Establishing and Validating an Ichthyosis Severity Index

Nareh V. Marukian1, Yanhong Deng2, Geliang Gan2, Ivy Ren1, Wisblaude Thermidor1, Brittany G. Craiglow1,3, Leonard M. Milstone1,6 and Keith A. Choate1,4,5,6

We designed and validated a Visual Index for Ichthyosis Severity for scale and erythema that provides (1) written descriptions of the features characteristic of each level of severity, (2) visual standards for four body sites, and (3) two distinct standards to account for different types of scale. We tested the Visual Index for Ichthyosis Severity for reliability and reproducibility using two different settings: one that utilized scoring of 60 test photographs by 10 dermatologists, and one with in-person evaluations on 85 subjects by 12 dermatologists at the Foundation for Ichthyosis and Related Skin Types conference. The validation process revealed high reliability and reproducibility for both scale and erythema. The interrater and intrarater intraclass correlation coefficients for scale were consistently near or greater than 0.7 in both settings. By contrast, the intrarater reliability for erythema was higher during in-person validation compared with validation on test photographs. Our analysis indicates that the Visual Index for Ichthyosis Severity performs better in person than with photographs, an important consideration in the design of clinical trials. Power analysis predicts that a 1-point difference on this 5-step scale would be detectable with 12 subjects in each of two defined groups. This index provides a tool for clinical phenotyping and assessment of therapeutic response for many disorders of keratinization.

INTRODUCTION

Ichthyoses comprise a clinically and genetically heterogeneous group of disorders of keratinization, characterized by the presence of a generalized or localized scale, as well as a variable degree of erythema. Careful examination of patients who have ichthyosis reveals that the amount of scale on specific sites is quite variable, and that there are subtle differences in the quality of scale between different genotypes. Although textbooks often suggest a monomorphous presentation for different types of ichthyosis, our experience from carefully photographing and genotyping over 800 individuals with ichthyosis shows significant patient-to-patient variability in amount of scale at any given body site and variability of scale quality on different sites of individual patients. Previous estimates of scale severity have been limited by focus on a single body site (Kuster et al., 1998), by examples of severity that utilize different body sites as comparators (Kamalpour et al., 2010), or by concentration on a single phenotype (Ganemo et al., 1999).

Erythema is often a clinically significant component of ichthyosis even though the relative roles of vasodilation and inflammation are not well understood. Erythema is an important component of many skin diseases, but there is no agreement as to how its severity should be measured. Strong arguments have been made for the clinical validity of visual analog scales (Haigh and Smith, 1991; Held et al., 1998), or the superiority and sensitivity of reflectance spectroscopy (Draaijers et al., 2004; Shah et al., 1989; Sterner et al., 2014). To our knowledge, all measurements of erythema in ichthyosis have used a visual analog scale.

A reliable method of assessing the clinical severity of subjects with ichthyosis is critical to evaluating the efficacy of new treatments. Although prior studies have used visual scales to measure the clinical severity of this rare group of disorders (Bodemer et al., 2011; Ganemo et al., 1999; Jennings et al., 1998; Kamalpour et al., 2010; Kuster et al., 1998; Paller et al., 2017), to date, only one has been validated.

The most comprehensive, but yet unvalidated, severity index is provided by Bodemer et al., who measure the degree and extent of scale, erythema, pruritus, skin pain, palmoplantar keratoderma, fissures, hand retraction, foot retraction, movement impairment, ectropion, and eclabium, all of which can be present in ichthyosis. In contrast, the Congenital Ichthyoses Severity Index (Kamalpour et al., 2010) is simpler tool that evaluates erythema, scale, and alopecia, and is the only index that has been validated to date. In attempting to implement the Congenital Ichthyoses Severity Index in our clinic, we found that the standard photographs did not effectively represent the severity spectrum in our
patients, and was limited by the need to put patients into one of four clinical categories. Finally, the Congenital Ichthyoses Severity Index was validated with a relatively small number of subjects, and few expert raters were included in the analysis. The most recent ichthyosis severity index published is the Ichthyosis Area Severity Index (Paller et al., 2017), which uses the area and severity of erythema and scale as measures of ichthyosis severity. Like the Bodemer scale, the Ichthyosis Area Severity Index has not been validated.

