The Birmingham Epidermolysis Bullosa Severity score: development and validation

C. Moss, A. Wong and P. Davies*

*Departments of Dermatology and *Research and Development, Birmingham Children's Hospital NHS Foundation Trust, Birmingham B4 6NH, U.K.*

Correspondence
Celia Moss.
E-mail: celia.moss@bch.nhs.uk

Accepted for publication
3 November 2008

Key words
epidermolysis bullosa, score, severity of illness index

Conflicts of interest
None declared.

This work was first presented as a poster at the British Association of Dermatologists 87th Annual Meeting, Birmingham, U.K., 10–13 July 2007 [Wong A, Ward R, Hadley N et al. *A scoring system for epidermolysis bullosa. Br J Dermatol 2007;* **157** *(Suppl. 1): 26–7 (Abstract)].*

DOI 10.1111/j.1365-2133.2009.09041.x

Summary

Background Objective severity scores facilitate clinical care and research. However, the rarity and heterogeneity of epidermolysis bullosa (EB) make scoring difficult.
Objectives To develop a severity score covering all subtypes of EB at all ages that is simple, valid, sensitive and reliable.
Methods Score items and weightings were generated by expert consensus, and refined for content and face validity. The Birmingham EB Severity (BEBS) score was tested on 97 patients aged 0–64 years.
Results Eleven items were scored: area of damaged skin, involvement of nails, mouth, eyes, larynx and oesophagus, scarring of hands, skin cancer, chronic wounds, alopecia and nutritional compromise. Area was allocated 50 points, and the 10 other items 5 points each, giving a maximum score of 100. Lowest BEBS scores occurred in Weber–Cockayne EB simplex (median 1·0; range 0·1–3·0; n = 12), highest scores in generalized non-Herlitz junctional EB (28·5; 5·0–62·0; n = 7), Hallopeau–Siemens recessive dystrophic EB (HS-RDEB) (22·9; 4·3–69·0; n = 23) and Herlitz junctional EB (H-JEB) (14·4; 2·5–49·3; n = 9), and intermediate scores in dominant dystrophic EB (5·3; 0·5–15·9; n = 19), Dowling–Meara EB simplex (DM-EBS) (6·3; 2·8–22·5; n = 16) and non-Hallopeau–Siemens recessive dystrophic EB (7·8, 2·8–27·8; n = 11). Intra- and interobserver correlations were high. With age, scores increased for H-JEB ($r = 0·9$, $P = 0·001$) and HS-RDEB ($r = 0·73$, $P = 0·001$) and decreased for DM-EBS ($r = -0·62$, $P = 0·01$), with positive but nonsignificant correlations for the other types.
Conclusions The BEBS score appears valid and reproducible, gives appropriate scores for different subtypes, and reflects changes with age.

Epidermolysis bullosa (EB) is a diverse group of hereditary blistering disorders encompassing four major forms, simplex (EBS), junctional (JEB), dystrophic (DEB) and Kindler syndrome, as well as numerous subtypes.[1] EB ranges in severity from occasional friction blisters to generalized and lethal disease. While objective scoring systems are well established for other chronic skin disorders, such as Psoriasis Area and Severity Index for psoriasis[2] and SCORAD index for eczema,[3] there is none available for EB. Objective severity scores are useful to assess response to interventions and to estimate resource needs. They may also help research into pathogenesis, by quantifying natural variability within and between subtypes. Therefore we set out to devise an objective scoring system for EB.

Developing and validating a scoring system requires large numbers of patients with all varieties of the disease. However, EB is rare: even the commonest type, Weber–Cockayne EBS (EBS-WC), affects only 1 in 17 000. In 2002 a National Health Service-funded EB Service was commissioned for England and Wales, based in London and Birmingham, with care being shared between specialist centres and periphery, coordinated by a team of outreach nurses. Already more than 700 patients with EB are managed by the Service. Diagnosis by immunofluorescence (IMF) analysis of skin biopsy followed by mutation analysis is available to all patients, and is routinely carried out in severe types. Unification of the EB service with large numbers of accurately diagnosed patients made this project feasible.

Heterogeneity as well as rarity makes EB challenging to score. EBS-WC is considered mild because blisters are generally limited to the palms and soles, and do not scar. By contrast, Hallopeau–Siemens recessive DEB (HS-RDEB) is severe because it affects mucosae as well as the whole skin surface, causes mutilating scarring and predisposes to skin cancer. Other types of EB are not only intermediate in severity but have different manifestations: this is not a simple linear

spectrum. For example, in the Dowling–Meara type of EBS (DM-EBS), nail dystrophy and palmoplantar keratoderma cause considerable morbidity, while scarring and mucosal involvement do not usually occur. Meanwhile, patients with the Herlitz type of JEB (H-JEB) may have limited areas of skin damage in early infancy and rarely have mutilating scarring, but usually die in infancy. Despite this heterogeneity, nurses and doctors experienced in EB can usually agree on whether a case is 'mild' or 'severe', even within diagnostic subtypes. This encouraged us to try to quantify clinical severity.

