

## Validity and Reliability of the Munsell Soil Color Charts for Assessing Human Skin Color

Anthony I. Reeder<sup>1</sup>, Ella Iosua<sup>2</sup>, Andrew R. Gray<sup>2</sup>, and Vanessa A. Hammond<sup>3</sup>

### Abstract

**Background:** Skin pigmentation is a key factor for ultraviolet radiation exposure-related cancers and can make a significant contribution to the patterns of other diseases. For surveys, and to appropriately target cancer control activities, valid and reliable measures of skin color are required.

**Methods:** Validity and reliability of the Munsell Soil Color Charts were investigated for skin color assessment. The unexposed skin color of 280 university students was measured by spectrophotometer to calculate an Individual Typology Angle (ITA) value, and categorized by two independent raters according to the Munsell system (the latter was repeated after a 7-day interval).

**Results:** Interrater and intrarater reliability for the Munsell charts was found to be acceptable [intraclass correlation coefficients (ICC) of 0.85 and 0.86, respectively]. When ITA values were converted to the six Del Bino skin color categories, weighted  $\kappa$  for agreement between raters, within rater, and between Munsell chip and spectrophotometer were 0.63, 0.60, and 0.61, respectively. A tendency toward overestimation of the extremes of skin pigmentation was evident, particularly for the "brown" and "dark" skin types.

**Conclusions:** Study findings suggest that the Munsell Soil Color Charts represent a reliable and valid measurement strategy when assessing skin type.

**Impact:** The Munsell Year 2000 Soil Color Charts may provide a useful instrument for fieldwork contexts. Subsequent classification of individuals into skin cancer risk categories, rather than the use of precise ITA values, may be sufficient for targeting public health messages for skin cancer prevention and other health risks. *Cancer Epidemiol Biomarkers Prev*; 23(10); 2041–7. ©2014 AACR.

### Introduction

Skin color makes a significant contribution to the patterns of certain diseases observed in populations. In particular, light skin color is associated with increased skin cancer risk, whereas dark skin color is associated with a greater risk of vitamin D deficiency (1). Among dermatologists, it has been acknowledged that "a new era for studying human cutaneous diversity" has begun (2). For the study of such issues, acceptable, valid, and reliable instruments for measuring skin color are required.

In an earlier publication, we reported how a self-reported skin color measure used in telephonic questionnaire surveys of large, randomly selected samples was a valid and acceptably reliable assessment tool, although there was a bias toward overestimation of skin pigmentation (3)—a pattern also observed among Caucasians in Australia (4). Such a bias may lead individuals to underestimate the vulnerability of their skin to damage from ultraviolet radiation (UVR) exposure and misinterpret the personal relevance of skin cancer prevention messages.

Although sophisticated spectrophotometry and other technologies are available for use, making it possible to obtain a direct assessment of skin color, the potential usefulness of other fieldwork procedures, especially when working with sizeable, free-living populations, is worthy of investigation. This may apply, in particular, to studies in which precise measures are not required, but validly identifying skin color categories is necessary, for example, for the targeting of appropriate public health messages for skin cancer prevention and other health risks.

The Munsell Year 2000 Revised Washable Edition of the Soil Color Charts which, as the name implies, was originally designed to assist the classification of soil samples, is described as also being suitable "for the evaluation of skin, hair and eye color in anthropology, criminology, pathology and forensic medicine" (5). The Munsell color

<sup>1</sup>Cancer Society of New Zealand Social and Behavioural Research Unit, Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand. <sup>2</sup>Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand. <sup>3</sup>Public Health South, Southern District Health Board, Wanaka, New Zealand.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

**Corresponding Author:** Anthony I. Reeder, Cancer Society of New Zealand Social and Behavioural Research Unit, Department of Preventive and Social Medicine, Dunedin School of Medicine, University of Otago, PO Box 913, Dunedin 9054, New Zealand. Phone: 64-3-479-7257; Fax: 64-3-479-7298; E-mail: [tony.reeder@otago.ac.nz](mailto:tony.reeder@otago.ac.nz)

doi: 10.1158/1055-9965.EPI-14-0269

©2014 American Association for Cancer Research.

