

# Assessing the uncertainty of treatment outcomes in a previous systematic review of venous leg ulcer randomized controlled trials: Additional secondary analysis

Kristen A. Eckert MPhil | Marissa J. Carter PhD, MA 

Strategic Solutions, Inc, Bozeman, Montana

## Correspondence

Marissa J. Carter, PhD, MA, Strategic Solutions, Inc., 37 Voyager Lane, Bozeman, MT 59718.

Email: [mcarter@strategic-solutions-inc.com](mailto:mcarter@strategic-solutions-inc.com)

## Abstract

In this secondary analysis of a previous systematic review, we assessed randomized controlled trials evaluating treatments of venous leg ulcers in terms of factors that affect risk of bias at the study level and thus uncertainty of outcomes obtained from the interventions. Articles that assessed the wound bed condition in venous leg ulcers and that were published in English between 1998 and May 22, 2018 were previously searched in PubMed, Embase, CINAHL, CENTRAL, Scopus, Science Direct, and Web of Science. Duplicates and retracted articles were excluded. The following data were extracted to assess the risk of bias: treatment groups; primary and secondary endpoints that were statistically tested between groups, including their results and  $p$  values; whether blinding of patients and assessors was done; whether allocation concealment was adequate; whether an intention-to-treat analysis was conducted; whether an appropriate power calculation was correctly done; and whether an appropriate multiplicity adjustment was made, as necessary. Pre- and post-study power calculations were made. The step-up Hochberg procedure adjusted for multiplicity. Results were analysed for all studies, pre-2013 studies, and 2013/post-2013 studies. We included 142 randomized controlled trials that evaluated 14,141 patients. Most studies lacked blinding (72.5–77.5%) and allocation concealment (88.7%). Only 49.3% of trials provided a power calculation, with 27.5% having an appropriate calculation correctly done. Adequate statistical power of the primary endpoint was found in 27.2% of trials. The lack of multiplicity adjustment in 98.6% of studies affected the uncertainty of outcomes in 20% of studies, with the majority of the secondary endpoints (67.7%) in those studies becoming non-significant after multiplicity adjustment. Recent studies tended to weakly demonstrate improved certainty of outcomes. Venous leg ulcer randomized controlled trials have a high degree of uncertainty associated with treatment outcomes. Greater attention to trial design and conduct is needed to improve the evidence base.

**ABBREVIATIONS:** EBM, evidence-based medicine; ITT, intention-to-treat; RCT, randomized controlled trial; SD, standard deviation; VLU, venous leg ulcer.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Wound Repair and Regeneration* published by Wiley Periodicals LLC on behalf of The Wound Healing Society.

**KEYWORDS**

randomized controlled trials, risk of bias, trial design, uncertainty of outcome, venous leg ulcers

## 1 | INTRODUCTION

There is a glaring gap between evidence and clinical practice in wound care, with many clinicians relying solely on their clinical experience and a traditional approach to care.<sup>1,2</sup> The application of evidence-based medicine (EBM) to wound care is further complicated by the diverse variation in wound types and treatment options. Consequently, many clinical practice guidelines and recommendations have been based on expert opinion.<sup>2</sup> Limited evidence produced from wound care randomized controlled trials (RCTs) is a result of poorly designed studies that are underpowered with small sample sizes, have too short follow-up periods to be able to properly assess wound outcomes, and employ poor analysis of endpoints.<sup>3-5</sup> The lack of a sound and applicable evidence base in wound care results in great clinical uncertainty that clouds clinical decision-making and can contribute to the use of suboptimal treatments, inequalities in care, and wasted resources.<sup>3,6-9</sup> Uncertainty in outcome effects has also been a focal point of the GRADE system,<sup>10</sup> which has now been extended to the concept of the threshold or ranges to rate certainty of the evidence for an individual outcome.<sup>11</sup>

