Agreement between Self-Assessment of Melanocytic Nevi by Patients and Dermatologic Examination

Petra G. Buettner1 and Claus Garbe2

The number of melanocytic nevi is the strongest risk factor for cutaneous melanoma. As pigmented skin lesions are visible to everybody, the question has been raised about whether people can identify themselves as being at risk for melanoma through self-counting of moles. In 1991, a total of 513 central European melanoma patients and 498 controls were asked to count the total number of nevi and the number of atypical nevi on the whole body. Whole-body examination by dermatologists followed. Agreement was assessed on categorized nevus counts by means of ordinal kappa values and log-linear modeling. Study subjects significantly underestimated the total number of melanocytic nevi ($p < 0.0001$). Chance-corrected overall agreement was rather poor (kappa = 0.14), and the ability to detect many existing nevi was low. Agreement was higher for atypical melanocytic nevi counts (kappa = 0.37), and the sensitivity to detect more than one atypical nevus was 0.48. Self-assessment of the number of melanocytic nevi was difficult to perform accurately, and people severely underestimated the actual number. Despite these results, people should be encouraged to perform regular skin self-examination for early detection of melanoma. Am J Epidemiol 2000;151:72-7.

The number of benign melanocytic nevi has been identified as one of the strongest predictors for the development of cutaneous melanoma (1). Studies that differentiated between common and atypical melanocytic nevi revealed both types to be very important risk factors for cutaneous melanoma (1–3). In fact, study results suggest that people at risk for developing melanoma may well be reliably identified by their pigmentary status alone, without further knowledge of their case history (2, 3).

Cutaneous melanoma is an obvious candidate for screening programs, since incidence rates are increasing in most white populations, and prognosis is excellent in early stages of the disease (4–6). Until now, the only screening test proposed was the visual examination of the skin, which seems to be a rather simple exercise. Yet, so far, the value of a screening program for cutaneous melanoma has not been formally assessed in a randomized trial in terms of mortality, test validity, or effectiveness. Many dermatologists recommend total skin examinations for all new dermatology patients and for patients who previously suffered from melanoma, and several studies have shown that total skin examinations are valuable in the detection of suspicious skin lesions (7–9). It has been proposed that physicians should be more aware of the importance of pigmented lesions for the identification of people at risk and thereby contribute to early detection of cutaneous melanoma.

On the other hand, cutaneous melanoma as well as benign pigmented lesions are visible to everybody. Therefore, the idea of individuals being able to assess their own skin lesions and, therefore, contributing to early detection of melanoma through self-screening is very attractive. One case-control study was able to show that skin self-examination will probably reduce mortality from melanoma, but results should be interpreted with caution because the follow-up time was short (10). Because the risk of developing melanoma increases with the number of melanocytic nevi, self-counting of nevi could be a cost-effective and promising way of self-assessment of risk and could be routinely performed as part of skin self-examination (1–3). Thus, it seems worthwhile to study the validity of self-counting of nevi in more detail. Within the framework of the central European risk factor study of cutaneous melanoma, subjects self-counted benign melanocytic nevi, and this study compares these self-counts with the findings of dermatologists (3, 11).
MATERIALS AND METHODS

Study subjects

In 1991, 283 (55.2 percent) female and 230 (44.8 percent) male patients with primary cutaneous melanoma were newly diagnosed at nine cooperating university departments of dermatology in Germany (Berlin, Mannheim, Luebeck, Hamburg, Bonn, Homburg, and Heidelberg), Austria (Graz), and Switzerland (Zurich). The median age of the patients was 55 years (range, 16–93 years). A total of 498 non-melanoma patients from the participating dermatologic clinics, matched to the melanoma patients according to age (±5 years) and sex, were chosen as controls. Further methodological details were published previously (3, 11).

