

# Infusion phlebitis assessment measures: a systematic review

Gillian Ray-Barruel RN BSN, BA(Hons) Grad Cert (ICU Nursing),<sup>1</sup> Denise F. Polit RN PhD,<sup>2</sup> Jenny E. Murfield BSc(Hons)<sup>3</sup> and Claire M. Rickard RN PhD<sup>2</sup>

<sup>1</sup>Senior Research Assistant, <sup>2</sup>Professor in Nursing, <sup>3</sup>Research Development Coordinator, NHMRC Centre for Research Excellence in Nursing, Centre for Health Practice Innovation, Griffith Health Institute, Griffith University, Brisbane, Queensland, Australia

## Keywords

assessment, measurement, peripheral intravenous catheter, phlebitis, psychometric assessment, scales

## Correspondence

Ms Gillian Ray-Barruel  
Centre for Health Practice Innovation  
Griffith Health Institute  
Griffith University  
Bldg N48  
Kessels Road  
Nathan  
Qld 4111  
Australia  
E-mail: g.ray-barruel@griffith.edu.au

Accepted for publication: 22 November 2013

doi:10.1111/jep.12107

## Abstract

**Rationale, aims and objectives** Phlebitis is a common and painful complication of peripheral intravenous cannulation. The aim of this review was to identify the measures used in infusion phlebitis assessment and evaluate evidence regarding their reliability, validity, responsiveness and feasibility.

**Method** We conducted a systematic literature review of the Cochrane library, Ovid MEDLINE and EBSCO CINAHL until September 2013. All English-language studies (randomized controlled trials, prospective cohort and cross-sectional) that used an infusion phlebitis scale were retrieved and analysed to determine which symptoms were included in each scale and how these were measured. We evaluated studies that reported testing the psychometric properties of phlebitis assessment scales using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines.

**Results** Infusion phlebitis was the primary outcome measure in 233 studies. Fifty-three (23%) of these provided no actual definition of phlebitis. Of the 180 studies that reported measuring phlebitis incidence and/or severity, 101 (56%) used a scale and 79 (44%) used a definition alone. We identified 71 different phlebitis assessment scales. Three scales had undergone some psychometric analyses, but no scale had been rigorously tested.

**Conclusion** Many phlebitis scales exist, but none has been thoroughly validated for use in clinical practice. A lack of consensus on phlebitis measures has likely contributed to disparities in reported phlebitis incidence, precluding meaningful comparison of phlebitis rates.

## Introduction

The insertion of a peripheral intravenous cannula (PIVC) for intravenous (IV) fluids and medications is the most common procedure in hospitalized patients worldwide [1]. A frequent PIVC complication is phlebitis, that is, inflammation of the vein, which may be mechanical, chemical or bacterial in origin [2,3]. Phlebitis causes a cascade of unwelcome repercussions – significant pain, failure of the PIVC, interruption to prescribed therapy and requirement for insertion of a new PIVC with associated increased equipment costs and staff time. Phlebitis compromises future venous access [4], and untreated bacterial phlebitis may lead to bloodstream infection [5]; therefore, early detection of complications and removal of the PIVC is crucial.

Phlebitis may be localized to the insertion site or travel along the vein. If extravasation (also called infiltration) of fluids in the interstitial space occurs [6], oedema may prevent recognition of phlebitis symptoms, such as induration (hardened tissue), because of difficulty in palpating the vein. Phlebitis may occur during catheterization or up to 48 hours after removal [7].

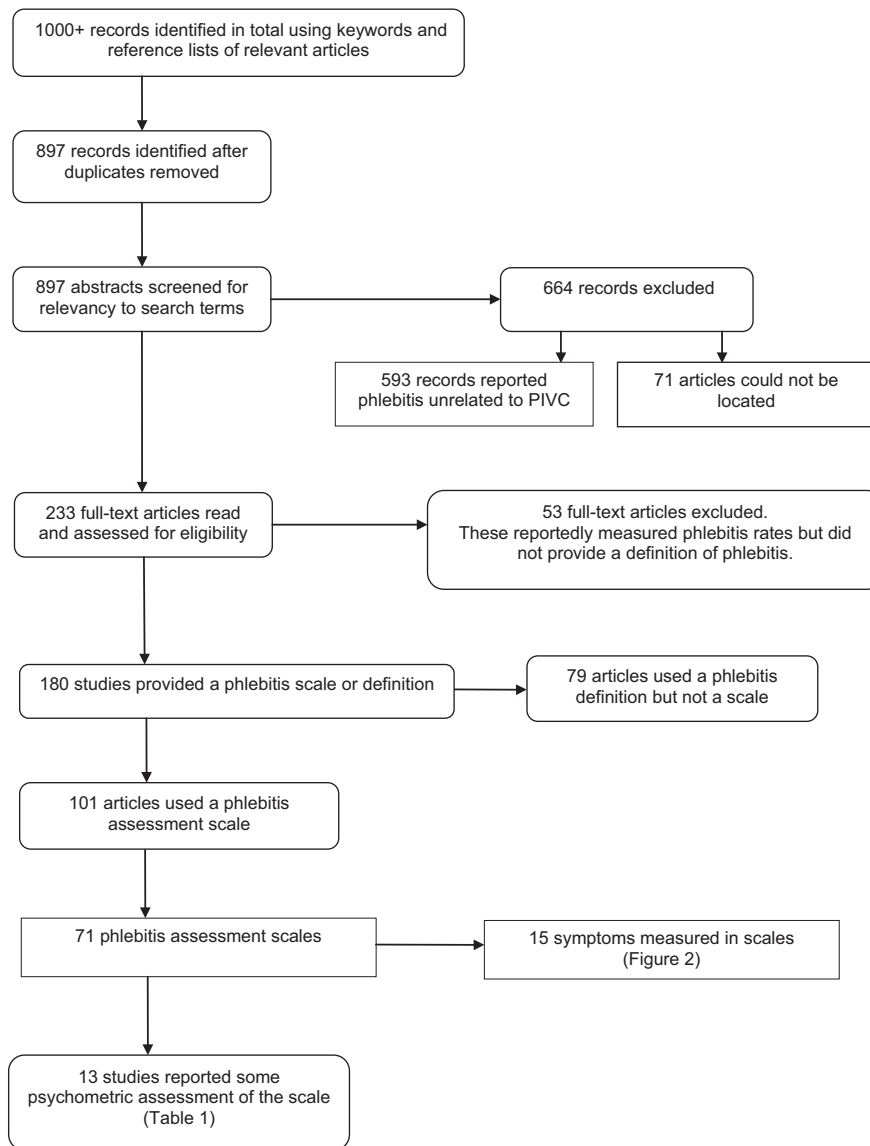
This systematic review sought to address the following questions:

- Which diagnostic criteria are used to determine infusion phlebitis in the clinical setting?
- Do any existing infusion phlebitis assessment scales have strong measurement properties, including reliability, validity, responsiveness and feasibility?

This review is intended to inform clinicians about existing methods of phlebitis assessment, based on evidence of the measurement quality of existing assessment scales.

## Methods

We searched the Cochrane library, Ovid MEDLINE and EBSCO CINAHL for research articles in English, using the following search terms: infusion phlebitis; thrombophlebitis; peripheral IV catheter; phlebitis score; phlebitis grade; and phlebitis assessment. Research studies (randomized controlled trials, prospective cohort and cross-sectional) that reported phlebitis incidence in adult patients with PIVCs or that evaluated a phlebitis scale were



**Figure 1** Process of selecting studies for the review.

included. No date limitations were applied, with citations published until September 2013 included. Titles and abstracts were initially screened for relevance. Full texts of potentially relevant articles were obtained and evaluated for inclusion. The reference lists of these articles were checked for other studies of potential relevance, and these were also retrieved.

