SUPPLEMENTARY MATERIALS AND METHODS

Development of archive of self-report measures

A conceptual model of itch was previously developed and suggested that a number of itch-specific health-related quality of life issues were not captured by existing questionnaires (Silverberg et al., 2018). The next step of the development of the Patient-Reported Outcomes Measurement Information System (PROMIS) Itch Questionnaire was to create a preliminary item pool based on the conceptual model that encompassed the impact of itch on emotional well-being, physical function, social function, and clothing decisions. For items that reflected concepts in other PROMIS domains, we used the wording of corresponding items from those existing PROMIS domains whenever possible and added the attribution to itch. The item pool went through several different iterations based on results from cognitive debriefing interviews with patients with itch, expert review, literacy review, and translatability review (to enable future translation into different languages). Items were reviewed by 12 additional patients (three patients per concept) with chronic itch through probing cognitive interviews (Willis, 2005). Sample demographics for the probing cognitive interviews included 6 males and 6 females, with a mean age of 61.8 years (range, 45–76 years). Most participants were Caucasian (91.7%), and there was representation across different levels of education (8.3% completed high school, 16.7% some college, 33.3% college, and 41.7% an advanced degree). Expert review included insight from patient-reported outcome development experts and professionals with clinical expertise in itch; experts provided feedback with regard to item overlap, appropriateness of the content to each subdomain, wording suggestions and/or changes, and content coverage (i.e., that all aspects of the specified domain were represented). Finally, a translatability review was conducted to minimize use of culture-specific wordings. We focused on Spanish translation for this review. Forward and backward Spanish translations were conducted by a Spanish-speaking translation scientist to identify potential concerns, such as items that contained wording or concepts that would be difficult to translate.

The initial item pool was comprised of 104 items; 23 items were deleted and six items were revised based on expert review, six items were revised based on translation review, and six items were deleted and three items were revised based on cognitive interview feedback. The remaining 75 items were field tested in 600 adults in the US with chronic itch. The temporal context for all items was seven days (e.g., “In the past seven days, because of itch, I was absent from work.”). Response options were limited to two sets: (i) never, rarely, sometimes, often, and almost always and (ii) not at all, a little bit, somewhat, quite a bit, and very much. The temporal period and response categories were examined in cognitive debriefing interviews to ensure that they were meaningful to and easily understood by individuals living with chronic itch.

Sample size considerations

Sample size consideration was determined based on the requirements of item response theory analysis, the primary method in this effort. Although sample sizes of 200–1,000 have been proposed when using a graded response model, in which a larger sample size can produce more stable parameter estimation (Samejima et al., 1996), rules of thumb dictate that a minimum of 5–10 individuals are needed for every item within an item pool (Bryant and Yarnold, 1995, Everitt, 1975). With 75 items in the pool, 600 individuals were needed for reliable item response theory calibration data.

Study participants

Participants were recruited in July 2016 using two screening questions:

1. Have you suffered from itch for more than 6 weeks?
    A. Yes
    B. No
2. Which statement best captures your experience with itch in the past 6 weeks?
    A. I have had no problem at all with itching.
    B. I have had an itch from time to time, but it has not had any effect on my life.
    C. I have had some minor problems with itching that have not interfered significantly with living a normal life.
    D. I have had some moderate problems with itching that have interfered to some degree with living a normal life.
    E. I have had severe problems with itching that have interfered significantly with living a normal life.

Participants were included if they provided response A to the first question and response B, C, D, or E for the second question. Sample quotas were established to recruit 150 participants for each response B, C, D, and E. In addition to the PROMIS Itch Questionnaire candidate items, participants completed numerical rating scales for worst and average itch over the past seven days (0 [no itch] to 10 [worst imaginable itch]). All participants completed all items; therefore, there were no missing data for any items.