Venous leg ulcers (VLUs) are among the most ubiquitous types of wound,<sup>12</sup> with an annual incidence rate estimated to be greater than 2%, costing the United States up to \$14.9 billion each year.<sup>13</sup> Considered to be the highest level of evidence,<sup>1,2</sup> systematic reviews are the most relevant vehicle to evaluate the certainty of RCT outcomes. In health care, they are now used to develop clinical practice guidelines and are often required as a prerequisite to research funding.<sup>14,15</sup> A well-conducted systematic review can produce more reliable, precise, and generalizable results with limited bias to be used by providers, payers, researchers, and policymakers for therapeutic advancements.<sup>14,16</sup> To be able to properly assess the bias, uncertainty of outcomes, and external validity of VLU RCTs, a systematic review must assess the randomization process, allocation concealment, blinding, power analysis, attrition rates, study group similarities, eligibility criteria, primary outcome measures, the inclusion of an intention-to-treat (ITT) analysis, and multiplicity adjustment of secondary endpoints.<sup>1,14,17,18</sup>

In 2019, Gethin et al published a systematic review of 144 RCTs involved in the treatment of VLUs to assess the quality of reporting of data related to their external validity.<sup>19</sup>

Their results showed there was inadequate reporting of factors that aid the clinician in determining the applicability of research findings to their patient population, despite the recommendations from CONSORT being available for over 20 years. Generalizability of studies is 1 of the 5 key domains of the GRADE approach to conducting systematic reviews.<sup>20</sup> The goal of our study was to assess the same RCTs in terms of other factors that could affect risk of bias at the study level and assess certainty of outcomes obtained from the

interventions. We therefore sought to determine the uncertainty of outcomes for patients with VLUs treated with any drug, biologic, or device compared to standard of care or placebo.

## 2 | MATERIALS AND METHODS

### 2.1 | Study selection

We included the same studies selected by Gethin et al in their 2019 systematic review.<sup>19</sup> They included 144 RCTs published in English between 1998 and May 22, 2018 that assessed the wound bed condition. They excluded any studies of non-venous wounds and any studies that they did not have open access to from their search. The following online databases were searched: PubMed, Embase, CINAHL, CENTRAL, Scopus, Science Direct, and Web of Science. We screened the original article selection to confirm there were no duplicates and no retracted articles.

### 2.2 | Data extraction

One of us (KAE) extracted the following data from the studies into Word files, which were used to assess the risk of bias at the study level: (1) treatment groups; (2) primary and secondary endpoints that were statistically tested between groups, including results of these endpoints for treatment groups and *p* values; (3) whether blinding of patients and outcomes assessment was done; (4) whether allocation concealment was adequate; (5) whether an ITT analysis was conducted for the primary endpoint; (6) whether a power calculation for the primary endpoint was reported that was appropriate and correctly done; and (7) whether an appropriate adjustment was made for multiplicity of statistical testing of secondary endpoints if more than 1 endpoint was tested. The other study author (M.J.C.) independently verified the data and collated it in Excel sheets for each study.

### 2.3 | Assessments and definitions

Unlike traditional systematic reviews, we scored assessments as binary (yes or no); for example, if an ITT analysis was performed, this assessment was scored as 'yes'. If any assessment was unclear, it was scored as a 'no'; for example, the authors stated patient blinding was done but did not provide supporting evidence. While we know that unclear assessments in bias components do not automatically equate to high risk assessments,<sup>21</sup> there has been no universal agreement on how to treat this issue. Consequently, it was decided to accept this loss of information in order to present the data in a simpler fashion.

**TABLE 1** Percentage of studies (*n*) with adequate blinding, allocation concealment, and ITT analysis conducted for the primary endpoint

Study characteristic	All studies ( <i>n</i> = 142)	Studies published before 2013 ( <i>n</i> = 97)	Studies published in 2013 and later ( <i>n</i> = 45)
Patients blinded	22.5% (31)	24.7% (23)	17.8% (8)
Study assessors blinded	27.5% (38)	24.7% (23)	33.3% (15)
Allocation concealment	11.3% (16)	9.3% (9)	15.6% (7)
ITT analysis	62.7% (89)	61.9% (60)	64.4% (28)

Abbreviation: ITT, intention-to-treat.

