

## CPD

# On the definition of dermatological disease. Part 2: approaches for defining dermatological diseases

Hywel C. Williams<sup>1</sup>  and Esther Burden-Teh<sup>1</sup> 

<sup>1</sup>Centre of Evidence Based Dermatology, Population and Lifespan Sciences, School of Medicine, University of Nottingham, Nottingham, UK

doi:10.1111/ced.15278

## Abstract

In Part 1 of this two-part review, conceptual frameworks for defining skin diseases were articulated. In this review, the main approaches that can be used to develop diagnostic criteria for skin disease are summarized, using atopic dermatitis (AD) as an example. Different frameworks for defining skin disease for research purposes are articulated, including statistical, prognostic, operational, clinical and epidemiological approaches. All share the common aim of attempting to develop criteria that enable meaningful comparisons between groups of people. The desirable attributes of a good definition are described: diagnostic criteria should measure what they are meant to measure; the results should be the same for different assessors; the criteria should be coherent with what is known about that disease; they should reflect some degree of morbidity and not pick up subclinical disease; they should be easy to administer; and they should be applicable to a range of people of different ages, sexes/genders and ethnicities. Consensus-based criteria are contrasted with epidemiological derivation methods that assess the performance of diagnostic criteria in relation to a reference standard. The sensitivity and specificity of a disease definition is explained, along with how the trade-off between these two properties can vary, depending on the purpose of the study and the study setting. The review closes with some reflections on when it is appropriate to consider splitting a disease into more than one category and how diagnostic criteria can be interpreted in the clinical setting.

## Introduction

Part 1 of this review emphasized the importance of disease definition in dermatology.<sup>1</sup> Conceptual frameworks were described, along with the hazards of underdiagnosis and overdiagnosis. What was not covered is *how* diagnostic criteria emerge – do they just appear as dermatologists declare an ‘entity’ based on a constellation of morphological and histological features?<sup>2</sup> This part of the review explores approaches for

defining skin diseases for the purpose of making comparisons for clinical practice and research.

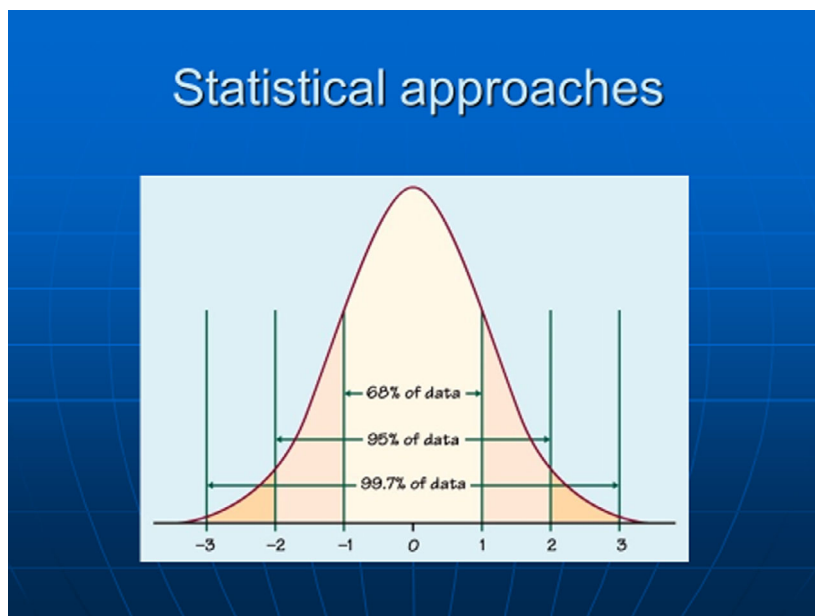
## Diverse approaches that can be used for defining skin disease