Receiver operating characteristic curves

Logistic regression models were constructed with verbal rating scale for severity of average itch in the past week (moderate vs. mild, severe vs. mild, severe vs. moderate, very severe vs. mild, very severe vs. moderate, and very severe vs. severe). The independent variables were item bank T-scores. Receiver operating characteristic curves were employed to determine optimal thresholds between severity groups. Separate receiver operating characteristic curves were developed for each severity cut point. Discriminant validity was determined using area under the curve. Area under the curve scores of ≥0.90 were considered excellent, 0.80–0.89 good, 0.70–0.79 fair, <0.70 poor, and <0.60 fail (Metz, 1978). We hypothesized there would be fair to excellent area under the curve scores for PROMIS Itch Questionnaire scores and itch severity. Optimal thresholds for each receiver operating characteristic curve were determined using the maximum conditional probability function *c*, which is equal to the product of sensitivity and specificity. Sensitivity and specificity are the probabilities of correctly being assigned and not being assigned to a particular severity group, respectively. A previous simulation study showed that the concordance probability method outperformed other commonly used methods (Rota and Antolini, 2014).

SUPPLEMENTARY RESULTS

Reliability and information function

In item bank 1, reliability approximately equals 0.9 when the information function is ≥10 (−1.65 ≤ theta ≤ 2.38) and reliability equals 0.95 when the information function is ≥20 (−1.35 ≤ theta ≤

2.1). In item bank 2, reliability approximately equals 0.9 when the information function is ≥10 (−2.10 ≤ theta ≤ 2.00) and reliability equals 0.95 when the information function is ≥20 (−1.90 ≤ theta ≤ 1.65). In item bank 3, reliability approximately equals 0.9 when the information function is ≥10 (−1.73 ≤ theta ≤ 2.15) and reliability equals 0.95 when the information function is ≥20 (−1.40 ≤ theta ≤ 1.90). In item bank 4, reliability approximately equals 0.9 when the information function is ≥10 (−1.25 ≤ theta ≤ 1.45). Because this scale consists of 5 items, it was not surprising that the maximum information was <20, that is, the reliability function is <0.95 across the continuum.

SUPPLEMENTARY REFERENCES

Bryant FB, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold RR, editors. Reading and understanding multivariate statistics. Washington, DC: American Psychological Association; 1995. p. 99e136.

Everitt BS. Multivariate analysis: the need for data, and other problems. Br J Psychiatry 1975;126:237e40.

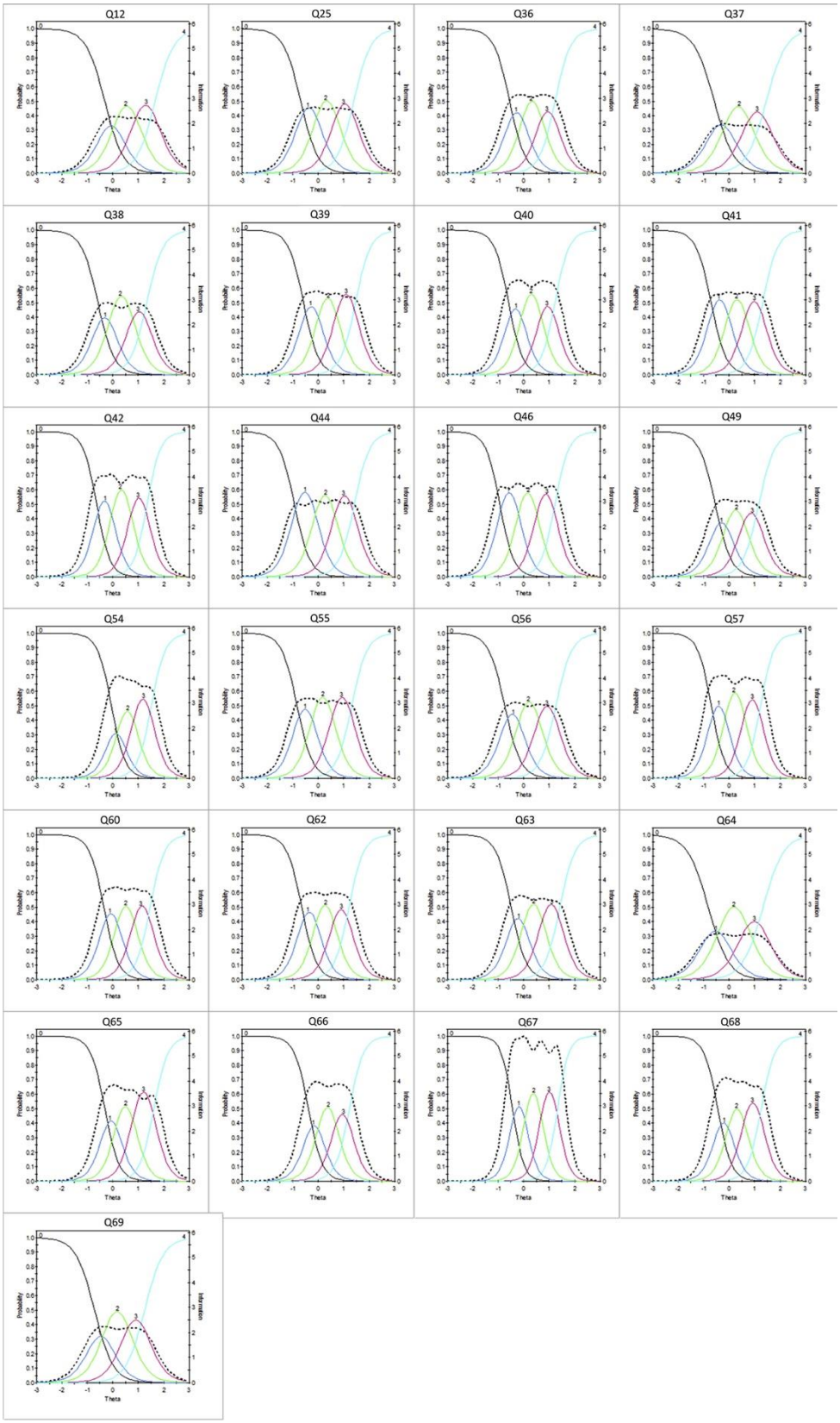
Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283e98.

