



Review Article

Validation studies of outcome measures in pemphigus ☆☆☆☆

Sarah Hanna^{a,b}, Minhee Kim, MBBS^{a,b}, Dedee F. Murrell, MA, BMBCh, FAAD, MD, FACD, FRCP(Edin)^{a,b,*}^a Department of Dermatology, St. George Hospital, Sydney, Australia^b Faculty of Medicine, University of New South Wales, Sydney, Australia

ARTICLE INFO

Article history:

Received 1 September 2016

Received in revised form 28 October 2016

Accepted 28 October 2016

Keywords:

autoimmune blistering diseases

pemphigus

outcome measures

validation

pemphigus disease area index

autoimmune bullous skin disorder intensity

score

ABSTRACT

Pemphigus is a group of rare and potentially fatal autoimmune blistering diseases that are associated with auto-antibodies that target intercellular adhesion molecules. Incidence of pemphigus varies among populations, with the lowest incidence in Switzerland and Finland at 0.6–0.76 per million per year and the highest in Jewish communities at 16.1–32 per million per year. Pemphigus is associated with devastating morbidity and despite advancements in our understanding of the disease and a widening array of therapeutic options, no cure exists. The delay in the development of a cure may in part be attributed to the absence of a standardized and completely validated severity outcome measures to allow for high-quality multicenter control studies. Such a tool is necessary to define the best practice in clinical studies, allow for accurate comparisons between study results, justify drug use within the clinical setting, and reduce the cost burden that is associated with the use of ineffective therapies. Utilizing outcome measures that are not validated provides an opportunity to synthesize outcome measures with the intent to favor particular treatments and thus produce false conclusions. According to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) group, a validation of these measurement instruments requires investigating their responsiveness, reliability, and validity. More than 116 outcome measures exist to assess pemphigus severity, of which the Pemphigus Disease Area Index (PDAI), Autoimmune Bullous Skin Disorder Intensity Score (ABSIS), and Pemphigus Vulgaris Activity Score (PVAS) are the most comprehensively corroborated measures. With regard to validity and reliability, PDAI was unsurpassed by ABSIS and PVAS. Data indicate that ABSIS is more reliable than PVAS, but PVAS seems to have greater validity although the results are not consistent. PDAI, ABSIS, and PVAS have not yet had their responsiveness analyzed, which should be the next step to completely validate the outcome measures and conclusively determine which measure is superior.

© 2016 The Authors. Published by Elsevier Inc. on behalf of Women's Dermatologic Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

This literature review will discuss the key features of pemphigus, illustrate the significance of validated scoring systems, outline previous responsiveness studies for other dermatological scoring systems, and discuss existing outcome measures for pemphigus. The purpose

is to understand which scores are better for use in studies and clinical practice and what research remains to be conducted in this area.

Background

Pemphigus is a group of autoimmune vesiculobullous diseases that are associated with auto-antibodies that target intercellular adhesion molecules. The majority of these auto-antibodies are immunoglobulin (Ig) G that target the ectodomain of desmosomal cadherins and in doing so cause loss of keratinocyte-to-keratinocyte adhesion (i.e., acantholysis). Acantholysis leads to blister formation in the epidermis and patients may develop cutaneous flaccid bullae, erosions, or pustules and/or mucosal erosions. The mechanism whereby IgG auto-antibodies induce keratinocyte detachment is still widely debated. The two principal theories include steric

☆ Conflicts of interest: D. F. Murrell was a co-author/senior author on some of the outcome measures listed (PDAI, BPDAl).

☆☆ Funding sources: Australian Blistering Diseases Foundation grant and Independent Learning Program of the University of New South Wales Australia.

★ IRB status: The study received ethical approval from the South Eastern Sydney Local Health District Human Research Ethics Committee – Northern Network (11/STG/134).

* Corresponding Author:

E-mail address: d.murrell@unsw.edu.au (D.F. Murrell).

<http://dx.doi.org/10.1016/j.ijwd.2016.10.003>

2352-6475/© 2016 The Authors. Published by Elsevier Inc. on behalf of Women's Dermatologic Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

hindrance, which implies direct interference with desmosomal adhesion, and triggering of intracellular signalling, which causes loss of keratinocyte adhesion (Evangelista et al., 2015). There is amassing evidence to support both theories and it is likely that both are significant to the pathogenesis of pemphigus.

There are multiple pemphigus subtypes that possess characteristic clinical, histological, and immunologic features, seemingly due to distinctive desmosomal protein targets. These subtypes include pemphigus vulgaris (PV), pemphigus foliaceus (PF), endemic pemphigus foliaceus or fogo selvage (FS), paraneoplastic pemphigus (PNP), and IgA pemphigus (Evangelista et al., 2015). A diagnosis of pemphigus is reliant on clinical features, findings via lesional and perilesional biopsy (histopathology and direct immunofluorescence respectively), and serology (indirect immunofluorescence and enzyme-linked immunosorbent assay [ELISA]; Table 1).

The incidence of pemphigus alters markedly between populations. This variability is due to the different genetic backgrounds and trigger factors that are associated with particular geographical locations. Most epidemiological studies concur that persons with Jewish ancestry are the most at risk to develop PV. However, the quality of these studies is impeded by their retrospective design and inability to ensure inclusion of all patients (Schmidt et al., 2015).

Mortality associated with PV and PF dropped dramatically from 75% to 30% with the introduction of corticosteroid treatment in the 1950s. Subsequently, adjuvant use of immunosuppressant drugs in the 1980s decreased mortality rates further to approximately 5% of the study populations. Most recently, studies in Taiwan and the United Kingdom have demonstrated that a patient's risk of death compared to a healthy control is 2–3 times greater, primarily because of infections and particularly pneumonia and septicemia (Huang et al., 2012; Langan et al., 2008; Schmidt et al., 2015).

Importance of scoring systems

Measurement in medicine is impeded by the absence of a consensus on the best instruments to utilize to characterize disease severity. Consequently, this results in non-comparable study outcomes, conceivably false conclusions, and non-evidence based practice. Scoring systems in dermatology are particularly challenging given the shortage of radiographic and laboratory findings that are known to correlate with disease severity (Gaines and Werth, 2008). Thus, generic instrumentation such as the Physician Global Assessment (PGA) are often utilized. The advantage of generic instrumentation is their versatility but their poor reproducibility, reliance on physician experience with the condition, and inability to capture the severity of illnesses that are localized to small areas (e.g., acne) are significant disadvantages (Albrecht and Werth, 2007). Disease-specific scoring systems provide superior accuracy and sensitivity compared with generic scoring systems, as has been proven with the Psoriasis Area and Severity Index (PASI) and Scoring Atopic Dermatitis (SCORAD) (Schram et al., 2012; Weisman et al., 2003).

