REVIEW ARTICLE

BJD
British Journal of Dermatology

Development of core outcome sets in hidradenitis suppurativa: systematic review of outcome measure instruments to inform the process

J.R. Ingram,[1] S. Hadjieconomou[2] and V. Piguet[1]

[1]Division of Infection and Immunity, Cardiff University and [2]Cardiff and Vale University Health Board, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, Wales, U.K.

Linked Comment: van der Zee et al. *Br J Dermatol* 2016; **175**:242.

Summary

The recent hidradenitis suppurativa (HS) Cochrane review identified outcome measure heterogeneity as an important issue to address when designing future HS trials. Our objective was to follow the Harmonising Outcome Measures for Eczema (HOME) roadmap, by performing a systematic review of HS outcome measure instruments to inform the development of an HS core outcome set. We performed a systematic review to identify validation evidence for outcome measure instruments used in HS randomized controlled trials (RCTs), and assessed the methodological quality of all HS outcome measure validity studies using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. The 12 RCTs included in the Cochrane review utilized 30 outcome measure instruments, including 16 physician-reported instruments, 11 patient-reported instruments and three composite measures containing elements of both. Twenty-seven (90%) of the instruments lacked any validation data. Two further instruments have been developed and partially validated. Of the seven studies meeting our inclusion criteria, six were of 'fair' or 'poor' methodological quality, in part because most of the studies were not primarily designed for instrument validation. The HiSCR instrument is supported by good-quality validation data, but there are gaps, including assessment of internal consistency, inter-rater reliability and minimal clinically important difference, and convergent validity fell below the acceptable range for some comparisons. Multiple, usually unvalidated, outcome measure instruments have been used in HS RCTs. Where validation evidence is available there are issues of low methodological quality or incomplete validity assessment and so, currently, no instruments can be fully recommended.

Correspondence

John R. Ingram.
E-mail: ingramjr@cardiff.ac.uk
http://orcid.org/0000-0002-5257-1142

Accepted for publication

2 February 2016

Funding sources

J.R.I. is funded by a Health Fellowship from Health and Care Research Wales. The views expressed in this publication are those of the authors and not necessarily those of Health and Care Research Wales.

Conflicts of interest

J.R.I. is a local principal investigator for an observational study sponsored by AbbVie. He has not acted as a consultant for the manufacturer or taken part in paid advisory boards. V.P. undertakes personal advisory work with Pfizer, AbbVie, Janssen, Novartis and Almirall. V.P.'s department receives financial support from the Dermatology Life Quality Index copyright.

DOI 10.1111/bjd.14475

What's already known about this topic?

- The recent hidradenitis suppurativa (HS) Cochrane review identified heterogeneity of outcome measure instruments as an important obstacle in the design of future HS trials.
- The Harmonising Outcome Measures for Eczema (HOME) initiative provides a roadmap for developing a core outcomes set in HS.

What does this study add?

- Twenty-seven of the 30 outcome measure instruments used in HS randomized controlled trials are not supported by any formal validation data.
- Where available, validation evidence is generally of relatively low methodological quality, or remains incomplete, and so no instruments can be fully recommended currently.

Hidradenitis suppurativa (HS), also known as acne inversa, is a chronic painful inflammatory skin disease affecting about 1% of the European adult population.[1,2] Despite being a relatively common disease, there is a lack of high-quality HS clinical trials, confirmed by a recent Cochrane review of both medical and surgical HS interventions, which was able to include only 12 randomized controlled trials (RCTs).[3] The authors of the review concluded that more HS RCTs are required, to increase the HS evidence base and improve patient care; however, it was also noted that many different outcome measures had been used in the trials. Outcome measure heterogeneity did not have a significant impact on the Cochrane review in terms of meta-analyses involving direct comparisons as only two RCTs investigated the same intervention. However, indirect comparisons were prevented and the heterogeneity is likely to have a greater impact on updates of the Cochrane review, when more trials have been published and larger meta-analyses may be possible. In addition, interpretation of the results of the review was hampered by many of the instruments lacking validation data to confirm that results were clinically meaningful, valid and reliable.

Clinical trial research in HS has therefore reached an important stage where there is an opportunity to establish a consensus regarding which outcome measures should be used and to determine whether further instrument validation studies are needed. Lessons from the Harmonising Outcome Measures for Eczema (HOME) initiative, conducted to ensure that future eczema clinical trials incorporate a core outcome set, demonstrate that outcomes consensus is best conducted prior to clinical trials being instigated.[4] The HOME roadmap sets out a methodological framework for developing outcomes consensus.[4] Once the medical condition and the potential interventions have been defined, a core set of outcome domains is identified, encompassing the relevant range of outcomes from both a patient and clinician perspective. A systematic review of outcome measure instruments is required to generate a long list of the instruments currently available within each domain and to assess the extent and quality of the validation evidence in each case, which is the aim of the current review. A consensus process can then be employed to determine whether HS outcome domains can be accurately assessed by current instruments, or whether further validation data or new instruments are required.