In considering these indices in our clinical practice and in evaluating our cohort of 832 subjects enrolled in the National Registry for Ichthyosis and Related Skin Types, we recognized the need for a user-friendly and rapid tool for the evaluation of ichthyosis severity. In our experience with these cohorts, we have found that scale and erythema are the only findings present in ichthyosis of every genetic cause, and occur either upon skin of normal thickness (lamellar subtypes) or upon thickened skin (keratoderma subtypes). As such, we set out to learn from the framework provided by prior indices and generate a reliable method to assess ichthyosis clinical severity using solely these two variables.

We expected that a simple, validated, and rapidly employed severity index that could be used across many types of ichthyosis with the capacity to rate overall or site-specific severity would advance ichthyosis research. Recognizing that quality of standards is crucial for ease of implementation, we employed images with uniform focus, magnification, positioning, erythema, skin tone, and background. In addition, written descriptions for the degree of scale and erythema were provided to highlight the cardinal findings for each photographic representation and ensure a uniform review. We provided separate standards for each of four different body sites that are less frequently or aggressively groomed by affected individuals to enhance the uniformity and applicability of our index.

We designed a prospective study to validate our instrument, the Visual Index for Ichthyosis Severity (VIIS), using two different settings: one that utilized scoring of test photographs and one with in-person evaluations. A total of 60 photographs were sent to 10 raters for baseline scoring (stage 1, round 1), and shuffled photographs were sent 4 weeks later to test reproducibility (stage 1, round 2). To evaluate performance in real-time assessment, a total of 85 subjects were evaluated by three or four raters in 5-minute sessions at a Foundation for Ichthyosis and Related Skin Types Family conference. Analyses of intraclass correlation coefficients (ICCs) and power were performed to evaluate the performance and reproducibility of the VIIS.

RESULTS

Validation of the VIIS on test photographs (stage 1)

Demographic and clinical details for the subjects whose photographs were used as the test images can be found in Table 1.

Interrater reliability for rounds 1 and 2. Descriptive statistics for interrater reliability (IRR) for scale and erythema can be found in Figure 1. For both rounds 1 and 2, the ratings were found to be significantly correlated for scale (ICCs near 0.7 or greater). However, the IRR for erythema was poor (ICCs < 0.7).

Agreement between raters in choosing scale standards. Raters were given two separate sets of scale standards (lamellar or “L” set of standards and keratoderma or “K” set of standards) to account for the two different types of scales most commonly seen in subjects with ichthyosis. Raters were consistent in choosing a set of standards (“L” or “K”) for a given patient, and there was excellent ICC for all body sites during both rounds of testing (Kuder-Richardson Formula 20 > 0.9) (Table 2).

Intrarater reliability. To determine intrarater reliability, we employed a test-retest approach with a 4-week interval between rounds 1 and 2. There was high intrarater reliability in choosing a set of standards (“L” or “K”) for a given patient, with Kappa near or greater than 0.7 for all body sites. Across all 10 raters, the overall intrarater ICCs for the four different body sites ranged from 0.72 to 0.81 for scale and 0.41 to 0.60 for erythema.

Live validation of the VIIS at the Foundation for Ichthyosis and Related Skin Types (FIRST) conference (stage 2)

Demographic and clinical details for all subjects evaluated by three or four raters in 5-minute sessions at a Foundation for Ichthyosis and Related Skin Types Family conference (stage 2) can be found in Table 1.