Severity may be measured in terms not only of clinical manifestations but also of disease impact. The latter is an important outcome measure in clinical trials, recorded as health-related quality of life (QoL), functional disability or economic burden. Although 'biological severity' and 'disease burden' often correlate, patients may differ markedly in the extent to which the same disease affects their lives, further confounding attempts to quantify severity. There are well-established disease impact scores for dermatology, particularly the Dermatology Life Quality Index (DLQI), with validated versions also for children (CDLQI), infants (IDQOL) and families (FDLQI).[4] Horn and Tidman[5] applied the DLQI and CDLQI to Scottish patients with EB and found that scores in those with HS-RDEB exceeded scores of any skin disorder previously assessed, while the effect on QoL of EBS and other subtypes of DEB was similar to that of moderately severe psoriasis and eczema. We sought an objective score of clinical severity which can be used alongside disease impact scores. Together they should give a holistic picture of disease burden; separately they should help dissect the many factors that contribute to suffering.

A Japanese group devised different scoring systems for different EB subtypes, incorporating a variety of parameters including laboratory tests.[6] However, this has significant limitations in practice, particularly in infants before the diagnosis has been made, or where laboratory results are unavailable. Furthermore, validation is hampered by the rarity of EB and the multiplicity of subtypes.

We wanted a method of scoring clinical severity, for use not only in clinic but also in patients' homes, with the following characteristics. It should be *simple*, that is quick and easy to learn, use and calculate; *practical*, i.e. not needing special equipment or blood tests; *comprehensive* in being applicable to all types of EB, at any age; *valid*, covering important manifestations of EB and giving scores for different types of EB that 'feel' correct to professionals; *sensitive* to changes with age and treatment; and *reliable*, giving consistent results for different observers.

Materials and methods

Item and score generation

Our perception of clinical severity is derived from numerous factors including extent of organ damage, number of organs affected, degree of functional impairment, and the occurrence of obviously 'bad' outcomes such as growth failure, malignancy and death. Relevant items for scoring emerged from informal discussion between EB professionals in the Birmingham clinics, who were urged to consider whether the items selected were all relevant to severity (face validity) and whether they covered all the relevant aspects (content validity).[7]

An initial score sheet was devised and piloted, using first 'virtual' EB patients with each of the classic types, and then case notes with clinical photographs. After several adjustments the score sheet was subjected to further refinement by the Delphi method. It was trialled using scoring episodes from 39 Birmingham patients with EB, and informally by members of the London and German EB teams. It was modified according to feedback, by adding new items and rationalizing the weighting. The score sheet was then piloted and adjusted over several more cycles using the same patient data until the scores of the different EB subtypes corresponded to our impression of severity. All attributes generated considerable debate.

Area of skin involvement

Area is difficult to estimate,[8] particularly where there are multiple small lesions (as in DM-EBS). The 'rule of nines'[9] and body diagrams were helpful in this respect. We did not attempt to subdivide area according to nature of skin involvement (apart from chronic wounds, see below), but instead used the term 'area of skin damage' to include blisters, erosions, scabs, healing skin, erythema and atrophic scarring, but to exclude skin changes not resulting directly from damage such as mottled pigmentation in EBS and poikiloderma in Kindler syndrome. We also excluded postinflammatory pigmentary changes, which vary according to background skin type, and well-healed scars as seen following neonatal skin loss in dominant DEB (DDEB). Area of skin affected was considered the most important item, deserving higher weighting than the rest. In general, those patients with most extensive areas of damaged skin are probably more likely to suffer severe pain, nutritional compromise, infections, malignancy and earlier death. Conversely, an effective treatment for EB should reduce the area of damaged skin. Such an important item was felt to merit 50% of the score.

No agreement could be reached about weighting of 10 other attributes so they were all allocated 5 points to give a convenient maximum score of 100.

Nail involvement

This was included because of an impression that early nail dystrophy and loss correlate with more rapid disease progression, particularly in JEB and RDEB. Also, nail score helps to differentiate different subtypes of EB; for example EBS-WC is not usually associated with nail involvement. Loss is usually preceded by dystrophy, thus reflecting nail disease progression, and we therefore scored these two states differently.

Mucosal involvement

This was initially a single item, scored according to number of mucosal surfaces affected. However, we chose later to subdivide it according to site (eyes, mouth, larynx, oesophagus) in order to reflect the perceived severity of conditions particularly affecting mucosal surfaces (H-JEB and LOGIC syndrome).

Scarring of hands

This was initially allocated 10% of the score because it is a very serious and progressive aspect of HS-RDEB which should reflect disease progression. However, its weighting was later downgraded because it is mainly confined to DEB.

Skin cancer

This is a major cause of death in EB. Lesions are often multiple. Therefore we used a score that could flexibly reflect number of primary lesions and metastatic spread. However, no cancers are present in the cohort studied here, so this component remains unvalidated.