### Statistical, prognostic and operational approaches

In the first part of this review, it was explained how many skin diseases, including tinea capitis and AD can be considered as a continuum rather than an abrupt yes/no dichotomy.<sup>3</sup> For disease states measured in units such as ‘hypertension’ or ‘hypercalcaemia’, statistical approaches using arbitrary cutoff points representing 2 standard deviations from a normal distribution are often used (Fig. 1). Such approaches define

Correspondence: Professor Hywel C. Williams, Centre of Evidence Based Dermatology, School of Medicine, University of Nottingham, Applied Health Research Building, University Park, Nottingham NG7 2RD, UK  
E-mail: hywel.williams@nottingham.ac.uk

Accepted for publication 25 May 2022



**Figure 1** Statistical cutoffs based on 2 standard deviations from the mean are often used for potential disease states such as hypercalcaemia in order to define deviant states.

deviants from the mean, but presume that the prevalence of all 'disease' is 5%. Cutoff values may not be associated with a sharp rise in end organ disease, the risk of which is rarely linear.<sup>4</sup> Population norms also vary, e.g. hypertension thresholds may differ for older people with comorbidities.

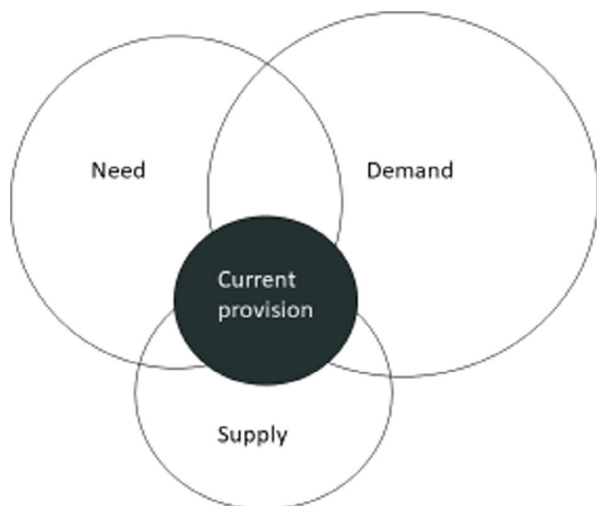
Prognostic approaches, i.e. states associated with impaired outcomes, can be also used to suggest disease definitions. Many mild forms of transient eczematous eruptions in infancy may not progress into AD and are therefore probably best not 'labelled' as such.<sup>5</sup> Prognostic classification is useful for staging established diagnoses such as melanoma<sup>6</sup> or cutaneous T-cell lymphoma,<sup>7</sup> both of which can exhibit a range of consequences from very low risk of spread to death. Frameworks for prognostic research are described elsewhere.<sup>8</sup>

Operational definitions, whereby action (in the form of cost-effective treatment) is preferred to inaction, are highly dependent on the available resources and competing needs. These definitions are likely to change over time as new treatments become available or costs of existing treatments change. Operational definitions are typically used for rationing access to costly medicines, such as the National Institute for Health and Care Excellence guidance for use of biologic therapies in psoriasis.<sup>9</sup> The relationship between dermatological need (ability to benefit from medical care), demand (that which is asked for) and supply (Fig. 2) varies by skin disease and has been discussed in detail in a UK dermatology healthcare needs assessment.<sup>10</sup>

### Clinical approaches

The visual nature of skin diseases facilitates simple descriptive approaches based on distribution, configuration and morphology. Such descriptive approaches, supplemented with histology findings or circulating blood abnormalities, are appropriate in defining disease for clinical work, especially if they become refined as more becomes known about that disease (progressive nosology, described in Part 1).<sup>1</sup> However, problems arise when groups of people need to be compared in clinical research as some clinical signs may not be present in some patients, or histology findings may be supportive but not diagnostic. Definitions using criteria that can consistently define similar groups of people are needed.