All articles that examined infusion phlebitis assessment in adults as a primary outcome measure were retrieved, but only those that used a phlebitis assessment scale were included in the final review. Each scale was examined to identify which signs and symptoms were included in the measurement of phlebitis. Figure 1 illustrates the study selection process. The role of the phlebitis assessor, how often assessment was performed and if training in phlebitis assessment had been provided were noted. Information regarding each scale’s psychometric properties, if provided, was also recorded.

This review used definitions of measurement properties and parameters consistent with those provided by the COnsensus-

based Standards for the selection of health Measurement INstruments (COSMIN) [8,9]. Relevant measurement properties for phlebitis assessment include reliability (inter-rater, intra-rater, test–retest), validity (content, face, criterion, construct) and responsiveness. Because phlebitis scales are formative indexes rather than reflective scales [10,11], neither internal consistency nor structural validity is relevant.

In addition, our review also considered attributes associated with excellence in clinimetrics. Feinstein’s [12] approach to developing clinical assessment tools, especially relevant for formative indexes like phlebitis scales, was taken, including evaluation of the ‘sensitivity’ of clinical instruments. Sensibility includes several properties covered in COSMIN (e.g. content validity, responsiveness), but also includes acceptability and feasibility, that is, ease of practical application of clinical instruments. Feasibility takes into consideration such issues as length of time to complete the scale, ease of administration and clarity of the items and instructions [12].

## Results

Although phlebitis incidence related to PIVCs was reportedly measured in 233 studies, 53 (23%) articles did not provide any definition of phlebitis. Of the 180 studies that described the method of phlebitis assessment, 101 (56%) reported using a scale and 79 (44%) used a definition alone. Seventy-one phlebitis assessment scales including 15 symptoms were identified.

The 15 symptoms included in phlebitis assessment scales were pain, tenderness, erythema or redness, oedema or swelling, palpable venous cord, induration or hardness, frank thrombosis, streak formation or red line, purulence or exudate, local warmth, local coolness, infusion slowed or stopped, fever or pyrexia, tissue damage and impaired function. The prevalence of these symptoms captured in phlebitis assessment scales is shown in Fig. 2.

### Phlebitis assessment scales

Large disparities were found among the 71 phlebitis assessment scales. Some authors used a previously published scale; others modified an existing tool or created their own. When a published tool such as the Visual Infusion Phlebitis (VIP) [13,14], Infusion Nurses Society (INS) [15–18], Maddox [19,20], Baxter [21], Lipman [22] or Dinley [23] scale was used, many authors did not state which version they had used, despite wide variations between different versions. Other authors did not report the source of their scale at all.

Assigning a phlebitis assessment score or grade was commonly performed in one of two ways. Phlebitis scores were either cumulative (assigning points for each symptom and adding them up) or progressive (based on more points for a specified progression of symptoms). Cumulative scales scored 0–2 points for each phlebitis symptom, depending on the presence, measured length (in centimetres), or severity, and their total potential scores ranged from 0–6 to 0–7, to 0–9 and to 4–16. Total phlebitis grading also varied considerably for progressive scales, ranging from 0–2 to 0–6.

The symptoms required for phlebitis varied considerably. Only erythema was reported as a phlebitis symptom in every scale.

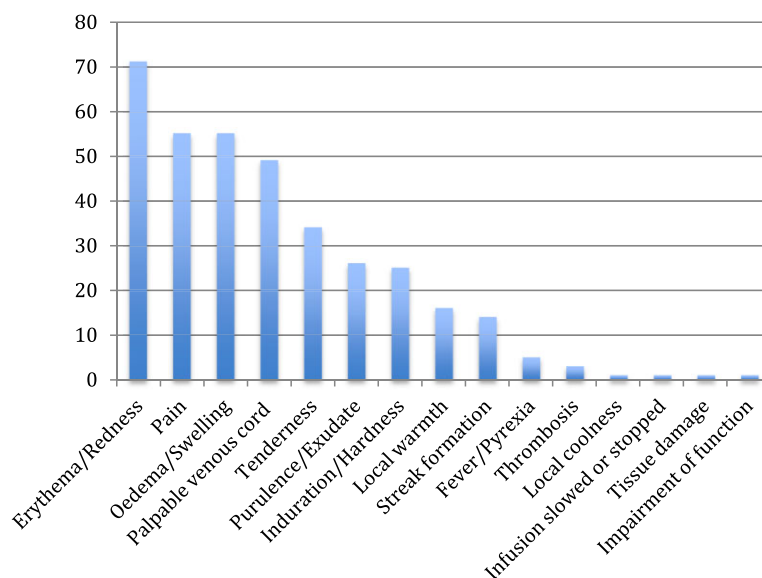
Several authors scored patients as positive for phlebitis with the finding of pain alone [24–28], erythema alone [29,30] or either [31–34]. Some authors considered a palpable venous cord alone to be sufficient for phlebitis [35–37], although the length of palpable cord required varied from 2.5 [7,38,39] to greater than 15 cm [40]. Exact measurement of symptoms, such as distance of erythema and oedema from insertion site, was undertaken in several studies, but the length or diameter required for concern varied considerably, from greater than 2 [41] to greater than 3 cm [36,37]. Some authors measured local warmth objectively, using a differential thermometer [41–43], but in most cases, temperature appeared to have been subjectively evaluated. Finally, some authors using progressive scores considered a patient had phlebitis when symptom severity met the criteria for a score of 1; others reported phlebitis only when severity scored as 2 or 3.

### Phlebitis incidence

Not all authors reported phlebitis in the same way. Some reported phlebitis incidence per patient (potentially including multiple PIVCs); others reported phlebitis incidence per PIVC. Reported phlebitis incidence varied dramatically for studies using a scale – from 0% [44] to 91% [45].

### The phlebitis assessment process

Frequency of reported assessment ranged from every PIVC access for medication or infusion, to twice daily, daily or second daily assessment. A handful of studies reported continued phlebitis assessment after cannula removal up to 24 hours [24], 48 hours [7,46] and 3 days [47]. One study reported follow up of patients until the phlebitis resolved; in one case of phlebitis, pain lasted for 5 months [48]. Assessors ranged from ward nurses, research nurses, experienced IV teams, medical students, doctors, to independent IV assessors. Some researchers reported providing phlebitis assessment training to staff, but the majority did not.



**Figure 2** Frequency of symptoms reported in 71 phlebitis scales.

## Psychometric evaluation of infusion phlebitis assessment scales

Although there are dozens of phlebitis assessment instruments, formal evaluations of their measurement properties are rare. Several scales were used in multiple studies, such as the Baxter scale [21,31,33], the Dinley scale [23,49,50] and the Lipman scale [22,51,52], but appear never to have been formally assessed. Thirteen articles reported evaluating some psychometric properties of their assessment scale (see Table 1), but only three provided detailed information. This section describes the psychometric adequacy of those three scales: VIP scale, INS phlebitis scale and PVC ASSESS.

### VIP scale/Jackson scale

As part of a randomized trial published in 1977, US pharmacists, Maddox and colleagues [19] created a phlebitis assessment instrument to grade phlebitis presence and severity using six symptoms: pain, erythema, swelling, induration, palpable venous cord and frank vein thrombosis. The scale ranged from 0 to 5+; a score of 1 was considered indicative of phlebitis. Their report included no evaluation of the scale's reliability or validity. During the 1980s and early 1990s, several researchers used the Maddox scale or a slightly modified version of it [27,47,53–61], but psychometric assessments were still not reported.

In the UK in 1998, Jackson [14] published guidelines for scoring phlebitis based on an adaptation of the Maddox method and a scale developed by Lundgren and colleagues in 1993 [48], which was relabelled the VIP score. This scale grades phlebitis progressively from 1 (no observable phlebitis symptoms) to 6 (advanced thrombophlebitis), and each grade is associated with a recommended action (e.g. cannula removal). The VIP scale assesses the presence/absence of six symptoms: pain, erythema, swelling, induration, palpable venous cord and pyrexia. Neither Jackson nor other researchers who subsequently used the scale [13,32,62–67] reported information about the scale's measurement properties.