Interrater reliability. Descriptive statistics for IRR for scale and erythema can be found in Figure 2. Across all rooms, the scores were found to be highly correlated for both scale and erythema, with ICCs near or greater than 0.7. The IRR for scale was generally greater than that for erythema. The scores in room 1 showed the highest degree of IRR, with ICCs

Table 1. Demographic and clinical characteristics for patients whose photographs were used for the first stage of validation (n = 60)

<table>
<thead>
<tr>
<th>Age</th>
<th>Upper back</th>
<th>Upper arm</th>
<th>Lower leg</th>
<th>Dorsal foot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥18 y</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>&lt;18 y</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Lower back</th>
<th>Upper arm</th>
<th>Lower leg</th>
<th>Dorsal foot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasian</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>45</td>
<td>75.0%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5.0%</td>
</tr>
<tr>
<td>Black</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3.3%</td>
</tr>
<tr>
<td>Unknown</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>16.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Lower back</th>
<th>Upper arm</th>
<th>Lower leg</th>
<th>Dorsal foot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ichthyosis type</th>
<th>Lower back</th>
<th>Upper arm</th>
<th>Lower leg</th>
<th>Dorsal foot</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>ARCI</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>Netherton's</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ichthyosis vulgaris</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Normal control</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Unclassified</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

Abbreviations: ARCI, autosomal recessive congenital ichthyosis; EI, epidermolytic ichthyosis.
greater than 0.7 for both scale and erythema for all four body sites.

**Agreement between raters in choosing scale standards.** Similar to the data from validation on test photographs discussed above, the raters were consistent in choosing the specific set of scale standards (“L” or “K”) for a given subject. The IRR was high with the Kuder-Richardson Formula 20 greater than 0.77 for all body sites (Table 4). In room 2, all of the raters unanimously agreed on “L” or “K” for all of the subjects seen (Kuder-Richardson Formula 20 of 1).

**Interrater reliability for combined and total scores.** An important feature of the design of our index is the ability to allow for a comprehensive assessment of ichthyosis severity that accounts for the severities at the four representative sites. We determined IRR for the combined scores that range from 0 to 16 (score of 0–4 for each of four body sites) for each category (erythema and scale). We also determined IRR for the final score (sum of combined scores for scale and erythema) that ranges from 0 to 32. As shown in Figure 3, the IRR was high, with ICCs consistently higher than 0.7.

Subanalyses for the adult and pediatric populations were performed to address the applicability of our index to these subgroups (Supplementary Figures S1 and S2 online). The number of adults (≥18 years old) and children (<18 years old) in each of the three rooms was as follows: 20 adults and 7 children in room 1, 14 adults and 13 children in room 2, and 9 adults and 22 children in room 3. Overall, smaller sample sizes in this analysis resulted in wider confidence intervals. Even for these small sample sizes, the subanalysis showed high ICCs for scale, with values ranging from 0.78 to 0.85 for adults and from 0.79 to 0.90 for children. For erythema, the index outperformed in children as compared with adults, with ICCs ranging from 0.53 to 0.86 in adults and...
from 0.75 to 0.85 in children. The scoring of erythema in adults may be complicated by the larger surface areas of body sites, as well as secondary skin changes.

Power analysis to predict a sample size for 1- and 2-point improvement. To determine the feasibility of using the VIIS in future clinical trials, we performed a power analysis to determine the number of subjects who would need to be enrolled to detect a single point improvement or decrement on our 5-step scale. For our cohort of 85 subjects, the mean scores for erythema and scale are 1.78 and 2.35, respectively, across all body sites. In many clinical trials, inclusion criteria require subjects to start with a severity of at least 2 to measure a 1- to 2-step improvement. Approximately 45% and 67% of subjects had scores greater than or equal to 2 for erythema and scale, respectively. For subjects with average scores greater than or equal to 2, the mean score was 2.8 with a standard deviation of 0.7, accounting for all body sites for both erythema and scale. The descriptive statistics for subjects with average scores greater than or equal to 2 can be found in Supplementary Table S1. Given the null hypothesis that the control and treatment groups have equal means of 2.8, the alternative hypothesis that the mean of the treatment group is 1.8 (1 unit difference), and an estimated group standard deviations of 0.7, sample sizes of 12 in the control group and 12 in the treatment group will achieve 92% power to detect a one unit difference. Eight subjects (four in each group) will achieve 92% power to detect a 2-point difference.