Consensus-based definitions, produced when a group of experts from a country or group of countries get together and suggest diagnostic criteria, are popular. Although the need for consensus implies lack of consensus,<sup>11</sup> it is a convenient first step in disease definition, especially when disease knowledge is scanty. Older consensus criteria often involved a round table of experts proposing cardinal features for a skin disease. The Hanifin and Rajka diagnostic criteria for AD are a good example of such criteria suggested by experts (Table 1). They represented a major milestone in describing the phenotype of AD for clinical practice, but the list of all possible markers is too cumbersome for comparing large groups of people, and features such as 'tendency towards cutaneous infection' are inadequately defined.



**Figure 2** The inter-relationship between dermatological need (that which can benefit from medical intervention), demand (what the public ask for) and supply (based on available state and private facilities) is a complex one that changes over time and by condition and country.

The concept of what a typical case of a common disease such as AD looks like may differ between countries, so it is essential that consensus is conducted at a global level if the emergent criteria are to be used globally. Methodologically rigorous consensus approaches using formal Delphi consensus methods and qualitative work with patients and carers are now used. Several such consensus criteria<sup>12–14</sup> have been developed for skin diseases and are generally a good starting point for making comparisons.

**Epidemiological approaches**

Epidemiological approaches offer a scientific approach to reducing a long list of all possible diagnostic features into a minimum list of reliable discriminators. Such an approach was used in developing the UK refinement of the Hanifin and Rajka criteria.<sup>15</sup> Using a case–control approach, the frequency, validity and reliability of symptoms and signs of AD in typical cases and controls with other skin diseases were evaluated in a blinded fashion<sup>16–18</sup> (Table 2). Other researchers have used similar methodological approaches successfully.<sup>19</sup>

**Sensitivity and specificity**

At this point, it is worth introducing the concept of sensitivity (proportion of people with the condition of

**Table 1** The Hanifin and Rajka diagnostic criteria for atopic dermatitis.

Major criteria (≥ 3 required)	Minor criteria (≥ 3 required)
1 Pruritus	1 Xerosis
2 Typical morphology and distribution: Flexural lichenification or linearity in adults Facial and extensor involvement in infants and children	2 Ichthyosis, palmar hyperlinearity, or keratosis pilaris
3 Chronic or chronically relapsing dermatitis	3 Immediate (type 1) skin-test reactivity
4 Personal or family history of atopy (asthma, allergic rhinitis, atopic dermatitis)	4 Raised serum IgE
	5 Early age of onset
	6 Tendency toward cutaneous infections (especially <i>Staphylococcus aureus</i> and herpes simplex) or impaired cell-mediated immunity
	7 Tendency toward nonspecific hand or foot dermatitis
	8 Nipple eczema
	9 Cheilitis
	10 Recurrent conjunctivitis
	11 Dennie–Morgan infraorbital fold
	12 Keratoconus
	13 Anterior subcapsular cataracts
	14 Orbital darkening
	15 Facial pallor or facial erythema
	16 Pityriasis alba
	17 Anterior neck folds
	18 Itch when sweating
	19 Intolerance to wool and lipid solvents
	20 Perifollicular accentuation
	21 Food intolerance
	22 Course influenced by environmental or emotional factors
	23 White dermographism or delayed blanch

**Table 2** The UK refinement of the Hanifin and Rajka criteria, which reduced the original long list of features to a minimal list of reliable discriminators.<sup>18</sup>

UK diagnostic refinement of Hanifin and Rajka criteria	
Must have:	
Itchy skin plus ≥ 3 of the following:	
1	History of flexural involvement
2	Concurrent asthma or hay fever
3	Early onset
4	Generally dry skin
5	Visible flexural dermatitis

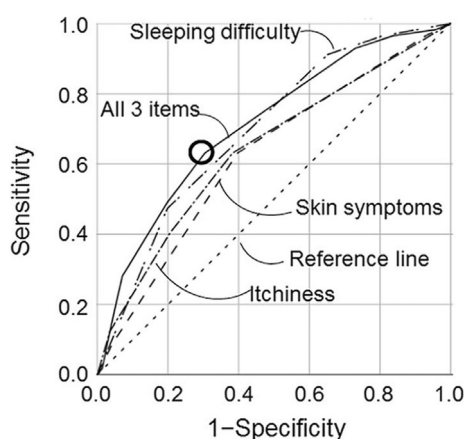
interest identified correctly) and specificity (proportion of unaffected people who are correctly identified) as shown in Fig. 3. Sensitivity and specificity are