A formal assessment of a modified version of the VIP scale was undertaken in the United States in 2006 by Gallant and Schultz [13]. They monitored 851 PIVCs in 513 cardiac surgical patients in one hospital. Jackson's original grading from 1–6 was recalibrated to 0–5; a score of 5 indicated purulent drainage, redness and a palpable cord greater than 7.6 cm. Other modifications were not described in detail, although pyrexia as a symptom was removed. Phlebitis was considered present if the VIP score was  $\geq 2$ , with associated recommendation for PIVC removal. Despite modifying the scale, the authors continued to use the label of VIP scale. Therefore, several versions of the VIP scale, including Jackson's original scale, are available and in use.

Staff nurses (number unreported) from two wards received training in the use of the Gallant and Schultz VIP scale, and then completed daily PIVC assessments. Inter-rater reliability was assessed by correlating each research nurse's VIP score with that of the principal investigator, a senior clinical nurse. The type of correlation (Pearson's *r*, Spearman's  $\rho$ , intraclass) was unreported. The number of PIVC assessments included in the inter-rater reliability checks was also unreported. Each nurse was said to achieve an acceptable inter-rater reliability correlation of  $\geq 0.85$ .

However, inter-rater reliability was not computed between the nurses themselves, which is a more standard approach. A key unanswered question that remains is whether rating consistency across similarly trained observers can be achieved with this scale.

In terms of validity, the report stated that expert nurses in the cardiac surgery unit 'established content validity for the modifications of the Jackson VIP scale' (p. 341). Data on content validity, using a quantitative assessment of agreement such as the content validity index [68,69], were not provided. The scale's criterion or construct validity was not discussed. However, assessment of inter-rater reliability could be construed as testing criterion validity. If the principal investigator was an 'expert' in phlebitis, then her scoring can be accepted as a 'gold standard' against which the nurses' ratings were tested. This study reported no analysis of specificity or sensitivity, which are standard parameters for criterion-related validity in scales such as the VIP that have a 'cut-point' for the presence/absence of an outcome. The researchers also did not specifically assess construct validity.

Gallant and Schultz concluded that their version of the VIP scale is a reliable and valid measure for assessing and determining the removal of a PIVC. However, the evidence for the scale's adequacy is extremely limited. The reliability assessments did not establish that nurses could be consistent in their evaluations of phlebitis symptoms with each other (inter-rater), nor with themselves (intra-rater). Test–retest reliability was not examined. The study yielded some information about criterion validity, but the VIP scale's specificity and sensitivity were not tested. Construct validity was not considered. Responsiveness – the ability to detect true changes in symptoms – was also not examined. Post-study, the hospital made a decision to adopt the VIP as a standardized assessment tool, which suggests they found it easy to use in clinical practice; however, no data regarding feasibility were provided.

### INS phlebitis scale

The INS in the United States developed the first INS phlebitis scale in 1998 [15–18]. The INS scale has changed over time, with the current version being a progressive score from 0 (no symptoms) to 4 (all symptoms present: pain, erythema, oedema, streak formation, palpable venous cord  $>2.54$  cm in length and purulent drainage) [16]. Any score of 1 or greater is considered phlebitis. Several studies included in the current review used an assessment tool based either on the INS scale or an adaptation [24,25,29,30,34,70–76]. Despite widespread use, the INS scale has had limited scrutiny for psychometric properties. Boyce and Yee [24] adapted the INS scale and consulted a panel of 18 experienced nurses to assess the revised scale's face validity, resulting in several further changes to the tool. Following pilot testing, the tool and instructions were modified to be 'more user-friendly' (p. 30). No other psychometric evaluation appears to have been undertaken by these authors. Dryburgh and Imlah [77] appeared to have adapted an early version; they assessed it for face validity and what they called 'test–test' reliability in 10 patients, without providing data. Washington and Barrett [74] reported assessing inter-rater reliability, but did not provide values. Powell *et al.* [30] reported agreement in rating phlebitis between two members of the IV team, but the ratings appear not to have been independent or blinded.

A more in-depth study by Groll and co-researchers [29] was undertaken in Canada to evaluate the psychometric properties of

**Table 1** Studies that reported measuring some psychometric properties of a phlebitis assessment scale

Study, year, design, country (alphabetical)	Population	Primary outcome	Phlebitis scale, definition	Source of scale	Symptoms measured	Reported phlebitis rates and grade	Assessor(s)	Assessment frequency	Measurement property assessment
Ahlgvist <i>et al.</i> , 2006 [44] Cross-sectional survey (pre/post) Sweden	2001: 107 PIVC in 67 medical and surgical patients; 2002: 99 PIVC in 63 medical and surgical patients	Effect of introducing guidelines for PIVC care on incidence of thrombophlebitis, nurses' care, handling and documentation	<b>Scale 0-3</b> Phlebitis defined as $\geq 1$	Lundgren <i>et al.</i> , 1993 [48]	Redness Tenderness Pain Swelling Increased temperature Palpable cord Pus	% PIVC 2001 survey: 39% grade 1 7% $\geq$ grade 2 2002 survey: 27% grade 1 0% $\geq$ grade 2	7 nurses not employed on study wards	Second daily	Reliability, validity, feasibility assessed in pilot study. No data provided.
Ahlgvist <i>et al.</i> , 2010 [86] Cross-sectional Sweden	67 PIVC <sup>§</sup>	Inter-rater reliability of phlebitis assessment using PIVC assessment tool	<b>PVC ASSESS</b> Points per symptom 1 or more symptom(s)	Hershey <i>et al.</i> , 1984 [7]	Pain Tenderness Erythema Oedema Induration Purulent exudate Streak formation Palpable cord	N/A	Nurses trained in phlebitis assessment	Group A (3 RNs) assessed at bedside Group B (3 RNs) assessed photos of same PIVCs 4 weeks later	Inter-rater and intra-rater reliability assessed; content validity informally assessed; acceptability and feasibility assessed <sup>†</sup>
Bostrom-Ezrati <i>et al.</i> , 1990 [53] Prospective cohort United States	514 medical and surgical patients at 4 hospitals <sup>¶</sup>	Incidence of IV site symptoms, and associated patient and practice factors	<b>Maddox scale 0-5</b>	Maddox, 1983 [20]	Erythema Swelling Induration Pain Palpable venous cord	% patients 22.6% grade 1 17.3% $\geq$ grade 2	Nurse data collectors	Twice daily	Inter-rater reliability assessed
Boyce & Yee, 2012 [24] Prospective cohort United States	24 PIVC in 12 patients	Incidence and severity of phlebitis in patients given peripherally infused amiodarone	<b>Modified INS scale 0-4+</b> Phlebitis defined as $\geq 0+$ (pain)	INS, 2006 [16]	Pain Erythema Oedema Streak formation Palpable venous cord Purulent drainage	% PIVC 50% $\geq$ grade 0+	Staff nurses	Every 4 hours until 24 hours after infusion ceased	Content validity and feasibility assessed. No data provided.
Campbell, 1998 [31] Prospective cohort Northern Ireland	90 medical patients from 13 wards <sup>¶</sup>	Incidence and severity of phlebitis, contributing factors, extended length of stay, IV complications	<b>Baxter scale 0-5</b>	Baxter, 1988 [18]	Pain Erythema Swelling Induration Palpable venous cord	% patients 26% grade 1-3	Staff nurses	None stated	Test-retest reliability reported as being assessed. No data provided.
Catney <i>et al.</i> , 2001 [40] Prospective cohort United States	411 medical and surgical patients <sup>¶</sup>	Relationship of dwell time to phlebitis and infiltration	<b>Authors' scale 1-3</b> Phlebitis defined as $\geq 2$	None stated	Pain or tenderness Erythema Swelling Palpable venous cord	% patients 7.3% $\geq$ grade 2	6 IV team staff	Twice daily	Inter-rater reliability, construct validity and feasibility reported; assessed in pilot study. No data provided.
Dibble <i>et al.</i> , 1991 [54]* Prospective cohort United States	514 patients in 4 hospitals <sup>¶</sup>	Frequency of IV site symptoms	Modified <b>DeLuca</b> and <b>Maddox scale</b> 1-5 used	DeLuca <i>et al.</i> , 1975 [101]; Maddox <i>et al.</i> , 1977 [19]	Pain Redness Swelling Induration Palpable cord	% patients 39.9%	66 research assistants (nurse educators, RNs, nursing students)	Twice daily	Inter-rater reliability assessed