Materials and methods

Long list of outcome measure instruments

Generation of a long list of outcome instruments is based on the Cochrane review of 'Interventions for hidradenitis suppurativa'. The methodology employed by the Cochrane review is published elsewhere and, briefly, involved searches in MEDLINE, Embase, LILACS, the Cochrane Skin Group Specialised Register, and Cochrane Central Register of Controlled Trials (CENTRAL), from inception until August 2015, for all RCTs conducted in adults with HS.[3] Searches, with the same final search date, were also conducted in five trial registries and the proceedings of eight dermatology conferences. Abstract screening and data extraction were performed by two authors working independently. All of the instruments contained within the included RCTs were documented by the current study's authors, including details of the items contained in the scale and the range of possible values.

Search for instrument validation data

MEDLINE and Embase were searched from inception to November 2015 for varying combinations of the subject terms 'hidradenitis suppurativa', 'acne inversa', 'severity' and 'severity of illness index', using similar search terms to Schmitt *et al.*[5] The reference lists of relevant articles were also searched, in addition to performing a separate free internet search for HS psychometric data. No language restrictions were applied, but the literature search was limited to articles with abstracts involving at least five human participants and containing original data. Two authors (J.R.I. and S.H.) independently screened the abstracts and full texts for eligibility, and data extraction was also performed by the two authors independently. Disagreements were resolved by discussion between the authors. Articles were included if they contained evidence pertaining to at least one of validity (content, convergent, divergent); reliability (internal consistency, interobserver reliability, test–retest reliability); interpretability [sensitivity to change, minimal clinically important difference (MCID), clinical severity banding]; or feasibility (time taken) of a HS outcome measure instrument. For each scale quality criterion we assigned a rating of 'acceptable' or 'adequate', based on parameters set out in a previous systematic review of eczema outcome measure instruments that informed the HOME consensus process.[5] For example, in the case of convergent validity, a correlation coefficient of 0·60–0·69 was designated 'acceptable', while a coefficient $> 0{\cdot}70$ was deemed 'adequate'.

We repeated our searches for HS instrument validation studies in MEDLINE and Embase using the sensitive search filter developed by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group, which has a sensitivity of 97·4%.[6] No additional HS instrument validation studies were identified from the extra searches.

Assessment of methodological quality of included studies

Methodological quality of the included studies was assessed using the COSMIN checklist.[7] The checklist applies a four-point rating scale of 'excellent', 'good', 'fair' or 'poor' for each measurement property, based on the lowest rating for each item assessed, including handling of missing data, adequate sample size and appropriate statistical analysis. Two authors (J.R.I. and S.H.) independently assessed methodological quality and resolved any differences by subsequent discussion.

Results

Outcome instruments included in Cochrane review studies

The 12 RCTs included in the HS Cochrane review contained a total of 30 outcome measure instruments, including 16 physician-reported instruments, 11 patient-reported outcomes and three composite measures containing elements of both (Table 1).[8–20] Grouping the instruments in terms of potential outcome domains, six of the physician-reported instruments involved a skin lesion count;[8–13] five were based on a physician's global assessment (PGA);[13–17] one assessed recurrence rate;[18] one measured cosmesis;[8] one examined duration of recovery from surgical treatment;[18] and two contained elements of the PGA combined with either a measure of cosmesis or a lesion count.[7,12] Of the 11 patient-reported outcome measure instruments, four involved a patient global self-assessment;[13–15,19] two measured overall satisfaction with treatment;[9,20] two measured pain;[8,12,13,15,16] one used a patient lesion count;[8] one was a quality-of-life scale;[8,12,15,16] and one measured impairment of function.[12]

In several cases, particularly within the lesion count domain, there were a number of closely related instruments, often involving modification of a previously published instrument. For example, the lesion count score published by Sartorius *et al.* in 2003 has been modified to produce five different instruments, including a modification by the original authors in 2009.[21,22] In some cases, the instrument was modified to permit a within-participant trial design in which each region, rather than the whole individual, is given a severity score.[9,10] Each instrument contains different items and so, although there is quite a lot of overlap between the instruments, their scores cannot be directly compared in meta-analyses. The potential for confusion is confounded by different names being assigned to the same outcome instrument. In particular, the instrument published by Sartorius *et al.* in 2009 was named the 'modified HS Score (modified HSS)' in the article,[22] but the name 'modified Sartorius Score' (MSS) has been used subsequently in several publications, and 'HS-LASI' ('HS lesion, area, and severity index') has also been used as another alternative name.