Sensitivity = proportion of true positives correctly identified =  $a/a+c$

Specificity is proportion of true negatives correctly identified =  $d/b+d$

Test	Disease reference standard		Totals
	positive	negative	
+ve	a	b	a + b
-ve	c	d	c + d
Totals	a + c	b + d	a + b + c + d

**Figure 3** Sensitivity and specificity of a diagnostic test.



**Figure 4** An example of a receiver operator characteristic curve, plotting the sensitivity of different combinations of symptoms against false-positive rates as part of development of a paediatric dermatology screening tool for atopic dermatitis.<sup>20</sup>

determined in relation to a reference or ‘gold’ standard. For many skin diseases such as AD, no reliable objective reference is available, so expert clinical diagnosis is usually relied upon as a reference standard. Such standards are prone to subjective variation, which is why good agreement between those acting as reference standards must be demonstrated.<sup>17</sup> There is always a trade-off between sensitivity and specificity, which is often depicted in a receiver operating characteristic curve that plots sensitivity against false-positive rates (Fig. 4) to explore optimum thresholds for diagnostic performance.<sup>20</sup> Although the ‘sweet spot’ combining the best sensitivity and lowest false-positive rate may appear to be the obvious threshold choice, the optimum trade-off between sensitivity vs. specificity varies by study purpose (Fig. 5). For a smartphone app that purports to identify possible melanoma, it is

imperative melanomas are not missed (false negatives), so high sensitivity is paramount.<sup>21</sup> Negative results of a highly sensitive test are good at *ruling out* that disease, whereas a positive result for a very specific test tends to *rule in* that disease.<sup>22</sup> Diagnostic test measures, including predictive values, likelihood ratios and diagnostic reasoning, have been reviewed elsewhere.<sup>23–25</sup>

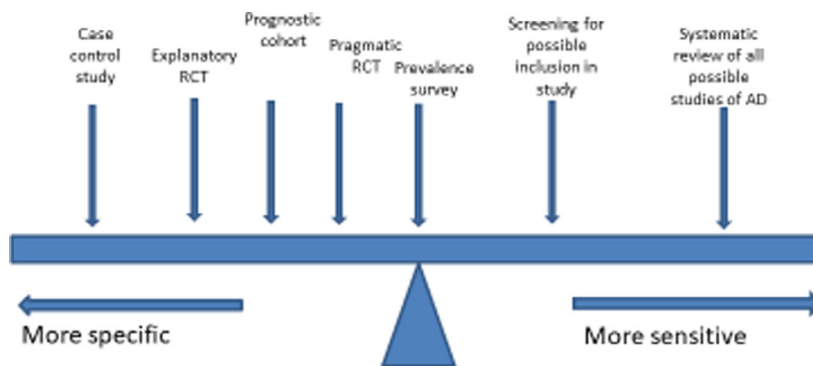
The attributes of good diagnostic criteria are shown in Table 3. Once criteria have been developed, it is important to replicate validation studies in different countries, and include people of different ages and ethnicities in order to ensure adequate external validity (Table 4).<sup>26</sup>

### Splitting diseases

Dividing skin diseases such as ‘atopic’ dermatitis into atopic and nonatopic types is of little value unless such divisions confer additional predictive value in the clinic and research.<sup>27</sup> Various methods such as latent class analysis are used to suggest that some skin conditions such as hidradenitis suppurativa are composed of several distinct subtypes.<sup>28</sup> Such splits need to be replicated<sup>29</sup> in order to minimize emergence of pseudoclusters. The utility of such splits requires evaluation before they are assimilated into practice, e.g. by conducting exploratory subgroup analyses within therapeutic trials.