DISCUSSION
We have designed and validated a visual severity index for ichthyosis, a critical tool for research and clinical management in this rare group of disorders. We found high intrarater and interrater ICCs for scale (near 0.7 or greater for all rounds of testing). Raters were highly consistent in deciding which set of scale standards to choose for a given patient (Kuder-Richardson Formula 20 greater than 0.77 for all rounds of testing), even though our subject population was heterogeneous with respect to clinical diagnosis (Tables 1, 4). Subgroup analysis revealed that high ICCs for scale were maintained even in the rare cases where raters disagreed on which set of standards to use for a given subject. This indicates that our two different sets of scale standards capture increments of severity in a similar fashion, possibly due to the detailed written descriptions that standardize the scoring regardless of which set of standards is used. This suggests that the standardized written descriptions may allow or even be critical for scoring of the many different clinical subtypes of ichthyosis—even those not represented in the photographs.

Although the IRR for erythema was high (ICCs near 0.7 or greater) during in-person validation at the FIRST conference, it was poor (ICCs less than 0.6) during validation on test photographs. This finding emphasizes the difficulty of capturing and reproducing erythema on photographic images. Many factors, including room lighting and color saturation settings, can impede a consistent evaluation of erythema on photographs. Much literature has been published on the evaluation of erythema, with compelling evidence for the use of both visual analog scales and more objective methods, such as reflectance spectrophotometry. Reflectance spectrophotometry calculates an erythema index by comparing the intensity of light absorbed by melanin (red light) with that of light absorbed by hemoglobin (green light). Several studies have provided evidence for the reliability of reflectance spectrophotometry (Draaijers et al., 2004; Sterner et al., 2014); however, others have shown lack of superiority of readings from reflectance spectrophotometry compared with visual analog scales (Conner et al., 1993; Held et al., 1998). Various operator factors can influence the reading obtained from reflectance spectrophotometry, including the pressure with which the instrument is held and the angle used (Fullerton et al., 1996). Furthermore, several individual factors can
Impact reflectance spectrophotometry readings, including blood pressure, smoking, caffeine intake, and skin phototype (Fullerton et al., 1996; Latreille et al., 2007). Latreille et al. (2007) showed that erythema induced by methyl nicotinate was not detected by reflectance spectrophotometry in the darker phototypes, possibly because the emitted light is predominantly absorbed by the melanin in the epidermis and does not reach the hemoglobin in the dermis. Given the time- and cost-effectiveness of visual observation, and lack of strong evidence for the superiority of reflectance spectrophotometry, we utilized a visual analog scale for the evaluation of erythema in our index.

During in-person evaluation, our index showed high IRR for both erythema and scale with ICCs near 0.7 or greater.

![Figure 2. Intraclass correlation coefficients (ICCs) for agreement and consistency for the three rooms during in-person evaluations at the Foundation for Ichthyosis and Related Skin Types conference (stage 2). For all rooms, the ratings were found to be highly correlated for scale and erythema (ICCs near 0.7 or greater).](image)

**Table 4. Kuder-Richardson Formula 20 (KR20) values for the three rooms during in-person evaluations at the FIRST conference (stage 2)**

<table>
<thead>
<tr>
<th>Room 1</th>
<th>Room 2</th>
<th>Room 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorsal foot 0.85 (0.72, 0.92) 1.00 (1.00, 1.00) 0.94 (0.89, 0.97)</td>
<td>Lower leg 0.85 (0.73, 0.92) 1.00 (0.99, 1.00) 0.92 (0.87, 0.96)</td>
<td>Upper arm 0.77 (0.59, 0.88) 1.00 (0.99, 1.00) 0.92 (0.87, 0.96)</td>
</tr>
<tr>
<td>Upper arm 0.83 (0.69, 0.91) 1.00 (1.00, 1.00) 0.92 (0.86, 0.96)</td>
<td>Lower leg 0.85 (0.73, 0.92) 1.00 (0.99, 1.00) 0.92 (0.87, 0.96)</td>
<td>Upper back 0.83 (0.69, 0.91) 1.00 (1.00, 1.00) 0.92 (0.86, 0.96)</td>
</tr>
</tbody>
</table>

Raters were consistent in choosing the specific set of scale standards (“L” or “K”) for a given patient, with KR20 greater than 0.77 for all body sites.