For pemphigus, there is a definite shortage of multicenter controlled studies that is widely attributed to the difficulty in objectively comparing therapeutic outcomes. A systematic literature review counted more than 116 different outcome measures for pemphigus severity that were used in 96 articles published during the preceding 25 years (Martin and Murrell, 2006). A standardized and validated scoring system is required to address this issue and used universally to: (1) quantify disease severity and progression for interventions in clinical studies (Gaines and Werth, 2008) and allow multidisciplinary discussion of cases (Loh et al., 2014); (2) justify drug use in clinical settings (Gaines and Werth, 2008); and (3) reduce financial costs by identifying and ceasing ineffective treatments (Gaines and Werth, 2008).

Importance of validation

Validation studies illustrate the responsiveness, reliability, and practicality of a tool with regard to its intended measure (Streiner et al., 2008). The use of unvalidated tools provides for the opportunity to produce incorrect study conclusions by utilizing scoring systems that are synthesized specifically in favor of particular treatments, as shown in a systematic review by Marshall et al. (2000) of schizophrenia scoring systems. The study uncovered that studies with unpublished scales were more likely to support a treatment over control.

To address the issue of unvalidated measurement tools, an international Delphi study was held from 2006 to 2007 with 43 health status measurement experts, collectively known as the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) group. The COSMIN group worked to establish a consensus on the measurement properties to include when validating Health-Related Patient-Reported Outcomes (HR-PROs) and their definitions, as well as standards and design requirements for the evaluation of the defined measurement properties.

The study resulted in the isolation of three quality domains: reliability, validity, and responsiveness. Within each quality domain, there were one or more measurement properties (Fig. 1; Mokkink et al., 2006; Mokkink et al., 2010a; Mokkink et al., 2010b). While the study was designed specifically for HR-PROs such as quality of life measures, the key principles, quality domains, and definitions are applicable to validate disease severity tools. However, some of the described measurement properties in the COSMIN study, for example cross cultural validity, are not applicable to disease severity tools and thus have been excluded from our discussion.

Quality domains

Reliability

Reliability ensures that the instrument is free from measurement error and provides insight in the inherent noise or variability of the score. The reliability quality domain contains three measurement properties: internal consistency (degree of association between the items), reliability (determined via consistency in score values between observers [inter-rater] and a single observer with sufficient delay between two scorings [intra-rater]), and measurement error (changes in score that are not reflective of true changes in the intended construct [e.g., Standard Error of Measurement or SEM]).

Validity

The validity quality domain also consists of three measurement properties. The first, content validity, determines whether the content of the instrument reflects the intended construct and includes the subjective measure of face validity. For pemphigus, a key question that concerns content validity is whether added weighting for site increases the validity or if lesion count alone is sufficient as site distribution plays a role in severity. The second measurement property is construct validity, which aims to conclude that the instrument results represent the intended construct by assessing internal relationships and relationships to other instruments for the same construct. It can also identify discrepancies between pertinent groups. Finally, the criterion validity measures accuracy by determining how much the outcome measure reflects a golden standard to ensure minimal systematic and random bias.

Responsiveness

The domain of responsiveness has only one eponymous measurement property to illustrate the capability of the instrument and perceive change over time. It aims to assess how well the outcome measure detects true changes in the disease state rather than measurement error. It is otherwise referenced as sensitivity to change

Table 1
Summary of clinical, histologic, and immunologic findings in pemphigus

	Epidemiology	Clinical Features	Histopathology	Direct immune-fluorescence	Indirect immune-fluorescence	ELISA	Variants
PV	<ul style="list-style-type: none"> • Most common (in most populations) • Mostly middle-aged; 50-60 years old, men and women affected equally 	<ul style="list-style-type: none"> • At presentation: mucosal erosions (oropharyngeal and/or genital) • Mucosal lesions → Pain when chewing and swallowing → poor alimentation, weight loss, and malnutrition • Flaccid blisters on normal-looking or erythematous skin; palms and soles usually spared • Pruritus often absent • Nikolsky sign can be provoked 	Suprabasilar split with acantholysis	Intercellular IgG deposition	Intercellular IgG deposition. Preferred substrate is monkey esophagus	<ul style="list-style-type: none"> • Dsg 3 auto-antibodies • Dsg 1 and Dsg 3 auto-antibodies 	<ul style="list-style-type: none"> • Pemphigus vegetans • Pemphigus herpetiformis
PF	<ul style="list-style-type: none"> • Second most common form • PV incidence > PF incidence except in vicinities with endemic form • Mostly middle-aged: 50-60 years old, men and women affected equally • Endemic form mainly in children and young adults 	<ul style="list-style-type: none"> • At presentation: superficial blisters + erosions on trunk and extremities • Common for blisters to rupture before presentation, thus examination reveals superficial crusting and erosions or erythematous patches • Usually seborrheic distribution • Nikolsky sign can be provoked • No mucosal involvement thus systemic symptoms absent 	Subcorneal split with acantholysis	Intercellular IgG deposition	Intercellular IgG deposition Preferred substrate: normal human skin or Monkey esophagus	Dsg 1 autoantibodies	<ul style="list-style-type: none"> • FS (believed to have environmental trigger) • Pemphigus erythematous • Pemphigus herpetiformis • Pemphigus vegetans
PNP	<ul style="list-style-type: none"> • Any age though mostly adults 	<ul style="list-style-type: none"> • In setting of malignancy • Extensive mucositis • Polymorphic cutaneous lesions e.g. blisters, erosions, lichenoid lesions which may resemble other autoimmune blistering diseases • Bronchiolitis obliterans 	Intraepidermal clefting with acantholysis Dense lichenoid infiltrate + interface dermatitis + necrotic keratinocytes	Intercellular and/or basement membrane zone deposition of C3 and/or IgG	IgG intercellular deposition Preferred substrate: rat bladder	<ul style="list-style-type: none"> • Dsg 1 and Dsg 3 auto-antibodies • Auto-antibodies to plakins (e.g., envoplakin and periplakin) 	
IgA pemphigus-subcorneal pustular dermatosis	Any age	<ul style="list-style-type: none"> • Grouped vesicles or pustules • Erythematous plaques with crusts • Annular, circinate, or herpetiform morphology • Trunk and proximal extremities most commonly involved • Mucosa usually spared 	Subcorneal clefting and pustules + nominal acantholysis Mixed dermal infiltrate	Intercellular IgA deposition	Negative in 50% Intercellular IgA deposition Preferred substrate: monkey esophagus	<ul style="list-style-type: none"> • Desmocollin 1 auto-antibodies • Target antigens likely non-desmosomal 	<ul style="list-style-type: none"> • IgA/IgG subtype (demonstrates intercellular deposition of both IgG and IgA); atypical clinical and histologic manifestations due to heterogeneity of auto antigens (desmocollins, Dsg 1 and Dsg 3), associated with internal malignancies (lung cancer)
IgA pemphigus – intraepidermal neutrophilic dermatosis	Any age	<ul style="list-style-type: none"> • Grouped vesicles or pustules • Erythematous plaques with crusts • Annular, circinate, or herpetiform morphology • Trunk and proximal extremities most commonly involved Mucosa usually spared 	Intraepidermal pustules + nominal acantholysis Mixed dermal infiltrate	Intercellular IgA deposition	Intercellular IgA deposition Preferred substrate: Monkey esophagus	<ul style="list-style-type: none"> • Desmocollin 1 auto-antibodies • Desmoglein 1 and Desmoglein 3 auto-antibodies 	<ul style="list-style-type: none"> • IgA/IgG subtype

Dsg = desmoglein; ELISA = enzyme-linked immunosorbent assay; FS = fogo selvage; IgA = immunoglobulin A; IgG = immunoglobulin G; PV = pemphigus vulgaris; PF = pemphigus foliaceus; PNP = paraneoplastic pemphigus.

