

Suggested methodology for longitudinal evaluation of nevi based on clinical images

Ofer Reiter^{1,2}  | Shenara Musthaq¹ | Nicholas R. Kurtansky¹  | Dulce M. Barrios¹ | Allan C. Halpern¹ | Michael A. Marchetti¹ | Ashfaq A. Marghoob¹ | Japbani K. Nanda¹ | Joseph Stoll¹ | Veronica Rotemberg¹

¹ Dermatology Service, Memorial Sloan Kettering Cancer Center, New York, USA

² Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

Correspondence

Ofer Reiter, Memorial Sloan Kettering Cancer Center, 454 E 74th St., New York, NY 10065, USA.

Email: ofer.rtr@gmail.com

Funding information

The National Cancer Institute Comprehensive Cancer Center Support Grant (CCSG) for Memorial Sloan Kettering Cancer Center, Grant/Award Number: P30 CA008748; Melanoma Research Alliance Dermatology Fellowship Award, Grant/Award Number: A # 654544

Abstract

Background: Melanoma screening includes the assessment of changes in melanocytic lesions using images. However, previous studies of normal nevus temporal changes showed variable results and the optimal method for evaluating these changes remains unclear. Our aim was to evaluate the reproducibility of (a) nevus count done at a single time point (method I) versus two time points (method II); and (b) manual and automated nevus diameter measurements.

Materials and methods: In a first experiment, participants used either a single time point or a two time point annotation method to evaluate the total number and size of nevi on the back of an atypical mole syndrome patient. A Monte Carlo simulation was used to calculate the variance observed. In a second experiment, manual measurements of nevi on 2D images were compared to an automated measurement on 3D images. Percent difference in the paired manual and automated measurements was calculated.

Results: Mean nevus count was 137 in method I and 115.5 in method II. The standard deviation was greater in method I (38.80) than in method II (4.65) ($p = 0.0025$). Manual diameter measurements had intraclass correlation coefficient of 0.88. The observed mean percent difference between manual and automated diameter measurements was 1.5%. Lightly pigmented and laterally located nevi had a higher percent difference.

Conclusions: Comparison of nevi from two different time points is more consistent than nevus count performed separately at each time point. In addition, except for selected cases, automated measurements of nevus diameter on 3D images can be used as a time-saving reproducible substitute for manual measurement on 2D images.

KEYWORDS

3-dimensional images, count, mole, nevus, size, total body photography

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Skin Research and Technology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

In recent years, medical photography has gained popularity among dermatologists.¹ One of the major cases used for imaging in dermatology is total body photography for nevus monitoring and melanoma screening. To understand the abnormal changes that should prompt suspicion for malignancy, one must first understand the normal and benign evolution of pigmented melanocytic nevi.

Although this topic has been studied for over 60 years, controversy still exists. Cross-sectional studies have shown that nevus counts are lower in older patient cohorts.^{2,3} This finding may be attributed to many factors, such as the disappearance of nevi with age or differences in sun exposure patterns across birth cohorts. While longitudinal studies have assessed the evolution of nevi, they found variable results and lacked standardization, evaluated nevi of varying numbers and anatomic sites, or had limited follow-up.⁴⁻⁷ Therefore, additional long-term longitudinal studies are required to produce results that will be applicable in clinical settings. However, the optimal method to evaluate images for changes in nevi, such as their appearance, disappearance, and changes in diameter, remains undetermined. In addition, manual measurements of nevi are laborious and time consuming and are not always reproducible. A need exists for an evidence-based, efficacious temporal nevus evaluation method to be used in long-term studies.

In recent years, several software systems have been developed that create three-dimensional (3D) reconstructions of cutaneous images.^{8,9} These systems show promising results for automated temporal measurements of skin area.¹⁰

The first aim of this study was to evaluate whether a single time point quantitative nevus count is as reproducible as two time point nevus count. The second aim was to examine the reproducibility of 2D nevus diameter measurements and whether they can be replaced with a 3D automated measurement tool, without compromising the results.

2 | MATERIALS AND METHODS

This study was approved by the Institutional Review Board of Memorial Sloan Kettering Cancer Center.

