INVESTIGATIVE REPORT

The Reliability, Validity and Responsiveness of Two Disease Scores (BPDAI and ABSIS) for Bullous Pemphigoid: Which One to Use?*

Asri WIJAYANTI[1,4], Cathy Y. ZHAO[1,4], David BOETTIGER[2,4], Yi Z. CHIANG[1], Norito ISHII[3], Takashi HASHIMOTO[3] and Dedee F. MURRELL[1,4]

[1]Department of Dermatology, St George Hospital, [2]The Kirby Institute, Sydney, Australia, [3]Kurume University Institute of Cutaneous Cell Biology, Fukuoka, Japan, and [4]Faculty of Medicine, University of New South Wales, Sydney, Australia

A significant obstacle in guiding evidence-based management of bullous pemphigoid (BP) is the lack of a standardised, validated scoring system for the condition. The aim of this study was to evaluate the suitability of the Bullous Pemphigoid Disease Area Index (BPDAI) and the Autoimmune Bullous Skin disorder Intensity Score (ABSIS) as outcome measures for BP in clinical trials. Thirty-two BP patients were repeatedly assessed over 4 years using Physician Global Assessment (PGA), anti-BP180 ELISA titres, BPDAI, ABSIS, BPDAI-Pruritus, Autoimmune Bullous Disease Quality of Life (ABQOL) and Treatment of Autoimmune Bullous Disease Quality of Life (TABQOL) questionnaires. The reliability, validity, responsiveness, and minimal clinically important differences (MCIDs) were calculated. For inter-rater reliability, the intra-class correlation coefficients (95% CI) were: BPDAI 0.957 (0.901–0.982) and ABSIS 0.881 (0.736–0.949). Compared to ABSIS, BPDAI was better correlated with PGA ($r = 0.875$, $p < 0.001$), BPDAI-Pruritus ($r = 0.632$, $p = 0.004$), ABQOL ($r = 0.521$, $p = 0.011$) and TABQOL ($r = 0.538$, $p = 0.008$). MCIDs for BPDAI were 4-points for assessing clinical improvement and 3-points for deterioration. ABSIS demonstrated less responsiveness with MCIDs at 8.6-points for improvement and 4-points for deterioration. These results indicate that BPDAI demonstrated excellent reliability, validity and responsiveness, while ABSIS had moderate to good reliability, validity and responsiveness.

Key words: bullous pemphigoid; outcome measures; BPDAI; ABSIS, ABQOL; TABQOL; validation; responsiveness; MCID.

Accepted May 24, 2016; Epub ahead of print May 31, 2016

Acta Derm Venereol 2017; 97: 24–31.

Corr: Prof. Dedee F. Murrell, Department of Dermatology, St George Hospital, Gray Street, Kogarah, Sydney, NSW 2217, Australia. E-mail: d.murrell@unsw.edu.au.

Bullous pemphigoid (BP) represents the commonest form of autoimmune bullous diseases (AIBDs), characterised by autoantibodies directed against two keratinocyte epitopes of the dermal-epidermal junction, BP180 and BP230 (1–4).

Although our scientific understanding of BP has progressed significantly in recent years, there is a scarcity of randomised-controlled trials to guide evidence-based management of BP (5–8). Pooling of data through meta-analysis of existing studies can provide valuable information but this is only achievable if uniform outcome measures are available. Most studies, however, employed a concoction of outcome measures based on non-validated, subjective, or non-specific ratings of disease activity (5–8). As such, a significant obstacle in comparing treatment modalities is the lack of a standardised, validated scoring system for BP. Clinical trials in BP require a standardised, well-validated scoring system.

The Bullous Pemphigoid Disease Area Index (BPDAI) (9) and the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) (10) are two independent disease severity assessments proposed to measure disease extent in BP. So far, there have been only two validation studies for BPDAI and ABSIS in patients with BP. A study conducted in Greece (11) revealed that both BPDAI and ABSIS were significantly correlated with anti-BP180 titers but not with anti-BP230 titres. Anti-BP180, but not anti-BP230, titres has been established to correlate well with disease activity (12–14). Another study, conducted in France (15), found good correlation of BPDAI with clinical parameters of disease activity and proposed a cut-off value of 56 for severe BP.

However, the intra-rater and inter-rater reliability and responsiveness to change for both tools have not been determined. These are crucial when conducting clinical trials because in actual practice, doctors assessing patients may change from visit to visit and if the outcome measure is still reliable between different doctors, it can be more easily applied in studies and in real life practice. In addition, patient-reported outcome (PRO) measures are becoming more and more relevant in clinical trials, and hence the correlation between these subjective PROs and the objective scores are important. Therefore, we evaluated the reliability, convergent validity, and responsiveness of BPDAI and ABSIS and their correlation to

*An oral presentation of the study findings was awarded first prize at the St George and Sutherland Medical Research Symposium in Sydney, Australia, October 2014. The study's findings were presented at the Australasian Society for Dermatology Research Cutaneous Biology Meeting in Queensland, Australia, September 2014; the Medical Dermatology Society Annual Meeting in San Fransisco, United States of America, March 2015; and 23rd World Congress of Dermatology in Vancouver, Canada, June 2015.

doi: 10.2340/00015555-2473
Acta Derm Venereol 2017; 97: 24–31

Journal Compilation © 2017 Acta Dermato-Venereologica. ISSN 0001-5555

validated PROs for bullous pemphigoid for use in clinical and research settings.

MATERIALS AND METHODS

Study design and ethics

This observational study received ethical approval from the South Eastern Sydney Local Health District Human Research Ethics Committee – Northern Network.

Patient recruitment and informed consent

All patients were older than 18 years and enrolled with informed consent. Inclusion criteria included clinical examination, histologic result from skin biopsy and immunofluorescence studies consistent with the diagnosis of BP. The patient sample included a range of disease severities (**Table I**).