Rota M, Antolini L. Finding the optimal cut-point for Gaussian and gamma distributed biomarkers. Comp Stat Data Anal 2014;69:1e14.

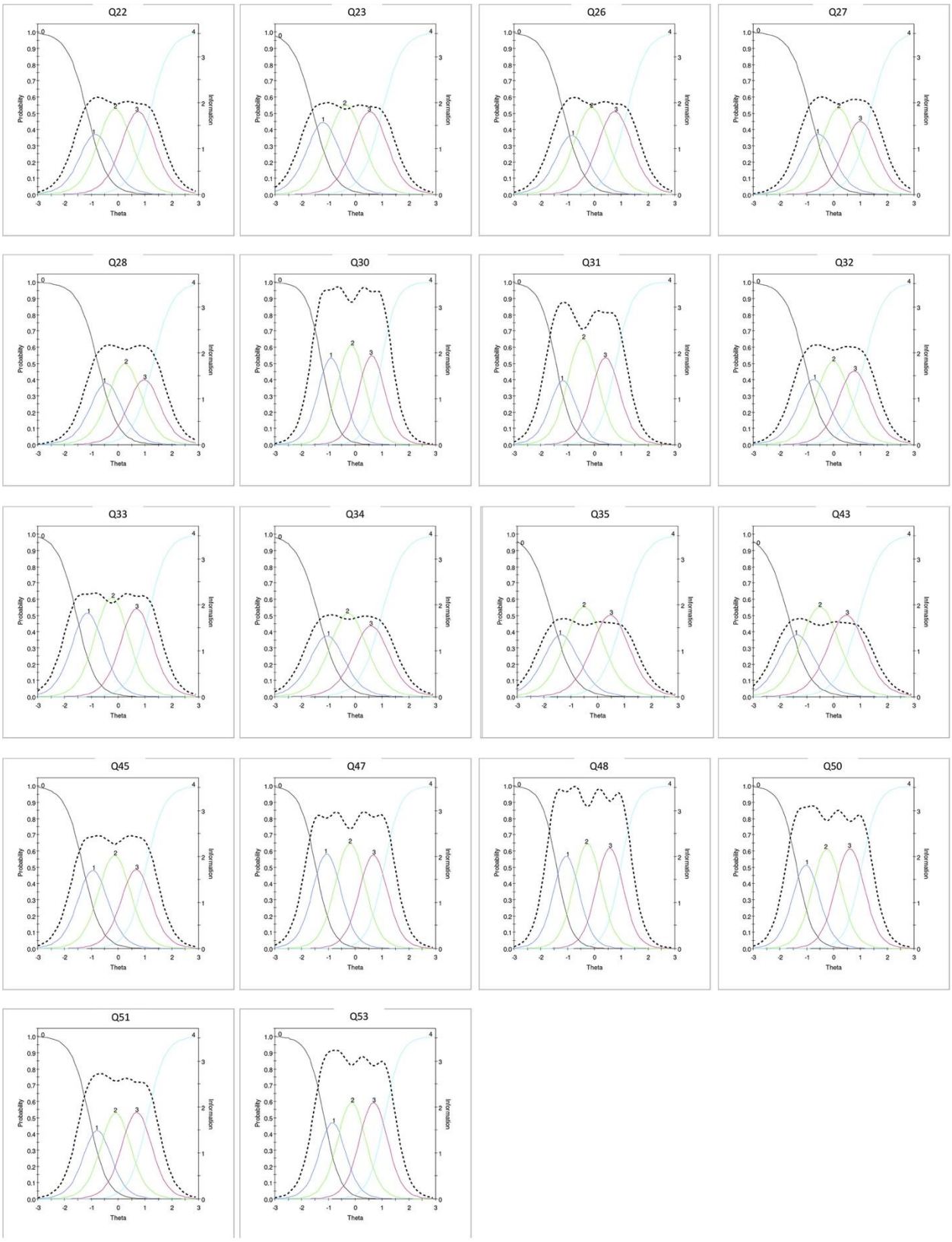
Samejima F, van der Liden WJ, Hambleton R. The graded response model. In: Van der Liden WJ, editor. Handbook of modern item response theory. New York: Springer; 1996. p. 85e100.