Pigmented lesions

The clinical characteristics of pigmented lesions were discussed by all examining dermatologists in a consensus workshop prior to the start of the study. The criteria agreed upon were illustrated by photodocumentation that served as a diagnostic guideline for the examiners throughout the study. Pigmented lesions with a diameter of greater than 2 mm were classified as common melanocytic nevi if they were 1) macular, brown to dark brown, and sharply bordered (junctional nevi), or 2) papular, regularly and sharply bordered as well as light to dark brown (compound nevi), or 3) papular to nodular and skin-colored to erythematous (dermal nevi). The clinical diagnosis of atypical melanocytic nevi was established when at least three of the following five characteristics were present: 1) diameter greater than 5 mm; 2) ill-defined border; 3) irregular margin; 4) varying shades in the lesion; and 5) simultaneous presence of papular and macular components. Actinic lentigines were defined and distinguished from melanocytic nevi as lesions with only a macular surface, light brown or grey-brown in color, and having well-defined borders with sometimes finely irregular margins. Examining dermatologists performed a whole-body examination on all study subjects. Common and atypical melanocytic nevi were recorded separately for 12 body sites (excluding scalp and genital region), and results were documented on standardized questionnaire forms. The results of the dermatologic assessment of the skin were regarded as the gold standard.

Self-counts of the number of melanocytic nevi

Patients counted nevi on their skin prior to the physical examination by a dermatologist. They received a short description in everyday language about the morphologic appearance of pigmented nevi, which are called "liver spots" in colloquial German language. They were informed about possible variations in size and pigmentation. In addition, the subjects were told about the criteria for diagnosing atypical melanocytic nevi (irregular border, ill-defined border, variation in color, flat macular part, and diameter of more than 5 mm) in colloquial German terms. No photographs of skin lesions were shown. Individuals were not instructed to follow a detailed procedure for counting, and they were not advised to use a mirror or a size gauge. However, mirrors were available in the changing rooms for use.

Patients were asked to count the total number of benign melanocytic nevi greater than 2 mm on their whole body and separately on their arms. In addition, they were asked to count the number of atypical melanocytic nevi on their whole body. For the total number of melanocytic nevi on the whole body, subjects distinguished between none, 1–9, 10–19, 20–39, 40–59, 60–100, and more than 100 nevi. To arrive at the number of melanocytic nevi on both arms, they differentiated between none, one, 2–4, 5–10, and more than 10 nevi. They were likewise instructed to count the number of atypical melanocytic nevi greater than 5 mm by distinguishing between none, one, and more than one atypical nevi. After patients had estimated the number of nevi, a dermatologist counted melanocytic nevi as part of their whole-body examination.

Statistical analyses

Overall percentages of agreement, uncorrected for agreement by chance, were calculated and given along with 95 percent confidence intervals. Chance-corrected ordinal kappa values calculated with linear weights were presented along with 95 percent confidence intervals (12). The proportions of correct estimates given by the patients were calculated within each class of nevi counts assessed by the dermatologists and were referred to as “sensitivities.” The percentages of correct estimates given by the patients in relation to total estimates within each category were designated as “predictive values.” Either exact or approximate 95 percent confidence intervals were given along with sensitivities and predictive values.

Log-linear models were used to investigate the structure of agreement in more detail (13). Three log-linear models were formulated: 1) the model of independence; 2) the model concerning diagonal agreement; and 3) that concerning diagonal agreement plus triangle parameters. The model of independence assumes that the assessments of the number of melanocytic nevi counted by patients and dermatologists are independent.
and was regarded as the baseline model. The deviance is a measure of the goodness-of-fit of the model to the cross-classified data. Calculations were performed in SPSS for Windows, release 6.1.3 (SPSS, Inc., Chicago, Illinois), and with GLIM release 4.1 (14). Throughout the analysis, p values below 0.05 were regarded as statistically significant.

RESULTS

Counts of melanocytic nevi on the whole body and on the arms

Whole-body counts of benign melanocytic nevi from both patients and dermatologists were available for 952 study subjects. In general, nevus counts of dermatologists were much higher compared with patients' self-assessments (table 1). Eighty-nine patients were found by dermatologists to be without any melanocytic nevi, whereas 397 patients assessed themselves as having no melanocytic nevi. Those patients who presented without melanocytic nevi showed a high sensitivity for recognizing this correctly (sensitivity = 0.79, 95 percent confidence interval (CI): 0.69, 0.87). However, in general, patients with many nevi had an extremely low sensitivity to judge themselves correctly. Predictive values were, in general, found to be rather low. Only the four patients who had estimated that they belonged to the highest category of nevus counts (more than 100 melanocytic nevi) classified themselves correctly. The uncorrected overall agreement reached 24.7 percent (95 percent CI: 22.0 percent, 27.4 percent), and the kappa statistic marked a rather low overall, chance-corrected concordance between self-assessments of patients and dermatologic counts of benign melanocytic nevi on the whole body (kappa = 0.14, 95 percent CI: 0.09, 0.18).