Sample size

Sample size was determined in consultation with a professional statistician from the University of New South Wales, Sydney, Australia. Calculations based on a review of similar studies showed that at least 30 patients and 100 separate visits would allow a sufficient intraclass correlation coefficient.

Setting

Thirty-two patients participated in a prospective-cohort validation study that was conducted over 4.5 years from February 2010 to September 2014. Several outcome instruments were completed at each routine clinic visit (baseline, week two, month one and subsequent 3-monthly visits). The principal clinician (D.F.M.) assessed patients using BPDAI, ABSIS, and Physician's Global Assessment of Severity (PGA). BPDAI measures skin and mucosal involvement with two activity components, one for extent of blistering, one for extent of urticarial/eczematous change and a separate damage component (9). Damage scores in BPDAI are not intended for assessment of disease severity; they are there to remind clinicians not to score post-inflammatory effects as activity (9). ABSIS assesses all lesions for extent by the rule of nines with a multiplication factor for degree of erosion; it also has a subjective component for oral mucosal involvement, which is often zero for BP patients (10). The PGA has been validated in studying pemphigus, psoriasis, dermatomyositis, and other dermatoses (16–18).

Table I. Distribution of scores of activity, damage and quality of life assessments

Outcome measure	Mean (SD, IQR)	Range
BPDAI total activity (skin + mucosal)	8.1 (12.1, 0–10)	0–79
BPDAI total skin activity	7.3 (12.1, 0–10)	0–77
BPDAI mucosa	0.8 (2.6, 0–0)	0–20
BPDAI damage	1.7 (2.7, 0–3)	0–11
ABSIS total	8.3 (12.6, 0.1–11.5)	0–65.5
ABSIS objective (skin + mucosal)	6.9 (11.5, 0.4–7)	0–53.5
ABSIS subjective (mucosal discomfort)	1.7 (5.2, 0–0)	0–23.5
Physician Global Assessment total	3 (2.8, 0.8–5)	0–9
BPDAI Pruritus component	5.7 (7.9, 0–10)	0–10
ABQOL	11.6 (8, 6–16)	0–40
TABQOL	14.7 (10, 7–22)	0–45
Anti-BP180 ELISA	8.4 (12.2, 0.9–10.6)	0–45.1
Anti-BP230 ELISA	14.8 (24.1, 1.3–21.6)	0.9–76.1

BPDAI: Bullous Pemphigoid Disease Area Index; ABSIS: Autoimmune Bullous Skin Disorder Intensity Score; PGA; ABQOL: Autoimmune Bullous Disease Quality of Life; TABQOL: Treatment of Autoimmune Bullous Disease Quality of Life; SD: standard deviation; IQR: interquartile range (25% to 75% of score).

Dermatology fellows, who changed at each visit, rated each BP patient independently, using BPDAI and ABSIS, for inter-rater reliability testing. For intra-rater reliability, repeated and blinded scoring was performed randomly after a mean of 30 min, in the period of March to September 2014. None of the assessors were informed beforehand about the rescoring activity to minimise recall bias.

To determine responsiveness and minimal clinically important differences (MCIDs), the principal clinician (D.F.M.) classified disease activity as improved, stable, or deteriorated since the last visit. This method of Physician's Subjective Assessment of Clinical Improvement (PSACI) has been used in multiple responsiveness studies of the Cutaneous Lupus Disease Area and Severity Index (CLASI) and Localised Scleroderma Cutaneous Assessment Tool (19–21).

Patients were asked to complete the BPDAI-Pruritus, validated Autoimmune Bullous Disease Quality of Life (ABQOL) (22) and Treatment of Autoimmune Bullous Disease Quality of Life (TABQOL) (23) questionnaires to measure the subjective element of disease progression. ABQOL measures AIBD-specific disease burden while TABQOL measures burden attributable to the side effects of AIBD treatment, with higher scores indicating worse QOL. Patients' sera were collected for anti-BP180 and anti-BP230 ELISA testing when routine safety bloods were taken at these visits.

Measurement tools

BPDAI. The BPDAI tool (9) computes 2 scores: total BPDAI activity and total BPDAI damage. The total BPDAI activity score is the arithmetic sum of the 3 subcomponents – cutaneous blisters/erosions, cutaneous urticaria/erythema, and mucosal blisters/erosions. The total BPDAI damage score is the arithmetic sum of the items rated regionally for damage caused by more permanent features such as post-inflammatory hyperpigmentation, scarring and other. BPDAI quantifies lesion number and size thresholds. Lesions are rated based on the regions affected. BPDAI gives additional weighting to areas of the skin primarily affected in BP, such as the limbs, and less emphasis to scalp and face, to better differentiate clinical response in BP. Scores can range from 0 to 360 for BPDAI total activity (maximum 240 for total skin activity and 120 for mucosal activity), and 0 to 12 for BPDAI damage, with higher scores indicating greater disease activity or damage. BPDAI also has a separate subjective measure known as BPDAI-pruritus.

ABSIS. ABSIS (10) is an instrument that is widely used for the measurement of disease activity in patients with autoimmune blistering diseases. ABSIS uses Rule of Nines and Rule of Palms methods where the palm of the patient's hand is set as 1% of total body surface area. The estimated percentage of body surface area involved is then multiplied with a weighting factor 1.5 (erosive, exudative lesions, blisters), 1 (erosive, dry lesions), and 0.5 (re-epithelialised lesions). The total score ranges from 0 to 206, with higher scores denoting more severe disease. It consists of subsections for skin involvement (maximum score of 150), mucosal extent (maximum score of 11), and patient-reported mucosal discomfort (maximum score of 45).

PGA. PGA is a ten-point Likert scale ranging from 0–10, indicating perfect health to worst skin condition. This scoring system was used to allow rating of disease activity by general overall impression. Patients were also categorised as having mild, moderate, or severe disease by the principle clinician.