Silverberg JI, Kantor RW, Dalal P, Hickey C, Shaunfield S, Kaiser K, et al. A comprehensive conceptual model of the experience of chronic itch in adults. Am J Clin Dermatol 2018;19:759e69.

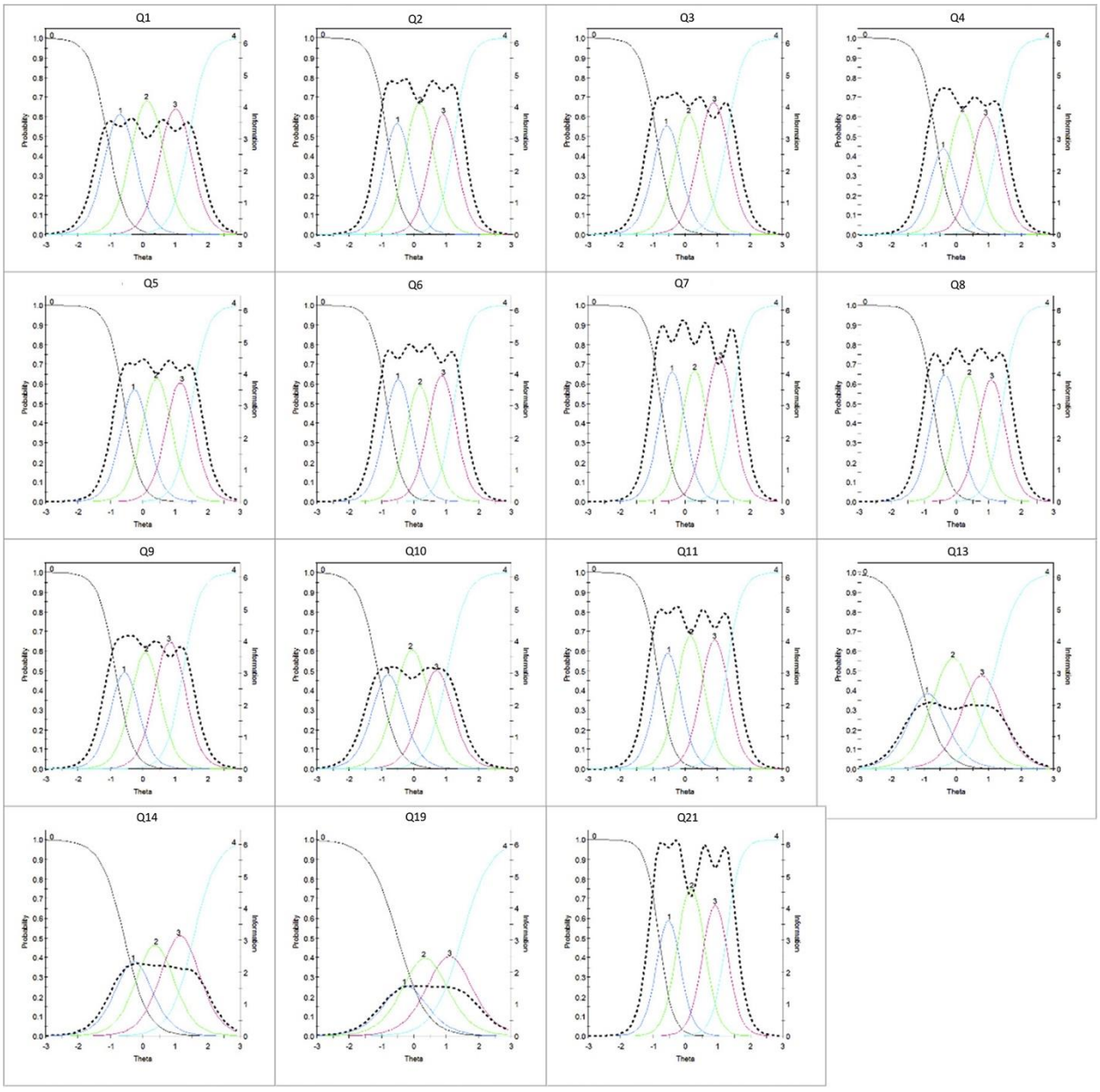
Willis GB. Cognitive interviewing : a tool for improving questionnaire design; xii. Thousand Oaks, CA: Sage Publications; 2005. p. 335.



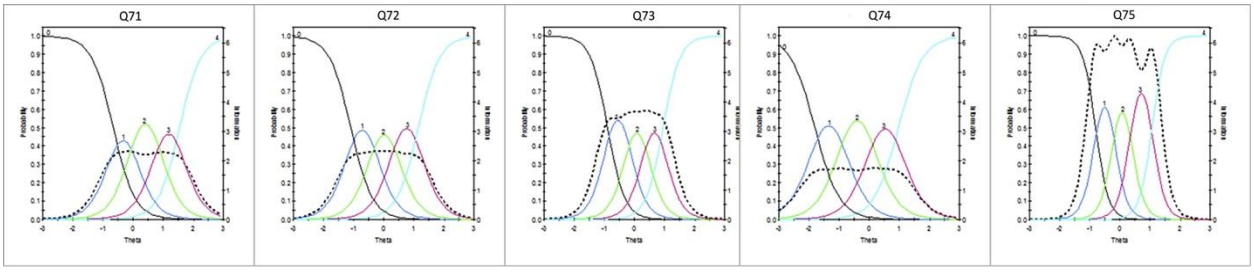
Supplementary Figure S1. **ICC curves of 25 items included in Factor 1.** ICC, item characteristic curve; Q, question.



