Article:
Coleman, S orcid.org/0000-0002-1437-421X, Smith, IL, McGinnis, E et al. (8 more authors) (2018) Clinical Evaluation of a New Pressure Ulcer Risk Assessment Instrument, the Pressure Ulcer Risk Primary or Secondary Evaluation Tool (PURPOSE T). Journal of Advanced Nursing, 74 (2). pp. 407-424. ISSN 0309-2402

https://doi.org/10.1111/jan.13444

Clinical Evaluation of a New Pressure Ulcer Risk Assessment Instrument, the Pressure Ulcer Risk Primary or Secondary Evaluation Tool (PURPOSE-T)

Abstract

Aim
To test the psychometric properties and clinical usability of a new Pressure Ulcer Risk Assessment Instrument including inter-rater and test–retest reliability, convergent validity and data completeness.

Background
Methodological and practical limitations associated with traditional Pressure Ulcer Risk Assessment Instruments, prompted a programme to work to develop a new instrument, as part of the National Institute for Health Research funded, Pressure UlceR Programme Of reSEarch (RP-PG-0407-10056).

Design
Observational field test.

Method
For this clinical evaluation 230 patients were purposefully sampled across four broad levels of pressure ulcer risk with representation from 4 secondary care and 4 community NHS Trusts in England. Blinded and simultaneous paired (ward/community nurse and expert nurse) PURPOSE-T assessments were undertaken. Follow-up retest was undertaken by the expert nurse. Field notes of PURPOSE-T use were collected. Data were collected October 2012-Jan 2013.

Results
The clinical evaluation demonstrated ‘very good’ (kappa) inter-rater and test–retest agreement for PURPOSE-T assessment decision overall. The percentage agreement for ‘problem/no problem’ was over 75% for the main risk factors. Convergent validity demonstrated moderate to high associations with other measures of similar constructs.

Conclusion
The PURPOSE-T evaluation facilitated the initial validation and clinical usability of the instrument and demonstrated that PURPOSE-T is suitable of use in clinical practice. Further study is needed to evaluate the impact of using the instrument on care processes and outcomes.

Summary Statements

Why is this research or review needed?

- There are methodological and practical limitations associated with the development and use of traditional Pressure Ulcer Risk Assessment Instruments.
- Recent work has been undertaken to develop a new Risk Assessment Instrument, incorporating 1) a systematic review, 2) a consensus study, 3) conceptual framework development, 4) Pre-test study.
- This phase 5) clinical evaluation paper, reports a subsequent fundamental step in the development of the Pressure Ulcer Risk Primary Or Secondary Evaluation Tool (PURPOSE-T) to confirm its suitability for use in clinical practice.

What are the key findings?

- The clinical evaluation of PURPOSE-T demonstrated the reliability, convergent validity and clinical usability of the instrument when used by expert and ward/community nurses in secondary care and community settings.
- The findings emphasise the importance of including skin status in the assessment process in order facilitate both primary and secondary prevention.
- The clinical evaluation aided refinement of the instrument and confirmed its suitability for use in clinical practice.

How should the findings be used to influence policy/practice/research/education?

- The clinical evaluation, was an important phase in the instruments development and the methods used should be considered by others developing health-related instruments.
- PURPOSE-T translates pressure ulcer risk factor evidence and expert opinion into a usable instrument that can facilitate the identification and management of pressure ulcer risk in practice.
- PURPOSE-T should be considered for clinical use in adult hospital and community populations.

Key Words

Pressure ulcer
Risk Assessment
Validity
Reliability
Usability
Tissue Viability
Nursing

Introduction

Pressure Ulcers (PUs) are 'localized injury to the skin and/or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear' (NPUAP/EPUAP/PPPIA, 2014). Skin sites susceptible to PUs are those exposed to pressure (e.g. buttocks and heels) in patients with very limited mobility where offloading is difficult. PUs are categorised numerically according to the tissue layers involved; category 1 indicates non-blanchable erythema and category 4 indicates full thickness tissue Loss (NPUAP/EPUAP/PPPIA, 2014).

PUs remain a considerable patient safety issue worldwide with prevalence in acute care settings being 11.9-15.8% and incidence being 2.8-9.0% (Briggs et al., 2013, Smith et al., 2016, Pieper, 2012). PUs cause undue burden on patients quality of life (Gorecki et al., 2009, Gorecki et al., 2012) and have a significant financial impact to healthcare organisations (Bennett et al., 2004, Severens et al., 2002, Schuurman et al., 2009, Berlowitz et al., 2011, Dealey et al., 2012). National and international guidelines agree that structured risk assessment is the cornerstone to PU prevention (Beeckman et al., 2013, NICE, 2014, NPUAP/EPUAP/PPPIA, 2014).

Background

In clinical practice structured risk assessment is underpinned by routine use of PU Risk Assessment Instruments (PU-RAIs). These assist nurses to identify those at risk with the aim of appropriately targeting preventative interventions. Since the 1960's over 40 PU-RAIs have been developed (Nixon and McGough, 2001), though their methodological development and conceptual and practical foundation are often limited. This was demonstrated by analysis of 14 PU-RAIs included in a recent NICE systematic review (NICE, 2014) which identified the following (Coleman, 2014):

- Lack of conceptual framework – only 2 PU-RAIs (Braden and Bergstrom, 1987, Suriadi et al., 2008) were underpinned by a conceptual framework.
- Inadequate development methods – only 3 PU-RAIs were developed from limited statistical modelling methods (Perneger et al., 2002, Suriadi et al., 2008, Page et al., 2011) with the remainder developed on the basis of clinical opinion and/or out-dated literature reviews or adaptations of original instruments.
- Limited evidence of target population involvement during development (SAC, 2002) –involvement of both clinical nurses and patient/carers is important as while nurses primarily use the instruments, assessment should involve the patient/carer and where possible lead to shared decision making about care (Coleman, 2014).
- Inconsistent risk factor inclusion – for example only 5 PU-RAIs include skin status (Andersen et al., 1982, Kwong et al., 2005, Cubbin and Jackson, 1991, Pritchard, 1986, Waterlow, 1985) yet a systematic review identified this as a key predictor of PU development (Coleman et al., 2013).

These issues undermine the content validity of PU-RAIs which is a fundamental property and raises concern about their ability to adequately identify risk (Nixon and McGough, 2001, Gould et al., 2002, Kottner and Balzer, 2010, Coleman, 2014). There are also practical limitations associated with their use (Coleman, 2014, Coleman et al., 2014a, Nixon et al., 2015):

- PU-RAIs are undertaken on all patients, including full assessment of those who are obviously not at risk, which diverts time away from other important care activities.
- Failure to distinguish between those with and without an existing PU which is important as those with a PU require intensified secondary prevention/treatment.

- Use of condensed numerical scores as a basis for care interventions which do not facilitate consideration of individual risk profiles in care-planning.

To address these conceptual, methodological and practical limitations we developed the Pressure Ulcer Risk Primary Or Secondary Evaluation Tool, PURPOSE-T as part of a NIHR funded PU Programme Of Research (PURPOSE: RP-PG-0407-10056). PURPOSE-T development drew on principles of the MRC complex intervention framework (MRC, 2000, MRC, 2008) and incorporated adapted 'gold standard' instrument development methods (FDA DHHS, 2009, SAC, 2002, Steyerberg, 2010, Mokkink et al., 2012), in a structured five phase approach:

i) Systematic review (Coleman et al., 2013)
ii) Consensus study (Coleman et al., 2014a)
iii) Conceptual framework development (Coleman et al., 2014c)
iv) Design and pre-testing (Coleman et al., 2016)
v) Clinical evaluation (Nixon et al., 2015)

The first four phases of this work were concerned with providing evidence of content validity which was indicated along with usability and acceptability in the phase iv) design and pre-test (Coleman et al., 2016). This led to the development of the preliminary PURPOSE-T for clinical evaluation.

Aims

The aim of the study was to test the psychometric properties and clinical usability of PURPOSE-T including inter-rater and test–retest reliability, convergent validity and data completeness.

Design

PURPOSE-T clinical evaluation comprised a field test of hospital and community patients using observational descriptive methods (Nixon et al., 2015). The psychometric properties assessed included reliability defined as 'the extent to which scores for patients who have not changed are the same for repeated measurement under several conditions' (Mokkink et al., 2012). In this study we considered this over time (test-retest) when assessed by the same nurse and on the same occasion by different nurses (inter-rater). Convergent validity was assessed for expected correlations between the items and overall assessment decision of PURPOSE-T and other PU-RAIs to demonstrate construct validity, that is 'evidence that relationships among items, domains, and concepts conform to a priori hypotheses concerning logical relationships that should exist with measures of related concepts or scores produced in similar or diverse patient groups' (FDA DHHS, 2009). In addition the extent to which scale items were completed and used to allocate risk and the experience of using PURPOSE-T in clinical practice were captured.

Participants

Nurses

Expert nurses were trained how to use PURPOSE-T via presentation, the use of vignette case studies, user manual provision and researcher (SC) support. The expert nurses used the same material to cascade PURPOSE-T training to participating ward/community nurses (Nixon et al., 2015).