**Table 1** *Continued*

Study, year, design, country (alphabetical)	Population	Primary outcome	Phlebitis scale, definition	Source of scale	Symptoms measured	Reported phlebitis rates and grade	Assessor(s)	Assessment frequency	Measurement property assessment
Dryburgh & Imlah, 2002 [77] RCT Canada	38 outpatients receiving IV antibiotics for cellulitis <sup>§</sup>	Incidence and severity of phlebitis with two antibiotics	Modified Phlebitis Rating Scale 1–4	American IV Nursing Standards, 1982 <sup>†</sup>	Erythema Tenderness Pain Swelling Induration Purulence	% patients 27% cefazolin 59% cloxacillin	Community nurses	Daily	Test-retest, face validity and construct validity assessed. No data provided.
Gallant & Schultz, 2006 [13] Prospective cohort United States	851 PIVC in 513 adult cardiac surgical and cardiothoracic patients	Reliability of a phlebitis scale	VIP scale 0–5 Phlebitis defined as VIP ≥ 2	Jackson, 1998 [14]	Pain Redness Warmth Oedema Purulence Palpable venous cord > 7.6 cm	% PIVC 6.2% VIP ≥ 2	Research IV team nurses trained in VIP scale	Daily	Authors reported testing inter-rater reliability, but actually assessed criterion validity. Content validity informally assessed. <sup>‡</sup>
Groll <i>et al.</i> , 2010 [29] Cross-sectional study and chart audit Canada	416 PIVC observations in 182 patients	Psychometric properties of phlebitis and infiltration scales in an acute and community care setting	INS phlebitis scale 0–4 Phlebitis defined as ≥ 1	INS, 2006 [16]	Pain Erythema Edema Streak formation Palpable venous cord Purulent drainage	% patients 18.3% ≥ grade 1 observed 7.7% episodes of phlebitis documented in chart	Two research nurses	None stated	Inter-rater reliability, acceptability and feasibility assessed. Authors reported concurrent (criterion) validity, but actually tested convergent validity. <sup>‡</sup>
Larson <i>et al.</i> , 1984 [58] Prospective cohort United States	876 PIVC in 707 medical-surgical patients	Relationship between selected risk factors and incidence of phlebitis	Maddox scale 0–5 Phlebitis not defined	Maddox <i>et al.</i> , 1977 [19]	None stated	% PIVC** 25.6%	Quality assurance research nurse	Daily	Inter-rater reliability assessed in pilot trial
Powell <i>et al.</i> , 2008 [30] Retrospective review United States	679 PIVC in inpatients <sup>§</sup>	Relationship between dwell time and phlebitis	INS phlebitis scale 0–4 Phlebitis defined as ≥ 1	INS, 2006 [16]	Erythema Pain Streak formation Palpable vascular cord Oedema Purulent drainage	% PIVC 3.7% ≥ grade 1	3 IV team nurses	Daily	Inter-rater reliability assessed. No data provided.
Washington & Barrett, 2012 [74] Point prevalence United States	188 PIVC in 169 medical-surgical patients	Point prevalence of phlebitis rates	INS phlebitis scale ≥ 2 Phlebitis defined as ≥ 2	INS, year not stated	Pain Erythema Oedema Streak formation Palpable cord	9.5% ≥ grade 2	10 data collectors	One assessment only	Inter-rater reliability and feasibility assessed. No data provided.

\*Bostrom-Ezrati *et al.*, 1990 and Dibble *et al.*, 1991 reported on the same study.  
<sup>†</sup>No reference given by authors and unable to locate reference.  
<sup>‡</sup>Measurement property values are shown in the text of the paper.  
<sup>§</sup>Number of patients not stated.  
<sup>¶</sup>Number of PIVC not stated.  
<sup>\*\*</sup>Phlebitis grade not reported.  
 INS, Infusion Nurses Society; IV, intravenous; PIVC, peripheral intravenous cannula; RCT, randomized controlled trial.

the most recent (2006) version of the INS phlebitis scale [16]. In the study, adults with a PIVC were recruited from a community hospital and a visiting home nursing agency. Pairs of independent research nurses who were not providing direct patient care undertook 392 observations of 176 patients. No information regarding the training of the research nurses was provided, nor did the report state how many pairs of nurses performed ratings. The study aimed to yield evidence regarding the INS scale's reliability (inter-rater), validity, acceptability and feasibility.

For inter-rater reliability, two nurses simultaneously scored the INS scale for each patient. The kappa statistic was used for the reliability index; proportion in agreement was not reported. It was not reported whether the kappa statistic was calculated based on agreement for the full scale's 0–4 range (i.e. a weighted kappa), or on a simpler dichotomous rating of phlebitis presence ( $\geq 1$ ) or absence (0). Furthermore, although different pairs of raters assessed different sets of patients (i.e. the design was not fully crossed), it is unclear whether the appropriate statistic – Fleiss's kappa [78,79] rather than Cohen's kappa [80] – was used. In any event, the reported kappa was 0.45, which is considered 'moderate' using Landis and Koch's [81] standards (kappas of 0.21–0.40 are 'fair,' 0.41–0.60 are 'moderate,' 0.61–0.80 are 'substantial,' and 0.81 and greater are 'almost perfect'). Standards for kappa are controversial [82,83], but few would argue that a kappa of 0.45 offers strong evidence of assessor agreement.

In terms of validity, Groll and colleagues assessed what they called 'concurrent validity'. Concurrent validity, a form of criterion validity, requires a 'gold standard,' which in this case was documentation of phlebitis in the patients' charts. The Spearman correlation between the number of times observers said phlebitis occurred based on the INS scale and the number of times phlebitis was documented in the chart was a modest 0.39. Research nurses using the scales identified more than twice as many cases of phlebitis as were recorded in patient charts. An entry in a patient's chart is a questionable choice for a 'gold standard.' Indeed, the authors noted that the discrepancy between the charts and the scale 'underscores the need for the use of validated tools' (p. 389). Within COSMIN's classification, the procedure would best be described as convergent validation (i.e. evidence that two separate measures of a construct are correlated) rather than criterion validation.

The INS scale was also assessed for acceptability and feasibility. The nurses completed the instruments relatively quickly, with a mean completion time of 1.3 minutes (range 1–15 minutes, SD 0.9 minutes) to complete both the phlebitis scale and the INS infiltration scale (the INS infiltration scale is not covered in this review). Feedback from six research nurses indicated that the phlebitis scale was acceptable for the purpose of identification and measurement of phlebitis, the instructions were clear, and the scale was deemed easy to use and clinically appropriate. Acceptability was further supported by the fact that there were only limited amounts of missing data.

The researchers concluded that the scale was 'valid and reliable in both the acute care and community settings' (p. 390). However, the values of both kappa for inter-rater reliability (0.45) and the correlation coefficient for the validity evaluation (0.39) are modest. The researchers perhaps interpreted statistically significant differences as evidence of the scale's good properties. However, statistical significance is of limited interest in assessments of measurement properties, and indeed they are seldom

reported in inter-rater reliability studies [84,85] because the focus is on how close the reliability coefficient is to 1.00, not whether it is different from 0.00.

