rate</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Menopause</td>
<td>Anxiety Mood</td>
<td>40</td>
<td>39.0</td>
<td>63</td>
<td>RCT</td>
<td>Computerized random number table</td>
<td>4</td>
<td>Simulated foot massage</td>
<td>Unit</td>
<td>Reflexologist</td>
<td>10</td>
<td>45</td>
<td>NA</td>
<td>None</td>
<td>Histamine</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Reference no.</th>
<th>Setting and focus</th>
<th>Setting</th>
<th>Focus</th>
<th>Sample</th>
<th>Mean age or range</th>
<th>Female (%)</th>
<th>Demographic</th>
<th>Methodologic</th>
<th>Intervention fidelity</th>
<th>Psychometric outcome</th>
<th>Biometric outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 Cancer</td>
<td>Mood Symptoms</td>
<td>17</td>
<td>74.0</td>
<td>76</td>
<td>RCT</td>
<td>Not given</td>
<td>Basic foot massage</td>
<td>4</td>
<td>Day care unit</td>
<td>Relaxation</td>
<td>Stress</td>
</tr>
<tr>
<td>39 Cancer</td>
<td>Quality of life</td>
<td>12</td>
<td>58–80</td>
<td>58</td>
<td>RCT</td>
<td>Not given</td>
<td>Sham foot massage</td>
<td>2</td>
<td>Reflexologist Unit</td>
<td>Quality of life Relaxation</td>
<td>None</td>
</tr>
<tr>
<td>40 Cancer</td>
<td>Immune and endocrine functioning</td>
<td>183</td>
<td>58.8</td>
<td>100</td>
<td>RCT</td>
<td>Computerized block randomization</td>
<td>3</td>
<td>Self-initiated support</td>
<td>Centre</td>
<td>Reflexologist Unit</td>
<td>None</td>
</tr>
<tr>
<td>41 None</td>
<td>Ovulation</td>
<td>49</td>
<td>18–38</td>
<td>100</td>
<td>RCT</td>
<td>Not given</td>
<td>No treatment</td>
<td>3</td>
<td>Hospital room Clinic</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>42 Pregnancy</td>
<td>Foot edema</td>
<td>55</td>
<td>28.2</td>
<td>100</td>
<td>RCT</td>
<td>Not given</td>
<td>No treatment</td>
<td>2</td>
<td>Resting</td>
<td>Reflexologist</td>
<td>NA</td>
</tr>
<tr>
<td>43 Cancer</td>
<td>Anxiety</td>
<td>30</td>
<td>54.2</td>
<td>100</td>
<td>Exp</td>
<td>Not given</td>
<td>No treatment</td>
<td>0</td>
<td>Attendees at clinic Unit</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>44 Multiple sclerosis</td>
<td>Parasthesia</td>
<td>53</td>
<td>54.4</td>
<td>64</td>
<td>RCT</td>
<td>Computerized block randomization</td>
<td>3</td>
<td>No treatment</td>
<td>Center</td>
<td>Reflexologist Unit</td>
<td>None</td>
</tr>
<tr>
<td>45 Osteoarthritis</td>
<td>Urinary frequency</td>
<td>120</td>
<td>56.2</td>
<td>100</td>
<td>RCT</td>
<td>Computer-generated random-number table</td>
<td>5</td>
<td>Non-specific foot massage</td>
<td>Center</td>
<td>Therapist</td>
<td>NA</td>
</tr>
<tr>
<td>46 Insomnia</td>
<td>Sleep</td>
<td>10</td>
<td>46.2</td>
<td>60</td>
<td>Exp</td>
<td>Computer-generated random-number table</td>
<td>4</td>
<td>Acupuncture/ music</td>
<td>Clinic</td>
<td>Reflexologist Unit</td>
<td>Perceived health status</td>
</tr>
<tr>
<td>47 Multiple sclerosis</td>
<td>Pain</td>
<td>71</td>
<td>51.5</td>
<td>83</td>
<td>RCT</td>
<td>Computer-generated random-number table</td>
<td>4</td>
<td>Shann reflexology</td>
<td>Own home</td>
<td>Reflexologist</td>
<td>Mood</td>
</tr>
<tr>
<td>48 Cold intolerance</td>
<td>Cold immersion</td>
<td>10</td>
<td>19.8</td>
<td>100</td>
<td>Exp</td>
<td>Not given</td>
<td>Opposite foot massage</td>
<td>Not given</td>
<td>Laboratory Unit</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>49 None</td>
<td>Cardiac functioning</td>
<td>16</td>
<td>38.0</td>
<td>31</td>
<td>RCT</td>
<td>Computer-generated random-number table</td>
<td>5</td>
<td>Non-specific foot massage</td>
<td>Unit</td>
<td>Reflexologist</td>
<td>None</td>
</tr>
<tr>
<td>50 None</td>
<td>Anxiety</td>
<td>25</td>
<td>26.9</td>
<td>48</td>
<td>Exp</td>
<td>Computer-generated random-number table</td>
<td>3</td>
<td>Relaxing foot holding</td>
<td>Not given</td>
<td>Reflexologist Unit</td>
<td>None</td>
</tr>
<tr>
<td>51 None</td>
<td>Pain threshold</td>
<td>15</td>
<td>37.7</td>
<td>73</td>
<td>Exp</td>
<td>Not given</td>
<td>No treatment</td>
<td>Not given</td>
<td>Laboratory Unit</td>
<td>Pain threshold</td>
<td>None</td>
</tr>
</tbody>
</table>
| 52 Multiple sclerosis | Quality of life | 20      | 56.4  | 55     | RCT              | Not given  | Non reflexology legfoot massage | 2 | Unit | None | Sleep quality | NA | NA | NA | None | None | (continued)
### Table 3. (Continued)