Instruments supported by validation data

Our search for instrument validation data identified 119 abstracts, of which 112 were excluded owing to a lack of our prespecified validation evidence (Fig. 1). The seven included studies relate to six outcome measure instruments, two of which, the Hidradenitis Suppurativa Clinical Response (HiSCR) and Acne Inversa Severity Index (AISI),[23,24] have only been published recently and have not yet been used in an RCT (Table 2). The HiSCR instrument was developed retrospectively from an RCT that used other outcome measures in the trial itself.[12] The HiSCR measure is designed to assess treatment response in a binary manner, rather than being a continuous or ordinal scale. It fits within the lesion count outcome domain, involving a count of the total number of abscesses and inflammatory nodules, designated 'ANs', as well as recording the number of sinuses draining purulent fluid present in an individual.[23] Treatment responders are defined as those who achieve at least a 50% reduction in ANs, with no increase in the number of abscesses or draining sinuses, relative to baseline. Because the instrument involves a percentage decrease in ANs, the patient population for the validation sample was limited to those with at least three ANs, to ensure that a reduction of one AN does not achieve the end point.

The AISI outcome measure is a composite instrument recently designed by a group of clinicians in Italy incorporating a physician-reported lesion count and a patient-reported visual analogue scale (VAS) measure of pain and disability.[24] The instrument designers assigned a score for each type of HS lesion (comedones, abscess/inflammatory nodule, sinus tract, keloid/fibrotic adherence, fibrosclerotic inflammatory plaque are scored 1, 2, 3, 4 or 5 points, respectively), which is then multiplied by the number of body sites where the lesion occurs (rather than the number of lesions observed). The score from 0 to 10 from the combined pain and disability VAS is then added to produce a final total. So far, one validation study has been performed in 46 patients with HS attending a secondary care dermatology clinic.[24]

Validation data and assessment of methodological quality of the studies

Despite the Hurley scoring system being one of the most recognized HS disease severity instruments,[25] very little formal outcome measure validity data are available. One recent study examining quality of life in HS provides divergent validity data in comparison with Dermatology Life Quality Index (DLQI) scores.[26] A correlation coefficient of 0·549 is reported; however, methodological quality was downgraded to 'fair' because it was not clear how missing items were handled (Table 2).

Two studies containing a total of 176 patients provide validation evidence for the MSS.[22,27] The Spearman rho correlation coefficient measuring divergent validity compared with DLQI scores was found to be 0·342 in one study and 0·48 in the other (Table 2).[22,27] Methodological quality was downgraded to 'fair' in each case because it was not clear how missing items were handled. The correlation coefficient for interobserver reliability was 0·95, indicating that this criterion was met adequately.[27] However, methodological quality was downgraded to 'poor' because, while 61 patients took part in some aspects of the study, only 23 patients were rated by more than one dermatologist independently.

An epidemiology study examining 302 patients with HS provides some convergent validity data between the Revuz version of the original Sartorius Score and the Hurley stage and degree of suppuration.[28] However, the study was not primarily designed to provide HS outcome measure validation

Table 1 Outcome measure instruments used in randomized controlled trials of interventions for hidradenitis suppurativa (HS)

Instrument	Potential domain	Range	Description	Trial reference
Physician reported				
Hurley stage[25]	Lesion count	I–III (worst)	Stage I: single or multiple abscesses without sinus tracts or scarring; stage II: recurrent, widely separated abscesses with sinus tracts and scarring; stage III: multiple interconnected sinus tracts and abscesses across entire region	Miller *et al.*[8]
Original version of Sartorius Score[21]	Lesion count	0–unlimited (worst)	Three points are assigned for each involved region and each scar, nodule and sinus are scored 1, 2 and 4 points, respectively; for each affected region, six points are added for lack of normal skin between lesions and the greatest distance between lesions is scored 2, 4 and 8 points for < 5 cm, 5–10 cm and > 10 cm, respectively	Miller *et al.*[8]
Highton version of Sartorius Score	Lesion count	0–unlimited (worst)	Each scar, nodule and sinus scored 1, 2 and 4 points, respectively; for each affected region, degree of erythema and discharge scored 0–3, 6 points are added for lack of normal skin between lesions, and the greatest distance between lesions is scored 2, 4 and 8 points for < 5 cm, 5–10 cm and > 10 cm, respectively	Highton *et al.*[9]
Tierney version 1 of Sartorius Score	Lesion count	0–unlimited (worst)	For a particular region, each scar, abscess, nodule and fistula are scored 1, 1, 2 and 4 points, respectively; 6 points are added for lack of normal skin between lesions and the greatest distance between lesions is scored 2, 4 and 8 points for < 5 cm, 5–10 cm and > 10 cm, respectively	Tierney *et al.*[10]
Modified Sartorius Score (Sartorius 2009)[22]	Lesion count	0–unlimited (worst)	Three points per region of involvement, 1 point per nodule, 6 points per fistula, Hurley III 9 points per region; greatest distance between two lesions in each region: < 5 cm, 1 point, 5–10 cm, 3 points, > 10 cm, 9 points	Fadel and Tawfik,[11] Kimball *et al.*[12]
Abscess and nodule count	Lesion count	0–unlimited	The number of nodules and abscesses are each counted	Kimball *et al.*,[12] Jemec and Wendelboe[13]
HS-PGA	PGA/lesion count	Clear–very severe	Ordinal scale: clear, minimal, mild, moderate, severe, very severe; based on total number of abscesses, fistulas, inflammatory and noninflammatory nodules	Kimball *et al.*[12]
Mortimer PGA scale	PGA	+9 to −6 (worst)	Relative to baseline, changes in disease activity are scored as clear (+3), much improved (+2), improved (+1), unaltered (0), worse (−1) and much worse (−2) for each of the number of inflamed/noninflamed nodules, degree of induration and tenderness, and presence of draining sinuses	Mortimer *et al.*[14]
SGA	PGA	0 (clear)–3 (severe)	SGA is a mean score derived from three domains: tenderness on palpation, erythema of lesions, and discharge, each measured from 0 to 3	Adams *et al.*[15]