**TABLE 2** Percentage of studies (*n*) for statistical power categories of the primary endpoint before and after the trial

Statistical power	All studies	Studies published before 2013	Studies published in 2013 and later
Initial power calculation			
≥90	18.8% (9)	21.2% (7)	13.3% (2)
80–89	53.1% (25)	48.5% (16)	66.7% (10)
<80	27.1% (13)	30.3% (10)	20.0% (3)
Final power calculation			
≥90	17.6% (23)	19.6% (18)	13.6% (6)
80–89	9.6% (13)	9.8% (9)	9.1% (4)
<80	72.8% (99)	70.7% (65)	77.3% (33)

If no primary endpoint(s) could be identified, the most relevant and/or prominent endpoint was chosen. Secondary endpoints were defined as any remaining endpoint that was tested statistically between treatment groups regardless of whether such endpoints were explicitly identified as such by the study authors. Evidence of successful blinding and allocation concealment had to be supported by detailed statements in the study reporting.

The ITT population was defined as all patients who were randomized to treatment groups. Exceptions were patients who were inappropriately randomized; that is, consent form not signed, or patient later found to be ineligible due to inclusion/exclusion criteria.

To be appropriate, a primary endpoint power calculation had to be congruent with the primary endpoint it supported, with reasonable assumptions, method(s), and sufficient data that a power calculation could be replicated. If the calculation was incorrectly performed by the study authors, the result was scored as a 'no'.

Any discrepancies between our initial independent assessments were resolved by consensus.

## 2.4 | Statistical analysis

Pre- and post-study power calculations were made using Pass13 (NCSS, LLC, Kaysville, UT) software with (a) the figures supplied in the study, if a power calculation was done by the study authors and (b) figures based on the results of the primary endpoint for the ITT population, if available, or another population, if that was used by the study authors. Results were categorized in terms of statistical power: (a) <80; (b) 81–89; and (c) ≥90.

Attrition rates for all treatment groups—that is, those patients whose outcomes became right-censored—were calculated based on the primary length of each study and expressed as a percentage of total patients in each treatment group. The overall attrition rate was calculated, as well as whether there was a difference of ≥20% between any treatment groups.<sup>22</sup>

Adjustment for multiplicity of statistical testing used the step-up Hochberg procedure and was executed in Excel. The *p* values of all endpoints that were statistically tested with the exception of primary endpoint(s) were entered into the adjustment calculation. If there were co-primary endpoints, these were entered in a separate calculation. If actual *p* values were not reported but it was clear from the text that a statistical test was carried out, the following conservative *p* value imputations were made: non-significant: 0.06; <0.05; 0.04; <0.01; 0.009; <0.001; 0.0009.

## 2.5 | Reporting

The percentage of studies in which patients were blinded was calculated for all studies, pre-2013 studies, and 2013/post-2013 studies. The same procedure was followed for blinded study assessment; adequate allocation concealment; ITT analysis (primary endpoint); reporting a study power calculation; appropriate power calculation; appropriate adjustment for multiplicity of statistical testing of secondary endpoints, if more than 1 endpoint was tested; the number of studies in which at least one secondary endpoint became statistically non-significant after adjustment; and the percentage of secondary endpoints that became statistically non-significant after adjustment.

The mean (standard deviation [SD]; range) of attrition rates across study breakpoints was also calculated.

**TABLE 3** Multiplicity adjustment analysis: (A) percentage of studies (*n*) in which one or more secondary endpoints became statistically non-significant; and (B) percentage of the affected endpoints that became statistically non-significant

Analysis	All studies	Studies published before 2013	Studies published in 2013 and later
(A) Percentage of studies ( <i>n</i> )	19.7% (27)	20.6% (20)	17.8% (8)
(B) Percentage of affected endpoints (SD; range)	67.7% (32.1; 1–100)	71.0% (30.1; 17–100)	59.5% (37.4; 1–100)

**TABLE 4** Mean attrition rates (%) by treatment group (SD; range)

Group no.	All studies	Studies published before 2013	Studies published in 2013 and later
All groups combined	10.9 (12.0; 0.0–73.7)	12.4 (13.0; 0.0–73.7)	7.6 (8.5; 0.0–35.0)
Group 1	9.0 (11.7; 0.0–89.5)	9.5 (12.5; 0.0–89.5)	8.2 (10.1; 0.0–39.1)
Group 2	10.4 (12.6; 0.0–57.9)	12.4 (13.8; 0.0–57.9)	6.7 (9.3; 0.0–40.0)
Group 3	4.5 (6.9; 0.0–22.2)	5.3 (7.2; 0.0–22.2)	0.0 <sup>a</sup>
Group 4	1.6 (4.2; 0.0–11.1)	2.2 (5.0; 0.0–11.1)	0.0 <sup>a</sup>
Group 5	1.6 (3.6; 0.0–8.0)	2.7 (4.6; 0.0–8.0)	0.0 <sup>a</sup>
Group 6–10 <sup>a</sup>	0.0	0.0	Not applicable

<sup>a</sup>*n* = 1 study.