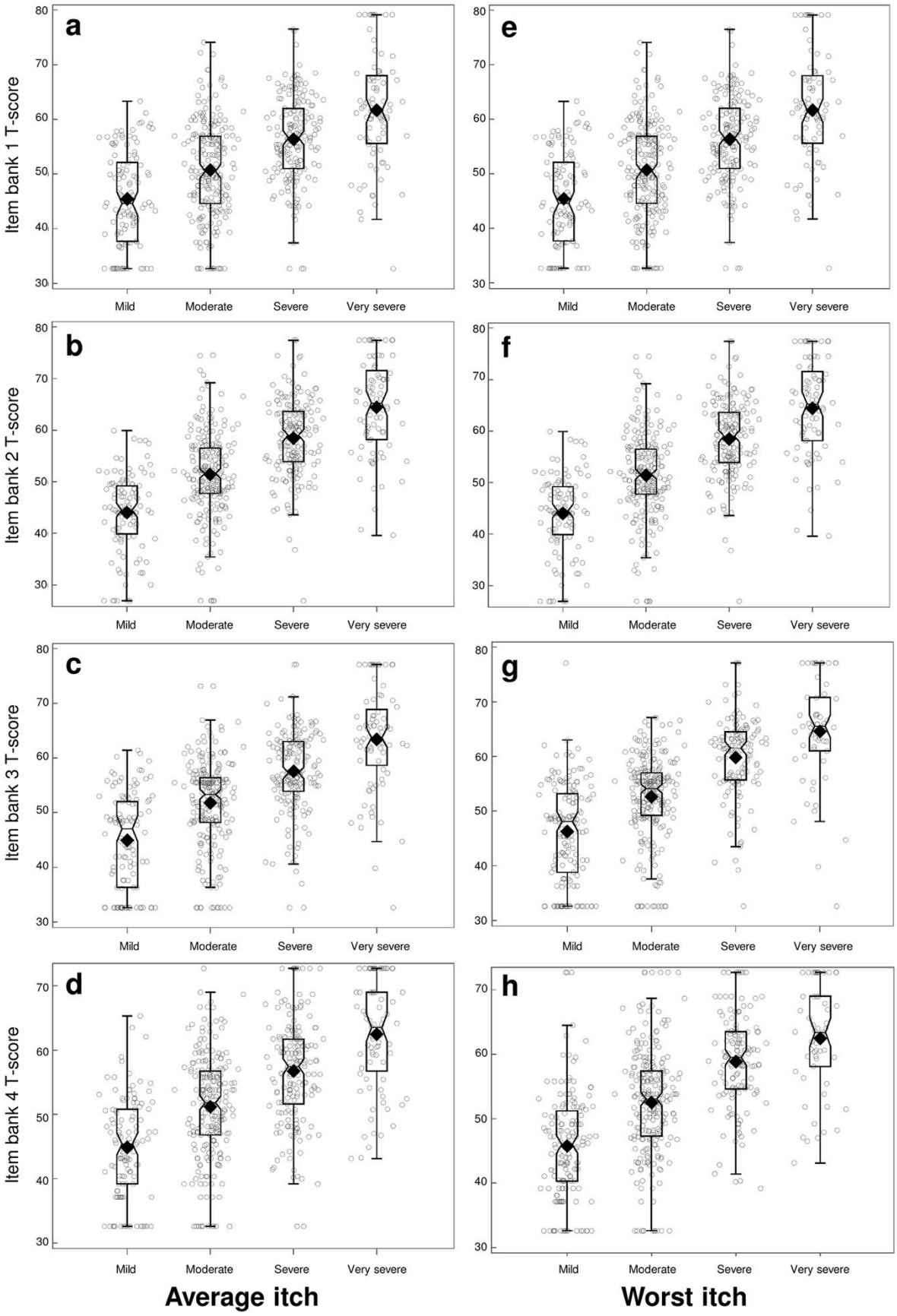
Supplementary Figure S2. **ICC curves of 18 items included in Factor 2.** ICC, item characteristic curve; Q, question.



Supplementary Figure S3. **ICC curves of 15 items included in Factor 3.** ICC, item characteristic curve; Q, question.



Supplementary Figure S4. **ICC curves of 5 items included in Factor 4.** ICC, item characteristic curve; Q, question.


a
b
c
d
e
f
g
h
Item bank 1 T-score
Item bank 2 T-score
Item bank 3 T-score
Item bank 4 T-score
80
70
60
50
40
30
Mild
Moderate
Severe
Very severe
Average itch
Worst itch

Supplementary Figure S5. **Box-whisker plots and overlaid jitter plots of T-scores for item banks 1, 2, 3, and 4.** Plots are presented for each level of (**a, b, c, d**) average or (**e, f, g, h**) worst itch severity.