Reliability of Counting Actinic Keratoses Before and After Brief Consensus Discussion

The VA Topical Tretinoin Chemoprevention (VATTC) Trial

Martin A. Weinstock, MD, PhD; Stephen F. Bingham, PhD; Gary W. Cole, MD; David Eilers, MD; Mark F. Naylor, MD; James Kalivas, MD; J. Richard Taylor, MD; Hayes B. Gladstone, MD; Daniel J. Piacquadio, MD; John J. DiGiovanna, MD

Objective: To assess the reliability of counts of actinic keratoses (AKs) and the effect of a brief joint discussion of discrepancies on that reliability.

Design and Intervention: Seven dermatologists independently counted AKs on the face and ears before and after a brief joint discussion of discrepancies.

Setting and Patients: A volunteer sample of 9 patients from the ongoing VA (Department of Veterans Affairs) Topical Tretinoin Chemoprevention (VATTC) Trial. All participating individuals are veterans and have had 2 or more keratinocyte carcinomas (basal or squamous cell carcinoma) in the 5 years before enrollment in the study.

Main Outcome Measure: Standard deviation of estimates of the Poisson regression parameter for the dermatologists.

Results: Substantial variation was found among the dermatologists in their AK counts. The SD of the parameter estimates for the dermatologists decreased from 0.45 to 0.24 after the brief joint discussion, a 47% decrease ($P = .076$). The variation attributable to the dermatologists also decreased substantially ($\chi^2$ decrease, 94 to 12).

Conclusions: Actinic keratoses are common, and there is a continuous spectrum of lesions that ranges from sun-damaged skin to squamous cell carcinoma in situ. Clinical distinguishing features may be difficult to delineate precisely. Counts of AK are commonly performed, but appear to be unreliable, even when performed by experienced dermatologists. Joint discussion of discrepancies may enhance the reliability of these counts, although substantial variation remains. Research that relied on these counts must be reevaluated in light of the marked variation among expert observers. Future studies should consider measures to assess and enhance reliability.

Arch Dermatol. 2001;137:1055-1058

ACTINIC KERATOSES (AKs) are the dysplastic nevi of the keratinocyte world. They are direct precursors of squamous cell carcinoma of the skin (some have called them incipient squamous cell carcinomas), and are important indicators of risk for all 3 common types of sun-related skin cancers and markers of UV exposure in sun-sensitive skin. Because they are very common, and commonly treated, they are responsible for substantial use of health care resources. Like dysplastic nevomelanocytic nevi, they excite controversy. The regulations governing payment for their treatment by Medicare in the United States have been particularly controversial, because AKs are much more commonly treated on diagnosis than are dysplastic nevi.

The AK count is also a secondary end point of the ongoing VA (Department of Veterans Affairs) Topical Tretinoin Chemoprevention (VATTC) Trial. Larger-than-expected variation among clinical sites of this trial in baseline AK counts provided the initial impetus for this investigation.

Large-scale investigations involving AK, including clinical trials and epidemiologic investigations, typically use as their primary measure the number of AKs as counted by a dermatologist or other suitably trained individual.1-5 Despite the use of this end point, the reliability and validity of these counts have been subject to very limited study. The present investigation seeks to evaluate the reliability of these counts in the context of the high-risk population enrolled in the VATTC Trial.

RESULTS

Five patients were recruited for the morning session, and 4 for the afternoon sess-
SUBJECTS AND METHODS

The context of this investigation was the VATTC Trial, a randomized, multicenter, double-blind vehicle-controlled trial of topical 0.1% tretinoin cream applied twice daily to the face and ears. The primary outcome of this trial is time to onset of new keratinocyte carcinoma (basal or squamous cell) on the face or ears. All participants in this trial have received care at 1 of the 6 participating sites (all Department of Veterans Affairs medical centers). All have had 2 or more keratinocyte carcinomas on the face and ears in the 5 years before enrollment in the study, and met the other requirements for the trial. Among the secondary outcomes of the VATTC Trial is AK count on the face and ears. For the purpose of this trial, AK was defined as a keratotic lesion with at least some erythema that did not fit an alternative diagnosis such as seborrheic keratosis or psoriasis. The present investigation was prompted by the observation that there were marked differences among the 6 clinical sites in the mean AK count at the baseline examination. These differences suggested discrepancies among the sites in AK recognition.