### Conclusion

This review describes approaches used to define skin disease. New approaches using artificial intelligence for refining diagnostic criteria and the widespread use of apps are evolving rapidly, and may influence the trade-off between sensitivity and specificity, depending



**Figure 5** Schematic illustration of the different requirements of the trade-off between sensitivity and specificity of diagnostic criteria for atopic dermatitis (AD) in relation to different study designs. RCT, randomized controlled trial.

**Table 3** Attributes of good diagnostic criteria for comparing groups in research.

Must be valid, i.e. must measure what they are meant to measure
Repeatable: one person using the criteria should come up with the same answer as when another person uses them
Reflect some degree of morbidity: needed to avoid subclinical disease such as defining acne on the basis of a few comedones, which is physiological during puberty
Coherent with clinical concepts: they should correspond with what dermatologists consider the condition to be
Simple to use: essential when describing large numbers of people
Good applicability to different people, including skin colour, older or very young skin, and that are culturally sensitive

**Table 4** Good practice recommendations for repeating validation studies of diagnostic criteria.<sup>26</sup>

The criteria questions and examination protocols should be used exactly as recommended and not modified
Translation should follow recommended guidance and the procedure should be documented
The criteria should be tested in the setting in which they are meant to be used
Those acting as assessors for criteria or who are acting as reference standards for the clinical diagnosis should be blinded to the criteria being tested
There should be some reassurance that the clinician(s) acting as the reference standard are comparable with other international clinicians in their concept of what constitutes that condition

on the setting in which they are used. In a clinical setting, diagnostic criteria are a *guide* to clinical diagnosis, whereas in research, they are *essential* in order to permit standardized comparisons. All diagnostic criteria should be developed using a clear and explicit methodology and replicated in independent studies. The methods used may progress over time from simple consensus<sup>30</sup> to epidemiological approaches.<sup>19</sup> The

utility of such criteria in predicting important implications to patients such as prognosis or treatment response should also be articulated before they are declared as being useful.

### Learning points

- Diagnostic criteria are only guides in clinical practice whereas they are essential for research purposes so that similar groups can be compared.
- Skin diseases can be defined by various approaches, ranging from simple consensus to case-control and cluster analysis methods.
- Consensus criteria for a skin disease run the risk of determining the 'eternal truth' by a show of hands, and need to be revisited periodically in the light of new evidence.
- Skin disease definitions must have demonstrable validity and applicability to a range of people of different ages, sex/genders and ethnicities.
- The trade-off between sensitivity and specificity of diagnostic criteria will vary according to their purpose.
- Skin diseases should not be split prematurely on epiphenomena unless the new divisions are shown to be stable and useful for clinical and research purposes.

### Conflict of interest

HCW led the development of the UK diagnostic criteria for atopic dermatitis. EB-T led the development of diagnostic criteria for psoriasis in children and young people.



## Funding

EB-T is funded as a National Institute for Health and Care Research Clinical Lecturer.

## Ethics statement

Ethics approval and informed consent not applicable.

## Data availability

Data are available on request from the corresponding author.