Patients

Patients were purposively sampled ensuring a similar number of hospital and community patients and representation of patients across four broad risk levels (Nixon et al., 2015). These comprised those without mobility restriction (i.e. low risk),

those with some mobility/activity limitations (i.e. at risk), those who were bedfast/chairfast (i.e. high risk) and those with an existing PU category 1 or above.

Eligible patients included those who were: ≥18yrs, an inpatient in an acute setting or nursing patient in a community setting, able to provide written informed consent/verbal witnessed consent/consultee agreement and expected to be available for PURPOSE-T retest. Patients were excluded if they were from obstetric, paediatric, day case surgery or psychiatric settings, deemed by the attending health-care professional to be too unwell to be approached or complete the study assessment schedule.

Sample size

The sample size was calculated using PASS software is based on a measure of reliability, specifically the kappa coefficient, κ, defined as the proportion of agreement after chance agreement is removed from consideration (Cohen, 1960). Assuming a null hypothesis of $\kappa = 0.6, 5\%$ significance level, and 15% withdrawal/non-compliance in paired assessments and 25% of patients are assessed as 'not at risk' and 75% of patients are assessed as 'at risk', 230 patients were required to be recruited to the study to detect a statistically significant value of $\kappa \geq 0.8$ (alternative hypothesis) with at least 90% power. For the evaluation of screening instruments, no examples of formal sample size estimation methods were identified in the literature (Nixon et al., 2015). Therefore, literature relating to the psychometric evaluation of rating scales was considered. The 'rule of thumb' recommendation of 5–10 patients for every item in a questionnaire was used to estimate the sample size of 115–230 patients (Blazeby et al., 2002, Nunnally and Bernstein, 1994). Therefore, the proposed sample size of 230 to assess inter-rater reliability of the instrument, with >95% expert nurse data compliance (based on previous research experience), was expected to provide a sufficient number of patients to assess the validity of the risk assessment instrument.

Data Collection

Hospital inpatients and community nursing patients were invited to participate. Ward/community-based nurses identified suitable patients from their area of practice. Attending clinical staff or a member of the tissue viability team provided the patient with a verbal explanation of the study and an information leaflet before they were invited to provide informed, written consent. Assessment of eligibility and informed consent/consultee agreement was undertaken by a member of the tissue viability team. Participants were registered centrally using the CTRU automated 24-hour telephone registration system (Nixon et al., 2015).

Participant baseline assessment was undertaken by an expert nurse (tissue viability nurse consultant/specialist/clinical research nurse) and incorporated the collection of demographic data including type of NHS facility, type of admission/referral (e.g. elective/acute), ward specialty (hospital patients), date of birth, gender, ethnicity and clinical assessment comprising subscales of established PU-RAIs, the Braden scale (Bergstrom et al., 1987) and Waterlow score (Waterlow, 1988). Braden and Waterlow were selected as they have undergone the most scrutiny in the literature, reflecting their widespread use in practice (Gould et al. 2002; Pancorbo-Hidalgo et al. 2006; NICE 2014). Waterlow incorporates mobility, moisture, perfusion, nutrition, gender, age, sensory perception, orthopaedic surgery/ below waist fracture, skin condition and medication while the Braden incorporates mobility, activity, friction and shear, moisture, sensory perception and nutrition. Both are ordinal scoring systems in which scores for each risk factor are added together to give the patients overall score. This overall score is then compared to a standard reference value to allocate a risk category (e.g. high risk, moderate risk, at risk).

In addition, a paired PURPOSE-T assessment was undertaken by the expert nurse and a ward/community nurse, incorporating detailed skin assessment and when applicable, PU classification (1-4 and unstageable categories) (NPUAP/EPUAP/PPPIA, 2014). These were conducted simultaneously by both assessors, but recorded separately with blinding maintained (Nixon et al., 2015).

The expert nurse also undertook a second visit and completed PURPOSE-T (blind to their initial assessment) and recorded significant clinical changes to the patient's condition since the baseline assessment (Nixon et al., 2015).The length of the test–retest interval was planned to be short enough to ensure that clinical change in PU risk was unlikely to occur but sufficiently long to be confident that the expert nurse did not recall responses from the first assessment. Nurses were asked to plan their retest visit 1 to 3 days and 1 to 7 days after baseline for hospital and community patients respectively, taking into account anticipated recovery/deterioration/stability of each patient's condition and, for hospital patients, length of stay. Expert nurses also kept field notes of their experience of using PURPOSE-T and comments from ward/community nurses (Nixon et al., 2015).

The Instrument

The preliminary PURPOSE-T incorporates instructions to support nurse decision making facilitated by the use of colour to weight risk factor items. This was based on the overall strength of epidemiological evidence and/or wider scientific evidence, its clinical resonance and its role in the PU causal pathway (Coleman et al., 2013, Coleman et al., 2014a, Coleman et al., 2014b). In PURPOSE-T blue indicates 'no problem'; yellow indicates a potential impact on PU risk; orange indicates risk and; pink indicates the patient has a PU or scar from a previous PU. This colour code is integrated throughout the 3 step assessment process:

- Step 1 – screening assessment comprises 4 mobility items (1 blue, 3 yellow) and 4 skin status items (1 blue, 2 yellow, 1 pink). This allows those who are clearly not at risk (with only blue items) to be quickly screened out and those with potential risk or actual PUs to proceed to the full assessment.
- Step 2 – full assessment comprises items for analysis of independent movement (5 items which include parameters relating to the extent and frequency of movement; 1 yellow, 4 orange); sensory perception (2 items; 1 blue, 1 orange), detailed skin assessment (13 skin sites each with 3 items relating to; normal skin (blue), vulnerable skin (orange) or PU (pink), and there is an option to add further skin sites assessed as vulnerable or with a PU in addition to the pre-specified list); previous PU history (3 items; 1 blue, 1 yellow and a potential additional pink, yellow or blue); perfusion (3 items; 1 blue 2 orange), nutrition (5 items; 1 blue, 4 orange), moisture (3 items; 1 blue, 2 yellow) and diabetes (2 items;1 blue, 1 yellow).
- Step 3 – requires consideration of step 2 responses to inform 1 of 3 assessment decisions comprising 'no PU not currently at risk' for those with only yellow or blue items ticked; 'no PU but at risk' for those with any orange (but no pink items) ticked or if yellow/blue boxes are ticked and the nurse assesses the patient to be at risk based on their overall risk profile; and 'PU category 1 or above or scarring from previous PU' for those with any pink items ticked.

The final version of PURPOSE-T is available at http://medhealth.leeds.ac.uk/accesspurposet.

Ethical Considerations

Patients at risk of PUs are often elderly, frail and considered vulnerable. NHS ethical approval for the study was sought through the Health Research Authority and the Integrated Research Application System and Research Ethics Committee.

Data Analysis

Study risk definitions

For the purposes of describing the study population and to assess the convergent validity of PURPOSE-T with other PU-RAIs, 'at risk' was defined as decision pathways 'No PU but at risk' and 'PU category 1 or above or scarring from previous PU' whilst 'not at risk' was defined as decision pathway 'no PU not currently at risk'. The cut-points used to identify patients at risk for the two other PU-RAIs were ≤ 18 for Braden (Bergstrom et al., 1998) and ≥ 10 for Waterlow (Waterlow, 1988).

Analysis methods

For each assessment data completeness was summarised for each element of PURPOSE-T including the percentage of missing item-level data and risk categories allocated. We produced the simple kappa statistic with 95% confidence interval (CI) and a weighted kappa statistic which incorporates the distribution of disagreements. We calculated weighted Kappa using (Cicchetti and Allison, 1971) weights for comparisons of outcomes with more than two levels. Prevalence-adjusted bias-adjusted kappa (PABAK) statistics were also produced to take into account the prevalence of risk status and bias between observers (Byrt et al., 1993). The final kappa statistic calculated to assess the inter-rater and test-retest reliability for overall risk status was the maximum value of kappa ($\kappa_{max.}$) statistic. Cross-tabulations of overall risk status by rater/retest were also produced. We examined the extent of agreement for individual PURPOSE-T items using cross-tabulations by rater/retest. In addition, we produced kappa (with 95% CI) and weighted kappa statistics to assess inter-rater reliability of the PURPOSE-T decision pathway and produced corresponding cross-tabulations by rater/retest. We used published benchmarks to interpret estimates of the kappa coefficient; Poor $\kappa < 0.20$, Fair $0.21 \leq \kappa \leq 0.40$, Moderate $0.41 \leq \kappa \leq 0.60$, good $0.61 \leq \kappa \leq 0.80$, very good $0.81 \leq \kappa \leq 1.00$ (Landis and Koch, 1977, Bland, 2008).