<table>
<thead>
<tr>
<th>Reference no.</th>
<th>Setting and focus</th>
<th>Demographic</th>
<th>Methodologic</th>
<th>Intervention fidelity</th>
<th>Psychometric outcome</th>
<th>Biometric outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sample Mean age or range</td>
<td>Female (%)</td>
<td>Methods Randomization method Jadad score Control strategy Where delivered? Delivered by whom? No. of treatments Length of treatment (min)</td>
<td>No significant effect on psychometric measure</td>
<td>Significant biometric outcomes</td>
</tr>
<tr>
<td>53</td>
<td>Labour</td>
<td>120 22.9 100</td>
<td>RCT</td>
<td>Random allocation software Not given 2</td>
<td>Spiritual, emotional support Ward Researcher Not given 40</td>
<td>Pain None Duration of labour None</td>
</tr>
<tr>
<td>54</td>
<td>Cancer</td>
<td>18 90.0 67</td>
<td>Randomized controlled trial</td>
<td>Swedish massage Residential care home Reflexologist 4 20</td>
<td>None</td>
<td>Pain Mood None Cortisol secretion</td>
</tr>
<tr>
<td>55</td>
<td>Renal</td>
<td>70 49.1 45</td>
<td>RCT</td>
<td>Sham contact on foot</td>
<td>Unit Certificated researcher 3 30</td>
<td>Pain Fatigue Cramps Pain intensity Pain characteristic Anxiety Mood Anxiety Pain None SBP DBP Cardiac index SV NA</td>
</tr>
<tr>
<td>56</td>
<td>Low back pain</td>
<td>50 40.9 40</td>
<td>RCT</td>
<td>Relaxing massage of lower leg Reflexology of abdominal area of foot Hospital room Trained researcher 12 30</td>
<td>None</td>
<td>Anxiety Pain None SBP DBP Cardiac index SV NA</td>
</tr>
<tr>
<td>57</td>
<td>Congestive heart failure</td>
<td>12 71.0 25</td>
<td>RCT</td>
<td>Computer-generated random-number table Hospital room Trained researcher 1 45</td>
<td>None</td>
<td>Anxiety Pain None SBP DBP Cardiac index SV NA</td>
</tr>
<tr>
<td>58</td>
<td>Cancer</td>
<td>243 54.8 100</td>
<td>RCT</td>
<td>Computerized minimization</td>
<td>Center Reflexologist 4 30</td>
<td>Dyspnoea Quality of life Mood Anxiety Pain Fatigue Anxiety on 4 consecutive days None SBP DBP Cardiac index SV NA</td>
</tr>
<tr>
<td>59</td>
<td>Coronary heart disease, postsurgical anxiety</td>
<td>80 58.8 50</td>
<td>RCT</td>
<td>Soft massage of left leg Hospital room Reflexologist 4 20</td>
<td>Anxiety None None SBP DBP Cardiac index SV NA</td>
<td></td>
</tr>
</tbody>
</table>