Sources: Evangelista et al., 2015; Hertl et al., 2006; Hertl and Sitaru, 2015; Mihai and Sitaru, 2007; Oiso et al., 2002; Santoro et al., 2013; Tsuruta et al., 2011; Yeh et al., 2003.

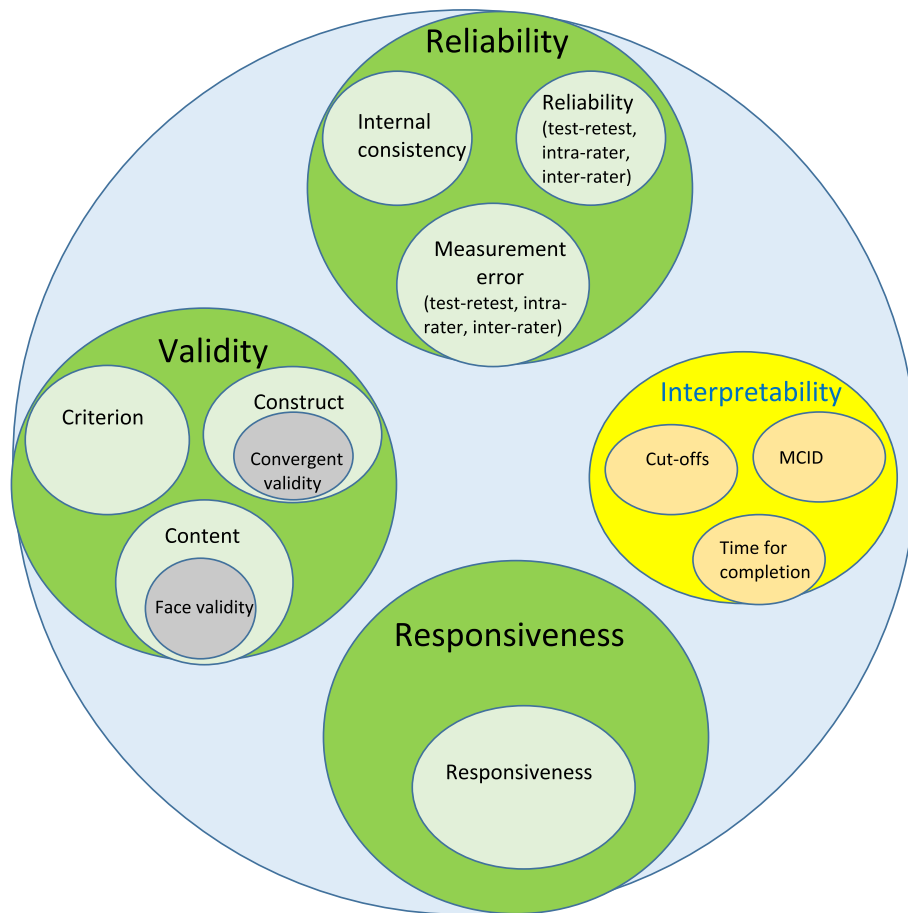


Fig. 1. Domains to decipher the quality of a disease severity outcome measure (Mokkink et al., 2010a).

or discriminant validity and has significant ramifications on conclusions that are drawn with regard to the efficacy of therapies in clinical studies.

Criteria in addition to those described by the COSMIN group are feasibility and cutoffs. Feasibility refers to the time taken to complete the scoring and the resources and/or costs needed to implement the instrument, which may have considerable implications on the outcome measure's practicality. Disease severity cutoffs allow for a differentiation between mild, moderate, and severe disease status and may have important implications in clinical practice when identifying appropriate therapies and within clinical studies when drawing meaningful comparisons.

Table 2 illustrates the degree of validation of commonly-used dermatological scoring systems by key measurement properties.

Previous responsiveness studies for dermatological scoring systems

Responsiveness is essentially proving the validity of the change score and whether the direction and magnitude correlate to the expected results. To determine responsiveness, a longitudinal study design is required with at least two measurements administered and some validated external reference to illustrate whether the patient's condition is improving, remaining stable, or deteriorating. This is significant in order to conclude whether the patient was truly stable, the instrument is not responsive, or whether the results are due to poor comparator instrument quality.

The COSMIN study divided the design requirements in cases with a golden standard (items 1-7, and 15-18) and cases without golden

standard (items 1-14; Fig. 2). In golden standard cases with dichotomous results, the preferred method to deduce responsiveness is a correlation between change scores using Receiver Operator Curve or sensitivity and specificity if the study instrument scores are also dichotomous (Mokkink et al., 2010a).

The COSMIN group also emphasize the need for specific hypotheses to be formulated beforehand and stated in the method section when a golden standard is not available. The hypothesis should predict direction (positive or negative) and magnitude (absolute or relative). A detailed hypothesis avoids bias by avoiding retrospective analysis and alternative explanations for weak correlations where the conclusion should be that the instrument is not responsive.

Table 3 outlines previous responsiveness studies on commonly-used dermatological scoring systems and the methods employed to establish responsiveness.

Minimal clinically-important difference

It is important to distinguish minimal clinically-important difference (MCID) from responsiveness. MCID is the smallest change in an instrument score that correlates to a meaningful clinical difference. It is an inappropriate measure of responsiveness because it is simply about the interpretation of a change score as opposed to validity (Mokkink et al., 2010a). MCID usually revolves around patient perception although variations can include MCID obtained through a clinical report, change in clinical parameter, and effect size. For disease severity outcome measures, it is standard for MCID to be derived from some form of physician global assessment or a related tool. There are up to nine methods to identify MCID, some that anchor

Table 2
Previous validation studies on commonly-used dermatological scoring systems

Instrument	Inter-rater reliability	Intra-rater reliability	Responsiveness	MCID	Content Validity	Construct Validity	Feasibility (mean to 1 decimal place)	Severity Cut-offs
PDAI	1	1			2	1	2.9mins ¹	Conflicting ^{3,4}
ABSIS	1	1			5	1	1.9mins ¹	³
PASI	Moderate, limited particularly in low-mod values ^{6, 27}	Limited ⁷	Acceptable ⁷		Used as gold standard ⁷			
EASI	Acceptable ⁹	Acceptable ⁹	25, 10	11	8	9,10	2.2mins ²⁸	
Sartorius	Low ¹²				12			
EBDASI	13	13	14	14	13	13	7.9mins ¹³	¹⁴
BPDAI	15	15	15, 17, 18	15	16	17, 18		
CLASI	20	20	19	19	20	19	5.3mins ²⁰	²¹
CDASI	22	22	22	23	22	22	4.8mins ²²	²³
SCORAD	24, 26	Acceptable ²⁴	25, 10	25	24	10, 26	<10mins ²⁴ 2.3mins ²⁰ 20 ²⁸	

Key:

Study Done
No previous studies
Equivocal

ABSIS = Autoimmune Bullous Skin Disorder Intensity Score; BPDAI = Bullous Pemphigoid Disease Area Index; CDASI = Cutaneous Dermatomyositis Disease Area and Severity Index; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; EBDASI = Epidermolysis Bullosa Disease Activity and Scarring Index; EASI = Eczema Activity and Scarring Index; MCID = minimal clinically-important difference; PASI = Psoriasis Area and Severity Index; PDAI = Pemphigus Disease Area Index; SCORAD = Scoring Atopic Dermatitis.