PSACI. To determine responsiveness to minimal clinically significant changes, the principle clinician (D.F.M.) also classified disease activity in each patient as improved, stable, or deteriorated since the last visit.

Serum markers. Anti-BP180 titres have been established to correlate well with disease activity while anti-BP230 levels appear to

be irrelevant to the extent of disease activity (12–14). Commercial ELISA kits (MBL) were used to measure anti-BP180 by bullous experts from Kurume University, Japan. The positive index for anti-BP180 level is ≥ to 9. We compared these levels to the scores of BPDAI and ABSIS.

BPDAI-Pruritus. Pruritus is a major symptom of BP that may signal disease onset and recurrence. A separate subjective component of the BPDAI, BPDAI-pruritus, was used to measure the severity of this. The intensity of pruritus was subjectively graded using a visual analog scale where 0 represented no itch and 10 as maximal itching. Patient marked an "x" to indicate severity of itching today, in the past week, and in the past month, producing a total score out of 30. In cases where patient was unable to reliably complete the grading, pruritus was inferred from the degree of excoriations, also scored out of 30.

ABQOL. Autoimmune bullous disease-specific quality of life (QOL) was measured using the ABQOL that has been shown to have good validity and reliability (22). This questionnaire consists of 17 items, which encompasses physical burden of the disease, psychiatric effects and effects to daily life functioning. Each question ranges from 0 to 3 points, with higher scores indicating worse QOL. The maximum ABQOL score is 51.

TABQOL. Side effects associated with the treatment of autoimmune blistering diseases can have significant risk of medical complications and impairment to QoL. TABQOL questionnaire was used in this study to measure the burden attributable to the effects of treatment specific to autoimmune bullous diseases (23). Similar to ABQOL, this questionnaire consists of 17 items, with higher scores indicating worse QOL due to treatment effects.

Statistical analysis

All statistical analyses were performed using software (SPSS, Version 22.0, IBM Corp, Armonk, NY; and Excel, Microsoft, Redmond, WA).

Reliability. Both inter- and intra-rater reliabilities were assessed by intraclass correlation coefficients (ICC) with 95% confidence intervals (CI), using Type A ICC with absolute agreement definition and two-way random effect analysis. An ICC value of above 0.81 is considered excellent (24). Visits were randomly selected using the Random Number Generator function on Excel. Selections and analyses were performed 3 times with different clinic visits to verify the results obtained. Bland-Altman plots were constructed to illustrate the spread of scores and inter-rater differences for the 2 scoring instruments. Linear regression analyses were conducted to detect any proportional bias in determining the inter-rater reliability.

Convergent validity. BPDAI activity/damage and ABSIS scores were examined against global severity rating, serum markers, and QoL measures. We used Pearson's correlation with the *p*-value to determine correlation coefficients between pairs of variables. *t*-test was performed to evaluate if positive anti-BP180 and anti-BP230 titres predict higher BPDAI and ABSIS scores. A *p*-value of ≤ 0.05 was considered significant. Visits were randomly selected using the Random Number Generator function and repeated three times to verify the correlation coefficients obtained.

Responsiveness. To validate and assess the responsiveness of BPDAI and ABSIS to clinically significant changes, paired *t*-test was conducted to determine if there was a statistically significant difference in score to the previous visit. For a tool to be responsive to clinically meaningful changes, there should be significant differences in visits with improvement and deterioration, and no significant difference in visits that remained stable.

MCIDs. Receiver Operating Characteristic (ROC) analyses were performed to determine the MCIDs of BPDAI and ABSIS. The sensitivity, specificity and percentage of patients correctly classified for each signed change were determined using this method. MCIDs were chosen based on the mean signed change in BPDAI and ABSIS scores derived by responders, which is then confirmed by assessing the ROC at and around that cut-off, focusing primarily on the highest percentage of patients correctly classified.

RESULTS

Baseline characteristics

Of the 32 BP patients, 40.6% were male and 59.4% were female, with a mean age of 78 years (range: 52–93 years). 65.6% of patients were Caucasian, 25% were Hispanic, and 9.4% were Asian.

All 32 patients were included in the convergent validity analysis. Twenty-two patients were assessed for inter-rater reliability and 15 patients had rescoring completed for intra-rater reliability. Twenty-seven patients had at least 2 visits with completed outcome measures, totalling to 108 visits included in the responsiveness analysis. 13 patients had anti-BP180 and anti-BP230 ELISA testing results from Japan (Table I).

Reliability

For inter-rater reliability of the 22 BP patients, the ICCs (95% CI) for all visits (including no activity) were: BPDAI total activity 0.957 (0.901–0.982), BPDAI damage 0.807 (0.592–0.915), and ABSIS total score 0.881 (0.736–0.949) (Table II). BPDAI was demonstrated in Bland-Altman plots to have less variability in scores compared to ABSIS, particularly as scores increase in magnitude. This is shown by the distribution of the scores close to the mean inter-assessor difference of scores (**Fig. 1**). Linear regression analyses for both BPDAI and ABSIS showed a T-value with no statistical significance, revealing that there was no proportional bias.

For intra-rater reliability of the 15 BP patients, the ICCs (95% CI) were: BPDAI total activity 0.993 (0.969–0.998), BPDAI damage 0.939 (0.828–0.979), and ABSIS total score 0.995 (0.987–0.998) (Table II). Analysis of BPDAI and ABSIS subsectional ICCs for various subcomponents were also calculated for inter- and intra-rater reliabilities (**Table II**).

Two repeat analyses demonstrated similar findings (**Table III**). Four patients had no visible evidence of disease activity, generating physician-assessed scores of 0. ICCs were recalculated with these 4 patients excluded to remove the floor effect. The resulting ICCs were only slightly reduced for inter-rater reliability and had close to no difference for intra-rater reliability (Table III). Perfect correlations were obtained for the subjective component of the ABSIS because these BP patients had no oral involvement and hence all scored zero for pain when eating different types of food. The same occurred with the mucosal component (extent) for BPDAI.