NA, not available; Exp, experimental; SBP, systolic blood pressure; DBP, diastolic blood pressure; BP, blood pressure; GP, general practitioner; FEV, forced expiratory volume; FVC, forced vital capacity; NK, natural killer; LAK, lymphokine-activated killer; HR, heart rate; SV, stroke volume.
STUDY DESIGN AND REFLEXOLOGY OUTCOMES

Most studies (n=19) used a control strategy that took the form of nonreflexology massage of parts of the foot (n=15), of which only 4 reported avoidance of the reflexology points under test, or of the head or lower limb. The control strategy in other studies was no treatment or passive relaxation (n=14) or involved a more active relaxation intervention, such as friendly discourse or a self-initiated strategy (n=4). For analytical purposes, control strategies were dichotomized as no treatment/relaxation (n=18) and sham/alternative massage (n=19). Control strategy was not significantly associated with any other methodologic fields, although a borderline significance (p=0.045; n=37) with methods suggested a potential link between use of a nonmassage strategy and non-RCT design, and of a massage strategy with RCT design.

Treatment fidelity. Table 3 identifies where the reflexology treatment was delivered, by whom, how often, and for how long at each session. In the 33 studies that reported the location of treatment, the majority (n=27) noted that the intervention took place in a room within a formalized clinical setting, such as a specialist center or unit, hospital ward, or clinic. The remainder were in nonclinical locations: residential home (n=2), own home (n=2), or laboratory (n=2). Reflexology was mainly delivered by qualified reflexologists (n=25 studies) or by partners, researchers, or students trained as part of the program of study (n=7). For analytical purposes, this field was dichotomized as practitioner/therapist and "Other.".

Interventions were delivered over 10.1 ± 13.6 weeks (n=32; median, 7 weeks). The number of reflexology treatments that each participant received averaged 6.3 ± 5.2 treatments per participant (median, 6; n=33 studies). The duration of each treatment also varied and averaged 35.4 ± 15.0 minutes (median, 30 minutes; n=33 studies). For analytical purposes, numbers of treatments were categorized as 1–4, 5–10, and 11 or more per participant, whereas treatment duration was categorized as 10–30, 31–45, and 46–60 minutes.

None of the treatment fidelity fields were significantly associated with each other, or with study setting or objective, demographic aspects, or any methodologic fields.

Study measures and outcomes

As noted, 25 of 37 study settings related to participants who had an underlying pathologic problem. Seven studies also related to symptoms of normal physiologic phenomena (e.g., menopause, labor, cold tolerance). Only 5 studies related to an absence of symptoms ("none"; see Table 2). Outcome measures therefore were mostly directed at evaluation of reflexology effect on biological or psychological measures specific to individual study objectives.

A total of 116 measures were identified from the 37 reports, an average of 3.1 measures per study. Sixteen studies reported application of 1 or 2 measures, 13 reported 3 or 4, 7 reported 5 or 6, and 2 reported 7 or more. Total measures applied were not associated with any reporting fields. Seventeen studies applied only psychometric (i.e., self-reported) measures, 6 applied only biometric measures, and 14 applied both types of measures (Table 3). Therefore, 31 studies applied at least 1 psychometric tool as a measure of emotional, cognitive, or behavioral indicators (e.g., questionnaire for pain, mood, or anxiety), and 20 applied at least 1 biometric tool as a measure of physical/biochemical functioning (e.g., diastolic/systolic blood pressure, hormone secretion, immune system components).

The total number of psychometric measures was 62 (2.0 per study on average), and the total number of biometric measures was 54 (2.7 per study on average). Type of measure (psychometric or biometric) was not associated with the methods or randomization fields, indicating no particular weighting in this respect or with most other fields. The exception was a not-unexpected significant association (p<0.001; n=37) of type of measure and study objective, primarily of psychometric measures with the behavioral/cognitive objective.