¹Rahbar et al., 2014; ²Murrell et al., 2008; ³Boulard et al., 2016; ⁴Shimizu et al., 2014; ⁵Pfütze et al., 2007; ⁶Gourraud et al., 2012; ⁷Puzenat et al., 2010; ⁸Tofte et al., 1998; ⁹Hanifin et al., 2001; ¹⁰Breuer et al., 2004; ¹¹Schram et al., 2012; ¹²Sartorius et al., 2010; ¹³Loh et al., 2014; ¹⁴Jain et al., 2016; ¹⁵Wijayanti et al., 2014; ¹⁶Murrell et al., 2012; ¹⁷Patsatsi et al., 2012; ¹⁸Lévy-Sitbon et al., 2014; ¹⁹Bonilla-Martinez et al., 2008; ²⁰Albrecht et al., 2005; ²¹Klein et al., 2011; ²²Goshi et al., 2012; ²³Anyanwu et al., 2015; ²⁴Stalder et al., 1993; ²⁵Schram et al., 2012; ²⁶Angelova-Fischer et al., 2005; ²⁷Langley and Ellis, 2004; ²⁸Zhao et al., 2015.

Box I. Responsiveness				
Design requirements		yes	no	?
1	Was the percentage of missing items given?	<input type="checkbox"/>	<input type="checkbox"/>	
2	Was there a description of how missing items were handled?	<input type="checkbox"/>	<input type="checkbox"/>	
3	Was the sample size included in the analysis adequate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	Was a longitudinal design with at least two measurement used?	<input type="checkbox"/>	<input type="checkbox"/>	
5	Was the time interval stated?	<input type="checkbox"/>	<input type="checkbox"/>	
6	If anything occurred in the interim period (e.g. intervention, other relevant events), was it adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
7	Was a proportion of the patients changed (i.e. improvement or deterioration)?	<input type="checkbox"/>	<input type="checkbox"/>	
Design requirements for hypotheses testing		yes	no	?
For constructs for which a gold standard was not available:				
8	Were hypotheses about changes in scores formulated a priori (i.e. before data collection)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> *
		yes	no	NA
9	Was the expected <i>direction</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Were the expected absolute or relative <i>magnitude</i> of correlations or mean differences of the change scores of HR-PRO instruments included in these hypotheses?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Was an adequate description provided of the comparator instrument(s)?	<input type="checkbox"/>	<input type="checkbox"/>	
12	Were the measurement properties of the comparator instrument(s) adequately described?	<input type="checkbox"/>	<input type="checkbox"/>	
13	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
Statistical methods		yes	no	NA
14	Were design and statistical methods adequate for the hypotheses to be tested?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Design requirement for comparison to a gold standard		yes	no	?
For constructs for which a gold standard was available:				
15	Can the criterion for change be considered as a reasonable gold standard?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Were there any important flaws in the design or methods of the study?	<input type="checkbox"/>	<input type="checkbox"/>	
Statistical methods		yes	no	NA
17	for continuous scores: Were correlations between change scores, or the area under the Receiver Operator Curve (ROC) curve calculated?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	for dichotomous scales: Were sensitivity and specificity (changed versus not changed) determined?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 2. Responsiveness checklist (with permission; open access; Mokkink et al., 2010a).

solely on external criteria while others utilize internal values. The results can vary enormously based on the method used (Cook, 2008). Table 4 illustrates previous studies and methods utilized to establish MCID.

Scoring systems for pemphigus

Pemphigus Disease Area Index

PDAI was published by the International Pemphigus Definitions Group (IDPG) in 2008 (Fig. 3). The IDPG held five consensus meetings

between 2006 and 2008 to establish consensus definitions and develop a scoring system for pemphigus that was molded from the Cutaneous Lupus Erythematosus Disease Area and Severity Index. The IDPG panel consisted of experts on autoimmune blistering diseases, led by Victoria Werth and Dedee Murrell (Murrell et al., 2008).

PDAI scores can range from 0 to 263, comprised of 250 points for disease activity (120 for skin, 10 for scalp, and 12 for mucosa) and 13 points for damage. For activity, the size and number of lesions in each area play a role in the calculation of points assigned. The damage score reflects post-inflammatory hyperpigmentation. A considerable

Table 3
Previous responsiveness studies in dermatology

Instrument	Authors	Year	Sample Size	Method for Responsiveness
EBDASI	Jain et al.	2016	36	Utilized distribution and anchor-based methods. Distribution-based: Mean change of scores, standardized response mean, and standardized effect size utilized to illustrate the magnitude of change in activity and damage scores. Anchor-based: Pearson's correlation coefficient utilized to determine the degree of correlation between change in EBDASI score and Likert scale of change.
BPDAI	Wijayanti et al.	2016	32	Physician subjective assessment: improved, stable, deteriorated. Paired t test with BPDAI to note statistical significance. To be responsive → statistically significant between improved and deteriorated, not statistically significant when stable.
BPDAI	Patsatsi et al.	2012	39	Correlated BPDAI to BP180 titers at baseline, 3-month, and 6-month interval using Spearman's rho correlation.
CLASI	Bonilla-Martinez et al.	2008	8	Utilized: correlations, linear regressions, and Wilcoxon rank sum and 1-sided signed rank exact tests The difference between baseline score and day 56 (change scores) were recorded for CLASI activity and damage, and each correlated via Pearson correlation coefficient with the change score of 1) Physicians assessment of patient's global skin health 2) Patients self-assessment of global skin health 3) Pain 4) Itch
CDASI	Goreshi et al.	2012	35	Included the two consecutive visits with the greatest variance in PGA-activity for analysis. Responsiveness was measured via Standardized Response Mean (SRM), SRM = ratio of the mean differences (i.e., CDASI score before and after clinical change was noted) to the standard deviation of the differences
EASI	Breuer et al.	2004	Pimecrolimus (n = 129) Control (n = 66)	Following treatment with 1% pimecrolimus, EASI, IGA, SCORAD dropped significantly compared to the vehicle/control group as was depicted by t-test. Between-group comparisons established via Cochran-Mantel-Haenzel test.
EASI, SCORAD	Schram et al.	2012	143	Close correlation between pairs of EASI, IGA, and SCORAD depicted via Pearson test. Mean scores of EASI and SCORAD were correlated to mean scores of IGA and PGA within each treatment group per time point. Then ROC was utilized.