### 3 | RESULTS

We included 142/144 (98.6%) articles analysed in the Gethin et al systematic review in our analysis (Supporting Information S1). One study was excluded, because it was retracted by the journal (No. 40), and another article was excluded for being a duplicate study (No. 1). One reference from the original systematic review (No. 89) was replaced with another article (Iglesias et al, 2004) published in the same year that had more complete information and data for the purposes of our analysis. There were 14,141 patients evaluated in the included studies. Complete study characteristics and data are provided in Supporting Information S1. Among the included studies, 97 (68.3%) were published before 2013 and 45 (31.7%) were published in 2013 or later.

#### 3.1 | Risk of bias assessment

While blinding and adequate allocation concealment was not reported in the majority of studies (72–89%), almost 63% of studies conducted an ITT analysis on the primary endpoint (Table 1). Better trends were evident for all study characteristics for later year studies with the exception of patient blinding, which was lower; study assessor blinding increased by almost the same amount.

Seventy (49.3%) studies reported a power calculation; 48 (49.5%) were studies published before 2013 and 22 (48.9%) were studies published in 2013 or later. Among all studies, 27.5% (39/142) had a power calculation that was appropriate and correctly done. Among all studies published before 2013, 28.9% (28/97) had a power calculation that was appropriate and correctly done compared to 24.4% (11/45) of all studies published in 2013 or later. Forty-eight (33.8%) of all studies, 33 (34%) studies published before 2013, and

15 (33.3%) published in 2013 or later reported an initial power calculation of the primary endpoints. There were 136 (95.8%) total studies, 92 (94.8%) studies published before 2013, and 44 (97.8%) studies published in 2013 or later in which a post-study calculation could be done regarding the primary endpoint. In the majority of trials, pre-trial calculations were optimistic in terms of statistical power, with post-trial results showing only 27.2% of trials having an adequate power (Table 2). Overall, there was a positive trend for pre-trial power but a negative trend for post-trial power in respect of newer versus older studies.

Only two studies (1.4%) had an appropriate adjustment for multiplicity of statistical testing of secondary endpoints (if more than 1 endpoint was tested). Both studies were published in 2013 or later (4.4%). Nearly one fifth of studies were negatively affected by multiplicity of statistical testing with over two thirds of endpoints becoming statistically non-significant (Table 3). Although the results varied little with respect to time, the percentage of endpoints that became statistically non-significant was lower in newer versus older studies.

Of all the studies, 127 (89.4%) had 2 treatment arms, 14 (9.6%) had 3 treatment arms, 7 (4.9%) had 4 treatment arms, 5 (3.5%) had 5 treatment arms and 1 (0.7%) had 10 treatment arms. Studies published in 2013 or later only had 2 treatment arms. The mean attrition rate for all 142 studies regardless of number of treatment arms was 10.9%, with attrition rates much better in later studies compared to more recent ones (Table 4). There were only eight studies (5.6%) in which there was greater than 20% difference reported between any treatment groups, equally divided into the two time periods. There were 29 studies (20.4%) with a total attrition rate of 20% or greater. Twenty-five of these studies (25.8%) were published before 2013, and 4 studies (8.9%) were published in 2013 or later.

## 4 | DISCUSSION

### 4.1 | Risk of bias and outcome uncertainty

Assessing the certainty (or uncertainty) of treatment outcomes in a given population for a chronic disease is an important criterion needed by clinicians and payers to determine whether a specific treatment should be employed. Many factors can influence that degree of certainty and our assessment of several of those factors shows that the majority of the RCTs involving treatment of VLU included in this study have one or more issues that affect the certainty of results. While meta-analysis can provide higher statistical power through pooling of outcome data, heterogeneous trial design frequently limits this approach.<sup>1,23-28</sup> Moreover, meta-analysis ignores the many factors that increase risk of bias so the final result itself (point estimate and 95% confidence intervals) can still be biased.<sup>24-28</sup> Metaregression or multilevel meta-analysis can potentially adjust for study parameter differences but can be limited by small sample sizes and other issues, and it requires considerable expertise.<sup>29</sup>