order system provides a visual standard using color chips arranged logically, according to a three-dimensional expression based around the concepts of hue, value, and chroma. *Hue* describes the relation of a color to red, yellow, green, blue, and purple (for presentation on separate sheets), *value* indicates lightness (arranged vertically on each sheet in visually equal steps with the darkest at the bottom; black = 0 and white = 10), and *chroma* indicates strength (arranged horizontally increasing from left to right). The sheets are presented in a loose-leaf album to facilitate color matching. The color of an object can be identified either descriptively (e.g., by the term "reddish brown" from the color name diagram facing each sheet) or by the Munsell notation, for example, 5YR 5/4, where 5YR is the hue/sheet (Y standing for yellow and R for red), 5 the value, and 4 the chroma.

Other versions of the Munsell system have been used to classify color for a range of human tissues, including the skin over hand veins (6) and gastrointestinal mucosa (7), as well as tooth enamel (8). However, we could identify only one study (of urine color) that specifically reported using the Soil Color Charts (9), and none that used these charts for assessing skin color. We were also unable to find any study that used the more recent, revised, washable editions, which would be especially appropriate for repeated fieldwork. Two studies investigating validity in geological usage have reported acceptable correlations with criterion instruments (10, 11). Of another two geological studies that investigated interrater variance, one reported 92% agreement between investigators (12), but the other was unable to replicate that (13).

A number of potential limitations in using the Munsell system need to be considered. In the introduction to the manual, three "principal difficulties" encountered when using the charts are identified: selecting the appropriate hue card, determining colors intermediate between the hues in a chart, and distinguishing between value and chroma where the latter are strong (5). However, practical training in the use of the system should help overcome these issues. In addition, two "masks" are included in the binder pocket: a black one for use with dark samples and a gray one for intermediate and light samples. The aperture in each mask permits four adjacent color chips to be viewed simultaneously while other, potentially distracting, chips remain covered.

Apart from the possible risk of colors fading as a result of extended exposure to light (14), other limitations relate to potential biases due to variation in (i) the light source, (ii) observer discrimination (color perception variation between people), and (iii) observation angle. Variation arising from each of these factors can potentially be minimized by following standardized data collection procedures.

A range of factors support the potential usefulness of the Munsell Soil Color Charts (hereafter referred to as the Munsell) for assessing skin color in the field. First, they use standard, internationally comprehensible, visual categories so that no translation of color names is required.

Second, the instrument is relatively economical, currently costing approximately US \$200. Third, the charts and binder are physically compact, portable, and robust. Fourth, the Munsell requires no external power source, calibration, or processing software. Fifth, it is minimally invasive and likely to be acceptable to vulnerable populations, for example, children, among whom other technologies may seem frightening or threatening. Sixth, because of its simplicity, use of the Munsell requires no technical knowledge and minimal training. Finally, no pressure is applied when a Munsell chart is held against the skin, so skin color is not affected by the measurement process. Overall, the Munsell may be particularly useful in those more remote and/or less economically developed settings where acceptability, cost, portability, and a lack of reliable power sources and technical backup may be especially important.

The present study was designed to answer two specific questions, namely, whether the Munsell provides *in vivo* measures of skin color that are both reliable (i.e., exhibit good interrater and intrarater test-retest reliability) and valid, compared with a spectrophotometer criterion measure.

## Materials and Methods

### Sample selection and follow-up

Volunteers were recruited from the population of students domiciled in the 14 University of Otago (Dunedin, New Zealand) residential colleges. These colleges included one which catered for 75 postgraduate residents and 13 which catered for between 152 and 518 (mean, 248) of both sexes, mostly in single study bedrooms. This student population was relatively easy to access and follow up, allowing an efficient process for obtaining and retaining a sample of sufficient size. Toward the end of 2007, well in advance of student recruitment scheduled for early 2008, the Heads/Masters of all residential colleges were simultaneously invited to permit us to carry out the study on site. Sample size calculations were based on the resulting precision of  $\kappa$  statistics assuming two categories. For test-retest reliability, 280 participants would allow a two-category  $\kappa$  to be estimated to within  $\pm 0.15$  using 95% confidence intervals assuming that  $\kappa$  was 0.6 or higher and where the smaller category contained at least 15% of the respondents, and allowing for up to 30% loss due to invalid responses and loss to follow-up. Cluster effects within residential colleges were assumed to be negligible for this calculation.