Table 1 (continued)

Instrument	Potential domain	Range	Description	Trial reference
PGA	PGA	Cleared to worse	7-point ordinal scale relative to baseline severity (% improvement): cleared (100%), excellent (75–99%), good (50–74%), fair (25–49%), slight (1–24%), unchanged, worse	Grant *et al.*[16]
PGA scale	PGA	Improved, same, worse	Ordinal scale with three levels for change from baseline in terms of ability of lesions to drain spontaneously without surgery and number of acute inflammatory lesions	Angel *et al.*[17]
Physician global evaluation	PGA	0–100 (worst)	VAS	Jemec and Wendelboe[13]
PGA scar scoring	Cosmesis/PGA	ND	ND	Miller *et al.*[8]
Manchester Scar Proforma[33]	Cosmesis	Higher score worse	Scar number, size, margins, surface, colour and texture are used to compile a score	Miller *et al.*[8]
Local recurrence rate	Recurrence rate	0–100% (worst)	Recurrence at surgical site following surgery (no specified maximum distance from scar)	Buimer *et al.*[18]
Time to complete wound healing	Duration of recovery	0–no limit	Time of complete wound healing in days	Buimer *et al.*[18]
Patient reported				
Patient global assessment of HS lesions	Patient global self-assessment	0–5 (severe)	Ordinal scale from 0 (good) to 5 (severe)	Adams *et al.*[15]
Patient global assessment VAS	Patient global self-assessment	0–100 (worst)	VAS	Jemec and Wendelboe[13]
Change in patient global assessment VAS	Patient global self-assessment	0–100 (better)	Baseline disease severity defined as a score of 50; a score of 100 denotes 'completely better' and 0 denotes 'worst it has ever been'	Mortimer *et al.*[14]
Patient global assessment	Patient global assessment	−2 to +2 (better)	Ordinal scale relative to baseline severity: +2, much improved, +1, improved, 0, unaltered, −1, worse, −2, much worse	Clemmensen[19]
Treatment satisfaction	Overall satisfaction	Excellent to worse	Likert scale relative to baseline: excellent, good, fair, unchanged, worse	Highton *et al.*[9]
Treatment satisfaction questionnaire	Overall satisfaction	ND	Recorded patients' views of treatment benefit both in absolute terms and relative to other interventions, adverse effects of therapy and overall satisfaction	Mahmoud *et al.*[20]
Pain	Pain	0–5 (severe)	Ordinal scale from 0 (none) to 5 (severe)	Adams *et al.*[15]
Pain VAS	Pain	0–100	VAS	Miller *et al.*,[8] Kimball *et al.*,[12] Jemec and Wendelboe,[13] Grant *et al.*[16]
Patient report of lesions	Lesion count	0–unlimited (worse)	Patient report of number of days with active lesions between clinician reviews	Miller *et al.*[8]
DLQI	Quality of life	0–30 (worst)	10 quality-of-life domains scored from 0 (no effect) to 3 (very large effect)	Miller *et al.*,[8] Kimball *et al.*,[12] Adams *et al.*,[15] Grant *et al.*[16]
TWPI score	Impairment of function	0–100 (worst)	Score obtained from WPAI-SHP questionnaire	Kimball *et al.*[12]
Composite scales				
HSSI	Lesion count and pain	0–19 (worst)	0–4 points for each of number of affected regions, BSA involved, number of active lesions, number of dressing changes per day, pain VAS score	Grant *et al.*[16]

Table 1 (continued)