Cross tabulations of overall risk status for PURPOSE-T,Braden and Waterlow were produced to explore convergent validity. Cross-tabulations of corresponding items between PURPOSE-T, Braden and/or Waterlow were produced and correlation coefficients were calculated to assess convergent validity. The Spearman rank correlation coefficient was used when each of the items being compared had more than two levels and the phi correlation coefficient was calculated when dichotomous variables were compared. For exploratory purposes, the following hypotheses were used as guides to the magnitude of correlations: high correlation $r > 0.7$; moderate correlation $0.3 \leq r \leq 0.7$; low correlation $r < 0.3$ (Burnand et al., 1990, Cohen, 1960). Moderate to high correlations ($r \geq 0.3$) were predicted for comparison of PURPOSE T with relevant Braden and Waterlow items. SAS 9.2 was used for the analysis.

Validity and reliability

To maintain blinding between assessors (expert and ward/community nurses) and assessments (baseline and follow-up), special adhesive data collection forms were used that were sealed on completion, only being opened by CTRU for analysis (Nixon et al., 2015).

To ensure the study population was representative of the clinical population assessed in the course of usual care, approval was sought and gained for witnessed

consent (for patients who were capable of giving consent but physically unable to complete the consent form) and consultee agreement (for patients who lacked capacity).

Results

In total, 230 of 394 patients screened were registered to the study between 3 October 2012-25 January 2013 (Figure 1), from four acute hospital (108(47.0%) patients) and four community NHS trusts (122(53.0%) patients) in England, with numbers of patients registered at each centre ranging from 14-54. The 230 patients recruited were assessed in part or full using PURPOSE-T uat baseline, providing a total of 230 paired assessments undertaken by 11 expert nrses and 73 ward/community nurses. Table 1 shows the characteristics of participants, indicating a mainly Caucasian population with good representation from each of the four broad levels of risk. Of the 230 patients registered, 217(94.3%) had retest assessments completed by the expert nurse (Figure 1).

Based on the baseline PURPOSE-T assessment conducted by expert nurses, there were 60(26.1%) patients who presented with a category 1 or above PU (Table 1). There were a total of 96 PUs across the 60 patients including 21(21.9%) category 1, 56(58.3%) category 2, 6(6.3%) category 3, 6(6.3%) category 4 and 7(7.3%) unstageable ulcers. PURPOSE-T identified 183(79.6%) patients as ‘at risk’, Waterlow identified 193(83.9%) patients as ‘at risk’ and Braden identified 85(37.0%) patients as ‘at risk’. Of the 145 patients identified as ‘not at risk’ on Braden 25(17.2%) had an existing PU (Table 1).

Data completeness and usability

Compliance with the completion guidelines for steps 1-3 were quantified together with a review of the field notes from the expert nurses (Table 2). At least 94.9% data completeness for each construct was observed with the exception of previous PU details (54.7% and 66.7%) and the decision pathway allocation by the ward/community nurses for patients who should not have progressed to step 2 (85.7%) (Table 2).

Appropriate completion of Step 1 (i.e. in line with the recommended assessment flow) by the expert nurses was similar for baseline (83.5%:192/230) and follow up (82.9%:180/217) respectively compared to 72.6%(167/230) of ward/community nurses assessments (Table 2). Progression/non-progression to step 2 was completed in line with the recommended assessment flow for 95.7% (220/230) of patients assessed by an expert nurse at baseline and 98.6%(214/217) assessed at follow up. For ward/community nurses this was 94.3%(217/230)(Table 2). A step1/step 3 decision pathway was allocated to all patients by the expert nurses at baseline and follow up and just one patient (0.4%) was not allocated a decision pathway by the ward/community nurses (Table 2).

Patients were allocated (Table 2,Table 3) to the correct decision pathway (i.e. in line with PURPOSE-T decision rules) by expert nurses at baseline (98.3%:226/230) and follow-up (99.1%: 215/217) and by ward/community nurses (95.2%:219/230). Expert and ward/community nurses allocated the majority of patients to the ‘not at risk’ decision pathway when they completed only yellow and blue boxes with 95.6% (43/45),95.9%(47/49) and 97.7%(42/43) for expert nurses at baseline, ward/community nurses and expert nurses at follow up respectively (Table 3).

Inter-rater reliability
At baseline there were 230 paired assessments by the expert nurse and the ward/community nurse for evaluating inter-rater reliability. The patient population included patients for whom the assessment was completed by both raters regardless of compliance with the recommended assessment flow (Table 2).

There was agreement in the 3 way decision pathway between expert and ward/community nurses for 81.7% (187/229) of paired assessments (Table 4). Under the assumption that the expert nurse and ward/community nurse would complete PURPOSE-T in a similar manner, the corresponding simple kappa statistic of 0.71(95% CI 0.63 to 79) and weighted kappa statistic of 0.76(95% CI 0.69 to 0.83) indicate good inter-rater reliability. When classified dichotomously as 'at risk'/'not at risk' there was agreement between the expert nurse and ward/community nurse for 93.4% (214/229) paired assessments (Table 4). The corresponding simple kappa statistic of 0.81 (95% CI 0.71 to 0.90), PABAK of 0.87 and κ_{max}. of 0.94 indicate very good inter-rater reliability.

Table 5 shows the level of agreement between expert nurses and ward/community nurses for specific risk factor items. In terms of complete agreement the lowest level was 59.2% (113/191) for the analysis of independent movement and the highest was 94.2%(180/191) for diabetic status. We also looked at the levels agreement in terms of the presence of a problem or not and the levels of agreement; ranged from 72.8% (139/191) for perfusion status to 94.2% for diabetic status.

Test–retest reliability
There were up to 217 paired assessments by the expert nurse at baseline and at follow-up for evaluating test-retest reliability however 4 were excluded due to a change in the patients clinical condition providing an analysis population of 213. The median number of days between the baseline and the retest expert nurse assessment was three (range 1–7). As with the inter-rater reliability, the patient population included patients for whom the assessment was completed at both time points regardless of compliance with the recommended assessment flow.

There was agreement in the 3 way decision pathway between the baseline and follow-up assessments for 92.0% (196/213) of paired assessments (Table 4). The corresponding simple kappa statistic of 0.87 (95% CI 0.81 to 93) and weighted kappa statistic of 0.89 (95% CI 0.84 to 0.94) indicate good test-retest reliability. When classified dichotomously as 'at risk'/'not at risk' there was agreement between the baseline and follow-up assessment for 204 (95.8%) paired assessments (Table 4). The corresponding simple kappa statistic of 0.87(95% CI 0.78 to 0.95), PABAK of 0.92 and κ_{max}. of 0.99 indicate very good test-retest reliability.

Table 5 shows the level of agreement between expert nurse's baseline and follow up assessments for specific risk factor items. In terms of complete agreement the lowest level was 64.4% (114/177) for the analysis of independent movement and the highest was 96.0% (170/177) for low BMI. We also looked at the levels of agreement in terms of the presence of a problem or not and levels of agreement ranged from 87.0% (154/177) for perfusion status to 96.0%(170/177) for low BMI.

Convergent validity
The overall risk status on PURPOSE-T was compared with Waterlow for all 230 patients PU free at baseline. A moderate association was observed between PURPOSE-T and Waterlow with a phi correlation coefficient of 0.63 (Table 6). A moderate association was also observed between PURPOSE-T and Braden for 169 PU-free patients, as assessed by the expert nurse at baseline, with a phi correlation

coefficient of 0.40. Individual constructs on PURPOSE-T were compared with relevant constructs on Braden and Waterlow with moderate to high correlations observed in each case (Table 7, Table 8).

Summary of expert nurse field notes

The field notes described positive and problem aspects of using PURPOSE-T in practice (Table 9). More general issues associated with all PU-RAIs were reported (Nixon et al., 2015):

- lack of knowledge of PU classification
- difficulty assessing:
    - mobility when the patient is unable to communicate and after only a short assessment period
    - sensory perception
    - medical history in community setting
    - poor nutritional intake
    - BMI in community setting

Discussion

The PURPOSE-T clinical evaluation involved 230 patients, assessed by both expert and ward/community nurses. Overall the level of data completion for each construct on PURPOSE-T was high at > 90%, with the exception of previous PU history. The inter-rater and test–retest reliability as determined by the kappa statistic was 'very good' for the assessment decision overall. The observed percentage agreement for the assessment of 'problem/no problem' for the eight risk factors (mobility, skin, previous PU, sensory perception, perfusion, nutrition, moisture and diabetes) was high for both inter-rater reliability and test–retest reliability. The lowest levels of absolute agreement was for the analysis of independent movement, a matrix item. This is likely related to the increased number of assessment options available (i.e. doesn't move/doesn't move, moves occasionally/slight position changes, move occasionally/major position changes, move frequently/slight position changes and moves frequently/major position changes) when compared to the other items. Indeed when agreement was considered on presence/absence of a problem, agreement was in line with other assessment items. As predicted moderate to high associations were demonstrated for convergent validity, assessed by comparison with the same or similar constructs on other PU-RAIs (Braden and Waterlow) (Nixon et al., 2015).