### Procedures

All participants were treated following a standard protocol. Procedures were pretested for practicality and participant acceptability and the final protocol was incorporated into a printed manual for use by assessors. Data collection was carried out, April 29 to May 21, 2008, during evening, indoor sessions under standard Philips 840 fluorescent strip lamps using pretrained, senior university

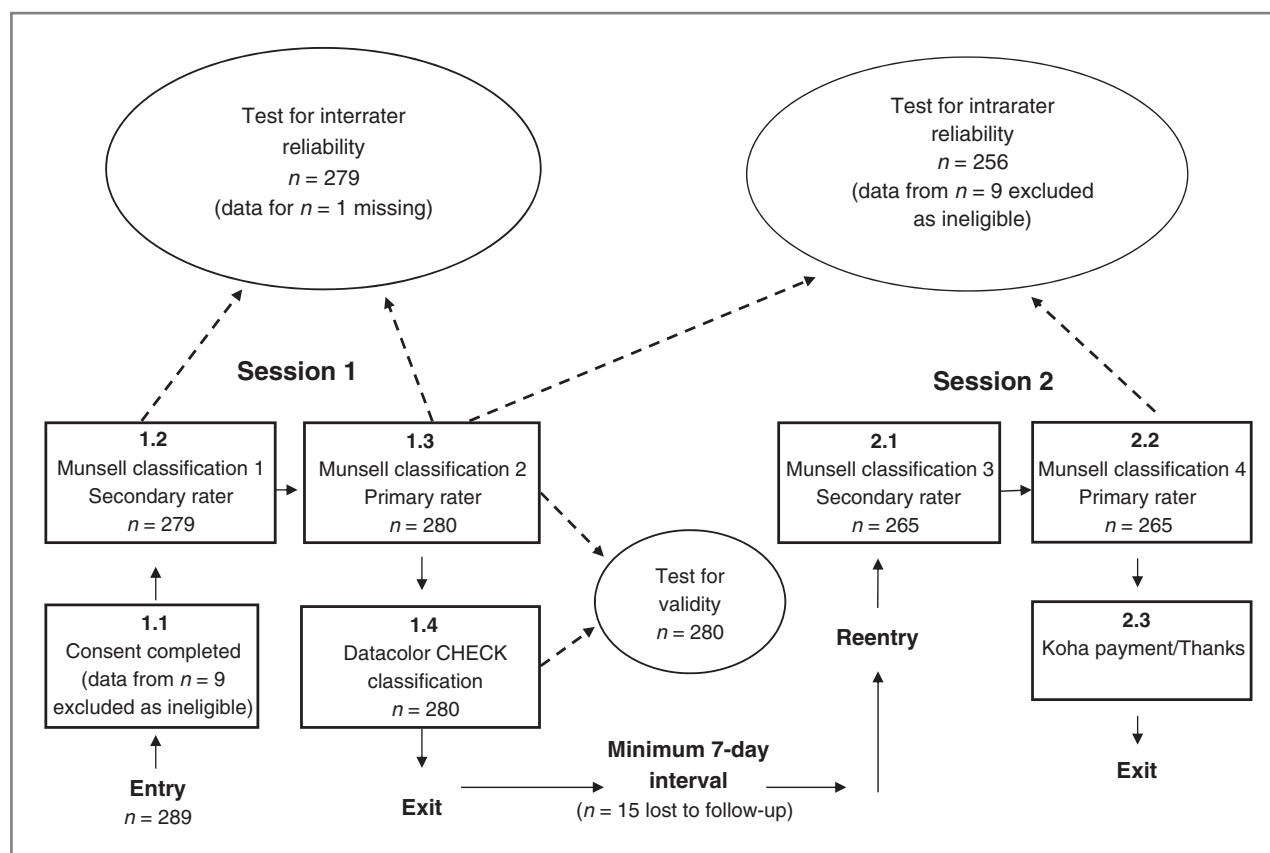


Figure 1. Flow chart of study procedures.

students as research assistants. Participants' skin types were assessed using the Munsell charts on two separate occasions (session 1 and session 2) to allow examination of test-retest reliability using the Munsell. These two sessions were separated by a minimum period of 7 days to minimize any recall effects on the part of assessors from the first occasion, while also minimizing the likelihood of any changes in the participants' skin reaction to sun exposure during the between-test period. Spectrophotometer readings were taken only during session 1. Figure 1 presents a flow diagram of participant movement through study assessments.

