Pruritus Assessment in Clinical Trials: Consensus Recommendations from the International Forum for the Study of Itch (IFSI) Special Interest Group Scoring Itch in Clinical Trials

Sonja STÄNDER1, Matthias AUGUSTIN2, Adam REICH1, Christine BLOME2, Toshiya EBATA4, Ngoc Quan PHAN1 and Jacek C. SZEPIETOWSKI1
1Department of Dermatology and Competence Center Chronic Pruritus (KCP), University Hospital Münster, Münster; 2Center for Dermatological Research (CeDeF), Health Economics and Qol. Research Group, German Center for Health Services Research in Dermatology (CVderm), University Clinics of Hamburg, Hamburg, Germany; 3Department of Dermatology, Venereology and Allergology, Wroclaw Medical University, Wroclaw, Poland, and 4Department of Dermatology, Jikei University School of Medicine, Chitofuna Dermatology Clinic, Tokyo, Japan

Chronic pruritus is a common symptom and there is an urgent need to test new anti-pruritic substances in high-quality clinical trials. However, no widely accepted standardized and validated method for objectively measuring pruritus is yet available. A special interest group of the International Forum for the Study of Itch has been established to assess scoring methods and questionnaires for use in clinical trials. This paper presents our current recommendations. The set of measures we recommend includes pruritus intensity scales, instruments for assessment of scratch lesions, chronic pruritus course, quality of life and patient benefits. Key words: itch; pruritus; visual analogue scale; patient benefit index; quality of life; data bank.

Accepted Feb 14, 2013; Epub ahead of print Apr 25, 2013
Sonja Ständer, Competence Center Chronic Pruritus, Department of Dermatology, University Hospital Münster; Von-Esmarch-Str. 58, DE-48149 Münster, Germany. E-mail: sonja.staender@uni-muenster.de

Novel anti-pruritic therapies are urgently needed, especially for certain types of pruritus, such as aquagenic pruritus, prurigo nodularis, atopic dermatitis and cutaneous T-cell lymphoma-related pruritus. In addition to the generation of new hypotheses regarding clinically relevant therapy targets in these disease entities, investigators have to face another challenge. Pruritus is a subjective symptom and cannot be measured objectively. The assessment of the anti-pruritic effect is, to date, based solely on patient reports on the course of itch or measurements of scratch movements (4). The aim of the Special Interest Group (SIG) Scoring Itch in Clinical Trials of the International Forum for the Study of Itch (IFSI; www.itchforum.net) is to define a robust set of valid measures to assess pruritus in clinical studies. We aim to improve the quality of the outcome measures, fulfill the quality criteria of the European Medicines Agency (EMEA) and Food and Drug Administration (FDA) regarding patient-reported outcomes, and make it possible to compare results across studies.

To avoid the development of duplicate instruments and to achieve a broad consensus on the choice and employment of available instruments, the SIG was founded in 2009. It comprises a panel of international experts and collaborates with the Japanese Society of Itch and the German working group on pruritus research (some members are the same in both groups) (4–7). Some of the available instruments have already been validated (4, 5) and a questionnaire for the assessment of pruritus in clinical trials has been proposed (8). In this consensus paper, currently available instruments are presented and our first recommendation on a valid set of assessment tools for clinical trials is provided. The SIG will continue to study this subject and will welcome criticism and input during what we expect will be a continuous process of revision and improvement. We are aware of the fact that, for many tools, no validation studies have been performed and versions in other languages are not available yet.

LEGAL REQUIREMENT FOR STUDY DESIGN: INCLUSION OF PATIENT-REPORTED OUTCOMES

It is strongly recommended by nearly all health authorities to include patient-reported outcomes (PRO) into...
the study design of randomized controlled trials (RCT). The aim of including PRO is to demonstrate a clinical treatment benefit of the new substance from the patient’s perspective (9). PRO are defined as being “any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else” (9). This includes, for example, satisfaction with the therapy, subjective benefit or QoL improvement (9). Thus, PRO comprise single and multi-dimensional measures of symptoms, health-related quality of life (HRQoL), health status, adherence to treatment, and satisfaction with treatment (10). The outcome, as reported by the patients, can be measured in 2 ways: in absolute terms (e.g. intensity scales) or as a change from a previous measure (e.g. comparison of QoL scores) (9). Some single dimensional PRO, such as assessment of change of the core symptom (e.g. by itch intensity scales) may serve as efficacy end-points of the RCT (10). The EMEA has published a reflection paper on regulatory guidance for the use of HRQoL measures in the evaluation of medicinal products (10). According to the EMEA, HRQoL assessment is optional and, because of its multi-dimensionality, cannot serve as the primary end-point. In chronic life-threatening conditions, such as CP, however, HRQoL information may be important for detection of any impact of the symptom on the daily life and social functioning of patients and changes in these parameters during the RCT. Furthermore, such information would enable the attending physician to make a choice between the medicinal products available (10). In this consensus paper, we have taken into account the various health authority requirements and have included several PRO and HRQoL questionnaires into our recommendations.