Although the researchers provided initial data on the psychometric properties of the widely used INS scale, the research did not comprehensively examine reliability and validity. There were no assessments of reliability over time (test–retest and inter-rater), and the validation efforts did not generate sufficient evidence of validity. Furthermore, the scale's responsiveness was not evaluated. The assessments of feasibility and acceptability were the most comprehensive published to date, but the findings would have been of greater value with a larger sample than six nurses. It would be advantageous to replicate this research in additional centres and with a better 'gold standard' phlebitis criterion, such as evaluation of patients by an infusion expert.

## PVC ASSESS

Ahlqvist and a team of Swedish co-researchers [86] developed a 45-item tool called PVC ASSESS to assess the management, documentation, signs and symptoms associated with PIVC use and complications. Only 11 of the items measure phlebitis symptoms – 5 based on patient reports (pain, tenderness, communicating) and 6 based on nurse observation. All observation items are dichotomous, indicating presence or absence of erythema, oedema, purulent exudate, induration at insertion site, streak formation and palpable cord. In the methodological paper describing the instrument, there were no guidelines for combining scores from the 11 symptom items into an overall score, nor any discussion about a discrete 'cut-off' score for phlebitis.

Reliability was assessed at the item level and only for the six items that required nurses' observations. Inter-rater reliability was estimated with 3 nurses and 66 patients. The researchers calculated proportion of agreement and kappa, using multi-rater kappa. Proportion of agreement ranged from 0.77 (erythema) to 0.95 (exudate and palpable cord). Because of low prevalence of most symptoms, kappa was computed only for one item (erythema), a modest 0.40 [95% confidence interval (CI) = 0.18–0.62]. Inter-rater reliability was assessed for three items that could be evaluated via colour photographs for 67 patients, using a different set of three nurse assessors. Proportion of agreement ranged from 0.76 to 0.89, and the only kappa value – again for erythema – was 0.58 (95% CI = 0.44–0.72).

The researchers also assessed intra-rater reliability (which they incorrectly called test–retest reliability) using photographs. Three nurses examined colour photographs of 67 patients. They rated the presence or absence of three signs (erythema, exudate and streak formation) on two occasions, 4 weeks apart. Commendably, the order of presentation of the photos was altered at the second viewing. Across the three raters, intra-rater kappas ranged from 0.49 (nurse 1, streak formation) to 0.76 (nurse 2, purulent exudate). The median intra-rater kappa was 0.59. This study was the only one in which intra-rater reliability (constancy of assessment by the same rater over time) was evaluated.

In terms of validity, only content validity was considered. The report indicates that the research group 'confirmed content validity . . . through comparisons with guidelines and published scientific literature in the field' (p. 1109). It does not appear that a formal content validity assessment was performed. The team did

undertake an assessment of acceptability and feasibility. A sample of 27 nurses and 93 nursing students informally used the instrument with nearly 600 patients, and then provided feedback about the clarity and content of items, and the usefulness and layout of the tool. A few changes were made after this feedback, but results of the feasibility assessments were not provided.

Although the researchers considered their tool as 'reliable,' kappa values for the inter-rater and intra-reliability of nurse-observed phlebitis items were modest. No information about the reliability for the five patient-reported items was provided, and one of these items ('Communicating') was not defined. Test-retest reliability (short-term stability of scores across different assessments) was not evaluated. In terms of validity, no evidence was offered regarding the criterion validity (e.g. comparison to a 'gold standard') or construct validity, nor was responsiveness of the index evaluated. Feasibility information was limited. Göransson and Johansson [87] also used the PVC ASSESS tool, but did not report any psychometric evaluation.

## Discussion

In this systematic review of research studies using phlebitis as the primary endpoint, we found numerous definitions of phlebitis, 71 different phlebitis assessment scales, a wide variation in assessment techniques and reported phlebitis rates, and very little psychometric evaluation of the existing scales. While it was surprising to find such an array of confounding factors in phlebitis assessment, of even greater concern was the fact that many studies reported phlebitis as a primary endpoint without providing *any* definition of phlebitis at all.

Among the 180 studies that explained how they determined phlebitis, either by scale or definition alone, we found a broad range of definitions. The Centers for Disease Control and Prevention [88] defines phlebitis as warmth, tenderness, erythema or palpable venous cord, citing Maki and Ringer [89], although this is not the definition used by those authors. Other commonly used descriptors include pain, swelling, induration and purulent drainage.

With cumulative scales, no uniformity exists as to how many signs must be present to qualify as phlebitis and/or warrant the removal of the PIVC. Many tools consider the presence of two or more symptoms as phlebitis, with others requiring only one sign, and others several signs. Furthermore, differentiating phlebitis from extravasation may be difficult when tenderness and oedema are the predominant signs [90].

Numerous progressive scales with grading according to symptom severity have been developed over the past 40 years, but persistent limitations include the following: (1) not all 'required' symptoms may be present, yet the PIVC is not working properly [91]; and (2) a patient may not develop the signs in the particular sequence outlined by the scale, and thus does not meet the threshold for phlebitis despite patient/staff concerns that trigger PIVC removal [60,91].

Phlebitis rates ranged widely in this review. This can be attributed in part to the absence of a universally accepted scale with strong demonstrated reliability. The INS [16,17] recommends a phlebitis rate of 5% or less as acceptable, but differences in definition and assessment procedures, study design (prevalence versus incidence), casemix of research trials and rate calculation methods make comparison difficult. The INS also recommends that phle-

bitis should be calculated as the number of phlebitis incidents per total number of PIVC multiplied by 100 [15,17,18], but this review found that reporting methods varied considerably: per patient, per PIVC and per 1000 catheter days.

The regular clinical use of a phlebitis tool is believed to provide a trigger, alerting nurses to take action if problems occur [92]. The review found that the most commonly used tools were the INS, VIP, Jackson, Baxter and Maddox scales; however, all of these have been modified by various authors and several versions of each scale exist, with some researchers continuing to use older versions. Typical modifications include the addition or removal of phlebitis symptoms and variations in the scoring process, including the number of symptoms required for diagnosis and changes to the numerical scale. The INS phlebitis scale is a popular tool, but several variations exist [15–18], and we found that many authors further modified the tool for their own purposes. The UK Royal College of Nursing recommends the VIP scale [93] because specific actions, such as PIVC removal, are given as severity of phlebitis increases. However, the VIP scale exists and continues to be used in multiple modified versions. In the United States, the INS currently recommends using either the INS tool [15,16], as evaluated by Groll and colleagues [29], or the VIP scale, as per Gallant and Schultz [13].

Frequency of phlebitis assessment ranged from every cannula access, to twice daily, daily or even second daily assessment. Accessibility or visibility of the PIVC site was not mentioned in the majority of studies, although presumably some used gauze and tape dressings, which are acceptable [94] but preclude visual inspection of some symptoms.

Assessors ranged from student nurses and ward nurses to experienced IV teams, and medical and nursing researchers. Although some authors reported providing education on phlebitis assessment, the majority did not. Inter-rater reliability of phlebitis assessment has proved to be problematic. A 2002 epidemiological literature review [60] reported that no diagnostic criteria for phlebitis had been proven valid or reproducible. Since then, several authors have reported measuring inter-rater reliability, but none has addressed the full psychometric properties of the scale used, as discussed earlier. The studies reviewed suggest that it is extremely difficult to use existing scales with confidence, given the modest inter-rater reliability values.

With the current state of knowledge about scale quality, we cannot recommend a particular phlebitis assessment scale. None of the existing scales has been subjected to rigorous and thorough psychometric testing. For example, sensitivity and specificity have not been calculated for any scale. With the current evidence, no scale stands out as being of particularly high quality. In particular, inter-rater reliability estimates tend to be quite modest.