BPDAI = Bullous Pemphigoid Disease Area Index; CDASI = Cutaneous Dermatomyositis Disease Area and Severity Index; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; EBDASI = Epidermolysis Bullosa Disease Activity and Scarring Index; EASI = Eczema Acvitivity and Scarring Index; IGA = investigator global assessment; PGA = physician global assessment; ROC = receiver operating characteristic; SCORAD = Scoring Atopic Dermatitis.

Table 4
Previous MCID studies in dermatology

Instrument	Authors	Year	Sample Size	Method for MCID
EBDASI	Jain et al.	2016	36	Pearson correlation coefficient >0.3 between Likert scale and EBDASI, thus sufficient to determine MCID. - MCID derived via linear regression analysis setting a responder score of 3 on Likert Scale. - MCID also calculated via ROC curves with responder score of 3 on Likert scale. - To account for baseline severity MCID analysis was also conducted utilizing ROC on percentage change in activity scores (change in activity score divided by baseline activity score)
BPDAI	Wijayanti et al.	2016	32	Average signed change in BPDAI of responders (determined by physician subjective assessment: improved, deteriorated, stable). Confirmed via ROC at/around this cut-off value
CLASI	Bonilla-Martinez et al.	2008	8	Clinical cut points which represent minimal clinically meaningful change (responders) were determined - PGA-VAS: change of 2 points - Pain: change of 2 points - Itch: change of 2 points - Patients global skin health rating: change of 3 points For each measure, Wilcoxon rank sum tests were used to compare CLASI change in responders vs. non-responders
CDASI	Anyanwu et al.	2015	128	Utilized PGA-VAS with a clinical cut point of 2 for responders and less than 2 for non-responders. Used ROC curve to determine the change score which correlated with responders
EASI, SCORAD	Schram et al.	2012	Data from three randomized control studies on atopic eczema treatments n = 143	Responders were defined as in improvement or decline greater than or equal to 1 in PGA and IGA. ROC utilized. >0.7 = fair, >0.8 = good, >0.9 = excellent responsiveness

BPDAI = Bullous Pemphigoid Disease Area Index; CDASI = Cutaneous Dermatomyositis Disease Area and Severity Index; CLASI = Cutaneous Lupus Erythematosus Disease Area and Severity Index; EBDASI = Epidermolysis Bullosa Disease Activity and Scarring Index; EASI = Eczema Acvitivity and Scarring Index; IGA = investigator global assessment; MCID = minimal clinically-important difference; PGA = physician global assessment; ROC = receiver operating characteristic; SCORAD = Scoring Atopic Dermatitis; VAS = visual analogue scale.

Skin		Activity		Damage	
Anatomical Location	Erosion/Blisters or new erythema		Number lesions if ≤ 3	Post-inflammatory hyperpigmentation or erythema from resolving lesion	
	0 absent 1 1-3 lesions, up to one >2 cm in any diameter, none > 6 cm 2 2-3 lesions, at least two > 2 cm diameter, none > 6cm 3 >3 lesions, none > 6 cm diameter 5 >3 lesions, and/or at least one >6 cm 10 >3 lesions, and/or at least one lesion >16 cm diameter or entire area			0 absent 1 present	
Ears					
Nose					
Rest of the face					
Neck					
Chest					
Abdomen					
Back, buttocks					
Arms					
Hands					
Legs					
Feet					
Genitals					
Total skin	/120			/12	
Scalp					
Scalp	Erosion/Blisters or new erythema		Number lesions if ≤ 3	Post-inflammatory hyperpigmentation or erythema from resolving lesion	
	0 absent 1 in one quadrant 2 two quadrants 3 three quadrants 4 affects whole skull 10 at least one lesion > 6 cm			0 absent 1 present	
Total Scalp (0-10)	/10			/1	
Mucous membrane					
Anatomical Location	Erosion/Blisters		Number lesions if ≤ 3		
	0 absent 1 1 lesion 2 2-3 lesions 5 >3 lesions or 2 lesions >2 cm 10 entire area				
Eyes					
Nose					
Buccal mucosa					
Hard palate					
Soft palate					
Upper gingiva					
Lower gingiva					
Tongue					
Floor of mouth					
Labial bucosa					
Posterior pharynx					
Anogenital					
Total Mucosa	/120				
Total Activity Score:		<input type="text"/>		Total Damage Score	
		<input type="text"/>			
©2007 by the authors					

Fig. 3. Pemphigus Disease Area Index (with permission; license number 3921300270111; Rosenbach et al., 2009).

advantage of this scoring system is its sensitivity to small lesion numbers, which increases inter-rater reliability (Zhao and Murrell, 2015). Furthermore, PDAI does not take body surface area (BSA) or lesion type into account, which are both arduous to evaluate, cannot capture mild amounts of disease activity, and can potentially exacerbate small variations between raters.

Autoimmune Bullous Skin Disorder Intensity Score

ABSIS is a generic, AIBD outcome measure produced by the German Blistering Group (Pfütze et al., 2007). The scores can vary from 0 to 206, of which 150 points represent skin involvement, 11 points for oral involvement, and 45 points for subjective discomfort. ABSIS

PVAS = [T _s (N _s +D _s +S _s)] + [T _m (N _m +D _m)]			
T _s :	Type of skin lesion	1 point: when there is blister or bulla, 0.5 point: crusted lesions, 0 point: when there are only pigmentation change	max:1
N _s :	Number of skin lesion	2 point: more than 20 bulla (average diameter size of 1 cm), 1 point: twenty or less blisters (≤20)	max:2
D _s :	Distribution of skin lesion. One point for each anatomical area	1 point: scalp, 1 point: face, 1 point: neck , 1 point: trunk, 1 point: each limb(0-4 point for no to four extremities involvement)	max:8
S _s :	Nikolsky's sign: pressure induced blister	1 point: On the unaffected skin, 0.5: around the lesions, 0: none.	max:1
T _m :	Type of mucosal lesion	1 point: when there is blister or bulla, 0.5 point: Ulceration, 0 point: none.	max:1
N _m :	Number of mucosal lesion	2 point: more than 2 bulla (>2), 1 point: one or two blister	max:2
D _s :	Distribution of mucosal lesion. One point for each anatomical area	1 point: oral cavity and/ or pharynx, 1 point: eyes , 1 point: upper airways, 1 point: anus , 1 point: genital area	max:5
Maximum skin score: 11, maximum mucosa score: 7, maximum total score [T _s (N _s +D _s +S _s)] + [T _m (N _m +D _m): 18			

Fig. 4. Pemphigus Vulgaris Activity Score (with permission; open access; Chams-Davatchi et al., 2013)

uses the rule of nines and rule of palms to establish BSA, and BSA and lesion type are weighting factors.