Chronic wounds present for > 6 months

This was a late addition. Initially we thought that this would be covered by area of skin damage, but the consensus was that such lesions deserve additional points to reflect defective wound healing and malignant potential as well as skin fragility. Number of unhealed wounds was later revised to incorporate area of unhealed wounds: it was argued that a chronic erosion affecting the whole back was more serious than two or three smaller lesions confined to bony prominences.

Alopecia due to epidermolysis bullosa

This was included at a late stage, following discussion with colleagues in Germany who regularly record this feature, in order to retain some equivalence between our scoring systems. Alopecia is a feature of both RDEB and JEB.

Nutritional compromise

This was not included initially because of two perceived problems. Firstly, the definition of nutritional compromise differs between adults (reduced body mass index) and children (growth failure), so we devised a simple 5-point score (where 0 is normal and 5 is cachectic) that would cover all ages. Secondly, there was uncertainty about how to score seriously affected patients whose nutritional status had been corrected by gastrostomy. On the basis that our scoring system should detect improvement following intervention we agreed that the score should be based on current status.

The final scoring system was given the name Birmingham EB Severity (BEBS) score. A single-page A4-sized record sheet was devised for clinical use with the scoring table on one side and brief instructions on the other (Appendix 1). A body diagram was included to facilitate calculation of surface area affected, and this required a different form for adults and children.

Table 1 Patients included in the study: number and age for each subtype of epidermolysis bullosa (EB)

Subtype of EB	n	Age (years), median (range)
Weber–Cockayne EBS	12	17·4 (5·6–60·3)
Dowling–Meara EBS	16	10·2 (0·2–64·0)
Herlitz JEB	9	0·1 (0·0–0·8)
Non-Herlitz JEB (generalized)	7	10·2 (0·0–27·9)
Dominant DEB	19	16·5 (0·1–60·1)
Non-Hallopeau–Siemens recessive DEB	11	11·1 (0·8–64·0)
Hallopeau–Siemens recessive DEB	23	3·7 (0·0–34·0)
Total	97	

EBS, EB simplex; JEB, junctional EB; DEB, dystrophic EB.

Patient sample

Over the next 6 months the BEBS score was recorded in 97 patients, comprising 65 children (< 16 years) and 32 adults, with all major subtypes of EB, including the original 39 in whom the system had been developed (Table 1). Diagnosis of patients with severe EB was almost always based on IMF, usually backed up by mutation analysis particularly in sporadic cases. Diagnosis was usually clinical in milder cases, sometimes supported by mutation analysis or IMF.

Scoring episodes

Because of the rarity and severity of EB, and to avoid burdening patients with additional physical examinations, scoring was carried out opportunistically. Patients were scored by dermatologists (C.M. and A.W.) and EB nurses during routine clinical assessments at Birmingham Children's Hospital and Solihull Hospital, and in the community. All scorers participated in both development and testing of the system.

Score properties

The behaviour of the score was assessed by considering its association with known determinants of severity, namely type of EB and age.

Reliability testing

Intraobserver variability (test–retest)

Fifteen patients with stable DEB were scored by a single observer on two occasions 2–26 weeks apart and the coefficient of variation was calculated.

Interobserver variability

Forty patients were scored independently by more than one observer on the same episode of EB. Only one multirater episode per patient was included. The observers normally scored on the same day, but scorings within 1 month were allowed except in the case of patients with H-JEB for whom change is more rapid so only same-day scorings were allowed. Because of the variable numbers and differing identities of scorers used for each episode, most of the standard approaches for assessing inter-rater agreement/variability (for example, kappa coefficients, Cronbach's alpha) were not applicable. In order to analyse these data, we ignored the identity of the raters, just looking at the variability between scorers for each episode, as if raters occurred in random sets. The key notion is 'intraclass correlation coefficient' and its derivation from analysis of variance and variance components, 'intraclass' meaning 'inter-rater'.

Practical aspects

The final score sheet is shown in Appendix 1. The BEBS score was quick and easy to use and calculate. Training took only 15 min, scoring took < 5 min, and it could be administered in patients' homes.

Ethics

The Central Office of Research Ethics Committees confirmed that formal approval was not needed, as this work was classed as service development, not research. Informed consent was obtained from patients and/or parents.

Results

Development and test groups

Initially it was our view that the scoring system should be evaluated using only the episodes assessed after the scoring system had finally been decided (test group) and not to employ scores from the 39 patients used earlier to develop the score (development group). However, the distributions of scores in the two groups were very similar over all EB types (Fig. 1) so for all subsequent analyses we combined results for the first episode of all 97 patients.

Score distribution for different types of epidermolysis bullosa

The median scores for the different subtypes were: EBS-WC 1·0 (range 0·1–3·0); DDEB 5·3 (0·5–15·9); DM-EBS 6·3 (2·8–22·5); H-JEB 14·4 (2·5–49·3); non-Herlitz JEB (NH-JEB) 28·5 (5·0–62·0); non-Hallopeau–Siemens RDEB (NHS-RDEB) 7·8 (2·8–27·8), HS-RDEB 22·9 (4·3–69·0) (Fig. 1, Table 2). The range of scores shows that the BEBS score covers a continuum of disease severity, with no ceiling or floor effects. The distribution of scores is skewed by the patients with EBS-WC, most of whom scored 0–5. All scores were under 70, allowing for more severely affected patients than were seen in our population. The three lowest scores in NHS-RDEB occurred in siblings recognized as having a mild familial phenotype.