This review highlights priorities for future psychometric evaluations of phlebitis scales. The most critical measurement properties to assess are inter-rater and intra-rater reliability, as well as criterion validity (although other properties, such as responsiveness, would be of interest). With respect to inter- and intra-rater reliability, it is statistically unlikely that any scale will show high kappa values due to the generally low prevalence of phlebitis among a group of hospital patients at one moment in time. Future evaluations of reliability should provide the actual proportions of phlebitis assessments with positive agreement and negative agreement [95], to assist in interpretation of kappa estimates. Byrt



*et al.*'s formula, which corrects for unbalanced prevalence, to present kappa values may also be useful [96]. Although Hoehler [97] has argued against Byrt *et al.*'s formula replacing Cohen's kappa formula, it would be very useful to present both kappa estimates, so that users could see the potential degree of agreement [85,98].

In terms of criterion validity, evaluators need to select a suitable criterion, that is, 'gold standard,' such as rating by a phlebitis expert. Most existing scales grade severity, which implies the need for analysis using a receiver operating curve that establishes the appropriate 'cut-off' value for phlebitis diagnosis, and to ascertain that area-under-the-curve values are acceptable (commonly 0.70 or higher is desirable [83]). It is also essential to calculate the scale's sensitivity and specificity (how often will it correctly test negative in those who *do* and *do not* have phlebitis?).

Lastly, it would be extremely useful to compare two or more scales for their psychometric adequacy in the same study. A direct comparison of reliability and criterion validity using the *same* sample of patients and raters would make it much easier for clinicians to select a phlebitis assessment scale with optimal properties.

### Limitations

Our study has several limitations. Firstly, we only retrieved studies published in English that assessed infusion phlebitis in adults, so we cannot extrapolate the findings to paediatrics or non-English-speaking countries. We did not contact study authors to request potentially unpublished psychometric data. We were unable to locate several older articles (pre-1985) that reported phlebitis, so it is possible that we missed some older phlebitis tools. It is also possible that there are newer phlebitis scales in use and as yet unpublished.

The extreme number and variation of measurement options for phlebitis, combined with the paucity of evidence for reliability and validity, is of great concern. Up to 80% of all hospital patients require IV therapy with about 330 million PIVCs sold each year in the United States alone [5,99]. Although phlebitis scales are quick to complete [29], the number of PIVCs used multiplies to significant nursing time and paperwork. In the United States, if 100 million PIVCs are used for an average of 3.5 days, and nurses assess PIVCs once each 8-hour shift, this accounts for about 23 million hours of skilled nursing time being used with questionable value each year in that country alone.

### Conclusion

The selection of appropriate measurement tools is essential to clinical practice [100]. Yet, it is unclear how best to assess phlebitis because no existing scale has undergone rigorous psychometric testing. This likely contributes to the wide variation in reported phlebitis incidence, which precludes meaningful comparison of studies. The current state of the evidence underlying phlebitis scales holds serious implications for PIVC assessment internationally.

### Acknowledgement

We would like to thank Nicole Marsh for reviewing the paper and providing expert clinical knowledge on assessing phlebitis.

## References

1. Webster, J., Clarke, S., Paterson, D., Hutton, A., van Dyk, S., Gale, C. & Hopkins, T. (2008) Routine care of peripheral intravenous catheters versus clinically indicated replacement: randomised controlled trial. *British Medical Journal*, 337, a339.
2. Campbell, L. (1998) I.v.-related phlebitis, complications and length of hospital stay: part 1. *British Journal of Nursing*, 7 (21), 1304–1306, 1308–1312.
3. Collin, J. & Collin, C. (1975) Infusion thrombophlebitis. *Lancet*, 306 (7932), 458.
4. Hawes, M. L. (2007) A proactive approach to combating venous depletion in the hospital setting. *Journal of Infusion Nursing*, 30 (1), 33–44.
5. Hadaway, L. C. (2012) Short peripheral catheters and infections. *Journal of Infusion Nursing*, 35 (4), 230–240.
6. Hecker, J. F. (1989) Failure of intravenous infusions from extravasation and phlebitis. *Anaesthesia and Intensive Care*, 17 (4), 433–439.
7. Hershey, C. O., Tomford, J. W., McLaren, C. E., Porter, D. K. & Cohen, D. I. (1984) The natural history of intravenous catheter-associated phlebitis. *Archives of Internal Medicine*, 144 (7), 1373–1375.
8. Mokkink, L. B., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D. L., Bouter, L. & DeVet, H. C. W. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63, 737–745.
9. Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R., Bouter, L. M. & DeVet, H. C. W. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, 21, 651–657.
10. Edwards, J. R. & Bagozzi, R. P. (2000) On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.
11. Streiner, D. L. (2003) Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80, 217–222.
12. Feinstein, A. R. (1987) *Clinimetrics*. New Haven, CT: Yale University Press.
13. Gallant, P. & Schultz, A. A. (2006) Evaluation of a visual infusion phlebitis scale for determining appropriate discontinuation of peripheral intravenous catheters. *Journal of Infusion Nursing*, 29 (6), 338–345.
14. Jackson, A. (1998) Infection control – a battle in vein: infusion phlebitis. *Nursing Times*, 94 (4), 68, 71.
15. Infusion Nurses Society (2000) Infusion nursing: standards of practice. *Journal of Intravenous Nursing*, 23, S1–S88.
16. Infusion Nurses Society (2006) Infusion nursing: standards of practice. *Journal of Infusion Nursing*, 29 (1S), S58–S59.
17. Infusion Nurses Society (2011) Infusion nursing standards of practice. *Journal of Infusion Nursing*, 34 (1S), S1–S110.
18. Intravenous Nurses Society (1998) Intravenous nursing: standards of practice. *Journal of Intravenous Nursing*, 21, s34–s36.
19. Maddox, R. R., Rush, D. R., Rapp, R. P., Foster, T. S., Mazella, V. & McKean, H. E. (1977) Double-blind study to investigate methods to prevent cephalothin-induced phlebitis. *American Journal of Hospital Pharmacology*, 34 (1), 29–34.
20. Maddox, R. R., John, J. F., Jr, Brown, L. L. & Smith, C. E. (1983) Effect of inline filtration on postinfusion phlebitis. *Clinical Pharmacology*, 2 (1), 58–61.
21. Baxter Healthcare Ltd (1988) *Principles and Practice of IV Therapy*. Compton, Berks: Baxter Healthcare Ltd.