Fifty measures (43% of total used) demonstrated a significant response when compared with control data. Closer examination suggested a differential between responses involving each type of measure. Of the 30 studies that applied at least 1 psychometric tool, 25 (83%) reported 36 significant responses, whereas of the 20 studies that applied biometric tools, 10 (50%) reported 15 significant responses. The distribution of outcomes suggested a significantly greater likelihood (p<0.001) of psychometric outcomes following reflexology massage. The most frequent psychometric responses were pain reduction (n=9 studies), reduced anxiety/improved mood (n=9), and improved perception of well-being/health/quality of life (n=7).

Biometric responses mainly centred on cardiac/peripheral circulation (n=7 studies), airways/breathing (n=2), and muscle tone (n=2). Several studies, however, did not demonstrate any significant changes to psychometric (n=14 studies) or biometric (n=13) measures and so there was no discernible overall pattern of responses to reflexology.

For further analytical purposes, the studies were collated as demonstrating no effect in all measures or significant/mixed effect, in which all of the measures responded or at least some did when outcomes were inconsistent; the latter category therefore comprised studies in which at least 1 indicator outcome was significant. Studies reporting at least 1 significant biometric outcome (n=20) were not significantly associated with any of the reporting fields. For psychometric measures, there was a highly significant association of studies reporting a significant outcome with the control strategy (p=0.007; n=31), especially when this did not involve massage (no treatment/relaxation). Otherwise significant outcomes were not associated with the remaining reporting fields.

Jadad evaluation

Although not a reporting field per se, the Jadad score (0–5) provides a composite related to sample randomization, blinding, and the recording of noncompleting participants that is considered to provide a measure of rigor in methodologic design. Only 2 reporting fields demonstrated a significant association with the collated Jadad score subgroups. Unsurprisingly, because it is a key criterion, the more rigorous computerized randomization process was highly associated with the higher Jadad score subgroup (p<0.001; n=37). Likewise, an association of the methods field (p<0.040; n=37), primarily RCT, with the high-score
group was also to be expected because RCTs were predominant in that group. The only other design field to associate with Jadad score was number of treatments \((p = 0.015; n = 33)\), which indicated a link between 5 or more treatments per participant with high Jadad scores and 1–4 treatments per participant with low Jadad score studies. The occurrence of psychometric or biometric outcomes also were not significantly associated with the collated Jadad scores \((p = 0.484; n = 30)\) and \(p = 0.670; n = 20\), respectively).

Discussion

Reviews of RCTs identify a lack of evidence for consistent outcomes of reflexology but also identify that evaluation is complicated by the breadth of designs. In seeking to address criticisms from biomedical researchers, this study has explored the potential influence that interventional study design might have on biometric and/or psychometric study outcomes. In doing so it has not addressed the potential role of practitioner–client relationships and therapeutic contact, which are important aspects of holistic care but are difficult to evaluate in the context of an experimental design. In the present analysis, the reporting by studies on fidelity issues, such as who delivered the reflexology, the school of reflexology, and the duration of treatment, point toward holistic influences. However, to directly evaluate the practitioner–client relationship would necessitate incorporation of appropriate qualitative data. These were largely absent from the studies and in any case were outside the scope of this current review, which sought statistical evidence of impact by involving studies that provided quantitative analysis of outcomes against control or comparator data. The outcomes of this review should be interpreted in light of this limitation.

The validity of integrating studies that had applied RCT or non-RCT methods within a single database was generally supported through appraisal of reporting of design fields, and a lack of their specific association with Jadad quality criteria, although the randomization process was generally more rigorous in the RCTs. The methods field also was not significantly associated with study setting, study objective, sample demographics, intervention fidelity fields, or the number and type of measures applied, but RCTs appear weakly linked to a contact sham/alternative massage control strategy. The present review therefore largely supports the published view that there is no consensus on reflexology research and extends this by not finding associations regardless as to whether a design was an RCT, or according to Jadad quality criteria.