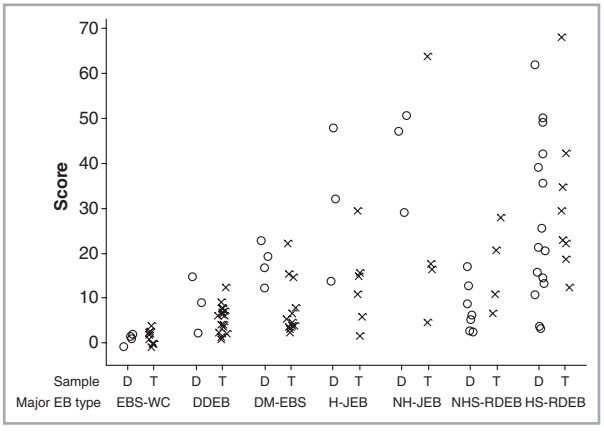


Fig 1. First score for all patients, against subtype of epidermolysis bullosa (EB), comparing development (D) (n = 39) and test (T) (n = 58) series. EBS-WC, Weber–Cockayne EB simplex; DDEB, dominant dystrophic EB; DM-EBS, Dowling–Meara EB simplex; H-JEB, Herlitz junctional EB; NH-JEB, non-Herlitz junctional EB; NHS-RDEB, non-Hallopeau–Siemens recessive dystrophic EB; HS-RDEB, Hallopeau–Siemens recessive dystrophic EB.

Score distribution for individual items

Scores for individual items were often skewed towards zero. In particular, none of our patients had skin cancer, and therefore this item scored zero in all cases. Some items always scored zero in a particular subtype, particularly EBS-WC. The distribution of scores for individual items is shown in Figures 2 and 3.

Association of score with age for different subtypes of epidermolysis bullosa

Figure 4 shows a scatterplot of total score with age at assessment with linear regression lines added for each EB type. There was a statistically significant correlation of score with age for H-JEB (r = 0·9, P = 0·001) and HS-RDEB (r = 0·73, P = 0·001), and positive but not significant correlation for EBS-WC (r = 0·21, P = 0·51), DDEB (r = 0·37, P = 0·11), NH-JEB (r = 0·67, P = 0·10) and NHS-RDEB (r = 0·23, P = 0·51). In the DM-EBS group there was a significant decline of score with age (r = −0·62, P = 0·01) (Table 3).

Reliability

Intraobserver variability

Fifteen patients with stable DEB were scored by a single observer on two occasions 2–26 weeks apart. Intraobserver coefficient of variation was 6·6%.

Table 2 Median (range) and distribution (%) of scores in different subtypes of epidermolysis bullosa (EB)

Score	EBS-WC	DDEB	DM-EBS	H-JEB	NH-JEB	NHS-RDEB	HS-RDEB	Overall
n	12	19	16	9	7	11	23	97
0–5	100	42·1	37·5	11·1	0	27·3	4·3	32·0
5–10	0	47·4	18·7	11·1	14·3	36·3	4·4	19·6
10–20	0	10·5	31·3	44·5	28·6	18·2	26·1	21·7
20–40	0	0	12·5	22·2	14·3	18·2	39·1	16·5
40–60	0	0	0	11·1	28·6	0	17·4	7·2
60–80	0	0	0	0	14·3	0	8·7	3·1
80–100	0	0	0	0	0	0	0	0·0
All	100·0	100·0	100·0	100·0	100·0	100·0	100·0	100·0
Median	1·0	5·3	6·3	14·4	28·5	7·8	22·9	8·4
(range)	(0·1–3·0)	(0·5–15·9)	(2·8–22·5)	(2·5–49·3)	(5·0–62·0)	(2·8–27·8)	(4·3–69·0)	(0·1–69·0)

EBS-WC, Weber–Cockayne EB simplex; DDEB, dominant dystrophic EB; DM-EBS, Dowling–Meara EB simplex; H-JEB, Herlitz junctional EB; NH-JEB, non-Herlitz junctional EB; NHS-RDEB, non-Hallopeau–Siemens recessive dystrophic EB; HS-RDEB, Hallopeau–Siemens recessive dystrophic EB.