Table II. Reliability analyses for Bullous Pemphigoid Disease Area Index and Autoimmune Bullous Skin Disorder Intensity Score

	ICC (95% CI) All visits	ICC (95% CI) Removing floor effect i.e. scores of 0
Inter-rater reliability		
Bullous Pemphigoid Disease Area Index (BPDAI)		
Skin erosions/blisters	0.939 (0.860–0.974)	0.930 (0.825–0.973)
Skin urticaria/erythema/other	0.929 (0.912–0.985)	0.925 (0.814–0.971)
Total skin activity	0.957 (0.900–0.982)	0.950 (0.873–0.981)
Mucosal activity	1.000	1.000
Total activity	0.957 (0.901–0.982)	0.949 (0.872–0.981)
Total damage	0.807 (0.592–0.915)	0.964 (0.690–0.996)
Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)		
Objective physician-assessed score (skin + mucosal score)	0.884 (0.741–0.950)	0.876 (0.698–0.952)
Subjective patient-reported score (mucosal discomfort)	0.947 (0.879–0.978)	0.947 (0.867–0.980)
Total score	0.881 (0.736–0.949)	0.871 (0.689–0.950)
Intra-rater reliability		
Bullous Pemphigoid Disease Area Index (BPDAI)		
Skin erosions/blisters	0.996 (0.989–0.999)	0.996 (0.984–0.999)
Skin urticaria/erythema/other	0.989 (0.955–0.997)	0.988 (0.915–0.997)
Total skin activity	0.994 (0.974–0.998)	0.993 (0.949–0.998)
Mucosal activity	0.996 (0.989–0.999)	0.996 (0.987–0.999)
Total activity	0.993 (0.969–0.998)	0.992 (0.931–0.998)
Total damage	0.939 (0.828–0.979)	0.932 (0.768–0.981)
Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)		
Objective physician-assessed score (skin + mucosal score)	0.995 (0.985–0.998)	0.995 (0.982–0.999)
Subjective patient-reported score (mucosal discomfort)	0.930 (0.794–0.978)	0.999 (0.999–1.000)
Total score	0.995 (0.987–0.998)	0.995 (0.981–0.999)

All values significant at $p < 0.001$. Inter-rater and intra-rater reliabilities for the two outcome instruments assessed using the ICC at 95% confidence level (CI). ICC: intraclass correlation coefficient.

Convergent validity

BPDAI activity had a substantial correlation with ABSIS total score (r=0.864, $p<0.001$) while, as expected, BPDAI damage had low correlation with ABSIS total score (r=0.362, $p<0.05$). BPDAI activity (r=0.875, $p<0.001$) showed higher correlation with PGA than ABSIS total score (r=0.752, $p<0.001$) with PGA. As with the Patsatsi study, both disease activity scores were significantly correlated with anti-BP180 levels ($p<0.05$) but not with anti-BP230 levels.

BPDAI activity had moderate and significant correlations with BPDAI Pruritus, ABQOL and TABQOL. BPDAI damage was not significantly correlated with any subjective measure. ABSIS total score was only significantly correlated with BPDAI pruritus (Table SI[1]) but not with ABQOL and TABQOL scores, which often remained high even when patients had no evidence of disease activity, due to anxiety about disease recurrence and concerns about their medications. When BPDAI and ABSIS were 0, mean (standard deviation, range, interquartile range) of ABQOL was 8.4 (5.7, 0–26, 4–12) and TABQOL 12.8 (8.2, 0–31, 6.5–18), with no correlation with QOL measures (Table III). All correlations were repeated two more times with different clinic visits and demonstrated similar findings to the first analysis (Table SI[1]).

Table III. Two repeat analyses for reliability of Bullous Pemphigoid Disease Area Index (BPDAI) and Autoimmune Bullous Skin Disorder Intensity Score (ABSIS)

	ICC (95% CI) First repeat	ICC (95% CI) Second repeat
BPDAI		
Skin erosions/blisters	0.962 (0.912–0.984)	0.933 (0.845–0.971)
Skin urticaria/erythema/other	0.752 (0.484–0.889)	0.858 (0.692–0.938)
Total skin activity	0.932 (0.843–0.971)	0.943 (0.870–0.976)
Mucosal activity	1.000	0.881 (0.738–0.948)
Total activity	0.935 (0.849–0.972)	0.928 (0.836–0.970)
Total damage	0.728 (0.455–0.877)	0.781 (0.499–0.907)
ABSIS		
Objective physician-assessed score (skin + mucosal score)	0.808 (0.599–0.915)	0.885 (0.742–0.950)
Subjective patient-reported score (mucosal discomfort)	1.000	1.000
Total score	0.808 (0.599–0.915)	0.886 (0.746–0.951)

All values significant at $p < 0.001$. Inter-rater reliability assessed using the intraclass correlation coefficient (ICC) at 95% confidence level. Intra-rater reliability analysis was not repeated as most patients were assessed randomly for intra-rater testing only at one visit. CI: confidence interval.

Responsiveness and minimal clinically important differences

There were 59 pairs of evaluable visits, including 18 instances (30.5%) of clinically significant improvement, 24 instances (40.7%) of remaining stable, and 17 instances (28.8%) of clinically significant deterioration.