This review of study designs identified four main findings in relation to reflexology outcomes. First, it identified a stronger likelihood of self-reported psychometric, rather than biometric, outcomes. Pain reduction, stress reduction, and improved perceptions of health and well-being were most prominent among the psychometric outcomes. That finding provides some support for a published review\(^5\) of RCTs that had a narrow clinical focus (setting) in which reflexology was found to significantly affect fatigue, sleep, and pain outcomes, and also published evidence for an acute reduction of perceived psychological stress following reflexology that is reproducible with repeat treatments.\(^21\) There is also some literature evidence \(^{21,22}\) for effect of reflexology on systolic/diastolic blood pressure and heart rate when this is specifically related to the stress response, but in this review no evidence was found to associate biometric responses with any particular measure. Present and related findings therefore support expressed views\(^1,3,5\) that less heterogeneity of pathologies and study objectives would aid interpretation of clinical outcomes of reflexology. However, efficacy was not the primary objective of this review; rather, the objective was to identify whether design issues might have influenced the identification of a significant outcome.

Second, the observation that significant outcomes following reflexology did not appear to be associated with use of RCT or non-RCT method suggests that these designs do not offer bias toward positive outcomes, at least with the sample sizes (mean, 59) involved in the selected studies. Likewise, there was no association with application of Jadad scoring criteria. A caveat is that Jadad quality criteria were low across the database, and so more studies are needed that also consider current findings against higher-rated studies, whether RCT or non-RCT.

Third, the likelihood of detecting a significant change in psychometric indicators appears to be increased when a (nonmassage) rest and relaxation control strategy is used. On one level this association looks highly promising as a recommendation for study designs, but it is also potentially problematic for reflexology research as it does not support a distinction of reflexology from the effects of other forms of massage. A further caveat here is that sham reflexology massage was applied in just four studies, and the “contact” data therefore are weighted toward massage that could not be considered a genuine placebo. Identifying a suitable placebo is problematic for CAM research.\(^9\) Present findings suggest that more research should be done to explore differential outcomes when sham reflexology and alternative massage are applied within the same study to help clarify the potential of interventional designs to evaluate efficacy.

Fourth, a lack of consensus regarding treatment fidelity has been noted in reflexology research,\(^12\) and current findings support this view by failing to identify associations between fidelity fields, or between them and other design fields. Additionally, no association was observed between fidelity fields and study outcomes despite the selected studies on average meeting recommendations\(^12\) that participants receive at least 6 treatments of 30 minutes each (studies in this review averaged 6.3 treatments of 35.4 minutes’ duration).

Conclusions

Published reviews of reflexology RCTs are critical of inconsistency and low rigor of study designs. The pooling of both RCT and non-RCT studies of variable methodologic rigor in this integrative review therefore enabled exploration of possible association of study outcomes with aspects of study design. The review identified that such inconsistencies extend to the design of any interventional research studies, whether RCT or non-RCT. A new finding was that an impact of reflexology is more likely on psychometric than on biometric indicators when responses are compared to nonmassage control data. Within the limits of this review, that finding does not support a distinction between reflexology and the effects of massage per se but it is possible that such
STUDY DESIGN AND REFLEXOLOGY OUTCOMES

distinction could have been masked by the range of contact
control designs applied.

It is important to identify specific control strategies in
order to evaluate the effectiveness of reflexology. Future
research therefore ought to look more closely at this issue by
comparing different control strategies within an otherwise
single interventional design, thereby enabling a stronger
analysis beyond that possible in this review. Additionally,
this review supports suggestions that effects of reflexology
are better explored when studies have a homogeneous
clinical or nonclinical setting and narrow study objectives.

Acknowledgments

This article arose after completion of an RCT involving
reflexology21 that was supported by funds allocated to au-
thors C.G. and C.E. by Anglia Ruskin University, United
Kingdom. A detailed paper of that study is in preparation.
Authors A.M., C.G., and C.E. are academic staff members
of Anglia Ruskin University and further acknowledge a
faculty undergraduate studentship awarded to author C.L. in
the undertaking of this review.

Author Disclosure Statement

No competing financial interests exist.