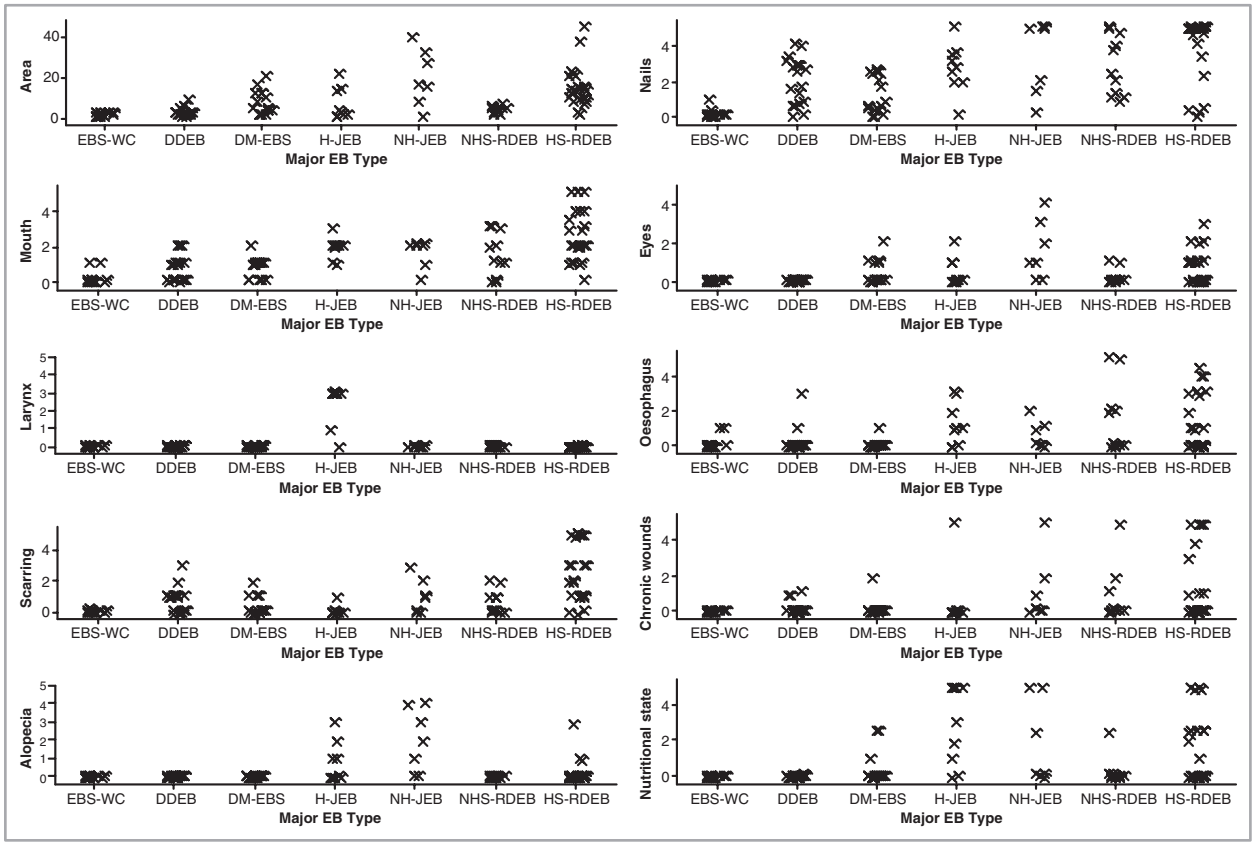


Fig 2. Individual item scores for first episodes of all patients, against subtype of epidermolysis bullosa (EB; n = 97). Cancer scores were 0 in all patients and are not shown. EBS-WC, Weber–Cockayne EB simplex; DDEB, dominant dystrophic EB; DM-EBS, Dowling–Meara EB simplex; H-JEB, Herlitz junctional EB; NH-JEB, non-Herlitz junctional EB; NHS-RDEB, non-Hallopeau–Siemens recessive dystrophic EB; HS-RDEB, Hallopeau–Siemens recessive dystrophic EB.

Interobserver variability

Eight individuals took part in scoring and there was substantial variability in the number and identity of raters used at each episode. Forty patients had episodes which were scored by two to five independent scorers, producing a total of 122 scores to assess inter-rater agreement. The raters used for each episode varied widely and have been treated as if occurring randomly. Analysis of variance was used to obtain the components of variance for between-patient variation and

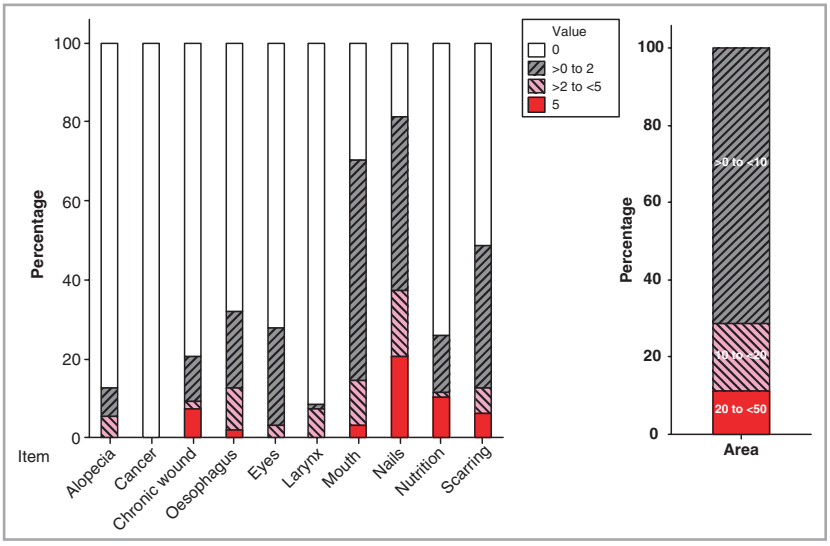


Fig 3. Percentage distribution of 97 scores for each item.