A limitation of our index is that the photographs included in the standards were from subjects with Fitzpatrick I–III skin type. As shown in Tables 1 and 3, the majority of the subjects during the validation process were Caucasian. Ideally, visual standards would have been created for a range of skin...
We aimed to increase the generalizability of our instrument while maintaining convenience by including written descriptions for each level for severity that are applicable to all skin types. Further studies will be necessary to determine whether our standards can be reliably used for Fitzpatrick IV–VI skin types, especially for erythema since the evaluation of erythema can be more challenging in individuals with darker phototypes.

We have designed a VIIS, and performed several rounds of rigorous validation to show high reliability and reproducibility. Such a standardized method of assessing disease severity has the potential to advance ichthyosis research. This user-friendly index was easily applied by dermatologists during a busy schedule of clinical evaluations at the FIRST conference with a restriction of 5 minutes for the scoring of each patient, and with high ICCs for the evaluation of scale and erythema. We expect this simple, time-efficient index to be useful in multiple settings, including in the outpatient setting, to monitor response to treatment and evaluate the effect of environmental factors and changes in daily skin routine, and in clinical trials.

Our analysis also indicates that to reliably capture all aspects of the index, it is essential that the evaluations be performed in person. This finding is critical to the study design of

Figure 3. Intraclass correlation coefficients (ICCs) for agreement and consistency for the combined scores for scale and erythema, as well as the final total scores during in-person evaluations at the Foundation for Ichthyosis and Related Skin Types conference (stage 2). For all rooms, the ratings were found to be highly correlated for scale and erythema (ICCs greater than 0.7).
clinical trials and emphasizes the importance of in-person evaluations rather than assessments of photographs, especially when erythema is one of the outcome measures. In addition, power analysis revealed that a sample size of 24 (12 in control group, 12 in treatment group) will achieve 92% power in detecting a unit difference, which is a feasible sample size. We expect this index to facilitate clinical phenotyping and to provide a measure of response for therapeutic trials that assess the cardinal features of this class of disorders.

**METHODS**

**Study design**

To design a user-friendly tool, we developed photographic standards for two characteristics of ichthyosis: scale and erythema. For scale, photographic standards for four representative body sites that are typically less aggressively groomed (upper arm, upper back, lower leg, and dorsal foot) were created. Each of the four representative body sites was carefully defined using anatomical landmarks. Two different sets of scale standards were developed: the lamellar (“L”) set of standards for the typical flat scales that are seen in most forms of ichthyosis and the keratoderma (“K”) set for the more columnar scales that are typical of epidermolytic ichthyosis and erythrokeratoderma. We employed 5-point Likert scales (0–4) with increasing clinical severity. The combined scores, therefore, range from 0 to 16 (score of 4 for all 4 body sites) for each category (erythema and scale). The final score accounts for both erythema and scale, and ranges from 0 to 32. Each photographic standard includes representative clinical photographs for each severity score, along with detailed written descriptions of features characteristic of the severity score represented. All of the clinical photographs were obtained at FIRST family conferences. The photographs included in the standards were limited to subjects with Fitzpatrick 1–3 skin type. Photographs were chosen with the aim of demonstrating severity alone. All of the photographic standards can be found in the Supplementary Material online.