High risk of bias featured prominently in both lack of blinding (about three quarters of trials) and lack of allocation concealment (approximately 9 out of 10 trials; Table 1). Although statistical power was consistent (Table 2), less than half of the trials evaluated (49.3%) provided a power calculation, and only 27.5% had an appropriate calculation that was correctly done, with sufficient information provided to replicate the calculation. More recent studies demonstrated a positive trend for pre-trial power and a negative trend for post-trial power compared to older studies. In particular, the finding that only 23% of studies had an adequately powered primary endpoint is disturbing. The lack of multiplicity adjustment in almost all studies (98.6%) ultimately affected the uncertainty of outcomes in 1 out of 5 studies, with the majority of the secondary endpoints (67.7%) in those studies becoming statistically non-significant after we conducted multiplicity adjustment (Table 3). These figures improved slightly in newer studies.

The strengths of these trials were that ITT analysis was performed in the majority (62.7%) (Table 1), and attrition rates were generally low (Table 4), with recent studies reporting much lower attrition rates than older studies. While there were no drastic differences between newer and older studies, recent studies tended to weakly demonstrate improved certainty of outcomes. However, our analysis also shows that VLU RCTs predominantly demonstrate a high risk of bias and low certainty of outcomes.

### 4.2 | Levels of evidence

Distilling these results into whether an individual RCT meets level 1 evidence requires a great deal of judgment. If we just take into account that an RCT has to be properly powered  $\geq 80\%$  and has  $< 20\%$  in regard to attrition (Oxford Centre for Evidence-based Medicine),<sup>30</sup> then only 13.7% of studies would be ranked at this level. Moreover, our analysis has demonstrated that many of these studies had serious design and/or reporting flaws that weakened their evidence and

external validity. We concur with Gethin et al, that these findings are at least partially due to study authors not reporting trial information in sufficient detail,<sup>19,31</sup> and we know from firsthand experience how authors are limited to journal word counts that can lead to shortened descriptions of the randomization and allocation concealment methods, as an example. Authors from the trials in the review tended to use the sealed envelope method for allocation concealment, but they glossed over the details to ensure that the envelopes were not tampered or manipulated prior to subject allocation. Perhaps, where word limitations are enforced, authors could include a Supporting Information where they further elaborate their allocation concealment methods to ensure trial integrity. Another issue is that many studies suffer from English language issues, which can lead to ambiguity or misinterpretation of key points. Nevertheless, there needs to be a subcategory ranking for RCTs with weak design, as systematic reviews of RCTs of wound care consistently demonstrate that RCTs are poorly designed.<sup>1,3,19,32</sup>

The clinical implication of our findings confirming uncertainty of outcomes in VLU RCTs is that the evidence base, which drives clinical practice guidelines, needs considerable improvement. While VLU guidelines universally promote high compression therapy for VLU management, beyond that recommendation, guidelines are marked by notable heterogeneity, lack of standardized definitions and criteria, conflicting recommendations, and poor applicability and uptake in clinical practice.<sup>33-35</sup> Thus, there remains a great need to improve EBM in venous wound care, and to do so, better RCT design and conduct is essential.

### 4.3 | Trial design

In wound care, robust RCT design is sometimes not possible, depending on the intervention or the condition studied. For example, in trials involving less common conditions (such as presence of an open abdomen in persons with a stoma) attaining the desired target can be problematic, while effective blinding can be hard to achieve in trials involving hyperbaric oxygen or negative pressure wound therapy, despite innovative approaches.<sup>1,36-38</sup> Ultimately, blinding in wound care RCTs may be best addressed by a third party blinded assessor,<sup>39</sup> but our analysis revealed only a little over a quarter of studies had a blinded assessor. In many other areas of medicine, including plastic surgery, there are similar EBM issues to wound care and difficulties exist in designing trials with appropriate randomization and blinding.<sup>3,18</sup> A review of 72 RCTs published in the *Journal of Bone and Joint Surgery* from 1988 to 2000 found that 60% of those trials were of low quality due to inappropriate randomization, blinding, and nebulous exclusion criteria.<sup>18,40</sup> Researchers have to be more creative in how they tackle these issues in their trial design; for example, could inactive bioresorbable materials be made to approximate the appearance of any of the numerous cellular and/or tissue-based products (CTPs) currently being used in wound care so that the current subject blinding issue is ameliorated? While this seems an outlandish suggestion (it would have to be demonstrated that any such material did not