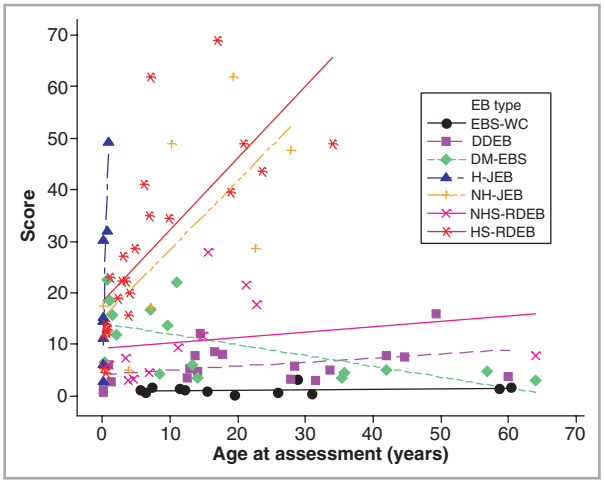


Fig 4. Scatterplot of score for different subtypes against age at assessment. EBS-WC, Weber–Cockayne EB simplex; DDEB, dominant dystrophic EB; DM-EBS, Dowling–Meara EB simplex; H-JEB, Herlitz junctional EB; NH-JEB, non-Herlitz junctional EB; NHS-RDEB, non-Hallopeau–Siemens recessive dystrophic EB; HS-RDEB, Hallopeau–Siemens recessive dystrophic EB.

within-patient variation which is here equivalent to inter-rater variation. Hence the intraclass correlation coefficient measuring inter-rater correlation was obtained. Table 4 also shows as a measure of inter-rater variability, the average range (maximum minus minimum) of raters' scores as a proportion of their median score. This was not an appropriate measure for cancer (which did not occur in our series), or chronic wounds and alopecia which occurred only once or twice in the multirater episodes. The large agreement on absence cannot meaningfully be compared with agreement on positive severity scores over the range.

Subjecting these 122 scores to analysis of variance, the components of variance and hence the correlation coefficients were derived for each item (Table 4).

Table 3 Correlation of score with age for different subtypes of epidermolysis bullosa (EB)

Subtype of EB	r-value	P-value
EBS-WC	0·21	0·51
DDEB	0·37	0·11
DM-EBS	−0·62	0·01
H-JEB	0·90	0·001
NH-JEB	0·67	0·10
NHS-RDEB	0·23	0·51
HS-RDEB	0·73	0·001

EBS-WC, Weber–Cockayne EB simplex; DDEB, dominant dystrophic EB; DM-EBS, Dowling–Meara EB simplex; H-JEB, Herlitz junctional EB; NH-JEB, non-Herlitz junctional EB; NHS-RDEB, non-Hallopeau–Siemens recessive dystrophic EB; HS-RDEB, Hallopeau–Siemens recessive dystrophic EB.

Table 4 Inter-rater correlation and variation for individual items scored

Item	Inter-rater correlation (%)	Inter-rater range/median
Nails	99·3	0·23
Mouth	86·0	0·61
Eyes	80·5	0·87
Larynx	92·3	0·10
Oesophagus	83·1	0·95
Scarring	81·7	1·10
Nutritional status	95·7	0·34
Area	88·7	0·49
Total score	95·5	0·35

All the inter-rater correlations are high. Also as most of the items are on a five-point scale, the fact that most of the inter-rater ranges as a proportion of their medians are < 1·0

means that on average raters are differing by < 1 scale point.

Discussion

The finding of Horn and Tidman, that severe EB impairs QoL more than any other skin disorder, suggests that the needs of these patients are not being met. Several interventions are already used routinely in EB, such as nutritional supplements, intravenous iron, amitryptyline and various topical agents and dressings. These appear to help individual patients although objective evidence of efficacy is lacking. New interventions are urgently required but must be assessed. The profound dearth of randomized controlled trials in EB is partly due to the lack of objective scoring systems. The rarity of EB is also a problem, but the new unified service for England and Wales now provides a sufficiently large cohort to test interventions. Another difficulty for EB clinical trials is recruitment: patients with a severe life-long disorder are naturally unwilling to make extra trips to hospital. A simple standard assessment that can be carried out in the patient's home is desirable and could facilitate patient monitoring and clinical trials, perhaps as a supplement to more specific outcome measures.