Pemphigus Vulgaris Activity Score

PVAS was created by Chams-Davatchi et al. (2013) and produces scores between 0 and 18 with 11 points for cutaneous and 7 points for mucosal involvement. Lesion type, lesion number, and distribution all contribute to the score (Fig. 4). Compared with PDAI, PVAS places less emphasis on the head and greater emphasis on the

limbs. PVAS also takes into account Nikolsky's sign and thus is more susceptible to variability based on the expertise of the rater.

Harman's scoring system

Harman's scoring system was created by Harman et al. (2001) in the United Kingdom and system scores are based on the number of skin and oral erosions (Fig. 5). Harman et al. related these scores with anti-desmoglein (Dsg) 1 and anti-Dsg3 ELISA and noted that there was a correlation between severity and Dsg antibody levels. The

<i>Oral Score</i>	<i>Skin Score</i>
0 = quiescent = no erosions	0 = quiescent = no lesions
1 = minor activity = 1-3 erosions	1 = minor activity = 1-5 discreet lesions
1 = moderate activity = 3-10 erosions or generalized desquamative gingivitis	2 = moderate activity = 5-20 discreet lesions
2 = severe = >10 erosions or extensive confluent erosions, or generalized desquamative gingivitis with discreet erosions at other oral sites	3 = severe = >20 discreet lesions or extensive confluent areas of eroded skin
Total Score = Oral score + Skin score	

Fig. 5. Harman's severity scoring system (Harman et al., 2001).

Table 5
Validation studies to date on commonly-used pemphigus measurement instruments

Instrument	Inter-rater reliability (95% CI)	Intra-rater reliability (95% CI)	Content Validity	Construct Validity (95% CI)	Responsiveness	MCI	Severity Cut-offs	Feasibility (mean [SD])
PDAI	0.76 (0.61-0.91) ¹ 0.98 (0.97-0.98) ²	0.98 (0.96-1.0) ¹	³	PGA q= 0.60 (0.49-0.71) ¹ . Anti-Dsg1 titers, ICC 0.67 (p<0.001) ² Anti-Dsg3 titers, ICC 0.35 (p=0.001) ²			Based on 25 th and 75 th centiles: 15/45 ⁶ Using physician's subjective impression of disease severity: 9/25 ⁸ Severity cut-offs calculated predict only anti-Dsg1 and anti-Dsg3 titers of 20 and 100 u/mL NOT necessarily disease severity ² .	20 seconds to 5.8 minutes (2.9 [1.3] minutes) ²
ABSIS (for pemphigus)	0.77 (0.63-0.91) ¹ 0.97 (0.96-0.98) ²	0.80 (0.65-0.96) ¹	⁴	PGA q= 0.43 (0.30-0.55) ¹ . Anti-Dsg1 titers, ICC 0.33(p=0.002) ² Anti-Dsg3 titers, ICC 0.38 (p<0.001) ²	Pfütze et al. stated a decrease in ABSIS correlated with a gradual decline in anti-Dsg1, and anti-Dsg3 titers, however, presented no correlation coefficients ⁴ .		Based on 25 th and 75 th centiles of scores: 17/53 ⁶ Severity cut-offs calculated predict only anti-Dsg1 and anti-Dsg3 titers of 20 and 100 u/mL NOT necessarily disease severity ² .	25 seconds to 5,6 minutes, (1.9 [1.1] minutes) ²
PVAS	0.93 (0.90-0.95) ²		⁵	PGA q= 0.751 ³ Anti-Dsg1 titers, ICC 0.52 (P < .001) ² Anti-Dsg3 titers, ICC 0.35 (p=0.001) ²			Severity cut-offs calculated predict only anti-Dsg1 and anti-Dsg3 titers of 20 and 100 u/mL NOT necessarily disease severity ² .	15 seconds to 3,9 minutes (1.1 [0.7] minutes) ²
Harman's			⁷					

¹Rosenbach et al (2009), ²Rahbar et al (2014), ³Murrell et al (2008), ⁴Pfütze et al (2007), ⁵Chams-Davatchi et al (2013), ⁶Boulard et al (2016), ⁷Harman et al (2001), ⁸Shimizu et al (2014).

Key:
Study Done
No previous studies
Equivocal

ABSIS = Autoimmune Bullous Skin Disorder Intensity Score; CI = confidence interval; Dsg = desmoglein; ICC = xxx; PDAI = Pemphigus Disease Area Index; PGA = physician global assessment; PVAS = Pemphigus Vulgaris Activity Score; SD = standard deviation;
¹Rosenbach et al., 2009; ²Rahbar et al., 2014; ³Murrell et al., 2008; ⁴Pfütze et al., 2007; ⁵Chams-Davatchi et al., 2013; ⁶Boulard et al., 2016; ⁷Harman et al., 2001; ⁸Shimizu et al., 2014.

use of this scoring system is limited by the lack of validation studies and the poor sensitivity to BSA involvement and anatomical distribution as scores are awarded irrespective of site and size.

We could not identify any studies to illustrate the reliability, validity, responsiveness, feasibility, or severity cutoffs for the Harman grading system.

Validation studies on pemphigus scoring systems

Rosenbach et al. conducted a study at the University of Pennsylvania to demonstrate the inter- and intra-rater reliability and convergent validity of PDAI and ABSIS. The study was conducted with the assistance of ten dermatologists who specialize in AIBD to score 15 patients with pemphigus using the PDAI, ABSIS, and PGA scoring systems. The study was limited because the majority of patients had stable disease. Nonetheless, the study demonstrated that PDAI has strong intra- and inter-reliability with an intra-class correlation coefficient (ICC) of 0.98 (95% confidence interval [CI]: 0.96–1.0) and 0.76 (95% CI: 0.61–0.91), respectively. ABSIS intra- and inter-rater reliability had an ICC of 0.80 (95% CI: 0.65–0.96) and 0.77 (95% CI: 0.63–0.91), respectively. Thus, inter-rater reliability for PDAI and ABSIS was almost indistinguishable. However, it is important to emphasize that despite of this, capturing low disease activity (both Dsg and ELISA negative) was different with PDAI. The difference in inter-rater reliability between PDAI and ABSIS became more apparent when only the objective skin activity scores were compared: ICC of 0.86 (95% CI: 0.76–0.95) for PDAI versus 0.39 (95% CI: 0.17–0.60) for ABSIS.

This study also illustrated good convergent validity between PDAI and PGA with a Spearman's rho correlation of 0.6 (95% CI: 0.49–0.71) compared with the poorer convergent validity of ABSIS of 0.43 (95% CI: 0.30–0.55; Rosenbach et al., 2009). Independent of this study, Chams-Davatchi et al. (2013) asked five experts to score 50 patients with PV to illustrate that PVAS has a superior convergent validity of 0.75 with PGA.