To address this issue, a meeting was held to conduct the present investigation. It included investigators from the VATTC Trial (all dermatologists) and a convenience sample of trial participants (the patients). All patients were receiving 0.1% tretinoin cream or a vehicle control; at that point in the study, the typical frequency of use was once daily. On the morning of the meeting, each dermatologist independently counted and recorded the number of AKs on the face and ears of each patient in the morning session. The dermatologists were asked to use the trial AK definition, and no further instruction was given. To facilitate counting and recording, the area of the face and ears was subdivided into 12 anatomic regions. These 12 regions were combined for all analyses. Once these data were collected, the dermatologists convened as a group around a single patient and specifically identified and discussed the lesions that were or were not counted as AK. This discussion was conducted for each of the remaining patients in the morning session. The goal of dermatologists reconvened after lunch to independently count the number of AKs on the face and ears of a second group of patients. None of the afternoon group of patients had participated in the morning session.

Seven dermatologists counted AKs (counters). These dermatologists included the primary investigators (G.W.C., D.E., N.F.N., J.K., and J.R.T.) at 5 of the 6 clinical sites of the VATTC Trial. Another was the collaborating investigator (H.B.G.) at the sixth site, and another was the study chairman (M.A.W.). The collaborating investigator was in fellowship training; the other 6 counters were board-certified dermatologists with a mean of 20 years of experience since certification (range, 7–30 years). All except the study chairman were actively involved in counting AKs for the VATTC Trial.

We used Stata version 6.0 to obtain maximum likelihood Poisson regression model estimates for each evaluation session. The model included patient- and counter-effect parameters. We used χ² goodness of fit tests to determine if the data fit the Poisson regression model. The difference between likelihood ratio χ² statistics was used to determine if there were significant counter effects. The SDs for the morning and afternoon counter parameters were compared using the F test. The VATTC Trial in which all of these participants were enrolled is a double-blind trial; that “blind” was not broken for these analyses.

Written informed consent was obtained from all patients participating in this study.

We have documented a substantial variation in results of AK counting by experienced dermatologists in our high-risk study population. We have also documented a substantial improvement in the consistency of AK counts after group discussion of discrepancies. The present investigation is limited in several respects. The numbers of patients and dermatologists were small, so statistical power was modest. All of the patients were participants in the VATTC Trial, so the results may not be generalizable to the broader popula-
tion, even if they apply to VATT C Trial participants. Generalization to other populations is limited by 3 important considerations. First, all of the patients were receiving the study medication or vehicle control, which might have altered the appearance of their AKs. Second, VATT C Trial participants are selected because of a diagnosis of 2 or more keratinocyte carcinomas during the 5 years before enrollment in the study. Hence, they had severe actinic damage and were at high risk for subsequent keratinocytic neoplasia. Finally, all patients were receiving care at the Miami Veterans Affairs Medical Center, so they may be different in relevant respects from populations that differ geographically, or by sex, age, or socioeconomic status. Generalizability from the study dermatologists to all practicing dermatologists also cannot be guaranteed.

Some have assumed that in experienced hands, dysplastic nevi are difficult to diagnose, but AKs are not. This assumption has not been demonstrated and may be false. Even among experts, both diagnoses may be difficult to define precisely. The increased scrutiny devoted to the limitations of accuracy of many clinical judgments in recent years is useful for properly understanding dermatologic diagnosis as well.

Prior investigations of the validity of clinical AK diagnoses have been limited. Predictive value of a positive clinical diagnosis of AK has generally been found to be reasonably good. One study confirmed the clinical diagnosis of 34 (94%) of 36 clinically typical AKs. A second study included a randomly chosen subset of a population-based clinical trial involving sunscreen use, and confirmed 39 (81%) of the 48 clinical diagnoses of AK. Other population-based studies have reported 80% and 91% histological confirmation of randomly chosen clinically diagnosed AKs. In all of these studies, the gold standard against which the clinical diagnosis was compared was a histological diagnosis, and the validity of the histological diagnosis has been questioned.

Studies of the clinical diagnoses of skin cancers, however, have turned up substantial numbers of misdiagnosed AKs in some but not all contexts. In a population-based study, 13% of suspected skin cancers were histologically diagnosed as AK (12% of 514 presumed basal cell carcinomas and 39% of 44 presumed squamous cell carcinomas); in another, 17% were so diagnosed; but in clinical practice setting, only 2% of suspected basal cell carcinomas were histologically confirmed as AK.

Reliability of AK diagnoses was assessed in a general population sample. Photographs were taken of suspected AKs, and a panel of 3 dermatologists agreed with the examiner’s diagnosis of AK in 86%, although the details are not entirely clear, including the extent of agreement among the 3 dermatologists. Other studies of classification of patients into categories of single AK, more than 1 AK, or no AKs in screening or clinic settings have observed interobserver k statistics ranging from 0.17 to 0.52 for a single AK and 0.48 to 0.78 for multiple AKs.