Instrument	Potential domain	Range	Description	Trial reference
Cumulated score	Lesion count and patient global self-assessment	Negative or positive (better), with no limit	Relative to baseline, 5 points assigned to each inflammatory nodule, abscess and change of one level in the patient global assessment scale, 1 point assigned to each pustule (for lesions, a positive score indicates resolution, while a negative score denotes a new lesion)	Clemmensen[19]
Tierney version 2 of Sartorius 2003	Lesion count and pain	0–unlimited (worse)	For a particular region, each scar, abscess, nodule and fistula are scored 1, 1, 2 and 4 points, respectively; 6 points are added for lack of normal skin between lesions and the greatest distance between lesions is scored 2, 4 and 8 points for < 5 cm, 5–10 cm and > 10 cm, respectively; up to 3 points are scored for each of erythema, oedema, pain and purulent discharge	Tierney *et al.*[10]

PGA, physician's global assessment; SGA, static global assessment; VAS, visual analogue scale; ND, no details; DLQI, Dermatology Life Quality Index; TWPI, Total Work Productivity Impairment; WPAI-SHP, Work Productivity and Activity Impairment – Specific Health Problem; HSSI, HS Severity Index; BSA, body surface area.

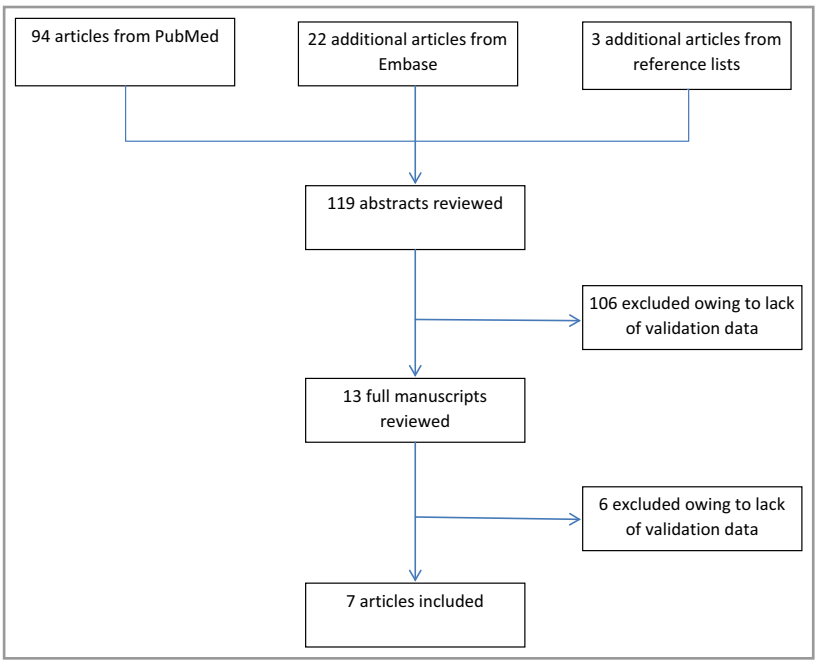

Fig 1. Validation studies included.

data and methodological quality was downgraded to 'poor' owing to use of linear regression rather than calculation of correlations or the area under the receiver operating curve.

An RCT of intense pulsed light for HS used the Highton version of the original Sartorius score and calculated a kappa statistic of 0·79 for inter-rater reliability.[9] Methodological quality was graded 'poor' because only 17 participants took part in the trial.

For the HiSCR instrument, acceptable convergent validity was met with regard to a PGA (correlation coefficient 0·61), while the correlation coefficient was below this level when compared with the Hurley Stage and MSS (coefficients of 0·49 and 0·51, respectively).[23] As expected, correlation was lower when compared with patient-reported outcome scores because these measure different constructs. Methodological quality was downgraded from 'excellent' to 'good' because, while the method for dealing with missing items was described in the original RCT publication, the percentage of missing items was not explicitly stated in the context that not all of the RCT participants were included. Test–retest reliability for HiSCR was adequate for both ANs and sinuses, with intraclass correlations of 0·91 and 0·95, respectively, and methodological quality was again rated 'good'. Using the original trial data,[12] more patients reached the threshold for improvement compared

Table 2 Validation studies for hidradenitis suppurativa (HS) instruments, including a COSMIN rating of methodological quality[a]