References

1. Wang MY, Tsai PS, Lee, PH, Chang WY, Yang CM. The
efficacy of reflexology: systematic review. J Adv Nurs
2008;62:512–520.
for the symptomatic treatment of breast cancer: a system-
3. Ernst E, Posadzki P, Lee MS. Reflexology: an update of a
systematic review of randomised clinical trials. Maturitas
2011;68:116–120.
C. Effects of foot reflexology on fatigue, sleep, pain: a
systematic review and meta-analysis. J Korean Acad Nurs
2011;41:821–833.
5. Yeung WF, Chung KF, Poon MMK, et al. Acupressure,
reflexology, and auricular acupressure for insomnia: a sys-
tematic review of randomized controlled trials. 2012;13:
971–984.
plementary and alternative medicine whole systems re-
search: beyond identification of inadequacies of the RCT.
7. Birch S. Clinical research on acupuncture: part 2. Con-
trolled clinical trials, an overview of their methods. J Alter
8. Caspar F, Grossman C, Unnussig C, Schramm E. Com-
plementary therapeutic relationship: therapist behaviour,
interpersonal patterns, and therapeutic effects. Psychother
9. Leach MJ. Rapport: a key to treatment success. Com-
10. Paterson C, Dieppe P. Characteristic and incidental (pla-
cebo) effects in complex interventions such as acupuncture.
11. Carter B. Methodological issues and complementary ther-
apies: researching interagues? Complement Ther Nurs
Midwife 2003:9:133–139.
Intervention fidelity. Aspects of complementary and al-
14. Steenkamp E, Scrooby B, van der Walt C. Facilitating
nurses’ knowledge of the utilisation of reflexology in adults
with chronic diseases to enable informed health education
during comprehensive nursing care. Health SA Gesondheid
2012;17:Article 567.
15. Whittemore R, Knafli K. The integrative review: updated
16. Ernst E. Is reflexology an effective intervention? A sys-
tematic review of randomised controlled trials. Med J
17. Liu R, Chang A, Reddy S, Hecht FM, Chao MT. Improving
patient-centered care: a cross-sectional survey of prior use
and interest in complementary and integrative health ap-
proaches among hospitalized oncology patients. J Altern
Extending the CONSORT Statement to randomized trials
of nonpharmacologic treatment: explanation and elabora-
Factors influencing intentions to stay and retention of nurse
quality of reports of randomized clinical trials: is blinding
controlled trial investigating the stress outcomes and cor-
tisol responses in breast cancer patients receiving reflex-
22. McCullough JEM, Liddle SD, Sinclair M, Close C, Hughes
CM. The physiological and biochemical outcomes associ-
ated with a reflexology treatment: a systematic review.
23. Stephenson N, Ndalton M, Dalton J, Keefe FJ, Engelke
M. Partner-delivered reflexology: effects on cancer pain
24. Stephenson N, Dalton J, Carlson J. The effect of reflexol-
ogy to improve quality of sleep amongst Taiwanese post-
25. McVicar AJ, Greenwood CR, Fewell F, D’Arcy V, Chand-
drasekharan S, Aldridge LC. Evaluation of anxiety, salli-
vary cortisol and melatonin secretion following reflexology
therapy: a pilot study in healthy individuals. Complement
26. Wilkinson ISA, Prigmore S, Rayner CF. A randomised-
controlled trial examining the effects of reflexology of
patients with chronic obstructive pulmonary disease
27. Mackereth PA, Booth K, Hillier VF, Caress AL. Reflex-
ology and progressive muscle relaxation training for people
with multiple sclerosis: a crossover trial. Complement Ther

Address correspondence to:
Andrew McVicar, PhD
Faculty of Health, Social Care and Education
Anglia Ruskin University
Chelmsford
United Kingdom

E-mail: Andy.McVicar@anglia.ac.uk
AUTHOR QUERY FOR ACM-2016-0015-VER9-MCVICAR_1P

AU1: Please confirm that all affiliations are correct.
AU2: Tables 1a and 1b have been combined into one table; Table 1c is now its own table (Table 2). The original Table 2 is now Table 3.
AU3: Please provide complete mailing address for corresponding author.
AU4: In the footnote for Table 1, please clarify what ex-1.0 means.
AU5: In Table 3, please define PS20.
AU6: In Table 3, reference 25 was listed twice. Please confirm change of second mention of this reference to 26.