Clinical trials and epidemiological studies involving AKs have typically used simple counts of clinically diagnosed lesions without histological confirmation (except occasionally on a small subset) or evaluation of reliability of the assessment. In light of our results, training and subsequent assessment of reliability would seem to be useful in future studies.

Counting of AKs is an imprecise business, even when performed by experts. Part of the difficulty is that there is a continuous spectrum of lesions that ranges from sun-damaged skin to squamous cell carcinoma in situ. Defining the difference between AK and sun-damaged skin on the one end, or the difference between AK and squamous cell carcinoma in situ on the other, is akin to defining the difference between green and blue: easy to recognize in the pure cases but less reliable in intermediate areas of the spectrum. Fortunately, observers working together can achieve a degree of reliability. We must be careful to interpret the existing literature involving AK with this in mind, and encourage assessments of AK in published studies to be accompanied by an estimate of the reliability of those assessments.

Accepted for publication April 11, 2001.

From the Dermatopidemiology Unit, Veterans Affairs (VA) Medical Center, Providence (Dr Weinstock), and Department of Dermatology, Rhode Island Hospital and Brown University, Providence (Drs Weinstock and DiGiovanna); the VA Cooperative Studies Program Coordinating Center, Perry Point, Md (Dr Bingham); Department of Dermatology, VA Medical Center Long Beach, Long Beach, Calif (Dr Cole); Department of Dermatology, University of

---

**Table: Measures of Central Tendency, Variability, and Skewness of Actinic Keratosis Counts**

<table>
<thead>
<tr>
<th>Patient Letter</th>
<th>Session</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>SD</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Morning</td>
<td>58.0</td>
<td>49</td>
<td>30</td>
<td>105</td>
<td>27.53</td>
<td>0.95</td>
</tr>
<tr>
<td>B</td>
<td>Morning</td>
<td>2.9</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>2.19</td>
<td>0.25</td>
</tr>
<tr>
<td>C</td>
<td>Morning</td>
<td>13.1</td>
<td>9</td>
<td>6</td>
<td>27</td>
<td>8.19</td>
<td>1.08</td>
</tr>
<tr>
<td>D</td>
<td>Morning</td>
<td>1.3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>E</td>
<td>Morning</td>
<td>6.7</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td>3.40</td>
<td>0.95</td>
</tr>
<tr>
<td>F</td>
<td>Afternoon</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.53</td>
<td>0.37</td>
</tr>
<tr>
<td>G</td>
<td>Afternoon</td>
<td>1.4</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1.13</td>
<td>-0.24</td>
</tr>
<tr>
<td>H</td>
<td>Afternoon</td>
<td>13.1</td>
<td>14</td>
<td>4</td>
<td>19</td>
<td>5.05</td>
<td>-0.93</td>
</tr>
<tr>
<td>I</td>
<td>Afternoon</td>
<td>17.6</td>
<td>18</td>
<td>11</td>
<td>26</td>
<td>5.19</td>
<td>0.48</td>
</tr>
<tr>
<td>Total*</td>
<td></td>
<td>12.7</td>
<td>5</td>
<td>0</td>
<td>105</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ellipses indicate not applicable.
California–Irvine (Dr Cole); Division of Dermatology, Hines VA Medical Center, Hines, Ill (Dr Eilers); Division of Dermatology, Loyola University, Chicago, Ill (Dr Eilers); Department of Dermatology, VA Medical Center Oklahoma City and University of Oklahoma, Oklahoma City (Dr Naylor); Department of Dermatology, Carl T. Hayden VA Medical Center and University of Arizona, Phoenix (Dr Kalivas); Department of Dermatology, University of Miami, and Dermatology Service, Miami VA Medical Center, Miami, Fla (Dr Taylor); Department of Dermatology, VA Medical Center West Los Angeles, Los Angeles, Calif (Dr Gladstone); and Department of Dermatology, University of California–San Diego (Dr Piacquadio). Dr Gladstone is now affiliated with the Division of Dermatologic Surgery, Stanford University, Stanford, Calif.

This project was supported by grant 402 from the Cooperative Studies Program of the Department of Veterans Affairs Office of Research and Development, Washington, DC.


We thank Margarita Givens, MD, Kimberly Marcolivio, MEd, and Gail Kirk, MS, for their assistance with this research.

Corresponding author and reprints: Martin A. Weinstock, MD, PhD, Dermatoepidemiology Unit, Veterans Affairs Medical Center–111D, 830 Chalkstone Ave, Providence, RI 02908-4799 (e-mail: maw@brown.edu).

REFERENCES