THE CHALLENGE OF MEASURING ITCH

CP is a symptom of different diseases with correspondingly different underlying mechanisms of cutaneous and extracutaneous pruritus induction. There is, therefore, no uniform clinical profile of itch perception that runs through different patient descriptions of CP. Variations can be observed in terms of localization, duration, quality, intensity, and course of CP. Patients’ perception of CP-related impact on QoL, sleep and psychosomatic reactions also differ, though often they demonstrate similar changes. In addition, all these parameters are subjective and vulnerable to confounders, such as current mood, environmental factors or stress. It is therefore important not to rely on only one parameter, such as the intensity of pruritus. Instead, a set of tools addressing different parameters (e.g. quality and course of CP) should accompany the assessment of the intensity. Furthermore, concerning the target indication of the trial, specific characteristics may have to be taken into account (for example scratch lesions or eczema formations) and additional assessment tools must be selected.

Thus, in RCT dealing with new antipruritic substances, selection of the proper pruritus parameters that may serve as PRO, and which reflect the real course of the symptom, is crucial. The current state is now presented here.

INSTRUMENTS FOR DATA COLLECTION ON PATIENT-REPORTED OUTCOMES

Questionnaires

For acquisition of data on CP, several questionnaires have been developed in the past and are currently used; for example, as the Eppendorf Itch Questionnaire (11). No international standardized questionnaire is, as yet, available (11). However, the use of the available questionnaires in the doctor’s office is recommended for structured and complete history-taking. In the future, a standard documentation procedure will be developed on behalf of IFSI (11). A questionnaire applicable in RCTs for gathering data on pruritus parameters, which will also allow comparison of parameters during a study, needs to be developed and validated.

Scales for gathering data on pruritus intensity

Gathering data on pruritus intensity is currently the most commonly used PRO in RCTs (4–6). Several scales are available for assessment of pruritus intensity, e.g. multidimensional scales, such as the Itch Severity Scale and the Pruritus Grading System, or monodimensional scales, such as the visual analogue scale (VAS). Multidimensional scales have methodological disadvantages (unequivocal classification is not always possible) and have not been adequately validated in clinical studies. Mono-dimensional scales represent a simple and rapid method for intensity assessment in RCTs and are therefore widely used. Among these, the most frequently employed instrument is the VAS, which was originally used for evaluating pain, but can also be employed for assessing pruritus intensity (4–6). The VAS is a horizontal 10-cm long line, on which patients make a vertical mark to indicate their subjective assessment of pruritus intensity. The end-points are described as follows: 0 = no itch and 10 = the worst imaginable itch (6). The VAS is usually used as a horizontal line; however, there is also a vertical version of VAS, from which it differs only in small and non-significant ways (5).

The numerical rating scale (NRS) is a similar method of pruritus measurement, with which patients can assess pruritus intensity on a scale of 0 (no pruritus) to 10 (the worst imaginable pruritus). Although the NRS is similar to the VAS, in the validating study it was found that fewer data were missing in the paper-based NRS. In particular, in patients older than 60 years, there was almost double the amount of missing data in the VAS compared with NRS (4). Furthermore, pruritus scores on the NRS were
slightly, although significantly, higher than on the VAS (4). It is still unclear whether these differences have any clinical relevance, for instance in the assessment of therapy effects. Nonetheless, it is advisable to carry out a test, asking patients to complete the VAS form, before it is employed in a clinical study, in order to make study participants acquainted with the instrument and thus minimize the amount of missing data. The verbal rating scale (VRS) is another instrument used in which the severity of pruritus is coded with graduated adjectives (from 0 = no pruritus; to 5 = very severe pruritus).