In instances of clinically significant improvement, the mean BPDAI activity and ABSIS total scores decreased by 11.5 ($p<0.001$) and 12.6 ($p<0.01$) points, respectively. In instances of clinically significant deterioration, the

[1]https://doi.org/10.2340/00015555-2473

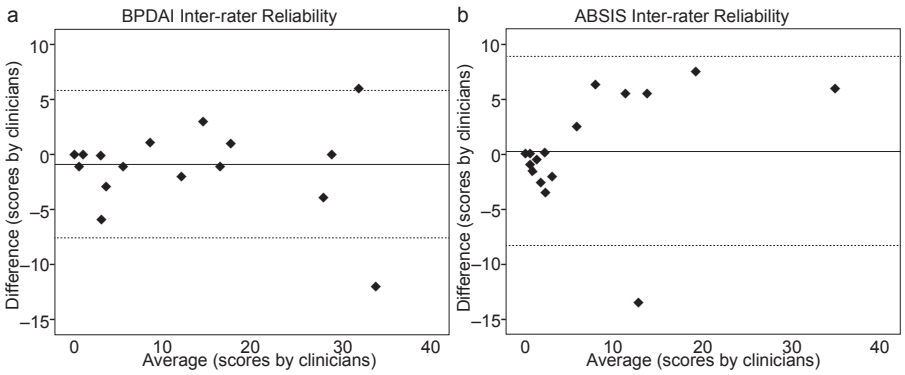

Fig. 1. Bland-Altman plots demonstrating the variation in patient scores, which is a measure of inter-rater reliability for Bullous Pemphigoid Disease Area Index (BPDAI) activity (a) and Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) total (b) scores. The middle horizontal line represents mean inter-assessor difference of scores. The dotted lines represent the 95% confidence limits.

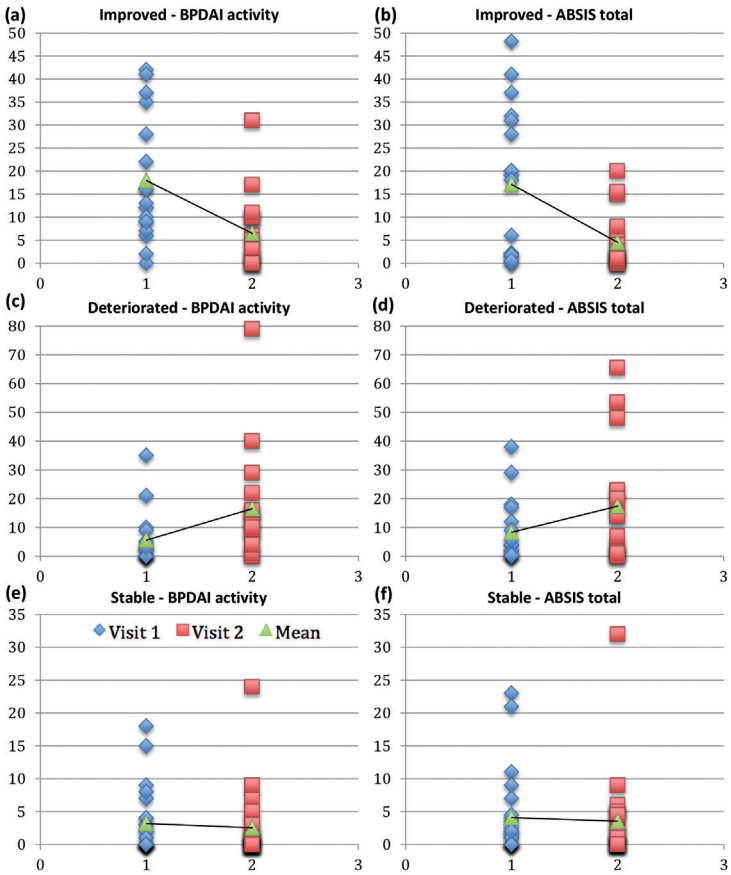

Fig. 2. Responsiveness of Bullous Pemphigoid Disease Area Index (BPDAI) activity and Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) total scores to clinical improvement (a, b), deterioration (c, d), and stability (e, f), respectively.

mean BPDAI activity and ABSIS total scores increased by 11.0 ($p<0.001$) and 9.1 ($p<0.05$) points, accordingly (**Fig. 2**). There were no statistically significant differences in visits that remained stable ($p=0.21$ and 0.57).

An ROC analysis indicated that a 4-point decrease in the BPDAI activity score was 82.4% sensitive and 91.7% specific for improvement, resulting in 87.8% of patients being correctly classified as having clinically improved or remained stable. The area under the curve was 0.913 (**Table IV**). Another ROC analysis indicated that a 3-point increase in the BPDAI activity score was 76.5% sensitive and 91.7% specific for deterioration, resulting in 85.4% of patients being correctly classified as having clinically worsened or remained stable (Table IV). Area under the curve was 0.924 (**Fig. 3**). These areas under the ROC curves are considered excellent, as they are greater than 0.9 (25).

In categorising improvement against remaining stable, a 4.75-point decrease in ABSIS total score was demonstrated to produce 58.7% sensitivity, 83.3% specificity, and 73.2% correct categorisation. This can be maximised by using 8.6-points to produce 78% correct categorisation. In classifying deterioration against remaining stable, a 4-point increase in ABSIS total score generated 52.9% sensitivity, 87.5% specificity, and 73.2% correct classification (Table IV). These sensitivities, specificities, and percentages of correct classifications were substantially lower than those with using the BPDAI activity score (Fig. 3). In addition, ROC analyses for ABSIS total score indicated a smaller area under the curve for both clinical improvement (0.799) and deterioration (0.729). These areas are considered fair as they lie between the range of 0.7 and 0.8 (25).

Table IV. Responsiveness of Bullous Pemphigoid Disease Area Index (BPDAI) and Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) to minimal clinically important differences

Clinical change	Δ	Sens. (%)	Spec. (%)	Corr. (%)	AUC	95% CI	*p*-value
BPDAI activity							
Improved	−2.5	82.4	87.5	85.4	0.913	0.817–1.000	<0.001
	−4.0	82.4	91.7	87.8			
	−5.5	76.5	95.8	87.8			
Deteriorated	+0.5	94.1	66.7	78	0.924	0.846–1.000	<0.001
	+1.5	82.4	83.3	82.9			
	+3.0	76.5	91.7	85.4			
ABSIS total							
Improved	−0.15	82.4	62.5	72.5	0.799	0.653–0.945	0.001
	−4.75	58.7	83.3	73.2			
	−8.60	52.9	95.8	78			
Deteriorated	+1.5	64.7	70.8	68.3	0.729	0.567–0.891	0.013
	+2.1	64.7	75	70.7			
	+4.0	52.9	87.5	73.2			

Δ: change score; Sens.: Sensitivity; Spec.: Specificity; Corr. Correctly classified; AUC: area under the curve CI: confidence interval.