An important feature of PURPOSE-T is the inclusion of skin status which was identified, as a key predictor of PU development (Coleman et al., 2013). This allows us to identify those who are at risk of PU development and require primary prevention and those who have an existing PU and require secondary prevention. Traditional RAIs were designed to identify 'at risk' patients, that is before they develop a PU, yet in practice they are used on all patients (those with and without PUs) and do not distinguish between these two groups (Coleman, 2014). Without skin assessment nurses may not identify the presence of an existing PU, and fail to initiate escalated interventions, leading to the progression of a severe PU (Pinkney et al., 2014). The results indicate that for Waterlow and PURPOSE-T which incorporate skin status, all patients with an existing PU were identified as 'at risk' (for PURPOSE-T the red 'Secondary prevention/treatment pathway' was allocated), while 17.2% of those assessed as 'not at risk' by Braden, which does not include skin status actually had an existing PU. It maybe that these patients were recovering and not considered 'at risk' of new PU development but they should still be considered a priority in clinical practice. It was beyond the scope of this study to assess whether their 'not at risk' status impacted the intensity of the interventions received.

The field notes recorded by the expert nurses highlighted positive and problem aspects of using PURPOSE-T in the clinical environment (Nixon et al., 2015). Negative aspects included difficulties in assessing some of PURPOSE-T items and concerns about reliability, but these were not evidenced in the formal evaluation of inter-rater and test–retest reliability. It is of note that where only yellow and blue data items were present both expert and ward/community nurses allocated the majority of patients (> 95%) to the 'not at risk' category, with clinical decision making reflecting systematic review evidence (Coleman et al., 2013) that there is a weaker relationship between these factors and PU development (Nixon et al., 2015).

The issues raised in the field notes and emerging evidence from the PURPOSE pain cohort study (Nixon et al., 2015) indicating pain as a predictor of PU development were considered in a post-clinical evaluation review of PURPOSE-T by the expert group and PU Research Service User Network (PURSUN UK http://www.pursun.org.uk/) involved in the earlier consensus study (Coleman et al., 2014a). The review resulted in final amendments to PURPOSE-T and the inclusion of pain within the stage 2 skin assessment section. PURPOSE-T and supporting literature is freely available for academic research and clinical use and can be downloaded following web-based registration http://medhealth.leeds.ac.uk/accesspurposet). It has since been implemented in community and acute NHS Trusts and early indications are positive, though further evaluation of its impact in practice is needed.

Clinical evaluation allowed construct validity (convergent, and discriminant), inter-rater and test retest reliability and clinical utility to be evaluated. These are important building blocks in the development of RAIs, yet have often been overlooked in the development of previous RAIs (Coleman, 2014) with only a few reporting the evaluation of reliability during instrument validation (Bergstrom et al., 1987, Lindgren et al., 2002, Suriadi et al., 2006) and a focus on establishing predictive validity (Coleman, 2014). While this is an important property, its evaluation is hampered by a number of important factors, including the subjective nature of PU risk factor measurement; the lack of a reference gold standard test (Kottner and Balzer, 2010)); the instigation of preventative interventions in routine practice impacting instrument performance (Deeks, 1996, Defloor and Grypdonck, 2004, Gould et al., 2002) and; studies often only assess the predictive validity of one instrument, rather than multiple RAIs in the same study population to allow comparison (Ferrante di Ruffano et al., 2012).

A major limitation of this traditional approach to PU-RAI validation and evaluation, has been a failure to consider them as complex interventions where their delivery contains several interacting components (MRC, 2008) including the assessment itself, the potential outcomes and decisions about care interventions set within the context of complex health care environments. Recent complex intervention guidance advocates the use of process evaluations allowing causal mechanisms, contextual factors and the quality of implementation to be considered alongside clinical outcomes (Moore et al., 2015, Richards and Hallberg, 2015, MRC, 2008). Therefore in addition to predictive validity testing (alongside other RAIs), the ongoing evaluation of PURPOSE-T will involve a realist process evaluation to identify and test theories and underlying assumptions regarding its use in practice to learn how it can be best used in different contexts, to enhance the probability of effectiveness. This will allow us to establish whether there is sufficient confidence that PURPOSE-T can 'reasonably be expected to have a worthwhile effect' (MRC, 2008) over a standard RAI and inform a potential future RCT.

Limitations

The sample included in this study were overwhelmingly Caucasian. The results, demonstrating the performance of PURPOSE-T should not be applied to other groups without consideration of the importance of skin assessment as a core (and novel) element of this instrument and the potential differences in skin assessment for people with non-white skin.

The level of 'training' in PURPOSE-T use, with expert nurses trained by the researcher which was then cascaded to local nurses was designed to replicate the roll-out of tools to NHS nurses. However, the focussed training of the experts and their direct access to PURPOSE-T development team may mean that the achieved level of competence might not necessarily be replicated in routine clinical practice.

Conclusion

PURPOSE-T evaluation facilitated the initial validation and clinical usability of the instrument. The results indicate very good inter-rater and re-test reliability, high levels of data completion and moderate to high associations for convergent validity, when the same or similar constructs of PURPOSE-T were compared with other PU-RAIs.

The expert nurse field notes allowed the positive and negative aspects of using the instrument to be captured. These along with the other results were reviewed and some finals amendments were made to PURPOSE-T. This culminated in the development of a new evidenced-based RAI for use in adult populations and is now being used in clinical practice. Further evaluation of its impact on care processes and patient outcomes is planned.

Table 1 Baseline Characteristics

Variable	PU at baseline[a] (N=60)	No PU at baseline[a] (N=169)	Missing PU status[a] (N=1[b])	Total[c]
Age (years)				
Mean (sd)	73.8 (15.9)	72.1 (18.3)		72.6 (17.6)
Median (range)	76 (29, 98)	78 (19,102)	78 (N/A)	77 (19, 102)
Sex, N(%)				
Male	27 (27.3%)	72 (72.7%)	0 (0.0%)	99 (43.0%)
Female	33 (25.2%)	97 (74.0%)	1 (0.8%)	131 (57.0%)
Ethnicity N(%)				
Causasian	58 (25.9%)	165 (73.7%)	1 (0.4%)	224 (97.4%)
Other	2 (33.3%)	4 (66.7%)	0 (0.0%)	6 (2.6%)
Setting, N(%)				
Community	37 (30.3%)	84 (68.9%)	1 (0.1%)	122 (53.0%)
Secondary care hospital	23 (21.3%)	85 (78.7%)	0 (0.0%)	108 (47.0%)
Mobility status, PURPOSE-T Step 1, N(%)				
Walks independently with or without walking aids	10 (12.7%)	69 (87.3%)	0 (0.0%)	79 (34.3%)
Needs help of another person to walk	6 (22.2%)	21 (77.8%)	0 (0.0%)	27 (11.7%)
Spends all/majority of time in bed/chair	16 (28.1%)	40 (70.2%)	1 (1.8%)	57 (24.8%)
Remains in same position for long periods	28 (42.4%)	38 (57.6%)	0 (0.0%)	66 (28.7%)
Not completed	0 (0.0%)	1 (100%)	0 (0.0%)	1 (0.4%)
Braden Score, N(%)				
At risk (≤18)	35 (41.2%)	50 (58.8%)	0 (0.0%)	85 (37.0%)
Not at risk (>18)	25 (17.2%)	119 (82.1%)	1 (0.7%)	145 (63.0%)
Waterlow total score				
At risk (≥10)	60 (31.1%)	132 (68.4%)	1 (0.5%)	193 (83.9%)
Not at risk (<10)	0 (0.0%)	37 (100%)	0 (0.0%)	37 (16.1%)
PURPOSE-T Risk Categorisation				
Secondary prevention/treatment pathway	60 (83.3%)	12 (16.7%)	0 (0.0%)	72 (31.3%)
Primary prevention pathway	0 (0.0%)	111 (100%)	0 (0.0%)	111 (48.3%)
Not currently at risk pathway	0 (0.0%)	46 (97.9%)	1 (2.1%)	47 (20.4%)

[a]Percentages in the PU status columns correspond to the proportion of patients within that characteristic who do (or do not) have a PU at baseline (e.g. 27.3% (27 out of 99) of the male population were observed to have a PU at baseline).

[b]There was one community patient for whom their PU status at baseline could not be determined as there were no skin assessments recorded by the tissue viability team member.

[c]Percentages in the total column correspond to the proportion of patients from the overall population with that characteristic (e.g. 43.0% of overall population were male; 57.0% female).