2.1 | Part I—Quantitative nevus counts that can be compared across time points

In this pilot study, four fourth-year medical students completed a 2-day course of basic clinical and dermoscopy skin lesion evaluation. They used a single time point (method I) and a two time point (method II) annotation method to evaluate the total number of nevi on the back of a patient with atypical mole syndrome. For this experiment, we selected a male patient who had more than 100 nevi with varying diameters on his torso and a relatively low number of seborrheic keratosis and lentiginos.

In each method, students were asked to differentiate between nevi that were smaller and greater than 4 mm in diameter and to only count nevi, excluding freckles, lentiginos, and seborrheic keratoses.

2.2 | Part II—Automated versus manual quantitative nevus size estimation

The four medical students additionally provided the largest diameter (mm) of each nevus they counted. Measurements were performed on 2D images using ImageJ software (National Institutes of Health, USA). Measurements on the 3D image were performed automatically on images taken 15 years after the first 2D images, using VECTRA software (Canfield Imaging, Parsippany, NJ, USA). Examples of 2D and 3D images can be found at the manufacturer's website (www.canfield.com).

In the second experiment, five atypical mole patients who had 2D and 3D images taken on the same day were randomly selected. A board-certified dermatologist (OR) performed measurements of nevi at least 4 mm in diameter on the 2D images using the same ImageJ software. The measurement of each nevus was compared to automated measurement on 3D images of the same patient taken on the same day, using VECTRA software (Canfield Imaging).¹⁰

2.3 | Statistical measurements

Nevus counts were compared between methods to assess whether inclusion of the second time point reduces the variability between readers. The nevus counts were assumed to follow a normal distribution. A Monte Carlo simulation using parameter estimates from method I was used to calculate the likelihood that the variance observed in the two time point annotation would have been observed under single time point annotation conditions.

For part II, lesions that were observed by all four readers were compared to assess the agreement of manual diameter estimation using intraclass correlation. The percent difference in the paired manual 2D and automated 3D measurements was calculated and bootstrapped to estimate systematic differences between the automated and manual measurements.¹¹ Percent difference between paired measurements proved to be the superior mode of quantification to combat the heteroscedastic variability between manual and automated estimates.

3 | RESULTS

3.1 | Part I—Quantitative nevus counts that can be compared across time points

A summary of results for the nevus count methods is presented in Table 1. The standard deviation (SD) in total nevus count was greater in method I (38.80) than in method II (4.65) ($p = 0.0025$). Similarly, the

TABLE 1 Summary of torso nevus count results

	Method I		Method II	
	\bar{x}	s	\bar{x}	s
Total nevi count	137.00	38.80	115.50	4.65
<4 mm	114.25	41.87	51.00	3.56
>4 mm	22.75	6.75	64.50	1.29

Note: Method I—Single time point annotation; method II—two time point annotation.

nevi count by size specifically less than 4 mm and greater than 4 mm had larger SDs under single time point annotation (41.87 and 6.75) compared to two time point annotation (3.56 and 1.29) ($p = 0.0009$ and 0.0099).

Using method II, the nevi count of nevus points that were at least 4 mm in diameter resulted in a variability represented by an SD of 1.29, as opposed to 3.56 for nevi smaller than 4 mm. This difference did not reach statistical significance ($p = 0.065$), which may reflect the limited number of participants.

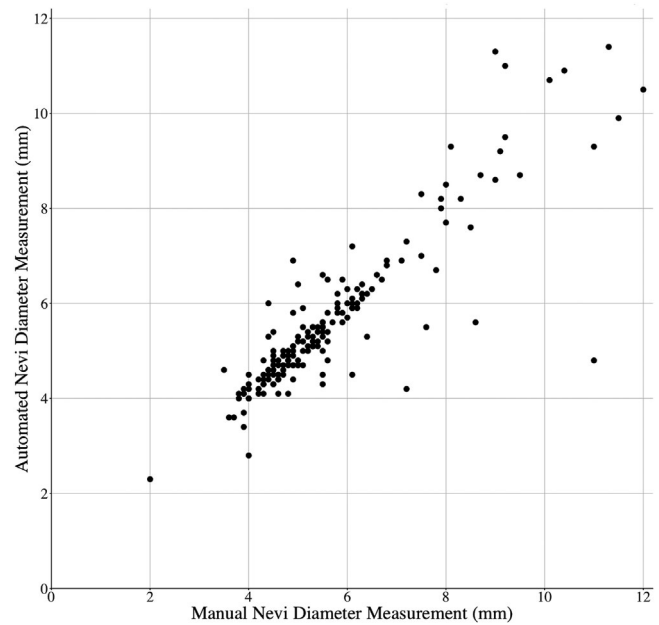
3.2 | Part II—Quantitative nevus size estimation