DISCUSSION

Thus far no clinical trials in BP have used a validated severity score. In order to choose a primary outcome measure for BP, the severity score needs to have content validity, demonstrate excellent inter and intra-rater reliability, and be sensitive to changes in clinical improvement and deterioration – i.e. responsiveness. This study was able to compare the reliabilities of the BPDAI and ABSIS for BP and determine the MCID to determine improvement and worsening and disease scores. The study found that the BPDAI had excellent inter-rater reliability and a superior ability to discriminate responsiveness to therapy than the ABSIS for BP.

BPDAI's better inter-rater reliability and precision are likely because it was developed in terms of content specifically for BP, whilst ABSIS was originally developed for pemphigus and then used for other AIBDs which rarely involve the mouth. BPDAI quantifies lesion number and size threshold, which has been shown in previous studies (16, 26–28) to be a more reproducible method of cutaneous assessment than percentage estimates of body surface area involvement used in ABSIS. The Rule of Nines and Rule of Palms methods adopted in ABSIS have been frequently illustrated to have poor accuracy,

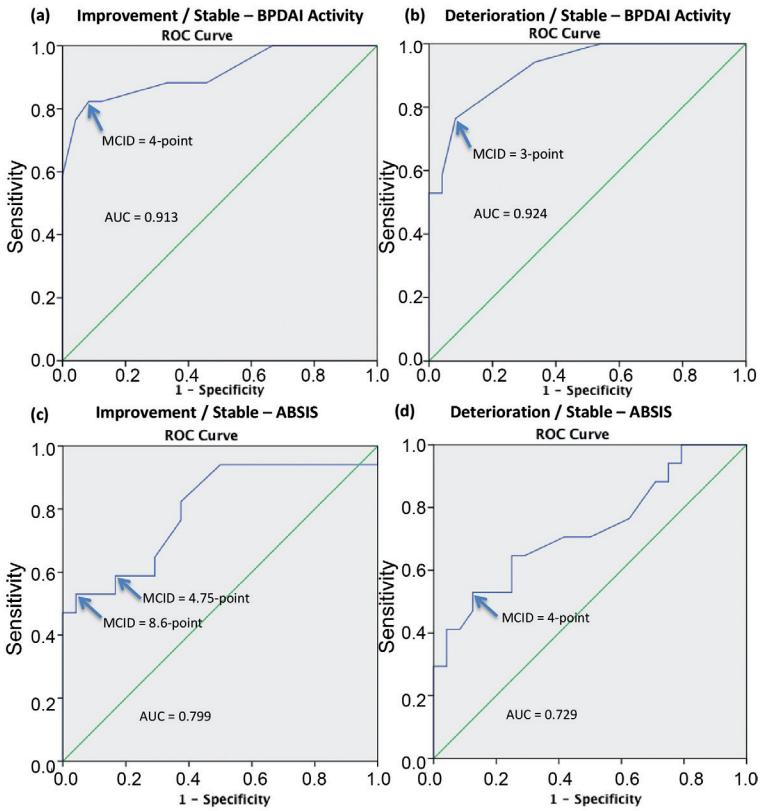

Fig. 3. Receiver Operating Characteristic curves for Bullous Pemphigoid Disease Area Index (BPDAI) (a, b) and Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) (c, d) for clinical improvement and clinical deterioration. AUC: area under the curve; MCID: minimal clinically important difference.

particularly at quantifying small areas with a tendency to overestimate the area involved (16, 29–31). The proportion of a patient's palm area remains relatively constant throughout life at 0.76–0.78% of total body surface area, thus not exactly equivalent to 1% (32, 33). In the original ABSIS publication (10), the authors acknowledged that estimating with these rules can be challenging, and may result in inter-rater disagreement when raters are untrained. In this study, the primary assessor is an experienced dermatologist with special interest in bullous diseases and all other assessors are clinicians training full-time in dermatology practice. Both of the measures were found to have superior inter-rater reliability than the inter-reliability of the PGA visual analogue score found by other studies examining other dermatological diseases (34).

The intra-rater reliabilities for BPDAI activity (ICC 0.993) and ABSIS total scores (0.995) in our study were equally excellent. Intra-rater testing was only performed from March to September 2014, recruiting patients with limited disease extent. Therefore, although this result represents excellent test-retest reliability for low disease activity, it cannot be extrapolated to patients with more severe disease. Compared to inter-rater reliability, intra-rater reliability may be less important in real practice in clinics, as clinicians change for the same patient.

BPDAI activity was found to have excellent convergent validity against PGA and anti-BP180 titre. ABSIS had only fair correlation with PGA, indicating that it is less accurate in measuring disease extent of BP than the BPDAI. Both BPDAI activity and ABSIS total scores were revealed to have moderate correlation with BPDAI pruritus component, implying that patients with worse disease activity are more likely to have prevalent symptoms of pruritus.

Patient-reported QOL measures correlated substantially better with BPDAI activity than BPDAI damage score. This denotes that patients experiencing severe BP activity reported worse QOL and that residual hyperpigmentation and other features were not particularly affecting their QOL. Accordingly, composite scores that included damage – ABSIS and PGA, were less likely to correlate with QOL. As the BPDAI gives greater weighting to the areas of the skin relevant to BP patients, perhaps that is an additional explanation of why the BPDAI correlated with QOL measures and yet ABSIS did not. It is also worth noting that the QOL in patients with BP remained poor despite resolving disease activity, suggesting the importance of holistic, long-term management of these patients.