Ethical approval was obtained (October 3, 2007) through the Department of Preventive and Social Medicine, following University of Otago protocols. As an incentive to return for the follow-up assessment, each participant was offered a gift (koha) of \$20 cash, to be paid after completion of their second assessment, thereby acknowledging their contribution and compensating for possible inconvenience. Each participant was also allocated a number that was entered into a random draw for an additional \$200 prize. All participants, whether or not they took part in a follow-up assessment, were offered feedback in the form of a brief printed personal report about their skin type, and current guidelines about recommended UVR exposure, under New Zealand (NZ) conditions (15).

## Measures

Skin reflectance was measured using a DATACHECK spectrophotometer, the study criterion measure, which compared the color of each participant's skin with a standard calibration reference by emitting a flash of white light to the surface under test, and measuring the spectral properties of the light reflected back to the instrument. In each case, reflectance was converted to an Individual Typology Angle (ITA) score and a Del Bino skin color category (see Table 1; ref. 16) for comparison with the results obtained by using the Munsell charts. The ITA is a measure of the reflectance of light from the skin where 100% reflectance represents pure white and no reflectance (−100%) represents pure black. Accordingly, high and positive ITA scores indicate the greater reflectance of light associated with lighter skin colors as compared with low and negative ITA scores that indicate darker skin colors. The ITA is calculated using the values of  $L^*$  (the difference along the lightness – darkness axis) and  $b^*$  (the difference along the yellow – blue axis). It was calculated according to Del Bino (16) as

$$ITA = \tan^{-1} \left( \frac{L^* - 50}{b^*} \right) \times \frac{180}{\pi}$$

Skin classification by ITA values has physiologic relevance to melanin content (16). Such methods of comparison have been used in previous research (17).

**Table 1.** Distribution of participants ( $n = 280$ ) by demographics and session 1 spectrophotometer-derived Del Bino skin color category and ITA range

Response category	Frequency	%
Sex		
Male	113	40.4
Female	167	59.6
Total self-report ethnicity <sup>a</sup>		
New Zealand European	177	63.2
Māori	11	3.9
Pacific	6	2.1
Asian	90	32.1
All other	6	2.1
Spectrophotometer-derived Del Bino skin color category (ITA range)		
Very light ( $\geq 55^\circ$ to $\leq 90^\circ$ )	61	21.8
Light ( $\geq 41^\circ$ to $< 55^\circ$ )	116	41.4
Intermediate ( $\geq 28^\circ$ to $< 41^\circ$ )	58	20.7
Tanned ( $\geq 10^\circ$ to $< 28^\circ$ )	23	8.2
Brown ( $\geq -30^\circ$ to $< 10^\circ$ )	20	7.1
Dark ( $\geq -90$ to $< -30^\circ$ )	2	0.7

<sup>a</sup>Total adds to more than 100% because multiple ethnicities could be reported.

The spectrophotometer was held against the skin of the upper, inner arm of each participant. At this recommended site (18), the natural (i.e., untanned) skin color can most readily be measured, because it is a site that is not usually exposed to sunlight (16) and access is relatively noninvasive. The surfaces of the instrument that came into contact with participants' skin were cleaned with alcohol wipes between assessments. Three replicate skin reflectance measurements were performed on each subject, and then each measure was averaged before being used to calculate a single ITA value as a quantitative measure of skin color.