## References

- Williams HC. On the definition of dermatological disease. Part 1: conceptual frameworks. *Clin Exp Dermatol* 2022. Online ahead of print. <https://doi.org/10.1111/ced.15279>
- Williams HC, Ashworth J, Pembroke AC, Breathnach SM. FACE – facial Afro-Caribbean childhood eruption. *Clin Exp Dermatol* 1990; **15**: 163–6.
- Diepgen TL, Fartasch M, Hornstein OP. Evaluation and relevance of atopic basic and minor features in patients with atopic dermatitis and in the general population. *Acta Derm Venereol Suppl (Stockh)* 1989; **144**: 50–4.
- Malyszko J, Muntner P, Rysz J, Banach M. Blood pressure levels and stroke: J-curve phenomenon? *Curr Hypertens Rep* 2013; **15**: 575–81.
- Endre KMA, Landrø L, LeBlanc M *et al*. Diagnosing atopic dermatitis in infancy using established diagnostic criteria: a cohort study. *Br J Dermatol* 2022; **186**: 50–8.
- Dickson PV, Gershenwald JE. Staging and prognosis of cutaneous melanoma. *Surg Oncol Clin N Am* 2011; **20**: 1–17.
- Scarisbrick JJ. Staging and management of cutaneous T-cell lymphoma. *Clin Exp Dermatol* 2006; **31**: 181–6.
- Kent P, Cancelliere C, Boyle E *et al*. A conceptual framework for prognostic research. *BMC Med Res Methodol* 2020; **20**: 172.
- National Institute for Health and Care Excellence. Systemic biological therapy for psoriasis. Available at: <https://www.nice.org.uk/guidance/ta134> (accessed 18 June 2022).
- Schofield J, Grindlay D, Williams HC. Skin conditions in the UK: a health care needs assessment. Available at: <https://www.nottingham.ac.uk/research/groups/cebdl/documents/hcnaskinconditionsuk2009.pdf> (accessed 6 March 2022).
- Skrabanek P. Nonsensus consensus. *Lancet* 1990; **335**: 1446–7.
- Simpson RC, Thomas KS, Leighton P, Murphy R. Diagnostic criteria for erosive lichen planus affecting the vulva: an international electronic-Delphi consensus exercise. *Br J Dermatol* 2013; **169**: 337–43.
- Zouboulis CC, Del Marmol V, Mrowietz U *et al*. Hidradenitis suppurativa/acne inversa: criteria for diagnosis, severity assessment, classification and disease evaluation. *Dermatology* 2015; **231**: 184–90.
- Maverakis E, Ma C, Shinkai K *et al*. Diagnostic criteria of ulcerative pyoderma gangrenosum: a Delphi consensus of international experts. *JAMA Dermatol* 2018; **154**: 461–6.
- Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol (Stockh)* 1980; (Suppl. 92): 44–7.
- Williams HC, Burney PG, Hay RJ *et al*. The U.K. Working Party's diagnostic criteria for atopic dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. *Br J Dermatol* 1994; **131**: 383–96.
- Williams HC, Burney PG, Strachan D, Hay RJ. The U.K. Working Party's diagnostic criteria for atopic dermatitis. II. Observer variation of clinical diagnosis and signs of atopic dermatitis. *Br J Dermatol* 1994; **131**: 397–405.
- Williams HC, Burney PG, Pembroke AC, Hay RJ. The U.K. Working Party's diagnostic criteria for atopic dermatitis. III. Independent hospital validation. *Br J Dermatol* 1994; **131**: 406–16.
- Burden-Teh E, Murphy R, Gran S *et al*. Identifying the best predictive diagnostic criteria for psoriasis in children (< 18 years): a UK multicentre case-control diagnostic accuracy study (DIPSOC study). *Br J Dermatol* 2022; **186**: 341–51.
- Sato H, Goto A, Murakami M, Kawabata Y. Development of a pediatric dermatology screening tool based on two parent-reported skin symptoms: comparison of parental recognition and physician diagnosis of skin symptoms of infants and toddlers. *J Prim Care Community Health* 2020; **11**: 2150132720974883.
- Freeman K, Dinnes J, Chuchu N *et al*. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ* 2020; **368**: m127.
- Power M, Fell G, Wright M. Principles for high-quality, high-value testing. *Evid Based Med* 2013; **18**: 5–10.
- Farkas J. PulmCrit – Mythbusting sensitivity and specificity. Available at: <https://emcrit.org/pulmcrit/mythbusting-sensitivity-specificity/> (accessed 25 February 2022).
- Brush JE Jr, Sherbino J, Norman GR. Diagnostic reasoning in cardiovascular medicine. *BMJ* 2022; **376**: e064389.
- Shreffler J, Huecker MR. Diagnostic testing accuracy: sensitivity, specificity, predictive values and likelihood ratios 2021. In: *StatPearls*. Treasure Island, FL: StatPearls Publishing, 2022. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK557491/> (accessed 7 June 2022).
- Williams HC. Diagnostic criteria for atopic dermatitis: where do we go from here? *Arch Dermatol* 1999; **135**: 583–6.
- Williams H, Flohr C. How epidemiology has challenged 3 prevailing concepts about atopic dermatitis. *J Allergy Clin Immunol* 2006; **118**: 209–13.