affect wound healing and might also need FDA approval) an industrial consortium approach in which all CTP manufacturers contribute could explore its feasibility. Finally, better use of adaptive designs or even hybrid designs would allow for more identification of more responsive subjects or more efficient determination of safety/efficacy, as well as better generalizability of outcomes with attendant lower uncertainty.<sup>39,41</sup> This last point refers to the inclusion of real-world data in trial design and generation of real-world evidence which can be patient-centric and better reflect the real-world clinical setting, by capturing patient preferences, perspectives, and harms versus treatment benefits that are not effectively captured by the traditional RCT design.<sup>39</sup> This can be achieved via a pragmatic, large, simple trial design, which enrolls a broader patient population than the typical controlled trial and target clinically meaningful outcomes that are more relevant and applicable to the real-world clinical setting.<sup>39,42</sup>

#### 4.4 | Trial financial resources: problems and solutions

We recognize that sponsors of wound care clinical trials in general, including those cited in our study, often do not have the deep financial pockets of most pharmaceutical and biologics manufacturers. This is especially true of medical devices being used to treatment of VLU, where such devices are often low tech and relatively cheap. This means that there are significant restrictions on the financial resources that companies are able to put into RCTs. Further, if devices follow the 510 (k) approval pathway in the US, there is no strict requirement to provide relative efficacy data in postmarket trials, which often leads to sample sizes that are not necessarily related to the effect size of the device but rather a sample size that fits the resources available for the trial. Clearly, this has the effect of increasing the chance of a type II error and can also lead to non-coverage by payers. We strongly recommend that authors spend a little time conducting sample size scenarios even when working with resource-limited trials. For example, in this review, the 2011 Brizzio et al trial (No. 11) had a small sample size calculated of 52 subjects who were allocated to two different groups. Nonetheless, the power calculation was adequate and based on a previously determined effect of 0.39 regarding healing with a power of 0.8 at a two-sided  $p < 0.05$ . Furthermore, working through different adaptive designs, developing best practice statistical analysis plans, and thoroughly documenting these processes, particularly in trial publications, would go a long way towards better trial evaluation by peers and perhaps provide more realistic expectations of the trial before subject recruitment starts. This is the proverbial 'an ounce of prevention is worth a pound of cure'.

Then, there are the problems of subject blinding. While it is often not practicable to make sham devices, especially for CTPs, it is encouraging to see that some manufacturers are becoming more creative in recent times. Among the articles selected in this review, there were several successful blinding examples using creative sham devices. Gupta et al (No. 46) evaluated the effect of low photon energy therapy versus a placebo therapy that used the same device, which applied light of the same colour to the VLUs (Supporting

Information S1). Jünger et al (No. 58) evaluated electrical simulation versus a placebo therapy that used the same pulse generator emitting electrodes but with a nonconductive power cable (Supporting Information S1). Similar shams can be developed for laser therapy and ultrasound therapy. It certainly costs very little to brainstorm the issue rather than dismissing it out of hand. In terms of blinding the clinical assessors, while one researcher could act as a blind assessor, this requires two people to treat and assess the patient, which is not practical in many clinical situations, although is recommended for *trial* situations. Gould and Li recommend that wound care trials implement a standardized wound assessment methodology that tackles blinding, by using a blinded, on-site assessor, who is not the treating clinician, and a blinded, remote adjudication panel of two to three wound care experts.<sup>43</sup> Some products, such as CTPs, leave telltale marks in the wound area, which immediately inform an experienced assessor that a subject was treated with the intervention. This kind of problem can automatically invalidate blind assessment and is probably the most challenging aspect to assessing VLU treatment, but the use of artificial intelligence, such as computerized planimetry, and remote assessors to assess wounds could overcome these limitations.<sup>43</sup>

Clearly this is not an ideal world, but one that many researchers still have to inhabit. Nevertheless, spending more time to develop new avenues to solving old problems before the trial starts, rather than ignoring them, is likely to pay off even under financial constraints.