Categorization according to the Munsell charts was assessed by both the Primary and a Secondary Rater at each of the two sessions. The Primary Rater's FM 100 color test score (11) was assessed as "good" because the mean total error score reported for the 30- to 39-year age group is 50 (19). The color chart was held against the skin of the upper inner arm, the same site used for taking spectrophotometer readings, at a 45° angle to the observer. The rear surfaces of the charts that came into contact with participants' skin were cleaned with alcohol wipes between assessments. Observations were recorded on a classification form, designed for the study, which permitted identification of one of the four yellow/red charts (2.5YR, 5YR, 7.5YR, and 10YR) and the respective "value" and "chroma" categories provided in the manual. The ITA values for these Munsell color chips were subsequently independently obtained from spectrophotometer readings

(Supplementary Methods and Materials; Supplementary Tables S1–S4). These ITA values derived from spectrophotometer readings of the Munsell chips were then used to assess agreement with the ITA values obtained directly from the spectrophotometric skin assessments.

Self-reported information on sex, age (in years), and ethnicity (responses were classified using the five equivalent high-level Statistics NZ categories: NZ European, Māori, Pacific, Asian and "all other") was obtained from a brief questionnaire. The latter are the level-3 ethnicity categories recommended by Statistics NZ and reported in the NZ Census (20). The NZ European category predominately comprises people of Northern European ancestry, following the pattern of colonial settlement from England, Scotland, and Ireland with fewer immigrants from Holland, France, and Germany.

The final questionnaire item sought a dichotomous (yes/no) response to the question: "In the past week (7 days) have you used a sunbed or spray-on tan lotion?" Participants who provided a positive response to this question at baseline were excluded from all analyses reported here and those who provided a positive response at follow-up only were excluded from the intrarater analyses.

### Statistical analyses

The interrater reliability, or degree of agreement between two different raters (Primary and Secondary) in their session 1 skin color classifications, was assessed using the intraclass correlation coefficient (ICC) between the respective ITA values of the selected Munsell chips. Test-retest reliability was investigated using the ICC for ITA values between the Primary Rater's classifications in session 1 and their corresponding classifications in session 2. Munsell chart validity (i.e., rater vs. spectrophotometer) for measuring skin color was examined using the mean difference (bias) between the selected Munsell color chip ITA from the Primary Rater and the spectrophotometer ITA in session 1. Limits of agreement (95%)—the reference interval of two standard deviations of the ITA differences either side of the bias for the bias—were also calculated. In addition, weighted  $\kappa$  (using linear weights) were used to investigate both the reliability and validity of the Munsell color chips when the Del Bino skin categories were applied. ICCs of 0.70 or greater were considered acceptable (21), and similar for  $\kappa$  of 0.6 or higher ("substantial" in Landis and Koch; ref. 22). Stata software version 12 was used for all statistical analyses (23).

### Results

At the point when the Heads/Masters of 6 of the 14 residential colleges had agreed to permit their resident students to participate, the involvement of further colleges was not sought because a sufficient number of students ( $n = 289$ ) had been recruited. Of the 289, who took part in session 1, 9 were excluded for using either a sunbed or spray-on tan within the past 7 days, leaving data from 280 participants for analysis. The demographic



characteristics of these 280 students and the sample frequency distribution for the six different skin classifications defined by Del Bino according to ITA values (16) are presented in Table 1.

The 280 participants had a median age of 18 years (range, 16–49 years). For 1 student, a session 1 secondary rater Munsell assessment was lacking, leaving 279 for the interrater reliability analysis. For session 2, 265 students (95%) were followed up, but only data for the 256 who reported neither sunbed use nor cosmetic tanning in either session were used for assessing intrarater reliability.

The ICCs for agreement between the Primary and Secondary Raters and the test–retest reliability of the Primary Rater suggest sufficient inter- and intrarater reliability of the Munsell color charts to offer *in vivo* measures of skin color. Bias between the Primary Rater and the criterion spectrophotometer was  $-0.09$  (with  $-26.71, 26.54$  being the corresponding 95% limits of agreement). The value of  $-0.09$  indicates that, on average, application of the Munsell color charts only marginally overestimates skin pigmentation. The limits, however, show that 95% of the students will have a difference in ITA value (spectrophotometer vs. Primary Rater) between  $-26.71$  and  $26.54$ . This does not preclude the potential for a two-category discrepancy according to the Del Bino skin classification system (e.g., "light" vs. "tanned"). Close examination of the spectrophotometer ITA versus the Primary Rater suggests that the latter may have been overly responsive to the extremes of skin pigmentation, with a pronounced tendency to overestimate the darkness of both "brown" and "dark" skin colors, and a less evident, subtle pattern of exaggerating "lightness" (Fig. 2).