Counts of benign melanocytic nevi on the arms showed a picture similar to the whole-body counts (table 2). Again, the proportion of patients who correctly estimated that they had no nevi on their arms was fairly high (sensitivity = 0.93, 95 percent CI: 0.90, 0.96). However, a total of 677 patients stated that they had no melanocytic nevi on their arms, while this assessment was confirmed by dermatologists for only 239 patients (predictive value = 0.35, 95 percent CI: 0.31, 0.39). Uncorrected overall agreement reached 35.9 percent (95 percent CI: 32.8 percent, 39.0 percent), and the chance-corrected overall agreement between the counts of melanocytic nevi on the arms was not substantially higher than that for whole-body counts (kappa = 0.18, 95 percent CI: 0.13, 0.23).

Self-assessments and dermatologic counts of the number of melanocytic nevi on the whole body were analyzed separately in patients with cutaneous melanoma and in nonmelanoma controls. For whole-body counts, the results in terms of sensitivity and predictive values (data not shown) as well as in kappa statistics were very similar (kappa = 0.12, 95 percent CI: 0.06, 0.18 for melanoma patients; kappa = 0.12, 95 percent CI: 0.06, 0.19 for nonmelanoma controls). Analysis was also stratified by age and gender, revealing no significant differences. Results were similar for the number of melanocytic nevi on the arms.

Counts of atypical melanocytic nevi

A different picture was obtained for the assessment of atypical melanocytic nevi (table 3). Both sensitivity and predictive value for the category of no atypical melanocytic nevi were high, with values of 0.84 (95 percent CI: 0.81, 0.87) and 0.89 (95 percent CI: 0.87, 0.91), respectively. Likewise, sensitivity and predictive value for the category of more than one atypical melanocytic nevi was fairly high, with values of 0.48 (95 percent CI: 0.37, 0.59) and 0.52 (95 percent CI: 0.41, 0.63), respectively. The uncorrected overall agreement was 76.4 percent (95 percent CI: 73.7 percent, 79.1 percent). The chance-corrected overall agreement was 70.0 percent (95 percent CI: 66.2 percent, 73.8 percent). The kappa statistic marked a rather low overall concordance between self-assessments of patients and dermatologic counts of atypical melanocytic nevi on the whole body (kappa = 0.12, 95 percent CI: 0.06, 0.18).

<table>
<thead>
<tr>
<th>TABLE 1. Concordance between self-assessment of patient and dermatologist on numbers of melanocytic nevi on the entire body (n = 952), central Europe, 1991</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. counted by dermatologist</td>
</tr>
<tr>
<td>-------------------------------</td>
</tr>
<tr>
<td>No. counted by patient</td>
</tr>
<tr>
<td>1–9</td>
</tr>
<tr>
<td>10–39</td>
</tr>
<tr>
<td>40–100</td>
</tr>
<tr>
<td>&gt;100</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Sensitivity</td>
</tr>
</tbody>
</table>

* Numbers in italics indicate agreement between the dermatologist and the patient.
agreement for the assessment of atypical nevi was higher than that for all melanocytic nevi (kappa = 0.37, 95 percent CI: 0.27, 0.47). When patients with cutaneous melanoma and nonmelanoma controls were analyzed separately, a reasonably high value for kappa was obtained only for melanoma patients (kappa = 0.39, 95 percent CI: 0.29, 0.50). In subjects, concordance was rather poor (kappa = 0.16, 95 percent CI: -0.04, 0.36). This difference in concordance between melanoma and nonmelanoma patients was probably related to the fact that dermatologists found only 13 persons with more than one atypical melanocytic nevus in the nonmelanoma group. The analysis was also stratified according to age and gender and found no significant differences.