#### 4.5 | Study strengths and limitations

Rather than select a body of studies to examine a given treatment approach for our systematic review, we chose an existing review that focused on a particular assessment—study generalizability to broader populations—so we could add our results to visualize a bigger picture for one very common wound type. Consequently, our secondary analysis inherits some of the same limitations described by Gethin et al.<sup>19</sup> We did not perform a new, updated search of trials from June 2018 onward. They only used English-language articles that they could access freely, so their initial search may not have been the most comprehensive, and there is some publication bias acknowledged. Further, we did not individually assess the risk of bias for each trial; our study design was based on assessing the overall risk of bias of these studies, including the newer versus older articles. We did not perform a comprehensive analysis of every factor that could potentially influence bias and uncertainty of outcomes in RCTs; for example, publication bias and consistency of treatment effects were not analysed in this review. However, allocation concealment, blinding, power analysis, attrition rates, primary outcome measures, the inclusion of an ITT analysis, and multiplicity adjustment of secondary endpoints are all major factors influencing uncertainty of outcomes and external validity that were not considered in the systematic review by Gethin et al.<sup>1,14,17-19</sup> Finally, given our bundling of unclear with negative assessment categories, we recognize that in some instances, our results may be seen as too conservative. Nevertheless, the large dataset compiled from 142 VLU RCTs is a major strength of our analysis, and the results of our study demonstrate that more critical analysis of



the uncertainty of outcomes in wound care is needed for other wound types and outcomes.

## 5 | CONCLUSIONS

VLU RCTs have high bias and poor uncertainty of outcomes incurred by lack of blinding and allocation concealment, insufficient statistical power associated with outcomes, and lack of multiplicity adjustment. Newer studies tend to very weakly demonstrate improved certainty of outcomes compared to older studies. Greater attention to the uncertainty of outcomes and trial design and conduct is needed to improve the evidence base in wound care.

### CONFLICT OF INTEREST

Kristen A. Eckert was a paid consultant of Strategic Solutions to this study; Marissa J. Carter: none to declare.