22. Lipman, A. G. (1974) Effect of buffering on the incidence and severity of cephalothin-induced phlebitis. *American Journal of Hospital Pharmacology*, 31 (3), 266–268.
23. Dinley, J. (1976) Venous reactions related to in-dwelling plastic cannulae: a prospective clinical trial. *Current Medical Research and Opinion*, 3 (9), 607–617.
24. Boyce, B. A. & Yee, B. H. (2012) Incidence and severity of phlebitis in patients receiving peripherally infused amiodarone. *Critical Care Nurse*, 32 (4), 27–34.
25. Kelsey, M. C. & Gosling, M. (1984) A comparison of the morbidity associated with occlusive and non-occlusive dressings applied to peripheral intravenous devices. *Journal of Hospital Infection*, 5 (3), 313–321.
26. May, J., Murchan, P., MacFie, J., Sedman, P., Donat, R., Palmer, D. & Mitchell, C. J. (1996) Prospective study of the aetiology of infusion phlebitis and line failure during peripheral parenteral nutrition. *British Journal of Surgery*, 83 (8), 1091–1094.
27. Nordenstrom, J., Jeppsson, B., Loven, L. & Larsson, J. (1991) Peripheral parenteral nutrition: effect of a standardized compounded mixture on infusion phlebitis. *British Journal of Surgery*, 78 (11), 1391–1394.
28. Scalley, R. D., Van, C. S. & Cochran, R. S. (1992) The impact of an i.v. team on the occurrence of intravenous-related phlebitis. A 30-month study. *Journal of Intravenous Nursing*, 15 (2), 100–109.
29. Groll, D. L., Davies, B., MacDonald, J., Nelson, S. & Virani, T. (2010) Evaluation of the psychometric properties of the phlebitis and infiltration scales for the assessment of complications of peripheral vascular access devices. *Journal of Infusion Nursing*, 33 (6), 385–390.
30. Powell, J., Tarnow, K. G. & Perucca, R. (2008) The relationship between peripheral intravenous catheter indwell time and the incidence of phlebitis. *Journal of Infusion Nursing*, 31 (1), 39–45.
31. Campbell, L. (1998) I.v.-related phlebitis, complications and length of hospital stay: part 2. *British Journal of Nursing*, 7 (22), 1364–1366, 1368–1370, 1372–1363.
32. do Rego Furtado, L. C. (2011a) Maintenance of peripheral venous access and its impact on the development of phlebitis: a survey of 186 catheters in a general surgery department in Portugal. *Journal of Infusion Nursing*, 34 (6), 382–390.
33. Panadero, A., Iohom, G., Taj, J., Mackay, N. & Shorten, G. (2002) A dedicated intravenous cannula for postoperative use effect on incidence and severity of phlebitis. *Anaesthesia*, 57 (9), 921–925.
34. Uslusoy, E. & Mete, S. (2008) Predisposing factors to phlebitis in patients with peripheral intravenous catheters: a descriptive study. *Journal of the American Academy of Nurse Practitioners*, 20 (4), 172–180.
35. Khawaja, H. T., Campbell, M. J. & Weaver, P. C. (1988) Effect of transdermal glyceryl trinitrate on the survival of peripheral intravenous infusions: a double-blind prospective clinical study. *British Journal of Surgery*, 75 (12), 1212–1215.
36. Monreal, M., Oller, B., Rodriguez, N., Vega, J., Torres, T., Valero, P., Mach, G., Ruiz, A. E. & Roca, J. (1999a) Infusion phlebitis in post-operative patients: when and why. *Haemostasis*, 29 (5), 247–254.
37. Monreal, M., Quilez, F., Rey-Joly, C., Rodriguez, S., Sopena, N., Neira, C. & Roca, J. (1999b) Infusion phlebitis in patients with acute pneumonia: a prospective study. *Chest*, 115 (6), 1576–1580.
38. Rypins, E. B., Johnson, B. H., Reder, B., Sarfeh, I. J. & Shimoda, K. (1990) Three-phase study of phlebitis in patients receiving peripheral intravenous hyperalimentation. *American Journal of Surgery*, 159 (2), 222–225.
39. Sherertz, R. J., Stephens, J. L., Marosok, R. D., *et al.* (1997) The risk of peripheral vein phlebitis associated with chlorhexidine-coated catheters: a randomized, double-blind trial. *Infection Control and Hospital Epidemiology*, 18 (4), 230–236.
40. Catney, M. R., Hillis, S., Wakefield, B., Simpson, L., Domino, L., Keller, S., Connelly, T., White, M., Price, D. & Wagner, K. (2001) Relationship between peripheral intravenous catheter dwell time and the development of phlebitis and infiltration. *Journal of Infusion Nursing*, 24 (5), 332–341.
41. Nichols, E. G., Barstow, R. E. & Cooper, D. (1983) Relationship between incidence of phlebitis and frequency of changing IV tubing and percutaneous site. *Nursing Research*, 32 (4), 247–252.
42. Madan, M., Alexander, D. J. & McMahon, M. J. (1992) Influence of catheter type on occurrence of thrombophlebitis during peripheral intravenous nutrition. *Lancet*, 339 (8785), 101–103.
43. Madan, M., Alexander, D. J., Mellor, E., Cooke, J. & McMahon, M. J. (1991) A randomised study of the effects of osmolality and heparin with hydrocortisone on thrombophlebitis in peripheral intravenous nutrition. *Clinical Nutrition*, 10 (6), 309–314.
44. Ahlqvist, M., Bogren, A., Hagman, S., Nazar, I., Nilsson, K., Nordin, K., Valfridsson, B. S., Söderlund, M. & Nordström, G. (2006) Handling of peripheral intravenous cannulae: effects of evidence-based clinical guidelines. *Journal of Clinical Nursing*, 15 (11), 1354–1361.
45. Bergeron, M. G., Bruschi, J. L., Barza, M. & Weinstein, L. (1976) Significant reduction in the incidence of phlebitis with buffered versus unbuffered cephalothin. *Antimicrobial Agents and Chemotherapy*, 9 (4), 646–648.
46. Rickard, C. M., Webster, J., Wallis, M. C., *et al.* (2012) Routine versus clinically indicated replacement of peripheral intravenous catheters: a randomised controlled equivalence trial. *Lancet*, 380 (9847), 1066–1074.
47. Kerin, M. J., Pickford, I. R., Jaeger, H., Couse, N. F., Mitchell, C. J. & Macfie, J. (1991) A prospective and randomised study comparing the incidence of infusion phlebitis during continuous and cyclic peripheral parenteral nutrition. *Clinical Nutrition*, 10 (6), 315–319.
48. Lundgren, A., Jorfeldt, L. & Ek, A. C. (1993) The care and handling of peripheral intravenous cannulae on 60 surgery and internal medicine patients: an observation study. *Journal of Advanced Nursing*, 18 (6), 963–971.
49. Fujita, M., Hatori, N., Shimizu, M., Yoshizu, H., Segawa, D., Kimura, T., Iizuka, Y. & Tanaka, S. (2000) Neutralization of prostaglandin E1 intravenous solution reduces infusion phlebitis. *Angiology*, 51 (9), 719–723.
50. Gupta, A., Mehta, Y., Juneja, R. & Trehan, N. (2007) The effect of cannula material on the incidence of peripheral venous thrombophlebitis. *Anaesthesia*, 62 (11), 1139–1142.
51. Bayer-Berger, M., Chiolero, R., Freeman, J. & Hirschi, B. (1989) Incidence of phlebitis in peripheral parenteral nutrition: effect of the different nutrient solutions. *Clinical Nutrition*, 8 (4), 181–186.
52. Falchuk, K. H., Peterson, L. & McNeil, B. J. (1985) Microparticulate-induced phlebitis. Its prevention by in-line filtration. *New England Journal of Medicine*, 312 (2), 78–82.
53. Bostrom-Ezrati, J., Dibble, S. & Rizzuto, C. (1990) Intravenous therapy management: who will develop insertion site symptoms. *Applied Nursing Research*, 3, 146–152.
54. Dibble, S. L., Bostrom-Ezrati, J. & Rizzuto, C. (1991) Clinical predictors of intravenous site symptoms. *Research in Nursing and Health*, 14 (6), 413–420.
55. Everitt, N. J. & McMahon, M. J. (1997) Influence of fine-bore catheter length on infusion thrombophlebitis in peripheral intravenous nutrition: a randomised controlled trial. *Annals of the Royal College of Surgeons of England*, 79 (3), 221–224.
56. Harrigan, C. A. (1984) A cost-effective guide for the prevention of chemical phlebitis caused by the pH of the pharmaceutical agent. *National Intravenous Therapy Association*, 7 (6), 478–479.
57. Jarrard, C., Goodner, W., Piazza, J. A. & Bomar, W. L. (1987) The syringe infusion pump system – its effect on phlebitis rates. *National Intravenous Therapy Association*, 10 (1), 29–33.