Results of log-linear models of agreement

Log-linear models were calculated for whole-body counts of benign melanocytic nevi, counts of melanocytic nevi on arms, and whole-body counts of atypical melanocytic nevi, as presented in tables 1-3. In all three cases, the model of independence showed a rather poor fit to the data. When diagonal parameters were additionally included in the three models, the goodness-of-fit significantly improved (p < 0.0001). In a second step, models with diagonal agreement plus triangle parameters were fitted to the data of counts of common melanocytic nevi on the whole body and on the arms. In general, this type of model is appropriate to prove that one observer (e.g., the patient) tends to judge systematically higher or lower than the other observer (e.g., the dermatologist). The additional inclusion of the terms of the upper triangle led to improvements that were again highly significant when compared with the latter models, statistically proving the underestimation of counts by patients (p < 0.0001).

When the data for atypical melanocytic nevi are analyzed in more detail, the improvement of the goodness-of-fit by inclusion of diagonal parameters could be confirmed for the first diagonal element (no atypical melanocytic nevi according to the patient’s estimate and the dermatologist’s count) and for the second element (more than one atypical melanocytic nevus according to patient’s estimate and dermatologist’s count) (p < 0.0001). The impact of the second diagonal element (one atypical melanocytic nevus according to the patient’s estimate and the dermatologist’s count) was not statistically significant, however, revealing people’s uncertainty about the definition of atypical nevi.

DISCUSSION

This study was part of a central European risk factor study on the development of cutaneous melanoma, and it is one of the few studies, and so far the largest study, assessing the validity of self-counts of melanocytic nevi. Approximately 1,000 study subjects participated, almost all with complete information about the self-assessment of benign melanocytic nevi on the whole body. Interviewers explained to the patients in colloquial German language and in detail that lesions were to be recognized as a common or a large atypical mole, and only after the self-assessment did dermatologists count the nevi. Taking the findings of dermatologists as the gold standard might be regarded as questionable, as some studies pointed out that agreement among trained clinicians was only limited, especially with respect to the morphologic features of atypical nevi (15-17). The apparent necessity of a standardized assessment of pigmented skin lesions was recognized during the design phase of this study, and a consensus workshop was held for all participating dermatologists, leading to a diagnostic guideline that was used.

Am J Epidemiol  Vol. 151, No. 1, 2000
throughout the study. Therefore, it can be assumed that the participating dermatologists classified pigmented lesions according to the same criteria.

The results of this study were based on a nonrandom sample of the adult German, Austrian, and Swiss population, since people included in the analysis were subjects of a case-control study of cutaneous melanoma and were patients of dermatology departments at that time. The potential bias resulting from the selection of hospitals should be of minor importance because the nine cooperating centers serve a large and varied population, including exclusively urban and more rural as well as higher and lower socioeconomic communities. On the other hand, melanoma patients are obviously not representative of the general population because, on average, they presented with higher numbers of nevi. The controls were also dermatologic patients, however, with other conditions not related to skin cancer or pigmented lesions. By choosing patients with dermatologic problems for our study, the bias most likely to occur is a detection bias, since dermatology patients might be more concerned about their skin lesions than is the general population. This detection bias should increase the likelihood of recognizing and counting nevi correctly, leading to an overestimation of the agreement of self-counts of nevi and counts by dermatologists in the general population. In addition, although the controls were not selected in a population-based manner, comparison of their age- and gender-stratified nevi distributions with previously published data for healthy Caucasian populations suggested that with respect to nevi they were fairly representative of a general Caucasian population (18, 19). Within our study, the inclusion of both a group with many moles and one with fewer moles seemed desirable because it allowed us to study high- as well as low-risk skin situations. Within the statistical analysis, being a melanoma patient or not showed influence only in the estimation of the number of atypical nevi.