	Validity			Reliability			Interpretability			Feasibility
Instrument and number of participants	Content	Convergent	Divergent: correlation coefficient	Internal consistency	IRR	Test–retest reliability	Sensitivity to change	MCID	Disease severity bands	Time taken (s)
Hurley, n = 55[26]	–	–	DLQI 0·549 (fair)	–	–	–	–	–	–	–
MSS, n = 115[22]	–	–	DLQI 0·342 (fair)	–	–	–	–	–	–	–
MSS, n = 61[27]	–	–	DLQI 0·48 (fair)	–	ICC 0·95 (poor)	–	–	–	–	–
Revuz version of original Sartorius Score, n = 302[28]	–	$P < 0.001$ for association with Hurley stage and suppuration (poor)	–	–	–	–	–	–	–	–
Highton version of original Sartorius score, n = 18[9]	–	–	–	–	Kappa statistic 0·79 (poor)	–	–	–	–	–
HiSCR n = 138[23]	–	Spearman rho correlation: Hurley 0·49, MSS 0·51, HS-PGA 0·61 (good)	Pain-VAS 0·33, DLQI 0·27, WPAI-TAI 0·47, WPAI-TWI 0·47 (good)	–	–	ICC: AN 0·91, sinuses 0·95 (good)	No correlation coefficient provided[29]	–	–	–
AISI, n = 46[24]	–	Spearman rho correlation: Hurley 0·71, Revuz score 0·97 (fair)	DLQI 0·83 (fair)	–	–	–	–	–	Mild < 10, moderate 10–18, severe > 18	AISI 46·4 ± 19·2, Revuz 83·2 ± 19·0

[a]Grading in parentheses indicates methodological quality using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist.[7] IRR, inter-rater reliability; MCID, minimal clinically important difference; DLQI, Dermatology Life Quality Index; MSS, modified Sartorius Score; ICC, intraclass correlation; HiSCR, Hidradenitis Suppurativa Clinical Response; HS-PGA, HS physician's global assessment; VAS, visual analogue scale; WPAI-TAI, Work Productivity and Activity Impairment Questionnaire – Total Activity Impairment; WPAI-TWI, Work Productivity and Activity Impairment Questionnaire – Total Work Impairment; AN, abscesses and inflammatory nodules; AISI, Acne Inversa Severity Index.

with the HS-PGA instrument;[29] however, a correlation coefficient was not calculated and so the sensitivity to change of HiSCR requires further evaluation.

The AISI instrument demonstrated adequate convergent validity with Hurley stage and the Revuz version of the Sartorius 2003 score (correlation coefficients of 0·71 and 0·97, respectively).[24] In terms of divergent validity, correlation with DLQI scores was relatively high but still within the acceptable range (with a coefficient of 0·83). Methodological quality for the validity assessments was downgraded to 'fair' owing to inadequate information regarding the handling of any missing items. The authors also used the distribution of AISI scores to determine three disease severity bands corresponding to mild, moderate and severe disease. The time required to complete the AISI score was about half that needed for the Revuz score (46·4 vs. 83·2 s).[24]

Across all seven studies, there was an absence of data covering the criteria of content validity, internal consistency and MCID.

Discussion

Our results confirm that outcome measure heterogeneity is a problem for HS research, with 30 instruments used in the 12 RCTs included in the HS Cochrane review, and will become an increasing problem if an outcomes consensus process is not undertaken in the near future. Of note, 27 of the 30 outcome measure instruments employed in HS RCTs to date have no validation data to support their use, representing 90% of the total. Seven papers were found to provide some validation data, but, in general, their methodological quality was relatively low. Using the COSMIN checklist, six of the seven studies were graded as 'fair' or 'poor' in quality, in part because instrument validation was not the primary purpose of the research.

Good-quality validation evidence is available for the recently described HiSCR instrument, which has acceptable convergent validity with the PGA and adequate test–retest reliability. However, the correlation coefficient falls below the acceptable range for comparisons with the MSS and Hurley staging. In addition, internal consistency, inter-rater reliability and MCID have not been assessed yet and so further evidence is still needed.

In some of the validation studies examined, confusion was apparent in the criteria of convergent and divergent validity. For example, some studies have compared a scale in the lesion count domain with a quality-of-life instrument to attempt to demonstrate convergent validity, which is comparing two different disease severity constructs and so instead falls within the divergent validity category. The relatively high correlation coefficient between AISI and the DLQI scale may, in part, be explained by inclusion of pain and disability, as well as a lesion score, in the AISI instrument, items that are also measured by the DLQI.

Our results are broadly in keeping with the findings of systematic reviews conducted for eczema outcome measures, to inform the HOME consensus process. In 2003, only 27% of eczema trials used a previously published severity scale and, from the 93 trials included, 56 different objective scoring systems were identified.[30] Assessment of validation data for eczema outcome measure instruments in 2007 found that only three instruments had undergone sufficient validation to recommend their use in trials and routine practice.[5]

One of the strengths of our systematic review is that we predefined 'acceptable' and 'adequate' parameters for each quality criterion relevant to assessment of measurement instruments. In addition, we assessed the methodological quality of the validity studies identified using the COSMIN checklist. These assessments will be helpful to inform the subsequent outcome measures consensus process. Our review of HS outcome measure instruments has been performed earlier in the research cycle than atopic eczema, at a time when fewer HS trials have been conducted. Hence, one limitation of our review is that it contains relatively few validation studies. It may be that the subsequent consensus process will have to identify and commission further validation studies for pre-existing instruments or recommend development of new instruments.