Table 2 Data completeness and completion

Construct	Data completeness					Completion of PURPOSE-T according to guidance			
	Number of items requiring completion	Expert nurse baseline assessment	Ward/Community nurse assessment	Expert nurse follow-up assessment	Denominator (i.e. number of items expected to have been completed)	Completion of PURPOSE-T according to guidance	Expert nurse baseline assessment	Ward/Community nurse assessment	Expert nurse follow-up assessment
Step 1 Screening									
Mobility	1 of 4	99.6% (229/230)	99.6% (229/230)	100.0% (217/217)	All patients, as all were required complete step 1 mobility	Completed	229 (99.6%)	229 (99.6%)	217 (100%)
						Not completed	1 (0.4%)	1 (0.4%)	0 (0.0%)
						Total	230 (100.0%)	230 (100.0%)	217 (100.0%)
Skin status	1 of 4 (if required)	98.7% (78/79)	96.6% (84/87)	100.0% (63/63)	All patients for whom only the blue box was ticked for step 1 mobility	Appropriate completion (no mobility limitation)	78 (33.9%)	84 (36.5%)	63 (29.0%)
						Inappropriate non-completion (no mobility limitation)	1 (0.4%)	3 (1.3%)	0 (0.0%)
						Appropriate non-completion (mobility limitation)	114 (49.6%)	83 (36.5%)	117 (53.9%)
						Inappropriate completion (mobility limitation)	36 (15.7%)	59 (25.7%)	37 (17.1%)
						Completed but no mobility assessment	0 (0.0%)	1 (0.4%)	0 (0.0%)
						Not completed and no mobility assessment	1 (0.4%)	0 (0.0%)	0 (0.0%)
						Total	230 (100.0%)	230 (100.0%)	217 (100.0%)
Decision pathway allocated	1	95.3% (41/43)	85.7% (36/42)	100.0% (38/38)	All patients for whom only the blue box was ticked for both step 1 mobility and step 1 skin status				
Progression to step 2		195	197	182		Appropriate progression to step 2	185 (80.4%)	185 (80.4%)	179 (82.5%)
						Inappropriate progression to step 2	9 (3.9%)	12 (5.2%)	3 (1.4%)
						Inappropriate non-progression to step 2	0 (0.0%)	1 (0.4%)	0 (0.0%)
						Appropriate non-progression to step 2	35 (15.2%)	32 (13.9%)	35 (16.1%)
						Step 2 completed but step 1 not completed	1 (0.4%)	0 (0.0%)	0 (0.0%)
						Total	230 (100.0%)	230 (100.0%)	217 (100.0%)
Step 2 Full Assessment									
Step 1 Mobility	1 of 4	99.5% (194/195)	99.5% (196/197)	100.0% (182/182)	All patients who progressed to step 2, as all were required to complete step 1 mobility				
Step 1 skin status	1 of 4 (if required)	97.7% (43/44)	96.3% (52/54)	100.0% (28/28)	All patients for whom only the blue box was ticked in step 1 mobility who progressed to step 2				

Analysis of independent movement	1 of 5	99.0% (193/195)	99.0% (195/197)	98.9% (180/182)	All patients who progressed to step 2				
Sensory perception and response	1 of 2	96.9% (189/195)	94.9% (187/197)	98.4% (179/182)					
Current detailed skin assessment	13	95.5% (2421/2535)	95.3% (2440/2561)	97.5% (2307/2366)	13 (number of main skin sites) x number of patients who progressed to step 2				
Previous PU history	1 of 2	99.0% (193/195)	95.9% (189/197)	98.4% (179/182)	All patients who progressed to step 2				
Previous PU details	At least 1	66.7% (40/60)	54.7% (29/53)	57.4% (35/61)	All patients reported to have a PU history				
Perfusion	At least 1	97.9% (191/195)	97.5% (192/197)	97.3% (177/182)	All patients who progressed to step 2				
Nutrition	At least 1	99.0% (193/195)	99.5% (196/197)	97.8% (178/182)					
Moisture	1 of 3	99.5% (194/195)	97.0% (191/197)	96.7% (176/182)					
Diabetes	1 of 2	99.0% (193/195)	95.9% (189/197)	96.7% (176/182)					
Decision pathway allocated	1 of 3	100.0% (194/195)	99.0% (195/197)	100.0% (182/182)					
Assessment Decision									
Pathway allocated at step 1 or step 3						Appropriate pathway	226 (98.3%)	219 (95.2%)	215 (99.1%)
						Inappropriate pathway	4 (1.7%)	10 (4.3%)	2 (0.9%)
						No pathway selected	0 (0.0%)	1 (0.4%)	0 (0.0%)
						Total	230 (100.0%)	230 (100.0%)	217 (100.0%)

Table 3 PURPOSE T Decision pathway by colour of boxes ticked

PURPOSE-T Decision pathway	Colour of boxes ticked N(%)			
	At least one pink box ticked	No pink boxes and at least one orange box ticked	Only blue and yellow boxes ticked	Total N(%)
Expert nurse baseline				
PU Category 1 or above or scarring	72 (31.3)	0 (0.0)	0 (0.0)	72 (31.3)
No PU, but at risk	0 (0.0)	109 (47.4)	2 (0.9)	111 (48.3)
No PU, not currently at risk	0 (0.0)	4 (1.7)	43 (18.7)	47 (20.4)
Total	72 (31.3)	113 (49.1)	45 (19.6)	230 (100.0)
Ward/Community nurse				
PU Category 1 or above or scarring	63 (27.4)	0 (0.0)	0 (0.0)	63 (27.4)
No PU, but at risk	5 (2.2)	107 (46.5)	2 (0.9)	114 (49.6)
No PU, not currently at risk	0 (0.0)	5 (2.2)	47 (20.4)	52 (22.6)
Missing	0 (0.0)	1 (0.4)	0 (0.0)	1 (0.4)
Total	68 (29.6)	113 (49.1)	49 (21.3)	230 (100.0)
Expert nurse follow-up				
PU Category 1 or above or scarring	68 (31.3)	0 (0.0)	0 (0.0)	68 (31.3)
No PU, but at risk	2 (0.9)	104 (47.9)	1 (0.5)	107 (49.3)
No PU, not currently at risk	0 (0.0)	0 (0.0)	42 (19.4)	42 (19.4)
Total	70 (32.3)	104 (47.9)	43 (19.8)	217 (100.0)

Table 4 Cross tabulation of expert nurse PURPOSE-T decision pathway at baseline by ward/community nurse decision pathway and expert nurse decision pathway at follow-up

Inter-rater				
Expert nurse baseline	**Ward/Community Nurse**			
	PU Category 1 or above or scarring	**No PU but at risk**	**Not PU, not currently at risk**	**Total**
PU Category 1 or above or scarring	54 (23.6)	18 (7.9)	0 (0.0)	72 (31.4)
No PU, but at risk	9 (3.9)	91 (39.7)	10 (4.4)	110 (48.0)
No PU, not currently at risk	0 (0.0)	5 (2.2)	42 (18.3)	47 (20.5)
Total	63 (27.5)	114 (49.8)	52 (22.7)	229 (100.0)
	At risk		**Not at risk**	
At risk	172 (75.1)		10 (4.4)	182 (79.5)
Not at risk	5 (2.2)		42 (18.3)	47 (20.5)
Total	177 (77.3)		52 (22.7)	229 (100.0)
Test retest				
Expert nurse baseline	**Expert nurse follow-up**			
	PU Category 1 or above or scarring	**No PU but at risk**	**Not PU, not currently at risk**	**Total**
PU Category 1 or above or scarring	64 (30.0)	5 (2.3)	0 (0.0)	69 (32.4)
No PU, but at risk	3 (1.4)	95 (44.6)	5 (2.3)	103 (48.4)
No PU, not currently at risk	0 (0.0)	4 (1.9)	37 (17.4)	41 (19.2)
Total	67 (31.5)	104 (48.8)	42 (19.7)	213 (100.0)
	At risk		**Not at risk**	**Total**
At risk	167 (78.4)		5 (2.3)	172 (80.8)
Not at risk	4 (1.9)		37 (17.4)	41 (19.2)
Total	171 (80.3)		42 (19.7)	213 (100.0)

Table 5 Levels of agreement between expert nurses and ward/community nurses and between expert nurses at baseline and at follow up for specific risk factor items

Item	Expert nurse vs. ward nurse	Expert nurse baseline vs. follow-up
Step 1: Mobility (absolute agreement)	156/228 (68.4%)	165/212 (77.8%)
Step 1: Mobility (agreement on presence/absence of problem)	207/228 (90.8%)	197/212 (92.9%)
Skin status [step 1 and 2 combined] (absolute agreement)	189/230 (82.2%)	191/213 (89.7%)
Skin status [step 1 and 2 combined] (agreement on presence/absence of problem)	204/230 (88.7%)	200/213 (93.9%)
Analysis of independent movement (absolute agreement)	113/191 (59.2%)	114/177 (64.4%)
Analysis of independent movement (agreement on presence/absence of problem)	165/191 (86.4%)	147/177 (83.1%)
PU History	160/191 (83.7%)	165/177 (93.2%)
Sensory perception	151/191 (79.1%)	154/177 (87.0%)
Nutrition (problem vs no problem)	156/191 (81.7%)	154/177 (87.0%)
Unplanned weight loss	159/191 (83.2%)	159/177 (89.8%)
Poor nutritional intake	163/191 (85.3%)	159/177 (89.8%)
Low BMI	176/191 (92.1%)	170/177 (96.0%)
High BMI	170/191 (89.0%)	165/177 (93.2%)
Diabetic status	180/191 (94.2%)	166/177 (93.8%)
Perfusion status (absolute agreement)	125/191 (65.4%)	138/177 (78.0%)
Perfusion status (agreement on presence/absence of problem)	139/191 (72.8%)	154/177 (87.0%)
Moisture status (absolute agreement)	145/191 (75.9%)	155/177 (87.6%)
Moisture status (agreement on presence/absence of problem)	155/191 (81.2%)	159/177 (89.8%)

Table 6 Cross tabulation of PURPOSE-T with the Waterlow and the Braden scales – overall risk.