This study revealed that both melanoma patients and controls severely underestimated the total number of melanocytic nevi as well as the number of atypical melanocytic nevi. Overall agreement between the dermatologist's counts and the patient's self-assessment was rather poor for total body counts of benign melanocytic nevi and was found to be only fair for the assessment of atypical nevi by melanoma patients. Calculated sensitivities were, in general, lower than previously reported results, but direct comparisons were hampered because of differences in study designs (20-22). Gruber et al. (20) focused on the number of palpable nevi on arms (sensitivity = 0.63 for detection of one or more) and on the number of nevi on the whole body that were 5 mm or greater (sensitivity = 0.68 for detecting one or more). Lawson et al. (21) targeted large nevi on the whole body and found that 79 percent of the self-counts agreed with the physician's count within plus or minus three nevi. Little et al. (22) compared self-counts of the number of nevi on the front of the trunk with physician's counts and found a sensitivity of 0.79, a positive of predictive value of 0.75, and a kappa value of 0.74 for more than seven nevi of 2 mm or greater. However, the cutoff point of seven nevi was optimized with respect to the data, while its meaning with respect to risk of melanoma is unclear, and the cited sensitivity was the highest Little et al. could establish.

In the light of these previous studies, the less encouraging results of our investigation were not surprising. It can be argued that taking into account flat or nonpigmented nevi will lower sensitivity, and, in fact, this was already shown by Lawson et al. (21). Focusing on large nevi may increase the sensitivity of self-counts for obvious reasons; however, risk factor studies of melanoma consistently found the total number of nevi to be most important (1-3). Thus, while people with several large nevi are undoubtedly at risk, it may be misleading to target solely the large nevi and miss the individuals with many smaller nevi.

Restricting the body site of self-counting with respect to accessibility and ease may inflict similar problems. It can be argued that validity of self-counts might be higher on the arms, the front of the trunk, or the legs, and although our own data could not substantiate this suggestion, Little et al. (22) did. On the other hand, it remains unclear whether counts on these sites are good indicators of the number on the whole body. In general, it can be expected that the body site with the highest number of nevi correlates best with the whole-body count. However, because nevi distribution is dependent on gender, age, and sun exposure, this body site may vary (11, 23, 24). Sites of interest could be the posterior trunk for men and the legs for women, where most melanomas as well as most melanocytic nevi grow (23). However, in contrast to the lower extremities, the back is clearly a difficult body site to count, decreasing the chances of valid counts for men. On the other hand, a whole-body count can be regarded a surrogate result of more or less correct counts.

Self-assessment of atypical melanocytic nevi showed a slightly different picture compared with that for all benign nevi. Although most subjects with atypical nevi did underestimate their number, many subjects with no atypical nevi (15.8 percent) did misclassify other skin lesions as atypical nevi. Individuals were obviously confused about the appearance of the lesions they were asked to count. So far, only one study investigated the validity of self-counting for atypical melanocytic nevi (25). This population-based
study of 400 middle-aged Swedish women found a sensitivity of 0.29 and a predictive value of 0.20 for the presence of atypical nevi. These values are very similar to our findings when analysis was restricted to women of the control group (sensitivity = 0.21, 95 percent CI: 0.06, 0.46; positive predictive value = 0.13, 95 percent CI: 0.04, 0.31). However, again the results were not completely comparable, since the previous study restricted nevi counts to the lower extremities, while our study dealt with whole-body counts.

Our findings were not totally unexpected, since recognizing and counting melanocytic nevi on the whole body is a tedious exercise and, as stated before, is known to be difficult even for dermatologists (16, 18). However, the findings should not discourage physicians from educating people to become familiar with their skin and with early warning signs of melanoma. From a public health point of view, it is very reasonable to educate people to examine their skin and to refer them to a specialist if they consider that they have a large number of skin lesions or if they notice a new lesion or changes in an old one. From this broader perspective, the true identity of the skin lesions or the exact number of nevi seems to be of minor importance. On the other hand, the revealed severe underestimation of numbers of nevi is of concern, since people who do not identify themselves as being at risk will not seek medical advice. Therefore, the findings emphasize the need for adequate information and training, leading to an improved identification of moles by laypersons performing skin self-examination.

We conclude that self-assessment of the number of melanocytic nevi on the whole body was difficult to perform accurately and that people tended to severely undercount their moles. Thus, counting nevi should remain a domain of medical professionals. Despite these findings, people should nevertheless be educated and encouraged to perform skin self-examinations, since awareness is probably the best form of early detection or even prevention of melanoma.

REFERENCES