Rahbar et al. (2014) conducted a study independent of the IDPG and German blistering group and produced unbiased results of ABSIS and PDAI in comparison with their PVAS scoring system. The study had a sizeable sample size of 100 patients with active lesions. The study produced higher values for inter-rater reliabilities, which may be due to the increased sample size in the study and perhaps demonstrated a learning curve by dermatologists in the application of scoring systems over the preceding 5 years. The results showed an ICC of 0.98 (95% CI: 0.97–0.98), 0.97 (95% CI: 0.96–0.98), and 0.93 (95% CI: 0.9–0.95) for PDAI, ABSIS, and PVAS inter-rater reliability, respectively, and thus illustrated that PDAI and ABSIS are the most reproducible with almost identical ICC rates. However, this study also considered ICC rates by range and the lower range (anti-Dsg1 and anti-Dsg3 negative; n = 10) was only statistically significant for PDAI with an ICC of 0.96 (95% CI: 0.93–0.98). This illustrates that PDAI is more reliable for low disease activity than PVAS and ABSIS.

Convergent validity against anti-Dsg1 titers was the highest for PDAI, producing a Spearman's rho correlation of 0.67 ($p < 0.001$), 0.33 ($p = 0.002$), and 0.52 ($p < 0.01$) for PDAI, ABSIS, and PVAS, respectively. Convergent validity as determined by anti-Dsg3 titers was poor for all three instruments with an ICC of 0.35 ($p = 0.001$), 0.33 ($p = 0.002$), and 0.35 ($p = 0.001$) for PDAI, ABSIS, and PVAS, respectively (Rahbar et al., 2014).

Two studies have investigated severity cutoffs for PDAI to divide patients based on disease severity into mild, moderate, and severe categories. The first study was conducted in Japan and utilized the physician's subjective impression of the disease state (mild, moderate, severe) and correlated this with the PDAI score to establish cutoffs using the Youden Index. The values obtained were identified as mild (0–8), moderate (9–24), or severe (≥ 25 ; Shimizu et al., 2014). In contrast, an independent French study by Boulard et al. (2016)

calculated severity cutoffs by identifying scores that correlated with the 25th and 75th percentiles of the scores, which resulted in significantly higher cutoffs. The results obtained were identified as mild (0–14), moderate (15–44), or severe (≥ 45). In their discussion, the researchers justify this discrepancy by sample selection, stating that Shimizu et al. recruited newly-diagnosed and previously-treated patients with minimal severe cases while they enlisted newly-diagnosed, non-treated cases and thus with greater severity (Boulard et al., 2016).

It is also possible that there are population-based variabilities or that the differing methods that were utilized make one study more vulnerable to become skewed. No reviews or studies exist that outline the best method to calculate severity cutoffs and it seems the method can have significant implications on the results, particularly if the population is concentrated in a particular subgroup. Boulard et al. also obtained cutoffs of 17 and 53 to distinguish the three groups for ABSIS using the same method.

The mean time to complete PDAI, ABSIS and PVAS is 2.9 minutes (standard deviation [SD] 1.3 min), 1.9 minutes (SD 1.1 min) and 1.1 minutes (SD 0.7 min), respectively, which makes PVAS the fastest instrument to complete (Rahbar et al., 2014). To date, there are no responsiveness studies on any of the pemphigus scoring systems (Table 5).

Conclusion

Despite the significant morbidity associated with pemphigus, there is a shortage of multicenter control studies to facilitate evidence-based practice due to the rarity of the disease and the inability to objectively compare therapeutic outcomes. PDAI and ABSIS are promising scoring systems that have proven to be valid and reliable; however, to undergo a complete validation, responsiveness must be assessed. MCID and reaffirmation of cutoff values would also provide significant information that may improve the utility of the instrument in clinical practice.

References

- Albrecht J, Werth VP. Development of the CLASI as an outcome instrument for cutaneous lupus erythematosus. *Dermatol Ther* 2007;20:93–101.
- Albrecht J, Taylor L, Berlin JA, Dulay S, Ang G, Fakharzadeh S, et al. The CLASI (Cutaneous LE Disease Area and Severity Index): an outcome instrument for cutaneous lupus erythematosus. *J Invest Dermatol* 2005;125:889–94.
- Angelova-Fischer I, Bauer A, Hipler UC, Petrov I, Kazandjieva J, Bruckner T, et al. The objective severity assessment of atopic dermatitis (OSAAD) score: Validity, reliability and sensitivity in adult patients with atopic dermatitis. *Br J Dermatol* 2005;153:767–73.
- Anyanwu CO, Fiorentino DF, Chung L, Dzuong C, Wang Y, Okawa J, et al. Validation of the Cutaneous Dermatomyositis Disease Area and Severity Index: Characterizing disease severity and assessing responsiveness to clinical change. *Br J Dermatol* 2015;173:969–74.
- 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;144:173–80.
- Boulard C, Duvert Lehenbre S, Picard-Dahan C, Kern JS, Zambruno G, Feliciani C, et al. Calculation of cut-off values based on the Autoimmune Bullous Skin Disorder Intensity Score (ABSIS) and Pemphigus Disease Area Index (PDAI) pemphigus scoring systems for defining moderate, significant and extensive types of pemphigus. *Br J Dermatol* 2016;175:142–9.
- Breuer K, Braeutigam M, Kapp A, Werfel T. Influence of pimecrolimus cream 1% on different morphological signs of eczema in infants with atopic dermatitis. *Dermatology* 2004;209:314–20.
- Chams-Davatchi C, Rahbar Z, Daneshpazhooch M, Mortazavizadeh SM, Akhyani M, Esmaili N, et al. Pemphigus vulgaris activity score and assessment of convergent validity. *Acta Med Iran* 2013;51:224–30.
- Cook CE. Clinimetrics corner: The Minimal Clinically Important Change Score (MCID): A necessary pretense. *J Man Manip Ther* 2008;16:E82–3.
- Evangelista F, Culton DA, Diaz LA. Desmosomal proteins as autoantigens in pemphigus. In: Murrell DF, editor. *Blistering diseases: Clinical features, pathogenesis, treatment*. Sydney: Springer; 2015. p. 55–65.
- Gaines E, Werth VP. Development of outcome measures for autoimmune dermatoses. *Arch Dermatol Res* 2008;300:3–9.