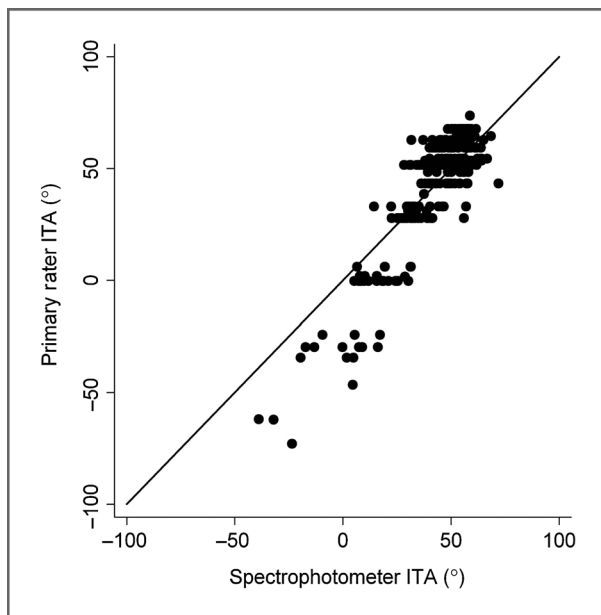


Figure 2. ITA recorded by the spectrophotometer compared with that reported by the Primary Rater.

When all three measures on the ITA scale were converted to Del Bino skin color categories, the weighted  $\kappa$  assessing agreement between raters (0.63), consistency within a rater (0.60), and convergence between the Munsell color chip and spectrophotometer (0.61), produced values that suggest acceptable interrater reliability, intrarater test–retest reliability, and criterion validity of the Munsell color charts in providing *in vivo* measures of skin color.

A sensitivity investigation was also performed for the intrarater test–retest reliability analyses. This involved re-running the analyses, but only including data where the difference in Del Bino skin classification for the comparison between sessions 1 and 2 for the Primary Rater was less than two categories, as this was quite large and could indicate actual changes in skin pigmentation between sessions (an example of an two or more category difference would be "light" vs. "tanned"). Data from 20 students were identified and removed. The subsequent inferences were comparable with those of the primary analyses with the ICC increasing from 0.86 to 0.93 and the  $\kappa$  from 0.60 to 0.70.

A more thorough inspection of the measurements taken by both the Primary and the Secondary Raters for the excluded participants indicated the possibility, in at least 10 of the 20 cases, of undisclosed tanning between sessions or collusion, whereby a different student with a different skin color presented in each session. This suggests that our conclusions about the intrarater reliability of the Munsell may be slightly conservative.

## Discussion

The objective of the present study was to answer two questions, namely, whether the Munsell Year 2000 Soil Color Charts can provide *in vivo* measures of skin color that are both reliable (i.e., exhibit acceptable interrater agreement and intrarater test–retest consistency) and valid, compared with a criterion measure (spectrophotometer).

Study findings indicate that the Munsell-derived ITA values and their related Del Bino (16) categories are acceptably reliable measures of skin color, including test–retest reliability when there is a 7-day interval period between assessments. With respect to validity, there was sufficient agreement between both the Munsell-derived constitutive skin color ITA values and associated Del Bino color categories for measurements obtained from the rater and from direct spectrophotometric assessment. These findings confirm that the Munsell Year 2000 Soil Color Charts could be considered an acceptable and appropriate alternative in situations where obtaining individual spectrophotometer readings is unlikely to be a practical option. Furthermore, our intrarater reliability conclusions may be conservative because we found even stronger agreement in the analysis where participants were removed if they could be suspected to have either not accurately

reported their tanning behaviors or colluded with colleagues to replace them as participants for session 2. However, some of these discrepancies may reflect substantial rater error rather than either of these alternative possibilities.