Precise diagnosis of EB, using IMF staining of a skin biopsy, is now routinely available for newborn affected babies in the U.K., facilitating accurate counselling and prognosis. However, there remains substantial clinical variability within diagnostic categories, even within families. A standardized severity score might help to characterize patients with unexpectedly mild or severe disease, and perhaps contribute to genotype–phenotype correlation.

This paper describes the development and validation of a severity score for EB. Overall, the BEBS score which we have devised meets many of the stated requirements. In particular, it is quick and easy to use and calculate and does not need special equipment or blood tests. It can be applied to all types of EB at any age. By consensus expert agreement it covers all the important manifestations of EB and gives scores that 'feel' correct to professionals (face and content validity). It appears reliable, giving consistent results within and between observers, but the heterogeneity of the data makes this difficult to demonstrate statistically. Changes in BEBS score with age reflect clinical observations, in particular disease progression with age in H-JEB and HS-RDEB, and improvement with age in DM-EBS. No patient scored higher than 70, reflecting the fact that our patient population could be considered young (mostly recruited from a children's hospital) and healthy (none had skin cancer): we can envisage worse affected, higher scoring patients than those included here. It remains to be seen whether the BEBS score will detect changes with treatment.

However, the BEBS score is far from perfect. Firstly, we acknowledge some circularity in our validation procedures. We devised a system which would give higher scores to severe subtypes, so it is not surprising that we found an association with subtype. Furthermore, we needed a substantial cohort of patients to develop the system and because of the rarity of EB we found it necessary to use the same group (with additional patients) to validate it.

The relative contributions of the different components (weighting) was essentially arbitrary. It could reasonably be argued that skin cancer, for example, is worth more points than loss of all fingernails, or that loss of sight from ocular damage should score more highly than total alopecia. However, for simplicity (an important practical consideration) all were given the same score. We cannot justify in statistical terms our allocation of 50% of the score to 'area affected', but this seems appropriate because area is given considerable weight in most other dermatological severity scoring systems, in EB it particularly contributes to the burdens of pain, dressings and morbidity, and area should be sensitive in the short to medium term to effects of treatment. The breadth of attributes scored, covering all types of EB, means that for most subtypes of EB certain attributes are redundant and will always score zero. For example, mutilating scarring of hands is usually redundant in all types other than RDEB, and most attributes other than area are redundant in EBS-WC. Equal contribution of individual items to the overall score is termed 'internal consistency' and is lacking in our system. Low internal consistency is inevitable for a score devised to cover such a heterogeneous disorder, and may limit its sensitivity for any one type of EB.

Reliability testing was also limited. Intraobserver variability was only tested for one observer. Interobserver variability has not been explored as systematically as would be desirable. Because of the extreme suffering of patients with severe EB, and the need to remove all dressings in order to score patients, the study was conducted in an opportunistic way, when the patient was being assessed for clinical purposes, often at home by their allocated nurse. All this means that it is essential to validate it at other centres in other patients.

However, on balance the BEBS scoring system, on preliminary testing, appears practical, valid and reliable. Further testing in more patients, with a wider range of age, severity and complications, in different specialist EB clinics, is planned. We have not yet shown that it is sensitive enough to detect changes with treatment, but we hope this scoring system will be a useful tool in clinical practice and research.

Acknowledgments

We are grateful to Professors Robin Eady and Leena Bruckner-Tudermann, and Dr Carol Cummins for encouragement and helpful discussion, and to numerous clinical colleagues for expert advice and for piloting various versions of the BEBS score, in particular, Doctors Jemma Mellerio, Anna Martinez, Adrian Heagerty and Hauke Schumann, and nurses Ruth Ward, Natalie Hadley, Emma McAndrew, Dawn James, Sheila Richards, Tracy Adni and DebRA staff. We also acknowledge the thoughtful and constructive comments of the *BJD* reviewers. Dr Alvin Wong was funded by DebRA.

References

1 Fine JD, Eady RA, Bauer EA *et al.* The classification of inherited epidermolysis bullosa (EB): Report of the Third International Consensus Meeting on Diagnosis and Classification of EB. *J Am Acad Dermatol* 2008; **58**:931–50.

2 Berth-Jones J, Grotzinger K, Rainville C *et al.* A study examining inter- and intrarater reliability of three scales for measuring severity of psoriasis: Psoriasis Area and Severity Index, Physician's Global Assessment and Lattice System Physician's Global Assessment. *Br J Dermatol* 2006; **155**:707–13.

3 Pucci N, Novembre E, Cammarata MG *et al.* Scoring atopic dermatitis in infants and young children: distinctive features of the SCORAD index. *Allergy* 2005; **60**:113–16.

4 Katugampola RP, Finlay AY. Impact of skin diseases on quality of life. *Eur J Dermatol* 2007; **17**:102–6.

5 Horn HM, Tidman MJ. Quality of life in epidermolysis bullosa. *Clin Exp Dermatol* 2002; **27**:707–10.

6 Tamai K, Hashimoto I, Hanada K *et al.* Japanese guidelines for diagnosis and treatment of junctional and dystrophic epidermolysis bullosa. *Arch Dermatol Res* 2003; **295** (Suppl. 1):S24–8.