Within the 37 nevi recorded unanimously by all 4 students, there was good agreement in diameter estimation between participants, summarized with intraclass correlation coefficient (ICC) = 0.88 (95% CI [0.77, 0.93]). In method II, reader agreement for changes in diameter between the two time points produced a strong ICC (0.85, 95% CI [0.73, 0.92]).

In the second experiment comparing manual and automated measurements of nevi, a total of 184 nevi on the torso of 5 patients were measured. Median patient age was 51 (range 49–79), one patient was female, and the median number of torso nevi at least 4 mm in diameter was 23 (range 10–82). Variability between manual and automated estimates increased with size (Figure 1). The observed mean percent difference between manual and automated measurements was 1.5% (95% CI [−0.5%, 3.8%]). The manual measurements were on average 1.5% (95% CI [−0.5%, 3.8%]) larger than their matching automated measurements. Nevi that were lightly pigmented or at the lateral edge of the torso had lower agreement between manual and automated estimates. The estimated manual diameters of these moles were 37.4% larger than their automated diameters. When excluding these lightly pigmented or lateral lesions from the calculation, the mean percent difference between manual and automated measurements improved to 0.3% (95% CI [−1.0%, 1.7%]).

4 | DISCUSSION

As different studies found variable results using different methods for temporal nevus evaluation on images,^{2–7} we sought to find a reproducible method for quantifying nevus count and size that will be comparable across different time points.

**FIGURE 1** Correlation between manual and automated nevus diameter estimates

4.1 | Part I—Quantitative nevus counts that can be compared across time points

The results of this experiment suggest that asynchronous counting of nevi at two time points is highly variable and may lead to incorrect conclusions regarding the appearance or disappearance of nevi. Synchronous counting of nevi via comparative image analysis of two time points is more accurate to evaluate the number of nevi, especially the number of new and disappearing nevi. Lightly pigmented nevi exemplify the advantage of a two time point comparison because they may be missed if some images are too bright. Additionally, a nevus that has become lighter may be missed at the second time point if a direct comparison of anatomic site is not made. Although imperfect, the direct comparative method is less likely to miss new or disappearing nevi as demonstrated in this experiment. Marking nevi ensured the same nevus was not counted twice.

The inherent challenges of a single time point measurement should be considered when evaluating the validity and reliability of nevus count assessments. For example, cross-sectional studies comparing nevus counts of individuals of different age groups may be subject to bias and did not capture new nevi or nevi that became lighter in older individuals, which may explain the differences between cross-sectional studies that suggested nevi tend to disappear with age^{2,3} and longitudinal studies that found the phenomenon of disappearing nevi to be much less common.^{4,12}

The SD of the number of counted nevi ≥ 4 mm was lower than for nevi < 4 mm. Similarly, previous studies that evaluated self-counts of nevi compared to physician counts found that counts of larger nevi (> 5 mm in diameter) were more reliable.^{13,14} This can be attributed to

the appearance of small freckles, lentigines, and nevi being more easily differentiated when larger.

4.2 | Part II—Quantitative nevus size estimation

Manual measurements of nevus diameter using ImageJ software were reproducible between different participants and correlate well with automated measurements made by VECTRA 3D software. In general, it seems that automated 3D measurements can be used as a reliable substitute for the time consuming and laborious manual measurements of nevi on 2D images. However, the major limitation of 3D automated measurement is automated lesion segmentation, which can be inaccurate for lightly pigmented lesions. The main limitation for 2D manual measurement is curvature of body sites, such as the lateral torso, where the 2D projection of the lesion on the image is smaller than its actual size. It is important that investigators be familiar with these limitations when drawing conclusions regarding temporal changes in nevus diameter.