PURPOSE-T overall risk status	Waterlow overall risk status N(%)			Correlation coefficient
	At risk (≥10)	Not at risk (<10)	Total	
At risk	175 (76.1)	8 (3.5)	183 (79.6)	Phi 0.63 - moderate
Not at risk	18 (7.8)	29 (12.6)	47 (20.4)	
Total	193 (83.9)	37 (16.1)	230 (100.0)	
PURPOSE-T overall risk status	**Braden overall risk status N(%)**			**Correlation coefficient**
	At risk (≤18)	**Not at risk (>18)**	**Total**	
At risk	50 (29.6)	73 (43.2)	123 (72.8)	Phi 0.40 - moderate
Not at risk	0 (0.0)	46 (27.2)	46 (27.2)	
Total	50 (29.6)	119 (70.4)	169 (100.0)	

Table 7 – Cross tabulations of dichotomised PURPOSE-T constructs with relevant constructs on the Waterlow and the Braden scales.

PURPOSE T mobility Step 1	Braden Mobility			Correlation coefficient
	No limitation	Slightly/very limited/completely immobile	Total	
No problem	69 (30.1%)	10 (4.4%)	79 (34.5%)	Phi 0.60 Moderate
Problem	37 (16.2%)	113 (49.3%)	150 (65.5%)	
Total	106 (46.3%)	123 (53.7%)	229 (100.0%)	
PURPOSE T mobility Step 1	**Braden Activity**			**Correlation coefficient**
	Walks frequently	**Walks occasionally, chairfast or bedfast**	**Total**	
No problem	56 (24.5%)	23 (10.0%)	79 (34.5%)	Phi 0.66 Moderate
Problem	11 (4.8%)	139 (60.7%)	150 (65.5%)	
Total	67 (29.3%)	162 (70.7%)	229 (100.0%)	
PURPOSE T Sensory response and Perception	**Braden Sensory**			**Correlation coefficient**
	No impairment	**Slightly, very or completely limited**	**Total**	
No problem	134 (70.9%)	7 (3.7%)	141 (74.6%)	Phi 0.74 High
Patient unable to feel and/or respond appropriately to discomfort from pressure	11 (5.8%)	37 (19.6%)	48 (25.4%)	
Total	145 (76.7%)	44 (23.3%)	189 (100.0%)	
PURPOSE T nutrition	**Braden Nutrition**			**Correlation coefficient**
	Excellent or adequate	**probably inadequate or very poor**	**Total**	
No problem	93 (48.2%)	1 (0.5%)	94 (48.7%)	Phi 0.58 Moderate
Problem	47 (24.4%)	52 (26.9%)	99 (51.3%)	
Total	140 (72.5%)	53 (27.5%)	193 (100.0%)	
PURPOSE T nutrition	**Waterlow malnutrition screening tool: Patient eating poorly or lack of appetite**			**Correlation coefficient**
	Yes	**No**	**Total**	
Problem	64 (33.7%)	35 (18.4%)	99 (52.1%)	Phi 0.60 Moderate
No problem	6 (3.2%)	85 (44.7%)	91 (47.9%)	
Total	70 (36.8%)	120 (63.2%)	190 (100.0%)	
PURPOSE T nutrition – poor nutritional intake	**Braden Nutrition**			**Correlation coefficient**
	probably inadequate or very poor	**Excellent or adequate**	**Total**	
Yes	50 (25.9%)	12 (6.2%)	62 (32.1%)	Phi 0.82 High
No	3 (1.6%)	128 (66.3%)	131 (67.9%)	
Total	53 (27.5%)	140 (72.5%)	193 (100.0%)	
PURPOSE T nutrition – poor nutritional intake	**Waterlow malnutrition screening tool: Patient eating poorly or lack of appetite**			**Correlation coefficient**
	Yes	**No**	**Total**	
Yes	57 (30.0%)	5 (2.6%)	62 (32.6%)	Phi 0.79 High
No	13 (6.8%)	115 (60.5%)	128 (67.4%)	
Total	70 (36.8%)	120 (63.2%)	190 (100.0%)	
PURPOSE T nutrition – low BMI	**Waterlow build or weight for height**			**Correlation coefficient**
	BMI<20	**BMI>=20**	**Total**	
Yes	21 (11.0%)	0 (0.0%)	21 (11.0%)	Phi 0.72 High
No	16 (8.4%)	154 (80.6%)	170 (89.0%)	
Total	37 (19.4%)	154 (80.6%)	191 (100.0%)	
PURPOSE T nutrition – High BMI	**Waterlow build or weight for height**			**Correlation coefficient**
	BMI>=30	**BMI<30**	**Total**	
Yes	22 (11.5%)	4 (2.1%)	26 (13.6%)	Phi 0.74 High
No	9 (4.7%)	156 (81.7%)	165 (86.4%)	
Total	31 (16.2%)	160 (83.8%)	191 (100.0%)	

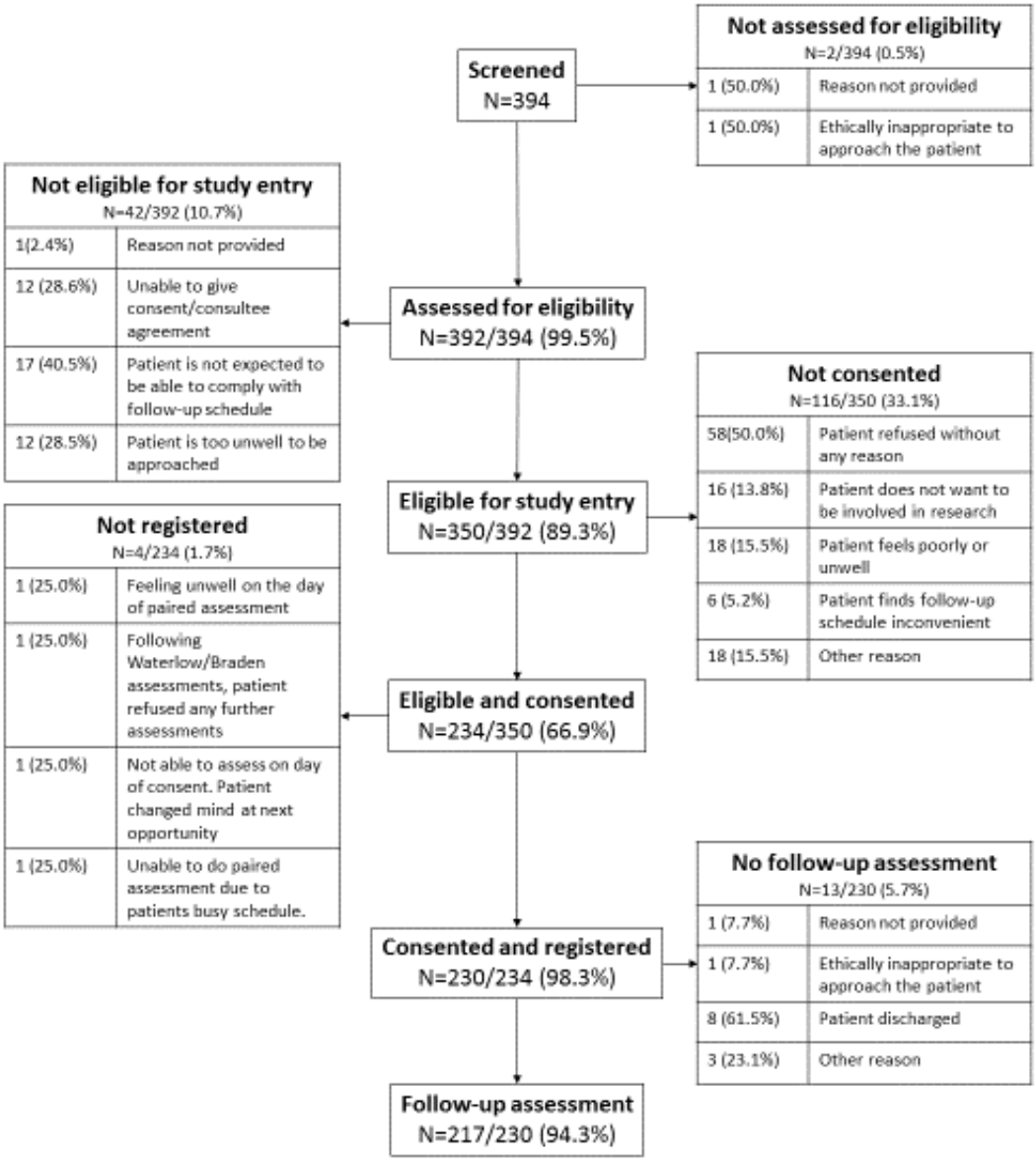
Table 8 – Cross tabulations of dichotomised PURPOSE-T constructs with relevant constructs on the Waterlow and the Braden scales.