For a long time, these scales for assessing pruritus remained unvalidated, but this issue has now been addressed in 2 recent studies (4, 5). In a study including patients with itchy dermatoses and patients with chronic pruritus of different aetiologies (a total of 781 European and Japanese patients) it was shown that data from VAS, NRS and VRS showed a high degree of reliability and correlation among each other (4–6). In a test-retest reliability analysis, a high degree of interclass correlation coefficients was found, confirming a good reproducibility of these instruments. Assessment of the areas addressed by VAS was carried out with the help of these studies (Fig. 1).

A short questionnaire for a poll among members of the IFSI was developed, asking the experts for their experience with VAS. Nineteen clinicians and IFSI members of 7 countries participated (Germany, n = 8; USA, n = 4; Japan, n = 2; Poland, n = 2; France, Turkey and Sweden, each n = 1). Inclusion of VAS into clinical trials was recommended by 94.7% of participants. The question as to how frequently data should be obtained via VAS in a RCT received varying answers: 75.0% voted for “during study visit and daily via diary”, 18.8% for “only via daily diary”, and 6.2% for “only during study visit”. The target parameters, as follows, were selected: assessment of mean itch (50.0% of participants), worst itch (68.8%), lowest itch (12.5%) and all (6.25%). There was also no consensus on how many times a day the VAS should be scored by patients: once daily (41.2%), twice daily (35.3%), three times daily (23.5%) were the answers. In order to obtain reliable data using the diary method, a preliminary study recommends that patients record their mean or worst pruritus intensity twice daily (35.3%), three times daily (23.5%) were the answers. In order to obtain reliable data using the diary method, a preliminary study recommends that patients record their mean or worst pruritus intensity twice daily (35.3%), three times daily (23.5%) were the answers. In order to obtain reliable data using the diary method, a preliminary study recommends that patients record their mean or worst pruritus intensity twice daily (35.3%), three times daily (23.5%) were the answers. In order to obtain reliable data using the diary method, a preliminary study recommends that patients record their mean or worst pruritus intensity twice daily (35.3%), three times daily (23.5%) were the answers. In order to obtain reliable data using the diary method, a preliminary study recommends that patients record their mean or worst pruritus intensity twice daily (35.3%), three times daily (23.5%) were the answers.
value (the PBI-P in a narrower sense) is calculated, which represents the patient-relevant therapy benefits. PNQ, as well as PBQ, contain a standardized list of 27 potential variables for capturing data on benefits of therapy, such as being able to sleep better or having to spend less time on daily treatment. These variables for reviewing therapy benefits were developed with the help of interviewing 50 patients with CP. PBI-P has been validated in 100 patients with CP. What the patients wanted most was to have less pruritus, to obtain a clear diagnosis and therapy and to have faith in the therapy. On average, 1.6% of the items were left unanswered. The correlation (convergent validity) between PBI-P and changes in CP measured by the VAS was $r=0.57$; correlation between improved QoL, as measured with the Dermatology Life Quality Index (DLQI) (14), and the PBI-P was $r=0.41$. The PBI-P can be used in clinical practice for gathering data on therapy goals and to evaluate the success of treatment from the point of view of patients. The PBI-P is also suitable for investigating patient-relevant benefits of pruritus therapy in clinical, health economic or healthcare studies.

**Supportive scales: quality of life, anxiety, depression and sleep**

CP is accompanied by a high level of psychiatric comorbidities (15), and marked sleep disturbances with considerable impact on the quality of life (16). Generic instruments for data gathering on QoL, such as the SF-36 (17) or its shorter version SF-12 (18), are widely used for comparing HRQoL in different diseases. For assessing differences in QoL between different dermatological diseases, the skin-specific instrument the DLQI is widely applied (14). DLQI has also been used to assess QoL in a number of itchy disorders (e.g. psoriasis, prurigo) documenting its usefulness for patients with CP (19). However, for non-dermatological itchy diseases (e.g. cholestatic pruritus) the DLQI may not be appropriate. Despite some disadvantages of DLQI (e.g. DLQI focuses mainly on patient functioning, while mental impairment is assessed to a lesser degree; item bias concerning gender and age, 20), this scale has also numerous advantages, e.g. is available in many languages, has a version adapted for children, and has defined cut-offs for the degree of QoL impairment (14).

*ItchyQoL*, the first pruritus-specific instrument for data collection on QoL, has recently become available (21). ItchyQoL comprises the 3 dimensions of symptoms, functions and emotions and enables comparison of impairments in QoL of patients with CP, irrespective of the underlying disease. ItchyQoL is currently only available in English and German (21, 22). Future studies and translation into more languages will be of high interest to evaluate the value in studies on CP.