Limitations of this experiment are the small numbers of participants that were mostly not board-certified dermatologists in addition to the small sample size, the inclusion of only patients with many moles and the analysis of only the torso and not the entire body.

5 | CONCLUSIONS

In conclusion, when evaluating nevus temporal evolution, comparison of nevi from two different time points is more consistent than nevus count performed separately at each time point. In addition, except for selected cases, automated measurements of nevus diameter on 3D images can be used as a time-saving and reproducible substitute for manual measurement on 2D images.

Therefore, we suggest that future studies evaluating the long-term evolution of nevi include images of each time point that will be compared synchronously, enabling reliable, unbiased, and reproducible measurements. These measurements may be assisted with an automated 3D measurement tool and preferably include larger nevi.

Further larger scale studies are required in order to examine these suggested methods and to form a gold standard for nevus temporal evolution assessment.

CONFLICT OF INTEREST

Allan C. Halpern—Canfield Scientific, Inc. (consultant), SciBase (advisory board), HCW LLC (equity), Skip Derm LLC (equity).

Veronica Rotemberg - Inhabit Brands, Inc. (expert advisor).

On behalf of all other authors, the corresponding author states that there is no conflict of interest.

ORCID

Ofer Reiter  <https://orcid.org/0000-0003-3360-2780>

Nicholas R. Kurtansky  <https://orcid.org/0000-0002-6745-0386>

REFERENCES

1. Milam EC, Leger MC. Use of medical photography among dermatologists: a nationwide online survey study. *J Eur Acad Dermatol Venerol*. 2018;32(10):1804–9.
2. Stegmaier OC. Natural regression of the melanocytic nevus. *J Invest Dermatol*. 1959;32(3):413–21.
3. Green A, Swerdlow AJ. Epidemiology of melanocytic nevi. *Epidemiol Rev*. 1989;11:204–21.
4. Halpern AC, Guerry D 4th, Elder DE, Trock B, Synnestvedt M, Humphreys T. Natural history of dysplastic nevi. *J Am Acad Dermatol*. 1993;29(1):51–7.
5. Schiffner R, Schiffner-Rohe J, Landthaler M, Stolz W. Long-term dermoscopic follow-up of melanocytic naevi: clinical outcome and patient compliance. *Br J Dermatol*. 2003;149(1):79–86.
6. Banky JP, Kelly JW, English DR, Yeatman JM, Dowling JP. Incidence of new and changed nevi and melanomas detected using baseline images and dermoscopy in patients at high risk for melanoma. *Arch Dermatol*. 2005;141(8):998–1006.
7. Abbott NC, Pandeya N, Ong N, McClenahan P, Smithers BM, Green A, et al. Changeable naevi in people at high risk for melanoma. *Australas J Dermatol*. 2015;56(1):14–8.
8. De Menezes M, Rosati R, Ferrario VF, Sforza C. Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. *J Oral Maxillofac Surg*. 2010;68(9):2129–35.
9. Rayner JE, Laino AM, Nufer KL, Adams L, Raphae AP, Menzies SW, et al. Clinical perspective of 3D total body photography for early detection and screening of melanoma. *Front Med*. 2018;5 (May):1–6.
10. Chung E, Marchetti MA, Scope A, et al. Towards three-dimensional temporal monitoring of naevi: a comparison of methodologies for assessing longitudinal changes in skin surface area around naevi. *Br J Dermatol*. 2016;175(6):1376–8.
11. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7(1):1–26.
12. Ribero S, Zugna D, Spector T, Bataille V. Natural history of naevi: a two-wave study. *Br J Dermatol*. 2021;184(2):289–95.
13. Buettner PG, Garbe C. Agreement between self-assessment of melanocytic nevi by patients and dermatologic examination. *Am J Epidemiol*. 2000;151(1):72–7.
14. Lawson DD, Schneider JS, Sagebiel RW. Nevus counting as a risk factor for melanoma: comparison of self-count with count by physician. *J Am Acad Dermatol*. 1994;31(3):438–44.

How to cite this article: Reiter O, Musthaq S, Kurtansky NR, Barrios DM, Halpern AC, Marchetti MA, et al. Suggested methodology for longitudinal evaluation of nevi based on clinical images. *Skin Res Technol*. 2022;28:71–74. <https://doi.org/10.1111/srt.13092>