PURPOSE T Step 2 Analysis of independent movement	Braden Mobility				Correlation coefficient
	Completely immobile	Very or slightly limited	No limitation	Total	
Does not Move	7 (3.6%)	4 (2.1%)	0 (0.0%)	11 (5.7%)	Spearman Rank 0.62 Moderate
Moves occasionally and slight or major position changes or moves frequently with slight position changes	1 (0.5%)	96 (49.7%)	26 (13.5%)	123 (63.7%)	
Moves frequently and major position changes	0 (0.0%)	12 (6.2%)	47 (24.4%)	59 (30.6%)	
Total	8 (4.1%)	112 (58.0%)	73 (37.8%)	193 (100.0%)	
PURPOSE T Step 2 Analysis of independent movement	**Braden Activity**				**Correlation coefficient**
	Bedfast	**Chairfast or walks occasionally**	**Walks frequently**	**Total**	
Does not Move	6 (3.1%)	5 (2.6%)	0 (0.0%)	11 (5.7%)	Spearman Rank 0.55 Moderate
Moves occasionally and slight or major position changes or moves frequently with slight position changes	15 (7.8%)	102 (52.8%)	6 (3.1%)	123 (63.7%)	
Moves frequently and major position changes	0 (0.0%)	31 (16.1%)	28 (14.5%)	59 (30.6%)	
Total	21 (10.9%)	138 (71.5%)	34 (17.6%)	193 (100.0%)	
PURPOSE T skin status	**Waterlow skin status**				**Correlation coefficient**
	Healthy	**Tissue paper dry oedematous clammy**	**Discoloured grade 1 or broken spots grade 2-4**	**Total**	
Normal Skin	47 (20.6%)	18 (7.9%)	0 (0.0%)	65 (28.5%)	Spearman Rank 0.83 High
Vulnerable skin	11 (4.8%)	79 (34.6%)	13 (5.7%)	103 (45.2%)	
PU Category	1 (0.4%)	0 (0.0%)	59 (25.9%)	60 (26.3%)	
Total	59 (25.9%)	97 (42.5%)	72 (31.6%)	228 (100.0%)	
PURPOSE T Moisture	**Braden Moisture**				**Correlation coefficient**
	Rarely or occasionally moist	**Very moist**	**Constantly moist**	**Total**	
No problem/ Occasional	154 (79.4%)	2 (1.0%)	0 (0.0%)	156 (80.4%)	Spearman Rank 0.67 Moderate
Frequent (2-4 times a day)	18 (9.3%)	12 (6.2%)	0 (0.0%)	30 (15.5%)	
Constant	0 (0.0%)	4 (2.1%)	4 (2.1%)	8 (4.1%)	
Total	172 (88.7%)	18 (9.3%)	4 (2.1%)	194 (100.0%)	

Table 9 Summary of Expert Nurse Field Notes

Characteristic	Positive Aspects of Using PURPOSE	Problem Aspects of Using PURPOSE T
Layout	• Easy to use and self-explanatory • Quick to use • Easier to use with familiarity • All on one page	• Tool looked 'busy' or 'complicated' • Font size small • Space for skin assessment too small
Format	• The RAG rating approach for assessment decision and use of colour made distinctive • Like the fact it didn't use a score like other risk assessment scales	• Form does not flow • Unclear whether to progress to Step 2 • Concern that exiting at Step 1 would miss assessment of important risk factors • Nurses wanted to complete full skin assessment at step 1.
Content	• Thorough and included important risk factors • Positive about the detailed skin assessment and suggested that this encouraged more careful skin assessment • Inclusion of pressure ulcer scar as a risk factor	• Reliability of assessment of skin vulnerability • Reliability of assessment of scarring • Difficulty establishing history of previous pressure ulcer: ▪ difficult and time consuming ▪ where information available was of poor quality (e.g. severity not clear) • Duration of weight loss not specified • Assessment of circulation items in patients with respiratory problems • Analysis of movement difficult to categorise
Usability	• Will be easy for nurses to remember and report red boxes at handover • Step 1 screening is efficient in allowing the quick identification of those who do not require a full assessment • Not having to visually inspect pressure areas when a patient was not at risk was appreciated	• Local production difficult if no colour printers available

Figure 1 Flow of participants



References

ANDERSEN, K. E., JENSEN, O., KVORNING, S. A. & BACH, E. 1982. Prevention of pressure sores by identifying patients at risk. British Medical Journal Clinical Research Ed., 284**,** 1370-1.

BEECKMAN, D., MATHEÏ, C., VAN LANCKER, A., VAN HOUDT, S., VANWALLEGHEM, G., GRYSON, L., HEYMAN, H., THYSE, C., TOPPETS, A., STORDEUR, S., VAN DEN HEEDE, K. A. & 2013. A National Guideline for the prevention of pressure ulcers: Good Clinical Practice (GCP), KCE Reports 193C.

. Belgian Health Care Knowledge Centre (KCE).

BENNETT, G., DEALEY, C. & POSNETT, J. 2004. The cost of pressure ulcers in the UK. Age Ageing, 33**,** 230-5.

BERGSTROM, N., BRADEN, B., KEMP, M., CHAMPAGNE, M. & RUBY, E. 1998. Predicting Pressure Ulcer Risk: A Multisite Study of the Predictive Validity of the Braden Scale. Nursing Research September/October, 47**,** 261-269.

BERGSTROM, N., BRADEN, B. J., LAGUZZA, A. & HOLMAN, V. 1987. The Braden Scale for Predicting Pressure Sore Risk. Nursing Research, 36**,** 205-10.

BERLOWITZ, D., VANDEUSEN LUKAS, C., PARKER, V., NIEDERHAUSER, A., SILVER, J., LOGAN, C. & AYELLO, E. 2011. Preventing Pressure Ulcers in hospitals: A toolkit for improving quality of care. Available: http://www.ahrq.gov/research/ltc/pressureulcertoolkit/putool1.htm [Accessed 15/10/2012].

BLAND, J. 2008. Measuring Health and Disease: Cohen's Kappa. . Department of Health Sciences, University of York.

BLAZEBY, J., SPRANGERS, M., CULL, A., GROENVOLD, M. & BOTTOMLEY, A. 2002. Guidelines for Developing Questionnaire Modules, 3rd Edition Revised. URL: http://groups.eortc.be/qol/sites/default/files/archives/guidelines_for_
developing_questionnaire-_final.pdf (accessed 9 March 2015). EORTC Quality of Life Group.

BRADEN, B. & BERGSTROM, N. 1987. A Conceptual Schema for the Study of the Etiology of Pressure Sores. Rehabilitation Nursing, 12**,** 8-16.

BRIGGS, M., COLLINSON, M., WILSON, L., RIVERS, C., MCGINNIS, E., DEALEY, C., BROWN, J., COLEMAN, S., STUBBS, N., STEVENSON, R., NELSON, E. A. & NIXON, J. 2013. The prevalence of pain at pressure areas and pressure ulcers in hospitalised patients http://www.biomedcentral.com/1472-6955/12/19. BMC Nurs, 12**,** 19.

BURNAND, B., KERNAN, W. N. & FEINSTEIN, A. R. 1990. Indexes and boundaries for "quantitative significance" in statistical decisions. Journal of Clinical Epidemiology, 43**,** 1273-1284.

BYRT, T., BISHOP, J. & CARLIN, J. B. 1993. Bias, prevalence and kappa. J Clin Epidemiol., 46.

CICCHETTI, D. & ALLISON, T. 1971. A New Procedure for Assessing Reliability of Scoring EEG Sleep Recordings. American Journal of EEG Technology, 11**,** 101-110.

COHEN, J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20**,** 37-46.

COLEMAN, S. 2014. The development of a pressure ulcer risk assessment framework and minimum data set. PhD, University of Leeds.

COLEMAN, S., GORECKI, C., NELSON, E. A., CLOSS, S. J., DEFLOOR, T., HALFENS, R., FARRIN, A., BROWN, J., SCHOONHOVEN, L. & NIXON, J. 2013. Patient risk factors for pressure ulcer development: Systematic review. International Journal of Nursing Studies, 50**,** 974-1003.

COLEMAN, S., NELSON, E. A., KEEN, J., WILSON, L., MCGINNIS, E., DEALEY, C., STUBBS, N., MUIR, D., FARRIN, A., DOWDING, D., SCHOLS, J. M. G. A., CUDDIGAN, J., BERLOWITZ, D., JUDE, E., VOWDEN, P., BADER, D. L., GEFEN, A., OOMENS, C. W. J., SCHOONHOVEN, L. & NIXON, J. 2014a. Developing a pressure ulcer risk factor minimum data set and risk assessment framework. Journal of Advanced Nursing, 70**,** 2339-2352.

COLEMAN, S., NIXON, J., KEEN, J., MUIR, D., WILSON, L., MCGINNIS, E., STUBBS, N., DEALEY, C. & NELSON, E. A. 2016. Using Cognitive Pre-Testing Methods in the Development of a New Evidenced-Based Pressure Ulcer Risk Assessment Instrument. BMC Research Methodology, 16.