There are a number of instruments that may be used to assess depression and anxiety in dermatological patients, including those with CP (e.g. Hospital Anxiety and Depression Scale (HADS), Beck’s Depression Inventory (BDI), Hamilton Rating Scale for Depression (HDRS)). The HADS (23) is a short, self-assessment questionnaire, and was developed as early as 1983 by Zigmond & Snaith. With the help of the scores of the HADS, the degree of current depression and anxiety level of the patients can easily be estimated. It may serve as a decision tool for inclusion of patients in RCT, as well as a monitoring tool. The BDI (24) is another self-assessment scale originally consisting of 21 questions assessing various aspects of depression. It is recommended that, in patients with somatic diseases (e.g. patients with AD), only the first 13 items concerning cognitive and affective symptoms of depression should be used in order not to overrate the depression level. Scores of >10 points indicate that a patient has depression and may require a psychiatric consultation. The HRSD (25), published for the first time in 1960, was developed to assess the severity of symptoms (mood, agitation, anxiety, sleep deprivation, somatic symptoms) in patients with major depression. Currently, it is one of the most widely used instruments for depression assessment in medical studies. Based on the psychiatric examination, the physician must choose 1 of 3–5 possible answers for each questionnaire item. Scores of <7 points indicates absence of depression, 8–12 suggest mild depression, 13–17 moderate, and >18 points severe depression.

Many patients with CP report problems with falling asleep or waking up frequently during the night. Many of them also regularly use sleep medication (26). It was demonstrated that patients with AD have significant reduction in sleep duration and quality (27). Thus, a measurement of sleeping problems may be a valuable aid in the valid assessment of pruritus. Various instruments have been developed to evaluate sleeping problems and sleepiness. The Stanford Sleepiness Scale (http://www.stanford.edu/~dement/sss.html) is a simple tool to assess diurnal sleepiness resulting mainly from insufficient sleep quality. Scores of more than 3 points may indicate significant sleep deficit and the need for more sleep. Another commonly used tool is the Epworth Sleepiness Scale (http://epworthsleepinessscale.com). With this questionnaire, patients indicate the probability of falling asleep in different situations using the scoring from 0 to 3. The scale consists of 8 questions. Scores totalling 10 or fewer points is considered as normal, a score of 11–15 may indicate mild to moderate level of daytime sleepiness, while >15 points may result from severe daytime sleepiness or even narcolepsy (28). The intensity of sleep-related problems may also be reliably evaluated with Athens Insomnia Scale, an 8-item questionnaire assessing both sleep quality and diurnal sleepiness (29). None of these scales was rigorously evaluated among patients who have chronic itch (although some studies are ongoing), and it is difficult to recommend which instrument should be chosen for clinical trials. However, based on the analysis
of questionnaire contents, the Athens Insomnia Scale seems to be the most promising one.

Physician assessment: scratching activity

Pruritus induces scratching, which in turn leads to skin changes. Therefore, it would appear that measuring scratch activity or the scratching-associated skin changes is adequate for assessing the symptoms and their course in RCTs. However, as soon as one tries to compare different patients with the same underlying origin of CP, it becomes clear that scratching activity shows inter-individual differences and is influenced by several external factors. Patients with few scratch lesions report that they use different strategies to avoid scratching the skin (self-control by not scratching, wearing gloves, using skin creams/cooling aids instead of scratching). There are different motivations behind this, such as social pressure (typically complaints from the partner regarding scratch activity and its consequences, such as blood stains on bed linen) or avoidance of stigmatization (visible scratch lesions tend to be noticed in professional and social environments). Some patients even report that scratching itself induces pruritus and its worsening; hence they try to avoid scratching. Patients with many scratch lesions may scratch in an automatic fashion and yet others who, as a matter of habit, scratch themselves even in the absence of pruritus. Thus, there is no uniform picture of scratch behaviour and scratching-induced skin changes. Therefore, measuring scratching activity as a parameter of the course of CP is prone to interference and can yield a distorted picture of the course of the symptom. In RCTs, various instruments for measuring scratching activity have been described (e.g. Accelerometer, Actigraphy, DigiTrac, ActiTrac) for a review, see 4, 7). There is, as yet, no validation of all these procedures for CP. An instrument for monitoring individual scratch lesions standardized against the body surface, the so-called Scratch Symptom Score (SSS), is currently validated and is probably suitable for measuring reduction in individual scratch lesions (7). A clinical methodological review of pruritus course and scratch behaviour is necessary in order to make a conclusive evaluation of assessment instruments and their value in clinical studies.