Validation of our instrument was performed in two stages the first utilized scoring of test photographs, and the second involved in-person evaluations. During the first stage, 10 dermatologists were provided with a series of 60 test photographs (15 photographs for each of the four body sites) and asked to independently use our index to score for scale and erythema (stage 1, round 1). The reshuffled photographs were sent to the same dermatologists 4 weeks later (stage 1, round 2) to determine intrarater reliability. This 4-week interval was chosen to reduce rater recall of prior scoring (Kamalpour et al., 2010).

The second stage of validation was performed at the FIRST family conference in San Diego in June 2016 (stage 2). Eligible participants included all subjects enrolled in the National Registry for Ichthyosis and Related Skin Types who participated in a clinical screening appointment at the conference. None refused to participate. Subjects were seen in one of three clinical evaluation rooms with four dermatologists assigned to each room. A total of 85 subjects (27 in room 1, 27 in room 2, 31 in room 3) were enrolled in the study. A total of 12 dermatologists participated, and they were asked to independently rate the clinical severity of the subjects seen in their room. The rating process was limited to the first 5 minutes of each 20-minute appointment. The discussion of ratings was not permitted and all raters attested that they adhered to this requirement. Four (two in room 2 and two in room 3) of the 12 dermatologists had previously participated as raters during the validation on test photographs.

All raters (for both stages 1 and 2) were experts in ichthyosis, and members of the FIRST Medical & Scientific Advisory Board. None of the raters reported systematic use or extended experience using other tools.

**Participants**

The study was approved by the Yale Human Investigation Committee, consistent with the Declaration of Helsinki guidelines, and written informed consent was obtained from the participants or their parents. The inclusion criterion was individuals who self-identified as having ichthyosis and attended a family conference for the FIRST. Each participant was asked to fill out a questionnaire, including self-reporting of gender, age, ethnicity, and ichthyosis type.

**Statistical analysis**

The reliability analysis of the use of the 5-point Likert scale was performed using ICCs. For interrater ICCs, two measures of ICCs were evaluated: ICC for absolute agreement, which is the more rigorous metric and requires absolute agreement among the raters, and ICC for consistency, which measures whether the raters provide scores that have the same rank order rather than requiring agreement on the absolute values. Specifically, interrater ICCs were estimated using the two-way random model. For intrarater ICCs, ICCs for absolute agreement were calculated using the one-way random model. Combined intrarater ICCs were estimated using the two-way mixed model controlling for raters. ICCs less than 0.7 were considered unacceptable, 0.7–0.79 fair, 0.8–0.89 good, and greater than or equal to 0.9 excellent (Shrout and Fleiss, 1979). The reliability of choosing the “L” or “K” standard by the same rater was estimated by the Cohen’s Kappa, which is a measure of intrarater reliability for a nominal variable. The reliability of choosing the “L” or “K” standard on the same subject across all raters was estimated by the Kuder-Richardson Formula 20, which is a measure of IRR for a nominal variable. All confidence intervals were calculated at 95% confidence level. All analyses were carried out using SAS 9.4.3 (SAS Institute, Cary, NC).

**CONFLICT OF INTEREST**

The authors state no conflict of interest.

**ACKNOWLEDGMENTS**

The Foundation for Ichthyosis and Related Skin Types (FIRST) supported this work through an award to the National Registry for Ichthyosis and Related Skin Types. We are grateful to Drs. Meise Levy, Mary Williams, John DiGiovanna, Susan Bayliss, Evelyn Lilly, Erin Mathes, John Browning, Latanya Benjamin, Heather Brandling-Bennett, Anne Lucky, Phil Fleckman, Mary Spraker, Lawrence Eichenfeld, William Rizzo, Joyce Teng, Anna Bruckner, Robert Silverman, Jonathan Dyer, Amy Nopper, Gabrielle Richard, Jennifer Hand, Ellen Milstone, Jonathan Leventhal, Mary Tomayko, and Amy Paller for their help with study design and analysis. NVM was funded by the National Institutes of Health-NCATS CTSA-TL1 Medical Student Research Fellowship.

**SUPPLEMENTARY MATERIAL**

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at http://dx.doi.org/10.1016/j.jid.2017.04.037.

**REFERENCES**