Other key new findings in this study were responsiveness and MCID. Responsiveness analysis of BPDAI and ABSIS helped in defining clinical improvement or deterioration, allowing trial investigators to objectively identify patients with improvement from treatment, or patients experiencing deterioration from relapse, from those undergoing no change. A lower MCID indicates better sensitivity at detecting a smaller degree of improvement or deterioration. MCIDs for ABSIS were higher and its sensitivity, specificity and area under the curve were substantively lower than BPDAI activity, resulting in a lower percentage of patients by approximately 10% classified in each category. BPDAI activity is therefore more capable of distinguishing minor differences in severity and correctly separating true responders from non-responders than ABSIS. BPDAI damage score, as expected, was not responsive to clinical change, emphasising the importance of separating activity and damage to avoid stability, or increase of the total score, while the active disease is improving.

In the responsiveness analysis, the sensitivity for the MCID selected was lower than specificity. Optimisation of sensitivity could have been achieved with lower change scores, but this would have caused specificity to decrease. For the objectives of a clinical trial, we agree with the CLASI responsiveness study for lupus (35) to consider specificity more important than sensitivity to

minimise the degree of false-positive responses. This reduces the inclusion of participants who have not experienced true clinical change. As it is critical to have high specificity, we based the analysis on the percentage of patients correctly classified.

Limitations of our study include the limited sample size, given that BP primarily occurs in older patients who may have more comorbidities and difficulties in mobilising to the outpatient clinic. To compensate for this, we scored patients longitudinally at repeated clinic visits. This sample size is larger than the number of patients recruited for other previous validation studies, the pemphigus study (15 patients) (16) and CLASI responsiveness study (8 patients) (19), and most of those patients also had limited disease extent. Moreover, the highest BPDAI total skin activity and ABSIS total scores in this study were 77/240 and 65.5/206 (32.1% and 31.8% of possible maximum scores), respectively. As with the PASI, the severe cases are generally in the 25% to 45% range of the maximum score (36). Additionally, the BPDAI and ABSIS should be validated in other settings in which it would be employed thus another limitation was that this is a single-centre study.

The comprehensively validated BPDAI and ABSIS provide simple and quantitative clinical tools that homogenise the way disease activity is quantified. Also, they provide guidelines for evaluating clinical improvement or deterioration in bullous pemphigoid. This analysis provides a foundation for the validated and practical use of BPDAI in clinical trials and routine practice.

ACKNOWLEDGEMENT

IRB status: The study received ethical approval from the South Eastern Sydney Local Health District Human Research Ethics Committee – Northern Network (HREC 06/94).

Presentation: An oral presentation of the study findings was awarded first prize at the St George and Sutherland Medical Research Symposium in Sydney, Australia, October 2014. The study's findings were presented at the Australasian Society for Dermatology Research Cutaneous Biology Meeting in Queensland, Australia, September 2014; the Medical Dermatology Society Annual Meeting in San Fransisco, United States of America, March 2015; and 23rd World Congress of Dermatology in Vancouver, Canada, June 2015.

The authors thank the Australasian Blistering Diseases Foundation for supporting this project and the Independent Learning Program of UNSW Australia.