RECOMMENDATIONS FOR THE MEASUREMENT OF CHRONIC PRURITUS IN CLINICAL TRIALS

To date, no finally validated set of instruments is available for acquiring data on pruritus in daily clinical routine and for use in studies. However, such instruments are in the process of development within the context of the SIG of the IFSI. The IFSI cooperates over this issue with a German working group (7) and Japanese Itch Society (6). At present, there is consensus among the participating experts that, in clinical trials, the above-presented questionnaires and scales (Table I) can be used for gathering data on PRO in CP. There are still several points that need to be addressed and weaknesses in the tools that need to be eliminated. Future research will provide a basis for revision and refining of our recommendations.

Table I. Assessment of chronic pruritus (CP) in clinical trials: recommendations for patient-reported outcome (PRO)*

<table>
<thead>
<tr>
<th>Tool</th>
<th>Recommendation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-reported outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pruritus intensity</td>
<td>Visual analogue scale (VAS) is not an optimal instrument, but currently it cannot be dispensed with. Alternatively, numerical rating scale (NRS) can be used. VAS or NRS should be used in combination with verbal rating scale</td>
<td>4, 5, 15</td>
</tr>
<tr>
<td>Pruritus course</td>
<td>The SD questionnaire claims to be a tool for assessment of pruritus course; however, due to limited experience, as yet it has not been conclusively evaluated. Three new tools (itch-free days, dynamic pruritus scale, patient global assessment) are currently being validated and will be available in the future. These demonstrated in preliminary studies robust data better than that provided by assessing subjective itch intensity.</td>
<td>14, 16, 17</td>
</tr>
<tr>
<td>Patient Benefit Index for pruritus (PBI-P)</td>
<td>PBI-P is an indispensable tool in RCTs. In private practice, for setting therapy goals by doctor and patient together, the PBI-P can be highly valuable.</td>
<td>7</td>
</tr>
<tr>
<td>Health-related quality of life (HRQoL)</td>
<td>Data on HRQoL are important as they enable assessment of impact on the daily life and social functioning. The Dermatology Life Quality Index is the instrument most frequently used. The value of the new ItchyQoL has not yet been conclusively evaluated.</td>
<td>18, 28</td>
</tr>
<tr>
<td>Anxiety and depression</td>
<td>Acquiring data on anxiety and depression using, for example, the HADS is reasonable both because of the close correlation of the data with the course of CP intensity as also for assessing the need for psychosomatic/psychiatric investigation.</td>
<td>29–32</td>
</tr>
<tr>
<td>Sleep</td>
<td>Several instruments to measure sleeplessness in patients are available, e.g. Stanford Sleepiness Scale, Epworth Sleepiness Scale or Athens Insomnia Scale. They may support the data on CP course.</td>
<td>33–37</td>
</tr>
<tr>
<td>PRO-independent assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scratching activity</td>
<td>Devices to capture scratch activity are employed, although methodological studies validating their use in CP are still missing.</td>
<td>Review in 4</td>
</tr>
<tr>
<td>Scratching-associated lesions</td>
<td>Currently, data on scratch-related skin lesions can be acquired only in a descriptive fashion. The Scratch Symptom Score is currently under development.</td>
<td>14, 17</td>
</tr>
</tbody>
</table>

*Before employing validated published measuring instruments, it is important to clear up all copyright issues.
Of greatest interest will be the development of tools to assess pruritus in an objective way. Further studies on actigraphy, biomarkers and neuroimaging, including functional magnetic resonance tomography (fMRI), are urgently needed. To date, various fMRI studies have been performed, but have not yet concluded a uniform specific encoding pattern of acute and chronic itch at the neural level (30). Ongoing studies may provide a future tool to evaluate an anti-pruritic response in proof of concept studies.

ACKNOWLEDGEMENT
The authors would like to thank Rajam Csdoras-Iyer for assistance in the preparation of the manuscript. The Special Interest Group “Scoring Itch in Clinical Trials” of the IFSI was supported by Beiersdorf AG, Germany. The launch symposium during the 5th World Congress on Itch, October 26th 2009, Tokyo was supported by the National Eczema Association (NEA), Schering-Plough KK and Shinonogi & Co. Ltd.

The authors declare no conflicts of interest.

REFERENCES