When we further examined our skin color category findings, a consistent pattern emerged of a tendency to overestimate the extremes of skin pigmentation, particularly for the darker ("brown" and "dark") skin types (there was a less evident, more subtle pattern of overestimating "lightness"). This is consistent with a pattern that we described as "the dark shift" in an earlier report (3). In that study, we also found some evidence that this tendency extended to the darker skin types, although the numbers involved were rather small.

Study findings suggest that it could potentially be useful to consider mailing out a high-quality color chart to survey participants when skin color is assessed, thereby permitting a direct comparison to be made between colors on the chart and personal skin color. At present, both mail and online population surveys depend, largely, on participant responses to self-report questionnaire items. Although lighting conditions are likely to vary, it could be useful to test whether this approach had the potential to provide more accurate data than the use of self-reported verbal color categories. With respect to online surveys, the provision of a standard color chart may be more challenging because of possible variation in color reproduction and difficulties in physical positioning for making comparison.

Our study had some limitations that future research should endeavor to address. First, although our sample included indigenous Māori, Pacific, and Asian participants, there were relatively few participants in the darker skin color range, which is characteristic of the NZ population. It would be useful to study whether overestimation of the extremes of skin pigmentation occurs among populations that include a greater number of participants with darker skin colors. We are planning such a study.

Second, the likelihood of collusion may be associated with the provision of cash rewards in our study. This was a possibility that we had not anticipated, but which should be guarded against in future studies by requiring authoritative photographic identification. Furthermore, we should have taken into account that young adult students have a reputation for challenging accepted practices and those domiciled in the same residential college may be a population particularly prone to taking advantage of the opportunities provided by the nonrequirement of authoritative photographic identification to collude, especially when there is a cash reward that can readily be shared. Nevertheless, it was useful to have been able to demonstrate the possibility of collusion as an example that could be used to help justify more rigorous identification procedures.

Finally, sun exposure was not assessed. However, the survey took place during April and May, which are

autumn months in NZ. During this period, the daily peak UVI level experienced in the Otago district does not exceed a value of 2 (24), below the level when the World Health Organization (WHO) recommends protection against the possibility of erythema. Furthermore, the mean air temperature for May is 9.3°C (25). Taken together, these factors contribute to there being a low likelihood of any significant tanning from sun exposure.

## Conclusions

The findings of this study suggest that the Munsell Year 2000 Soil Color Charts may be a reliable and valid measurement strategy when assessing skin type, although there is a modest rater bias toward overestimation of the extremes of skin pigmentation. Further research is needed with more diverse populations, in particular, including greater numbers in the darker skin color categories. It would also be useful to investigate the use of the Munsell by respondents themselves, for example, in postal surveys where it could be used instead of relying on verbal skin color categories that may strengthen the "dark shift" phenomenon. The present study, in testing Munsell reliability among trained researchers, represents a logical first step toward testing it among survey respondents.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Authors' Contributions

**Conception and design:** A.I. Reeder, A.R. Gray, V.A. Hammond  
**Development of methodology:** A.I. Reeder, A.R. Gray, V.A. Hammond  
**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** V.A. Hammond  
**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** E. Iosua, A.R. Gray  
**Writing, review, and/or revision of the manuscript:** A.I. Reeder, E. Iosua, A.R. Gray  
**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** A.I. Reeder, V.A. Hammond  
**Study supervision:** A.I. Reeder, V.A. Hammond

## Acknowledgments

The authors thank the student participants who gave their time and the research assistants who helped with data collection. The authors also thank Qa-t-a Amun for conducting an exploratory Munsell-related literature search and Maria Polak for obtaining the ITA values for the Munsell color chips from spectrophotometer readings (Supplementary Tables S1–S4).

## Grant Support

A.I. Reeder and the Cancer Society Social and Behavioural Research Unit received support from the Cancer Society of New Zealand Inc. and the University of Otago. A.I. Reeder and V.A. Hammond were supported by University of Otago Research Grant and E. Iosua was supported by University of Otago Department of Preventive and Social Medicine Post-doctoral Fellowship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 17, 2014; revised June 10, 2014; accepted June 25, 2014; published OnlineFirst July 13, 2014.