COLEMAN, S., NIXON, J., KEEN, J., WILSON, L., MCGINNIS, E., DEALEY, C., STUBBS, N., FARRIN, A., DOWDING, D., SCHOLS, J. M., CUDDIGAN, J., BERLOWITZ, D., JUDE, E., VOWDEN, P., SCHOONHOVEN, L., BADER, D. L., GEFEN, A., OOMENS, C. W. & NELSON, E. A. 2014b. A new pressure ulcer conceptual framework. Journal of Advanced Nursing, 70**,** 2222-34.

COLEMAN, S., NIXON, J., KEEN, J., WILSON, L., MCGINNIS, E., DEALEY, C., STUBBS, N., FARRIN, A., DOWDING, D., SCHOLS, J. M. G. A., CUDDIGAN, J., BERLOWITZ, D., JUDE, E., VOWDEN, P., SCHOONHOVEN, L., BADER, D. L., GEFEN, A., OOMENS, C. W. J. & NELSON, E. A. 2014c. A new pressure ulcer conceptual framework. Journal of Advanced Nursing, 70**,** 2222-2234.

CUBBIN, B. & JACKSON, C. 1991. Trial of a pressure area risk calculator for intensive therapy patients. Intensive Care Nursing, 7**,** 40-4.

DEALEY, C., POSNETT, J. & WALKER, A. 2012. The cost of pressure ulcers in the United Kingdom. Journal of Wound Care, 21**,** 261-6.

DEEKS, J. J. 1996. Pressure sore prevention: using and evaluating risk assessment tools. Br J Nurs, 5**,** 313-4, 316-20.

DEFLOOR, T. & GRYPDONCK, M. F. 2004. Validation of pressure ulcer risk assessment scales: a critique. Journal of Advanced Nursing, 48**,** 613-21.

FDA DHHS 2009. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (accessed March 2014)
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.

FERRANTE DI RUFFANO, L., HYDE, C. J., MCCAFFERY, K. J., BOSSUYT, P. M. M. & DEEKS, J. J. 2012. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ, 344.

GORECKI, C., BROWN, J. M., NELSON, E. A., BRIGGS, M., SCHOONHOVEN, L., DEALEY, C., DEFLOOR, T., NIXON, J. & EUROPEAN QUALITY OF LIFE PRESSURE ULCER PROJECT, G. 2009. Impact of pressure ulcers on quality of life in older patients: a systematic review. Journal of the American Geriatrics Society, 57**,** 1175-83.

GORECKI, C., NIXON, J., MADILL, A., FIRTH, J. & BROWN, J. M. 2012. What influences the impact of pressure ulcers on health-related quality of life? A qualitative patient-focused exploration of contributory factors. Journal of Tissue Viability, 21**,** 3-12.

GOULD, D., GOLDSTONE, L., GAMMON, J., KELLY, D. & MAIDWELL, A. 2002. Establishing the validity of pressure ulcer risk assessment scales: a novel approach using illustrated patient scenarios. International Journal of Nursing Studies, 39**,** 215-28.

KOTTNER, J. & BALZER, K. 2010. Do Pressure Ulcer risk Assessment Scales Improve Clinical Practice. Journal of Multidisciplinary Healthcare, 3**,** 103-111.

KWONG, E., PANG, S., WONG, T., HO, J., SHAO-LING, X. & LI-JUN, T. 2005. Predicting pressure ulcer risk with the modified Braden, Braden, and Norton scales in acute care hospitals in Mainland China. Applied Nursing Research, 18**,** 122-8.

LANDIS, J. R. & KOCH, G. G. 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics, 33**,** 159-174.

LINDGREN, M., UNOSSON, M., KRANTZ, A. M. & EK, A. C. 2002. A risk assessment scale for the prediction of pressure sore development: reliability and validity. Journal of Advanced Nursing, 38**,** 190-9.

MOKKINK, L., TERWEE, C., PATRICK, D., ALONSO, J., STRATFORD, P., KNOL, D., BOUTER, L. & DE VET, H. 2012. COSMIN checklist manual EMGO Institute for Health and Care Research

MOORE, G. F., AUDREY, S., BARKER, M., BOND, L., BONELL, C., HARDEMAN, W., MOORE, L., O'CATHAIN, A., TINATI, T., WIGHT, D. & BAIRD, J. 2015. Process evaluation of complex interventions: Medical Research Council guidance. BMJ, 350.

MRC 2000. A framework for development and evaluation of RCTs for complex interventions to improve health.

MRC 2008. Developing and evaluating complex interventions: new guidance. In: COUNCIL, M. R. (ed.). London.

NICE 2014. Pressure ulcer prevention: the prevention and management of pressure ulcers in primary and secondary care, Clinical Guideline 179, Methods, evidence and recommendations. National Clinical Guideline Centre.

NIXON, J. & MCGOUGH, A. (eds.) 2001. Principles of patient assessment: screening for pressure ulcers and potential risk in: Morison, M (Ed). The prevention and treatment of pressure ulcers, Edinburgh: Mosby.

NIXON, J., NELSON, E. A., RUTHERFORD, C., COLEMAN, S., MUIR, D., KEEN, J., MCCABE, C., DEALEY, C., BRIGGS, M., BROWN, S., COLLINSON, M., HULME, C., MEADS, D., MCGINNIS, E., PATTERSON, M., CZOSKI-MURRAY, C., PINKNEY, L., SMITH, I. L., STEVENSON, R., STUBBS, N., WILSON, L. & BROWN, J. M. 2015. Pressure UlceR Programme Of reSEarch (PURPOSE): using mixed methods (systematic reviews, prospective cohort, case study, consensus and psychometrics) to identify patient and organisational risk, develop a risk assessment tool and patient-reported outcome Quality of Life and Health Utility measures. Programme Grants Appl Res 3.

NPUAP/EPUAP/PPPIA 2014. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel and Pan Pacific Pressure Injury Alliance.

Prevention and Treatment of Pressure Ulcers: Clinical Practice Guideline. Emily Haesler (ed). Cambridhe Media: Obsborne Park Western Australia.
NUNNALLY, J. & BERNSTEIN, I. 1994. Psychometric Theory, New York, NY, McGraw-Hill.
PAGE, K. N., BARKER, A. L. & KAMAR, J. 2011. Development and validation of a pressure ulcer risk assessment tool for acute hospital patients. Wound Repair Regen, 19**,** 31-7.
PERNEGER, T. V., RAE, A. C., GASPOZ, J. M., BORST, F., VITEK, O. & HELIOT, C. 2002. Screening for pressure ulcer risk in an acute care hospital: development of a brief bedside scale. Journal of Clinical Epidemiology, 55**,** 498-504.
PIEPER, B. E. 2012. Pressure Ulcers: Prevalence, Incidence, and Implications for the Future. Washinton DC: NPUAP.
PINKNEY, L., NIXON, J., WILSON, L., COLEMAN, S., MCGINNIS, E., STUBBS, N., DEALEY, C., NELSON, A., PATTERSON, M. & KEEN, J. 2014. Why do patients develop severe pressure ulcers? A retrospective case study, http://bmjopen.bmj.com/content/4/1/e004303.full.pdf
BMJ Open, 4.
PRITCHARD, V. 1986. Pressure sores. Calculating the risk. Nursing Times, 82**,** 59-61.
RICHARDS, D. A. & HALLBERG, I. R. 2015. Complex Interventions in Health: An Overview of Research Methods, London, Routledge Taylor and Francis Group.
SAC 2002. Assessing health status and quality-of-life instruments: Attributes and review criteria. Quality of life research, 11**,** 193-205.
SCHUURMAN, J.-P., SCHOONHOVEN, L., DEFLOOR, T., VAN ENGELSHOVEN, I., VAN RAMSHORST, B. & BUSKENS, E. 2009. Economic evaluation of pressure ulcer care: a cost minimization analysis of preventive strategies. Nursing Economics, 27**,** 390-400.
SEVERENS, J. L., HABRAKEN, J. M., DUIVENVOORDEN, S. & FREDERIKS, C. M. A. 2002. The cost of illness of pressure ulcers in The Netherlands. Advances in Skin & Wound Care, 15**,** 72-7.
SMITH, I. L., NIXON, J., BROWN, S., WILSON, L. & COLEMAN, S. 2016. Pressure ulcer and wounds reporting in NHS hospitals in England part 1: Audit of monitoring systems. Journal of Tissue Viability, 25**,** 3-15.
STEYERBERG, E., W. 2010. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, New York, USA, Springer.
SURIADI, SANADA, H., SUGAMA, J., THIGPEN, B., KITAGAWA, A., KINOSITA, S. & MURAYAMA, S. 2006. A new instrument for predicting pressure ulcer risk in an intensive care unit. Journal of Tissue Viability, 16**,** 21-6.
SURIADI, SANADA, H., SUGAMA, J., THIGPEN, B. & SUBUH, M. 2008. Development of a new risk assessment scale for predicting pressure ulcers in an intensive care unit. Nurs Crit Care, 13**,** 34-43.
WATERLOW, J. 1985. Pressure sores: a risk assessment card. Nursing Times, 81**,** 49-55.
WATERLOW, J. 1988. The Waterlow card for the prevention and management of pressure sores: towards a pocket policy. Care, Science and Practice, 6**,** 8–12.