- 28 Canoui-Poitrine F, Le Thuaut A, Revuz JE *et al.* Identification of three hidradenitis suppurativa phenotypes: latent class analysis of a cross-sectional study. *J Invest Dermatol* 2013; **133**: 1506–11.
- 29 Marzano AV, Patrizi A, Atzori L *et al.* Characterization of hidradenitis suppurativa phenotypes: a multidimensional latent class analysis of the National Italian Registry IRHIS. *J Invest Dermatol* 2021; **141**: 1236–42.e1.

## CPD questions

### Learning objective

To gain up-to-date knowledge on methods and interpretation of diagnostic criteria for skin diseases.

### Question 1

What does the validity of diagnostic criteria mean?

- (a) The extent to which they come up with the same result when different patients are examined.
- (b) The extent to which they come up with the same result when different physicians examine the same patient.
- (c) The extent to which they measure what they are meant to measure.
- (d) That the criteria are very sensitive.
- (e) That the criteria can be used in infants.

### Question 2

Which of the following statements about diagnostic criteria for a skin disease is correct?

- (a) They must be fulfilled in order to treat patients with a suspected diagnosis of that condition.
- (b) They are usually 100% sensitive and specific.
- (c) They must only include physical signs.
- (d) One validation study is usually enough.
- (e) Valid criteria are essential for research purposes so that similar groups of people can be compared in different studies.

### Question 3

Which of the following statements describes the sensitivity of diagnostic criteria?

- (a) The extent to which the criteria are sensitive to patient's needs.
- (b) The proportion of borderline cases correctly identified.

- 30 Burden-Teh E, Thomas KS, Gran S, Murphy R. Development of clinical diagnostic criteria for plaque psoriasis in children: an electronic Delphi consensus study with the International Psoriasis Council. *Br J Dermatol* 2019; **181**: 856–7.

- (c) The probability that someone has a disease.
- (d) The proportion of true cases that are correctly identified.
- (e) The proportion of noncases that are correctly identified.

### Question 4

Which of the following statements describes consensus-based diagnostic criteria?

- (a) They are correct and permanent.
- (b) They should ideally involve international consensus from a range of professionals using formal Delphi methods and qualitative research with patients.
- (c) They should not change as new understanding of a disease emerges.
- (d) They are best done at a national level.
- (e) They are developed using a case-control analytical study.

### Question 5

In relation to validation studies for established diagnostic criteria, which of the following statements is correct?

- (a) Dermatologists acting as reference standards for the clinical diagnosis of the criteria should be blinded to the content and results of the criteria being tested.
- (b) It is OK to use just one expert dermatologist as the reference standard.
- (c) It is acceptable to modify the criteria slightly according to local needs.
- (d) It is acceptable to change the order of questions.
- (e) Translation of the criteria can be done by someone who speaks the language.

## Instructions for answering questions

This learning activity is freely available online at <http://www.wileyhealthlearning.com/ced>

Users are encouraged to

- Read the article in print or online, paying particular attention to the learning points and any author conflict of interest disclosures.
- Reflect on the article.
- Register or login online at <http://www.wileyhealthlearning.com/ced> and answer the CPD questions.
- Complete the required evaluation component of the activity.

Once the test is passed, you will receive a certificate and the learning activity can be added to your RCP CPD diary as a self-certified entry.

This activity will be available for CPD credit for 2 years following its publication date. At that time, it will be reviewed and potentially updated and extended for an additional period.