## References

1. Lucas RM, McMichael AJ, Armstrong BK, Smith WT. Estimating the global disease burden due to ultraviolet radiation exposure. *Int J Epidemiol* 2008;37:654–67.
2. Dadzie OE. Ushering in a new era for studying human cutaneous diversity. *Br J Dermatol* 2013;169:iii–iv.
3. Reeder AI, Hammond VA, Gray AR. Questionnaire items to assess skin color and erythmal sensitivity: reliability, validity, and "the Dark Shift". *Cancer Epidemiol Biomarkers Prev* 2010;19:1167–73.
4. Harrison SL, Buettner PG. Do all fair-skinned Caucasians consider themselves fair? *Prev Med* 1999;29:349–54.
5. Munsell Color. Munsell Soil Color Charts. Year 2000 revised washable edition. Michigan, USA: Munsell Color 4300 44th Street SE, Grand Rapids, MI 49512, USA; 2000.
6. Reisfeld PL. Blue in the skin. *J Am Acad Dermatol* 2000;42:597–605.
7. Tanaka M, Kidoh Y, Kamei M, Terasaki T, Watanabe A, Sakamoto T, et al. A new instrument for measurement of gastrointestinal mucosal color. *Dig Endosc* 1996;8:139–46.
8. Pizzamiglio E. A color selection technique. *J Prosthet Dent* 1991; 66:592–6.
9. Sivakumaran T. Use of a Munsell color chart to describe urine color (Letter). *Clin Chem* 1975;21:639.
10. Post DF, Levine SJ, Bryant RB, Mays MD, Batchily AK, Escadafal R, et al. Correlations between field and laboratory measurements of soil color. In: Bigham JM, Ciolkosz E, editors. *Soil Color*. 31st ed. Madison, WI: Soil Science Society of America; 1993. p. 35–49.
11. Leone AP, Escadafal R. Statistical analysis of soil colour and spectroradiometric data for hyperspectral remote sensing of soil properties (example in a southern Italy Mediterranean ecosystem). *Int J Remote Sens* 2001;22:2311–28.
12. Powers JM, Capp JA, Koran A. Color of gingival tissues of blacks and whites. *J Dent Res* 1977;56:112–6.
13. Heydecke G, Schnitzer S, Turp JC. The color of human gingiva and mucosa: visual measurement and description of distribution. *Clin Oral Investig* 2005;9:257–65.
14. Gibson IM. Measurement of skin colour *in-vivo*. *J Soc Cosmet Chem* 1971;22:725–40.
15. Cancer Society of New Zealand. Position statement: The risks and benefits of sun exposure in New Zealand 2008. p. 1–18.
16. Del Bino S, Sok K, Bessac E, Bernard F. Relationship between skin response to ultraviolet exposure and skin color type. *Pigment Cell Res* 2006;19:606–14.
17. Holman CDJ, Armstrong BK. Pigmentary traits, ethnic origin, benign nevi, and family history as risk factors for cutaneous malignant melanoma. *J Natl Cancer Inst* 1984;72:257–66.
18. Pershing LK, Tirumala VP, Nelson JL, Corlett JL, Lin AG, Meyer LJ, et al. Reflectance spectrophotometer: the dermatologists' sphygmomanometer for skin phototyping? *J Invest Dermatol* 2008;128: 1633–40.
19. Kinnear PR, Sahraie A. New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5–22 and for age decades 30–70. *Br J Ophthalmol* 2002;86:1408–11.
20. Statistics New Zealand. Ethnicity. [accessed 2014 June 5]. Available from: <http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx>.
21. De Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006;59: 1033–9.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
23. StataCorp LP. Stata statistical software: release 12.0. College Station, TX: StataCorp LP; 2011. Available from: <http://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/>.
24. McKenzie R. A climatology of UVI for New Zealand. Report commissioned by the Cancer Society of NZ. Lauder, New Zealand: NIWA; 2008 Jan. Report No.: LAU2007-02RLM.
25. NIWA. Mean monthly air temperature. [accessed 2014 May 25]. Available from: <https://www.niwa.co.nz/education-and-training/schools/resources/climate/meanairtemp>.