7 Bland JM, Altman DG. Statistics notes: validating scales and indexes. *BMJ* 2002; **324**:606–7.

8 Charman CR, Venn AJ, Williams HC. Measurement of body surface area involvement in atopic eczema: an impossible task? *Br J Dermatol* 1999; **140**:109–11.

9 Miller SF, Finley RK, Waltman M, Lincks J. Burn size estimate reliability: a study. *J Burn Care Rehabil* 1991; **12**:546–59.

Appendix 1

The Birmingham Epidermolysis Bullosa Severity Score Sheet (child version)

Patient's name..... DOB..... Type of EB.....

Scorer's name..... Date.....

**See overleaf for detailed instructions*

Score item	Measure	Max	Actual score
*Nails	Lost nails ÷ 4 Dystrophic nails ÷ 8	5	
*Area	½ × % damaged skin: blisters, erosions, scabs, healing skin, erythema, atrophic scarring; **not dyspigmentation, or well-healed scars**	50	
*Mouth	0 = no mucosal involvement 1 = occasional blisters/erosions 2 = frequent blisters 3 = persistent symptoms, early structural abnormality 4 = moderate structural abnormality 5 = severe structural abnormality (see over for detailed scoring for each site)	5	
*Eyes		5	
*Larynx		5	
*Oesophagus		5	
Scarring of hands	0 = no scarring 1 = milia and/or atrophic scars 2 = just detectable contractures or webbing 3 = obvious contractures, or proximal webbing 4 = between 3 and 5 5 = mitten formation with fingers all fused	5	
Skin cancer (SCC)	Number of skin cancers + 1 for local/regional/lymph node spread + 2 for distant metastatic spread, up to maximum score 5	5	
Chronic wounds present for > 6/12	0 = none 1 = < 1% body surface area (1% = palm size) 2 = 1–2% 3 = 2–5% 4 = 5–10% 5 = > 10%	5	
Scarring alopecia due to EB	0 = no alopecia 1 = 1–19% scalp involvement 2 = 20–39% 3 = 40–59% 4 = 60–79% 5 = 80–100%	5	
Nutritional compromise	0–5 (where 0 = normal and 5 = cachectic)	5	
Total score		**100**	

How to fill in the BEBS score sheet

Nails: enter number in each box and add up horizontally

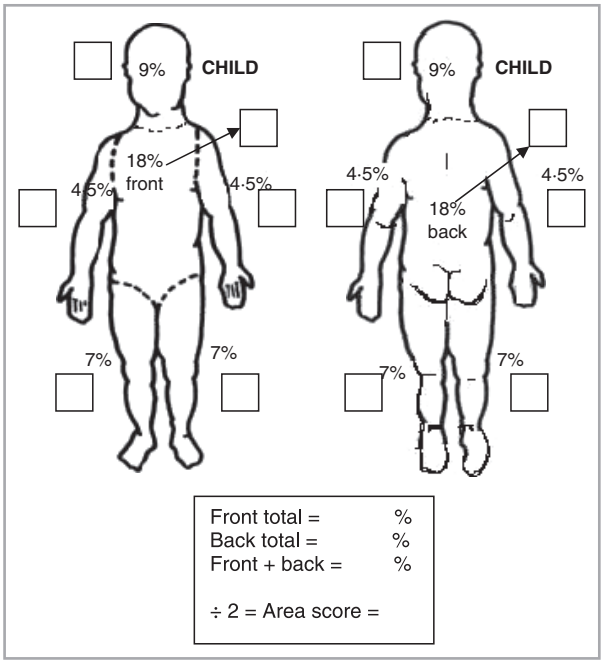
	R hand	L hand	R foot	L foot	Subtotals A	Subtotals B	Total score
Lost nails	+	+	+	+	=	÷ 4 =	} =
Dystrophic nails	+	+	+	+	=	÷ 8 =	
Normal nails							
Total	5	5	5	5			

Area:

Please shade in affected areas on the diagram, then work out percentage for each part and fill in the numbers in the adjacent boxes.

e.g. if half of the anterior trunk is affected, then put 9% in the box on anterior trunk.

Patient's palm size area corresponds to 1% of total body surface area.



Mouth, eyes, larynx, oesophagus: detailed scoring

	Mouth	Eyes	Larynx	Oesophagus
0	No problem from EB	No problem from EB	No problem from EB	No problem from EB
1	Occasional soreness	Occasional soreness	Occasional hoarseness	Occasional dysphagia
2	Frequent soreness	Frequent soreness	Frequent hoarseness	Frequent dysphagia
3	Persistent soreness Just detectable tongue tethering	Persistent soreness Early visible external eye disease	Persistent hoarseness	Persistent dysphagia
4	Between 3 and 5	Between 3 and 5	Between 3 and 5	Between 3 and 5
5	Severe tongue tethering and microstomia	Bilateral sight-threatening eye disease	Life-threatening laryngeal obstruction	Difficulty swallowing solids and liquids