### ORCID

Marissa J. Carter  <https://orcid.org/0000-0002-2265-6639>

### REFERENCES

- Carter MJ. Evidence-based medicine: an overview of key concepts. *Ostomy Wound Manage.* 2010;56(4):68-85.
- Ubbink DT, Santema TB, Stoekenbroek RM. Systemic wound care: a meta-review of Cochrane systematic reviews. *Surg Technol Int.* 2014;24:99-111.
- Christie J, Gray TA, Dumville JC, Cullum NA. Do systematic reviews address community healthcare professionals' wound care uncertainties? Results from evidence mapping in wound care. *PLoS One.* 2018;13(1):e0190045.
- Hodgson R, Allen R, Broderick E, et al. Funding source and the quality of reports of chronic wounds trials: 2004 to 2011. *Trials.* 2014;15:1-10.
- Serena T, Bates-Jensen B, Carter MJ, et al. Consensus principles for wound care research obtained using a Delphi process. *Wound Repair Regen.* 2012;20(3):284-293.
- Soares MO, Dumville JC, Ashby RL, et al. Methods to assess cost-effectiveness and value of further research when data are sparse: negative-pressure wound therapy for severe pressure ulcers. *Med Decis Making.* 2013;33(3):415-436.
- Palacios-Ceña D, Cachón-Peréz JM, Martínez-Piedrola R, Gueita-Rodríguez J, Perez-de-Heredia M, Fernández-de-las-Peñas C. How do doctors and nurses manage delirium in intensive care units? A qualitative study using focus groups. *BMJ Open.* 2016;6(1):e009678.
- Balsa AI, Seiler N, McGuire TG, Bloche MG. Clinical uncertainty and healthcare disparities. *Am J Law Med.* 2013;29(2-3):203-219.
- French B. Uncertainty and information need in nursing. *Nurse Educ Today.* 2006;26(3):245-252.
- Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol.* 2011;64(4):401-406.
- Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol.* 2017;87:4-13.
- Guest JF, Ayoub N, McIlwraith T, et al. Health economic burden that wounds impose on the National Health Service in the UK. *BMJ Open.* 2015;5:e009283.
- Rice JB, Desai U, Cummings AK, Birnbaum HG, Skornicki M, Parsons N. Burden of venous leg ulcers in the United States. *J Med Econ.* 2014;17(5):347-356.
- Gopalakrishnan S, Ganeshkumar P. Systematic reviews and meta-analysis: understanding the best evidence in primary healthcare. *J Family Med Prim Care.* 2013;2(1):9-14.
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.
- Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). *BMJ.* 1997;315(7109):672-675.
- Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest.* 1989;95(Suppl 2):2S-4S.
- Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* 2011;128(1):305-310.
- Gethin G, Ivory JD, Connell L, McIntosh C, Weller CD. External validity of randomized controlled trials of interventions in venous leg ulceration: a systematic review. *Wound Repair Regen.* 2019;27(6):702-710.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336(7650):924-926.
- Mhaskar R, Djulbegovic B, Magazín A, et al. Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. *J Clin Epidemiol.* 2012;65(6):602-609.
- Babic A, Tokalic R, Amílcar Silva Cunha J, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. *BMC Med Res Methodol.* 2019;19(1):76.
- Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ.* 1997;315(7121):1533-1537.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557-560.
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315(7109):629-634.
- Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA.* 1996;276(16):1332-1338.
- Villar J, Carroli G, Belizán JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet.* 1995;345(8952):772-776.
- Flournoy N, Olkin I. Do small trials square with large ones? *Lancet.* 1995;345(8952):741-742.
- Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. *Proc Natl Acad Sci U S A.* 2017;114(14):3714-3719.
- OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence 2. Oxford Centre for Evidence-Based Medicine. <https://www.cebm.net/index.aspx?o=5653>.
- Rothwell PM. Commentary: external validity of results of randomised trials: disentangling a complex concept. *Int J Epidemiol.* 2009;39(1):94-96.
- Beitz JM, Bolton L. Systematic reviews and meta-analyses: literature-based recommendations for evaluating strengths, weaknesses, and clinical value. *Ostomy Wound Manage.* 2015;61(11):26-42.
- Andriessen A, Apelqvist J, Mosti G, Partsch H, Gonska C, Abel M. Compression therapy for venous leg ulcers: risk factors for adverse events and complications, contraindications - a review of present guidelines. *J Eur Acad Dermatol Venereol.* 2017;31(9):1562-1568.
- Tan MKH, Luo R, Onida S, Maccatrozzo S, Davies AH. Venous leg ulcer clinical practice guidelines: what is AGREEd? *Eur J Vasc Endovasc Surg.* 2019;57(1):121-129.
- Couch KS, Corbett L, Gould L, Girolami S, Bolton L. The international consolidated venous ulcer guideline update 2015: process improvement, evidence analysis, and future goals. *Ostomy Wound Manage.* 2017;63(5):42-46.
- van Rijswijk L. It's all about the "E". *Ostomy Wound Manage.* 2009;55(6):4.

37. Lansdorp CA, van Hulst RA. Double-blind trials in hyperbaric medicine: a narrative review on past experiences and considerations in designing sham hyperbaric treatment. *Clin Trials*. 2018;15(5):462-476.
38. Peinemann F, Labeit A. Negative pressure wound therapy: a systematic review of randomized controlled trials from 2000 to 2017. *J Evid Based Med*. 2019;12(2):125-132.
39. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: the changing landscape of the regulatory approval process. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1201-1212.
40. Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the journal of bone and joint surgery from 1988 through 2000. *J Bone Joint Surg Am*. 2002;84(3):388-396.
41. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018;16(1):29.
42. Buch MH, Pavitt S, Parmar M, Emery P. Creative trial design in RA: optimizing patient outcomes. *Nat Rev Rheumatol*. 2013;9(3):183-194.
43. Gould L, Li WW. Defining complete wound closure: closing the gap in clinical trials and practice. *Wound Repair Regen*. 2019;27(3):201-224.

#### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Eckert KA, Carter MJ. Assessing the uncertainty of treatment outcomes in a previous systematic review of venous leg ulcer randomized controlled trials: Additional secondary analysis. *Wound Rep Reg*. 2021;29:327-334. <https://doi.org/10.1111/wrr.12897>