Our review has identified 10 potential efficacy outcome measure domains: quality of life; pain; lesion count; PGA; patient global self-assessment; recurrence rate; overall satisfaction with treatment; impairment of function; cosmesis; and duration of recovery. Most domains are relevant to both medical and surgical interventions. However, recurrence rate and duration of recovery are more specific for surgical treatments in the context that some procedures, such as extensive excision, aim to provide disease remission in the treated region, at the expense of prolonged wound healing. The subsequent consensus process will need to define the outcome domains in more detail, incorporating the views of all relevant stakeholders, namely patient representatives, dermatologists, surgeons, primary care physicians, regulatory authorities and journal editors, with predetermined definitions of consensus.[31]

Support from international dermatology outcomes consortia, such as the Cochrane Skin Group Core Outcomes Set Initiative (CSG-COUSIN) and the International Dermatology Outcome Measures (IDEOM) group has been sought, to ensure that the planned consensus process has a broad geographical base and methodological rigour. Several challenges are likely to be encountered in terms of scope, outcome domains and instrument validation. Regarding the scope of the process, debate will be needed as to whether medical and surgical interventions require separate outcome domains and, if so, into which category laser and light interventions should be placed. When defining recurrence rate as a potential domain, care will be required to determine whether recurrence is defined as any new HS lesions within a previously treated region, or whether only those lesions within a certain distance from the scar or treatment area should qualify. The domain of 'lesion count' may need broadening from the current paradigm, which is generally restricted to counting lesions at particular predetermined time points, potentially

missing menstrual or other exacerbations between clinician visits. The HOME group has agreed that long-term control of flares should be one of the core outcome domains in atopic eczema,[31] and it may be that the frequency of new lesions or recurrent activity in scar sites from previous lesions should be considered as a new outcome domain of 'flare frequency' in HS. Any new domain will, of course, require development and validation of new measurement instruments.

In summary, our systematic review of HS outcome measure instruments has demonstrated substantial heterogeneity, reinforcing the need for a consensus process. Ninety percent of the instruments that have been used in HS RCTs lack any validation evidence and most of the evidence that is available is of relatively low methodological quality. More validation studies are required to ensure that HS outcome measure instruments can satisfy the Outcome Measures in Rheumatology (OMERACT) filter,[32] which is designed to assess 'truth', 'discrimination' and 'feasibility'.

Acknowledgments

We wish to thank Professor Hywel Williams (Nottingham University), who kindly reviewed an earlier version of the manuscript.