- Goreski 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;132:1117–24.
- Gourraud PA, Le Gall C, Puzenat E, Aubin F, Ortonne JP, Paul CF. Why statistics matter: Limited inter-rater agreement prevents using the psoriasis area and severity index as a unique determinant of therapeutic decision in psoriasis. *J Invest Dermatol* 2012;132:2171–5.
- Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): Assessment of reliability in atopic dermatitis. *EASI Evaluator Group. Exp Dermatol* 2001;10:11–8.
- Harman KE, Seed PT, Gratian MJ, Bhogal BS, Challacombe SJ, Black MM. The severity of cutaneous and oral pemphigus is related to desmoglein 1 and 3 antibody levels. *Br J Dermatol* 2001;144:775–80.
- Hertl M, Eming R, Veldman C. T cell control in autoimmune bullous skin disorders. *J Clin Invest* 2006;116:1159–66.
- Hertl M, Sitaru C. Pathogenesis, clinical manifestations, and diagnosis of pemphigus. UpToDate. Available from: <http://www.uptodate.com/contents/pathogenesis-clinical-manifestations-and-diagnosis-of-pemphigus>; 2015.
- Huang YH, Kuo CF, Chen YH, Yang YW. Incidence, mortality, and causes of death of patients with pemphigus in Taiwan: A nationwide population-based study. *J Invest Dermatol* 2012;132:92–7.
- Jain SV, Harris AG, Su JC, Orchard D, Warren LJ, McManus H, et al. The Epidermolysis Bullosa Disease Activity and Scarring Index (EBDASI): Grading disease severity and assessing responsiveness to clinical change in epidermolysis bullosa. *J Eur Acad Dermatol Venereol* 2016 [Epub ahead of print].
- 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;147:203–8.
- Langan SM, Smeeth L, Hubbard R, Fleming KM, Smith CJ, West J. Bullous pemphigoid and pemphigus vulgaris—incidence and mortality in the UK: Population based cohort study. *BMJ* 2008;337:a180.
- Langley RG, Ellis CN. Evaluating psoriasis with Psoriasis Area and Severity Index, Psoriasis Global Assessment, and Lattice System Physician's Global Assessment. *J Am Acad Dermatol* 2004;51:563–9.
- Lévy-Sitbon C, Barbe C, Plee J, Goeldel AL, Antonicelli F, Reguiai Z, et al. Assessment of bullous pemphigoid disease area index during treatment: A prospective study of 30 patients. *Dermatology* 2014;229:116–22.
- Loh CC, 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.
- Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: A major source of bias in randomised controlled trials of treatments for schizophrenia. *Br J Psychiatry* 2000;176:249–52.
- Martin L, Murrell DF. Measuring the immeasurable: A systematic review of outcome measures in pemphigus. *Australas J Dermatol* 2006;47:A32–3.
- Mihai S, Sitaru C. Immunopathology and molecular diagnosis of autoimmune bullous diseases. *J Cell Mol Med* 2007;11:462–81.
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: Consensus-based Standards for the selection of health Measurement Instruments. *BMC Med Res Methodol* 2006;6:2.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res* 2010;19:539–49.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Murrell DF, Dick S, Ahmed AR, Amagai M, Barnadas MA, Borradori L, et al. Consensus statement on definitions of disease, end points, and therapeutic response for pemphigus. *J Am Acad Dermatol* 2008;58:1043–6.
- 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;66:479–85.
- Oiso N, Yamashita C, Yoshioka K, Amagai M, Komai A, Nagata Y, et al. IgG/IgA pemphigus with IgG and IgA antidesmoglein 1 antibodies detected by enzyme-linked immunosorbent assay. *Br J Dermatol* 2002;147:1012–7.
- Patsatsi A, Koletsa T, Sotiriadis D, Kartsios C, Papaconstantinou A, Kostopoulos I. Pyoderma gangrenosum in a patient with primary cutaneous peripheral T-cell lymphoma, not otherwise specified. *Eur J Dermatol* 2012;22:792–3.
- 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.
- Puzenat E, Bronsard V, Prey S, Gourraud PA, Aractingi S, Bagot M, et al. What are the best outcome measures for assessing plaque psoriasis severity? A systematic review of the literature. *J Eur Acad Dermatol Venereol* 2010;24(Suppl. 2):10–6.
- Rahbar Z, Daneshpazhooh M, Mirshams-Shahshahani M, Esmaili N, Heidari K, Aghazadeh N, et al. Pemphigus disease activity measurements: Pemphigus disease area index, autoimmune bullous skin disorder intensity score, and pemphigus vulgaris activity score. *JAMA Dermatol* 2014;150:266–72.
- Rosenbach M, Murrell DF, Bystryjn JC, Dulay S, Dick S, Fakhrazadeh S, et al. Reliability and convergent validity of two outcome instruments for pemphigus. *J Invest Dermatol* 2009;129:2404–10.
- Santoro FA, Stoopler ET, Werth VP. Pemphigus. *Dent Clin N Am* 2013;57:597–610.
- Sartorius K, Emtestam L, Lapins J, Johansson O. Cutaneous PGP 9.5 distribution patterns in hidradenitis suppurativa. *Arch Dermatol Res* 2010;302:461–8.
- Schmidt E, Borradori L, Joly P. Epidemiology of Autoimmune Bullous Diseases. In: Murrell DF, editor. *Blistering diseases: Clinical features, pathogenesis, treatment*. Sydney: Springer; 2015. p. 251–63.
- Schram ME, Spuls PI, Leeflang MM, Lindeboom R, Bos JD, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema: Responsiveness and minimal clinically important difference. *Allergy* 2012;67:99–106.
- Shimizu T, Takebayashi T, Sato Y, Niizeki H, Aoyama Y, Kitajima Y, et al. Grading criteria for disease severity by pemphigus disease area index. *J Dermatol* 2014;41:969–73.
- Stalder JF, Dutatre H, Laruche G, Litoux P. Photoprotection in children. *Ann Dermatol Venereol* 1993;120:485–8.
- Streiner DL, Norman GR, Cairney J. *Health measurement scales: A practical guide to their development and use*. 4th ed. Oxford: Oxford; 2008.
- Tofte SJ, Graeber M, Cherill R, Omoto M, Thurston M, Hanifin JM. Eczema area and severity index (EASI): a new tool to evaluate atopic dermatitis. *J Eur Acad Dermatol Venereol* 1998;11:S197.
- Tsuruta D, Ishii N, Hamada T, Ohyama B, Fukuda S, Koga H, et al. IgA pemphigus. *Clin Dermatol* 2011;29:437–42.
- Weisman S, Pollack CR, Gottschalk RW. Psoriasis disease severity measures: Comparing efficacy of treatments for severe psoriasis. *J Dermatolog Treat* 2003;14:158–65.
- Wijayanti A, Zhao CY, Boettiger D, Ishii N, Hashimoto T, Murrell DF. Reliability, validity, and responsiveness of Bullous Pemphigoid Disease Area Index (BPDAI). Paper presented at the Cutaneous Biologic Meeting; Australasian Society of Dermatological Research, North Stradbroke Island; 2014. <http://cutaneous-2014.pasnevents.com.au/days/2014-09-23/abstract/18743>.
- Wijayanti A, Zhao CY, Boettiger D, Chiang YZ, Ishii N, Hashimoto T, et al. The reliability, validity and responsiveness of two disease scores (BPDAI and ABSIS) for bullous pemphigoid: Which one to use? *Acta Derm Venereol* 2016 [Epub ahead of print].
- Yeh SW, Ahmed B, Sami N, Razaque AA. Blistering disorders: Diagnosis and treatment. *Dermatol Ther* 2003;16:214–23.
- Zhao CY, Murrell DF. Outcome measures for autoimmune blistering diseases. *J Dermatol* 2015;42:31–6.
- Zhao CY, Tran AQ, Lazo-Dizon JP, Kim J, Daniel BS, Venugopal SS, et al. A pilot comparison study of four clinician-rated atopic dermatitis severity scales. *Br J Dermatol* 2015;173:488–97.