REFERENCES

1. Stanley JR, Hawley-Nelson P, Yuspa SH, Shevach EM, Katz SI. Characterization of bullous pemphigoid antigen: a unique basement membrane protein of stratified squamous epithelia. Cell 1981; 24: 897–903.
2. Diaz, LA, Ratrie H, Saunders WS, Futamura S, Squiquera HL, Anhalt GJ, et al. Isolation of a human epidermal cDNA corresponding to the 180-kD autoantigen recognized by bullous pemphigoid and herpes gestationis sera. Immunolocalization of this protein to the hemidesmosome. J Clin Invest 1990; 86: 1088.
3. Sardy M, Kostaki D, Varga R, Peris K, Ruzicka T. Comparative study of direct and indirect immunofluorescence and of bullous pemphigoid 180 and 230 enzyme-linked immunosorbent assays for diagnosis of bullous pemphigoid. J Am Acad Dermatol 2013; 69: 748–753.
4. Nishie W, Sawamura D, Goto M, Ito K, Shibaki A, McMillan JR, et al. Humanization of autoantigen. Nature Med 2007; 13: 378–383.
5. Kirtschig G, Middleton P, Bennett C, Murrell DF, Wojnarowska F, Khumalo NP. Interventions for bullous pemphigoid. Cochrane Database Syst Rev 2010; CD002292.
6. Martin LK, Agero AL, Werth VP, Villanueva E, Segall J, Murrell DF. Interventions for pemphigus vulgaris and pemphigus foliaceus. Cochrane Database Syst Rev 2009; CD006263.
7. Martin LK, Murrell DF. Treatment of pemphigus: The need for more evidence. Arch Dermatol 2008; 144: 100–101.
8. Daniel BS, Hertl M, Werth VP, Eming R, Murrell DF. Severity score indexes for blistering diseases. Clin Dermatol 2012; 30: 108–113.
9. Murrell DF, Daniel BS, Joly P, Borradori L, Amagai M, Hashimoto T, et al. Definitions and outcome measures for bullous pemphigoid: recommendations by an international panel of experts. J Am Acad Dermatol 2012; 66: 479–485.
10. Pfütze M, Niedermeier A, Hertl M, Eming R. Introducing a novel Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) in pemphigus. Eur J Dermatol 2007; 17: 4–11.
11. Patsatsi A, Kyriakou A, Pavlitou-Tsiontsi A, Giannakou A, Sotiriadis D. Association of autoantibodies to BP180 with disease activity in Greek patients with bullous pemphigoid. Clin Dev Immunol 2012; 2012: 1–4.
12. Hofmann SC, Thoma-Uszynski S, Hunziker T, Bernard P, Koebnick C, Stauber A, et al. Severity and phenotype of bullous pemphigoid relate to autoantibody profile against the NH2- and COOH-terminal regions of the BP180 ectodomain. J Invest Dermatol 2002; 119: 1065–1073.
13. Schmidt E, Obe K, Bröcker EB, Zillikens D. Serum levels of autoantibodies to BP180 correlate with disease activity in patients with bullous pemphigoid. Arch Dermatol 2000; 136: 174–178.
14. Di Zenzo G, Thoma-Uszynski S, Fontao L, Calabresi V, Hofmann SC, Hellmark T, et al. Multicenter prospective study of the humoral autoimmune response in bullous pemphigoid. Clin Immunol 2008; 128: 415–426.
15. Lévy-Sitbon C, Barbe C, Plee J, Goeldel AL, Antonicelli F, Reguiaï Z, et al. Assessment of Bullous Pemphigoid Disease Area Index during treatment: A prospective study of 30 patients. Dermatology 2014; 229: 116–122.
16. Rosenbach M, Murrell DF, Bystryn JC, Dulay S, Dick S, Fakharzadeh S, et al. Reliability and convergent validity of two outcome instruments for pemphigus. J Invest Dermatol 2009; 129: 2404–2410.
17. Langley RG, Ellis CN. Evaluating psoriasis with Psoriasis Area and Severity Index, Psoriasis Global Assessment, ad Lattice System Physician's Global Assessment. J Am Acad Dermatol 2008; 58: 1043–1046.
18. Hundley JL, Carroll CL, Lang W, Snively B, Yosipovitch G, Feldman SR, et al. Cutaneous symptoms of dermatomyositis significantly impact patients' quality of life. J Am Acad Dermatol 2006; 54: 217–220.
19. Bonilla-Martinez ZL, Albrecht J, Troxel AB, Taylor L, Okawa J, Dulay S, et al. The Cutaneous Lupus Erythematosus Disease Area and Severity Index: a responsive instrument to measure activity and damage in patients with cutaneous lupus erythematosus. Arch Dermatol 2008; 144: 173–180.
20. Jolly M, Kazmi N, Mikolaitis RA, Sequeira W, Block JA. Validation of the Cutaneous Lupus Disease Area and Severity Index (CLASI) using physician- and patient-assessed health outcome measures. J Am Acad Dermatol 2012; 68: 618–623.
21. Kelsey CE, Torok KS. The Localized Scleroderma Cutaneous Assessment Tool: Responsiveness to change in a pediatric clinical population. J Am Acad Dermatol 2013; 69: 214–220.
22. Sebaratnam DF, Hanna AM, Chee SN, Frew JW, Venugopal SS, Daniel BS, et al. Development of a quality-of-life instrument for autoimmune bullous disease: the Autoimmune Bullous

Disease Quality of Life questionnaire. JAMA Dermatol 2013; 149: 1186–1191.
23. Tjokrowidjaja A, Daniel BS, Frew JW, Sebaratnam DF, Hanna AM, Chee S, et al. The development and validation of the Treatment of Autoimmune Bullous Disease Quality of Life questionnaire, a tool to measure the quality of life impacts of treatments used in patients with autoimmune blistering disease. Brit J Dermatol 2013; 169: 1000–1006.
24. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979; 86: 420–427.
25. Hanley JA, McNeil BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 1982; 143: 29–36.
26. Albrecht J, Taylor L, Berlin JA, Dulay S, Ang G, Fakharzadeh S, et al. The CLASI (cutaneous lupus erythematosus disease area and severity index): an outcome instrument for cutaneous lupus erythematosus. J Invest Dermatol 2005; 125: 889–894.
27. Goreshi R, Okawa J, Rose M, Feng R, Lee LA, Hansen CB, et al. Evaluation of reliability, validity, and responsiveness of the CDASI and the CAT-BM. J Invest Dermatol 2012; 132: 2171–2175.
28. Loh CCH, Kim J, Su JC, Daniel BS, Venugopal SS, Rhodes LM, et al. Development, reliability, and validity of a novel Epidermolysis Bullosa Disease Activity and Scarring Index (EBDASI). J Am Acad Dermatol 2014; 70: 89–97.e13.
29. Tiling-Grosse S, Rees J. Assessment of area of involvement in skin disease: a study using schematic figure outlines. Br J Dermatol 1993; 128: 69–74.
30. Wachtel TL, Berry CC, Wachtel EE, Frank HA. The inter-rater reliability of estimating the size of burns from various burn area chart drawings. Burns 2000; 26: 156–170.
31. Charman CR, Venn AJ, Williams HC. Measurement of body surface area involvement in atopic eczema: an impossible task? Br J Dermatol 1999; 140: 109–111.
32. Gehan EA, George SL. Estimation of human body surface area from height and weight. Cancer Chemother Rep 1970; 54: 225–235.
33. Amirsheybani HR, Crecelius GM, Timothy NH, Pfeiffer M, Saggers GC, Manders EK. The natural history of the growth of the hand, I: hand area as a percentage of body surface area. Plast Reconstr Surg 2001; 107: 726–733.
34. Goreshi R, Okawa J, Rose M, Feng R, Lee LA, Hansen CB, et al. Evaluation of reliability, validity, and responsiveness of the CDASI and the CAT-BM. J Invest Dermatol 2012; 132: 1117–1124.
35. Klein R, Moghadam-Kia S, LoMonico J, Okawa J, Coley C, Taylor L, et al. Development of the CLASI as a tool to measure disease severity and responsiveness to therapy in cutaneous lupus erythematosus. Arch Dermatol 2011; 147: 203–208.
36. Finlay AY. Current severe psoriasis and the rule of tens. Br J Dermatol 2005; 152: 861–867.