58. Larson, E. & Hargiss, C. (1984) A decentralized approach to maintenance of intravenous therapy. *American Journal of Infection Control*, 12 (3), 177–186.
59. Popovsky, M. A. & Ilstrup, D. M. (1986) Randomized clinical trial of transparent polyurethane i.v. dressings. *National Intravenous Therapy Association*, 9 (2), 107–110.
60. Tagalakis, V., Kahn, S. R., Libman, M. & Blostein, M. (2002) The epidemiology of peripheral vein infusion thrombophlebitis: a critical review. *American Journal of Medicine*, 113 (2), 146–151.
61. Williams, D. N., Gibson, J., Vos, J. & Kind, A. C. (1982) Infusion thrombophlebitis and infiltration associated with intravenous cannulae: a controlled study comparing three different cannula types. *National Intravenous Therapy Association*, 5 (6), 379–382.
62. do Rego Furtado, L. C. (2011b) Incidence and predisposing factors of phlebitis in a surgery department. *British Journal of Nursing*, 20 (14), S16–S18, S20, S22 passim.
63. Trinh, T. T., Chan, P. A., Edwards, O., Hollenbeck, B., Huang, B., Burdick, N., Jefferson, J. A. & Mermel, L. A. (2011) Peripheral venous catheter-related *Staphylococcus aureus* bacteremia. *Infection Control and Hospital Epidemiology*, 32 (6), 579–583.
64. Bertolino, G., Pitassi, A., Tinelli, C., Staniscia, A., Guglielmana, B., Scudeller, L. & Luigi Balduini, C. (2012) Intermittent flushing with heparin versus saline for maintenance of peripheral intravenous catheters in a medical department: a pragmatic cluster-randomized controlled study. *Worldviews on Evidence-Based Nursing*, 9 (4), 221–226.
65. Biswas, J. (2007) IV nursing care. Clinical audit documenting insertion date of peripheral intravenous cannulae. *British Journal of Nursing*, 16 (5), 281–283.
66. Nagata, K., Egashira, N., Yamada, T., Watanabe, H., Yamauchi, Y. & Oishi, R. (2012) Change of formulation decreases venous irritation in breast cancer patients receiving epirubicin. *Supportive Care in Cancer*, 20 (5), 951–955.
67. Yamada, T., Egashira, N., Watanabe, H., Nagata, K., Yano, T., Nonaka, T. & Oishi, R. (2012) Decrease in the vinorelbine-induced venous irritation by pharmaceutical intervention. *Supportive Care in Cancer*, 20 (7), 1549–1553.
68. Polit, D. & Beck, C. (2006) The content validity index: are you sure you know what's being reported? *Research in Nursing and Health*, 29 (5), 489–497.
69. Polit, D., Beck, C. & Owens, S. (2007) Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, 30 (4), 459–467.
70. Chee, S. & Tan, W. (2002) Reducing infusion phlebitis in Singapore hospitals using extended life end-line filters. *Journal of Infusion Nursing*, 25 (2), 95–104.
71. Gouping, Z., Wan-Er, T., Xue-Ling, W., Min-Qian, X., Kun, F., Turale, S. & Fisher, J. W. (2003) Notoginseny cream in the treatment of phlebitis. *Journal of Infusion Nursing*, 26 (1), 49–54.
72. Mestre, G., Berbel, C., Tortajada, P., *et al.* (2012) Successful multifaceted intervention aimed to reduce short peripheral venous catheter-related adverse events: a quasixperimental cohort study. *American Journal of Infection Control*, 41, 520–526.
73. Palefski, S. S. & Stoddard, G. J. (2001) The infusion nurse and patient complication rates of peripheral-short catheters. A prospective evaluation. *Journal of Intravenous Nursing*, 24 (2), 113–123.
74. Washington, G. T. & Barrett, R. (2012) Peripheral phlebitis: a point-prevalence study. *Journal of Infusion Nursing*, 35 (4), 252–258.
75. White, S. A. (2001) Peripheral intravenous therapy-related phlebitis rates in an adult population. *Journal of Intravenous Nursing*, 24 (1), 19–24.
76. Liu, F., Chen, D., Liao, Y., Diao, L., Liu, Y., Wu, M., Xue, X., You, C. & Kang, Y. (2012) Effect of Intrafix® SafeSet infusion apparatus on phlebitis in a neurological intensive care unit: a case-control study. *Journal of Internal Medicine Research*, 40 (6), 2321–2326.
77. Dryburgh, L. & Imlah, T. (2002) A comparison study: incidence and severity of clients diagnosed with severe cellulitis who develop phlebitis receiving cloxacillin vs. cefazolin in the community setting. *CINA: Official Journal of the Canadian Intravenous Nurses Association*, 18, 22–31.
78. Fleiss, J. L. (1971) Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382.
79. Fleiss, J. L., Nee, J. & Landis, J. (1979) Large sample variance of kappa in the case of different sets of raters. *Psychological Bulletin*, 86, 974–977.
80. Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
81. Landis, J. R. & Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
82. DeVet, H. C. W., Terwee, C., Mokkink, L. B. & Knol, D. L. (2011) *Measurement in Medicine: A Practical Guide*. Cambridge, UK: Cambridge University Press.
83. Gwet, K. (2012) *Handbook of Inter-Rater Reliability*. Gaithersburg, MD: Advanced Analytics.
84. Davies, M. & Fleiss, J. (1982) Measuring agreement for multinomial data. *Biometrics*, 38, 1047–1051.
85. Hallgren, K. (2012) Computing inter-rater reliability for observational data. *Tutorials in Quantitative Methods for Psychology*, 8, 23–34.
86. Ahlqvist, M., Berglund, B., Nordstrom, G., Kland, B., Wirén, M. & Johansson, E. (2010) A new reliable tool (PVC ASSESS) for assessment of peripheral venous catheters. *Journal of Evaluation in Clinical Practice*, 16, 1108–1115.
87. Göransson, K. E. & Johansson, E. (2012) Prehospital peripheral venous catheters: a prospective study of patient complications. *Journal of Vascular Access*, 13 (1), 16–21.
88. O'Grady, N. P., Alexander, M., Burns, L. A., *et al.* (2011) Guidelines for the prevention of intravascular catheter-related infections. *Clinical Infectious Diseases*, 52 (9), e162–e193.
89. Maki, D. G. & Ringer, M. (1991) Risk factors for infusion-related phlebitis with small peripheral venous catheters. A randomized controlled trial. *Annals of Internal Medicine*, 114 (10), 845–854.
90. Curran, E. T., Coia, J. E., Gilmour, H., McNamee, S. & Hood, J. (2000) Multi-centre research surveillance project to reduce infections/phlebitis associated with peripheral vascular catheters. *Journal of Hospital Infection*, 46 (3), 194–202.
91. Webster, J., Osborne, S., Rickard, C. & Hall, J. (2010) Clinically-indicated replacement versus routine replacement of peripheral venous catheters. *Cochrane Database of Systematic Reviews*, (3), CD007798.
92. Goddard, L., Clayton, S., Peto, T. E. & Bowler, I. C. (2006) The 'just-in-case venflon': effect of surveillance and feedback on prevalence of peripherally inserted intravascular devices. *Journal of Hospital Infection*, 64 (4), 401–402.
93. Dougherty, L., Bravery, K., Gabriel, J., *et al.* (2010) Standards for Infusion Therapy: The RCN IV Therapy Forum. London: Royal College of Nursing.
94. Gillies, D., O'Riordan, L., Carr, D., Frost, J., Gunning, R. & O'Brien, I. (2003) Gauze and tape and transparent polyurethane dressings for central venous catheters. *Cochrane Database of Systematic Reviews*, (4), CD003827.
95. Cicchetti, D. V. & Feinstein, A. R. (1990) High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, 43 (6), 551–558.
96. Byrt, T., Bishop, J. & Carlin, J. B. (1993) Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46 (5), 423–429.

97. Hoehler, F. K. (2000) Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology*, 53 (5), 499–503.
98. Sim, J. & Wright, C. C. (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85 (3), 257–268.
99. Dychter, S. S., Gold, D. A., Carson, D. & Haller, M. (2012) Intravenous therapy: a review of complications and economic considerations of peripheral access. *Journal of Infusion Nursing*, 35 (2), 84–91.
100. Gélinas, C., Loïselle, C. G., LeMay, S., Ranger, M., Bouchard, E. & McCormack, D. (2008) Theoretical, psychometric, and pragmatic issues in pain measurement. *Pain Management Nursing*, 9, 120–130.
101. DeLuca, P. P., Rapp, R. P., Bivins, B., McKean, H. E. & Griffen, W. O. (1975) Filtration and infusion phlebitis: a double-blind prospective clinical study. *American Journal of Hospital Pharmacology*, 32 (10), 1001–1007.