References

1 Jemec GB. Clinical practice. Hidradenitis suppurativa. *N Engl J Med* 2012; **366**:158–64.
2 Revuz JE, Canoui-Poitrine F, Wolkenstein P *et al.* Prevalence and factors associated with hidradenitis suppurativa: results from two case–control studies. *J Am Acad Dermatol* 2008; **59**:596–601.
3 Ingram JR, Woo PN, Chua SL *et al.* Interventions for hidradenitis suppurativa. *Cochrane Database Syst Rev* 2015; **10**:CD010081.
4 Schmitt J, Apfelbacher C, Spuls PI *et al.* The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol* 2015; **135**:24–30.
5 Schmitt J, Langan S, Williams HC; European Dermato-Epidemiology Network. What are the best outcome measurements for atopic eczema? A systematic review. *J Allergy Clin Immunol* 2007; **120**:1389–98.
6 Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009; **18**:1115–23.
7 Terwee CB, Mokkink LB, Knol DL *et al.* Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; **21**:651–7.
8 Miller I, Lynggaard CD, Lophaven S *et al.* A double-blind placebo-controlled randomized trial of adalimumab in the treatment of hidradenitis suppurativa. *Br J Dermatol* 2011; **165**:391–8.
9 Highton L, Chan WY, Khwaja N, Laitung JK. Treatment of hidradenitis suppurativa with intense pulsed light: a prospective study. *Plast Reconstr Surg* 2011; **128**:459–65.
10 Tierney E, Mahmoud BH, Hexsel C *et al.* Randomized control trial for the treatment of hidradenitis suppurativa with a neodymium-doped yttrium aluminium garnet laser. *Dermatol Surg* 2009; **35**:1188–98.
11 Fadel MA, Tawfik AA. New topical photodynamic therapy for treatment of hidradenitis suppurativa using methylene blue niosomal gel: a single-blind, randomized, comparative study. *Clin Exp Dermatol* 2015; **40**:116–22.
12 Kimball AB, Kerdel F, Adams D *et al.* Adalimumab for the treatment of moderate to severe hidradenitis suppurativa: a parallel randomized trial. *Ann Intern Med* 2012; **157**:846–55.
13 Jemec GB, Wendelboe P. Topical clindamycin versus systemic tetracycline in the treatment of hidradenitis suppurativa. *J Am Acad Dermatol* 1998; **39**:971–4.
14 Mortimer PS, Dawber RP, Gales MA, Moore RA. A double-blind controlled cross-over trial of cyproterone acetate in females with hidradenitis suppurativa. *Br J Dermatol* 1986; **115**:263–8.
15 Adams DR, Yankura JA, Fogelberg AC, Anderson BE. Treatment of hidradenitis suppurativa with etanercept injection. *Arch Dermatol* 2010; **146**:501–4.
16 Grant A, Gonzalez T, Montgomery MO *et al.* Infliximab therapy for patients with moderate to severe hidradenitis suppurativa: a randomized, double-blind, placebo-controlled crossover trial. *J Am Acad Dermatol* 2010; **62**:205–17.
17 Angel MF, Ramasastry SS, Manders EK *et al.* Beneficial effects of staphage lysate in the treatment of chronic recurrent hidradenitis suppurativa. *Surg Forum* 1987; **38**:111–12.
18 Buimer MG, Ankersmit MF, Wobbes T, Klinkenbijl JH. Surgical treatment of hidradenitis suppurativa with gentamicin sulfate: a prospective randomized study. *Dermatol Surg* 2008; **34**:224–7.
19 Clemmensen OJ. Topical treatment of hidradenitis suppurativa with clindamycin. *Int J Dermatol* 1983; **22**:325–8.
20 Mahmoud BH, Tierney E, Hexsel CL *et al.* Prospective controlled clinical and histopathologic study of hidradenitis suppurativa treated with the long-pulsed neodymium:yttrium–aluminium–garnet laser. *J Am Acad Dermatol* 2010; **62**:637–45.
21 Sartorius K, Lapins J, Emtestam L, Jemec GB. Suggestions for uniform outcome variables when reporting treatment effects in hidradenitis suppurativa. *Br J Dermatol* 2003; **149**:211–13.
22 Sartorius K, Emtestam L, Jemec GB, Lapins J. Objective scoring of hidradenitis suppurativa reflecting the role of tobacco smoking and obesity. *Br J Dermatol* 2009; **161**:831–9.
23 Kimball AB, Jemec GB, Yang M *et al.* Assessing the validity, responsiveness and meaningfulness of the Hidradenitis Suppurativa Clinical Response (HiSCR) as the clinical endpoint for hidradenitis suppurativa treatment. *Br J Dermatol* 2014; **171**:1434–42.
24 Chiricozzi A, Faleri S, Franceschini C *et al.* AISI: a new disease severity assessment tool for hidradenitis suppurativa. *Wounds* 2015; **27**:258–64.
25 Hurley HJ. Axillary hyperhidrosis, apocrine bromhidrosis, hidradenitis suppurativa, and familial benign pemphigus: surgical approach. In: *Dermatologic Surgery – Principles and Practice* (Roenigk RK, Roenigk HH, eds). New York: Marcel Dekker, 1989; 729–39.
26 Alavi A, Anooshirvani N, Kim WB *et al.* Quality-of-life impairment in patients with hidradenitis suppurativa: a Canadian study. *Am J Clin Dermatol* 2015; **16**:61–5.
27 Sartorius K, Killasli H, Heilborn J *et al.* Interobserver variability of clinical scores in hidradenitis suppurativa is low. *Br J Dermatol* 2010; **162**:1261–8.
28 Canoui-Poitrine F, Revuz JE, Wolkenstein P *et al.* Clinical characteristics of a series of 302 French patients with hidradenitis suppurativa, with an analysis of factors associated with disease severity. *J Am Acad Dermatol* 2009; **61**:51–7.
29 Kimball AB, Sobell JM, Zouboulis CC *et al.* HiSCR (Hidradenitis Suppurativa Clinical Response): a novel clinical endpoint to evaluate therapeutic outcomes in patients with hidradenitis suppurativa

from the placebo-controlled portion of a phase 2 adalimumab study. *J Eur Acad Dermatol Venereol* 2016; **30**:989–94.
30 Charman C, Chambers C, Williams H. Measuring atopic dermatitis severity in randomized controlled clinical trials: what exactly are we measuring? *J Invest Dermatol* 2003; **120**:932–41.
31 Schmitt J, Langan S, Stamm T *et al.* Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. *J Invest Dermatol* 2011; **131**:623–30.
32 Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998; **25**: 198–9.
33 Beausang E, Floyd H, Dunn KW *et al.* A new quantitative scale for clinical scar assessment. *Plast Reconstr Surg* 1998; **102**:1954–61.