Decision Tree Model vs Traditional Measures to Identify Patterns of Sun-Protective Behaviors and Sun Sensitivity Associated With Sunburn

Kasey L. Morris, PhD; Frank M. Perna, EdD, PhD

IMPORTANCE Understanding patterns of sun-protective behaviors and their association with sunburn can provide important insight into measurement approaches and intervention targets.

OBJECTIVE To assess whether decision-based modeling can be used to identify patterns of sun-protective behaviors associated with the likelihood of sunburn and to compare the predictive value of this method with traditional (ie, composite score) measurement approaches.

DESIGN, SETTING, AND PARTICIPANTS This cross-sectional study used a nationally representative sample of 31,162 US adults from the 2015 National Health Interview Survey, consisting of household interviews conducted in person and completed by telephone when necessary. Participants included civilian noninstitutionalized US adults. Data were collected from January 1 through December 31, 2015.

MAIN OUTCOMES AND MEASURES The associations among sun sensitivity, multiple sun-protective behaviors (ie, using sunscreen, seeking shade, wearing a hat, and wearing protective clothing), and sunburn were examined using a χ² automatic interaction detection method for decision tree analysis. Results were compared with a composite score approach.

RESULTS In our study population of 28,558 respondents with complete data (54.1% women; mean [SD] age, 49.0 [18.0] years), 20 patterns of sun protection were identified. Among 15,992 sun-sensitive individuals, those who used only sunscreen had the highest likelihood of sunburn (62.4%). The group with the lowest likelihood of sunburn did not report using sunscreen but engaged in the other 3 protective behaviors (24.3% likelihood of sunburn). Among 12,566 non–sun-sensitive individuals, those who engaged in all 4 protective behaviors had the lowest likelihood of sunburn (6.6%). The highest likelihood of sunburn was among those who only reported sunscreen use (26.2%). The decision tree model and the composite score approach correctly classified a similar number of cases; however, the decision tree model was superior in classifying cases with sunburn (44.3% correctly classified in the decision tree vs 25.9% with the composite score).

CONCLUSIONS AND RELEVANCE This innovative application of a decision tree analytic approach demonstrates the interactive and sometimes counterintuitive effects of multiple sun-protective behaviors on likelihood of sunburn. These data show where traditional measurement approaches of behavior may fall short and highlight the importance of linking behavior to a clinically relevant outcome. Given the scope of those affected and enormous associated health care costs, improving efforts in skin cancer prevention has the potential for a significant effect on public health.
Skin cancer represents an important public health concern, with nearly 100,000 Americans projected to be diagnosed with malignant melanoma in 2018. In addition, more than 5 million Americans are diagnosed with nonmelanoma skin cancers each year. Most skin cancers are caused by excessive exposure to UV radiation and could be prevented through the use of sun-protective behaviors. Recommendations for sun protection include sun avoidance, shade seeking, and use of sunscreen, as well as protective clothing, hats, and sunglasses. Despite intervention efforts, however, rates of skin cancers continue to rise.

In surveillance research, sun-protective behaviors are typically assessed in isolation or using an additive composite approach. Respondents are queried about whether they regularly engage in a variety of behaviors (eg, using sunscreen, wearing protective clothing, shade seeking, or wearing a hat), and engaging in more of these behaviors is considered to constitute better sun protection. However, this approach has several limitations. First, the additive approach does not appropriately capture the current recommendations that sunscreen use should be a supplement to other sun safety behaviors rather than a stand-alone practice. Second, the additive approach does not account for differential patterns of sun-protective behaviors; people may be more likely to engage in specific behaviors in conjunction with others, a nuance not captured with the additive approach. Finally, these assessments rarely link behavior to the outcome of sunburn or other indices of overexposure. This approach assumes that people are engaging in these behaviors for protective purposes and that doing so reduces UV exposure, but the link between behavior and a clinically relevant outcome is rarely made.

Recent estimates from the 2015 National Health Interview Survey (NHIS) showed that, although sun avoidance was associated with a decreased likelihood of sunburn, regular sunscreen users were more likely to report sunburn. This finding could suggest a failure in effective sunscreen application or a strategic motivation for sunscreen use (ie, to get a tan). However, the finding underscores the need to link these behaviors to relevant outcomes to assess their efficacy in reducing UV exposure and preventing skin cancers.

The importance of measuring behaviors and outcomes has implications for intervention research. A recent review of National Institutes of Health–funded grants related to skin cancer prevention found that less than half attempted to link sun-protective behavior to meaningful change in clinically relevant benchmarks. The ability to link behavior (and behavior in context) to sunburn is critical because no minimal effective dose of sun protection has been determined. That is, unlike other domains where a target level of behavior has been established (eg, 150 minutes of physical activity per week or 6% weight loss for obesity intervention), the minimal level of sun protection needed to reduce risk is unknown, and at a population level, sun-protective behaviors have not been reliably associated with sunburn.

The purpose of the present study was to address the limitations of measurements of sun protection using a decision tree approach and to compare it with an additive composite measure of sun protection. Decision trees are a data mining and classification tool used to develop a predictive algorithm that identifies the specific factors (ie, sun sensitivity, sun-protective practices) that differentiate the sample population on the outcome of a target variable (ie, sunburn).

The value of this approach is that it calculates the importance of a given behavior in estimating sunburn and tests for combinations of behaviors (ie, interactions between predictors) in their relevance to sunburn. This method overcomes the constraints of linear models, in which predictors are considered additive and independent and only predefined interactions are considered. Decision trees also offer a visual representation of prediction rules that can be more easily interpreted in clinical settings.

**Method**

**Participants**

Data were obtained from the cancer control supplement of the 2015 NHIS. The NHIS is an annual, cross-sectional survey of US noninstitutionalized adults conducted through in-person interviews. The NHIS was selected as the data source for this study because the survey items measuring sun protection constitute the behaviors assessed for national-level surveillance toward Healthy People 2020 objectives. The total sample consisted of 33,162 participants, and data were collected from January 1 through December 31, 2015. More information about the sample design and data collection procedures is found at the Centers for Disease Control and Prevention website. The institutional review board of the National Cancer Institute waived the need for approval of this study, and patient permission was not required because all data were deidentified.

**Measures**

**Sun Sensitivity**

Participants were asked to indicate what would happen to their skin if they were out in the sun for 1 hour after several months of not being exposed. Responses were coded as sun sensitive (ie, “get a severe burn with blisters,” “have a moderate sunburn with peeling,” and “burn mildly with some or no tanning”) or non–sun sensitive (ie, “turn darker without sunburn”
and “nothing would happen”). Responses of “other” and “do not go out in the sun” were excluded from the analysis.

Sunburn Participants indicated how many times they had experienced a sunburn in the past 12 months. Most participants (66.3%) reported having no incidents of sunburn in the past year. Among those who reported sunburn (33.7%), 16.3% reported 1 sunburn; 9.9%, 2 sunburns; and 5.8%, 3 to 5 sunburns. Less than 1% of the sample reported 6 to 360 sunburns. This variable was recoded to create a binary outcome variable for sunburn (none vs ≥1).

Sun Protection Six items assessed sun-protective behaviors.10 Respondents were asked to indicate “on a warm, sunny day, how often” they wore sunscreen, sought shade, wore a cap or visor, wore a wide-brimmed hat, wore long sleeves, and wore long pants. Answer choices included “always,” “most of the time,” “sometimes,” “rarely,” and “never.” Participants could also indicate that they do not go out in the sun. Binary variables were created to classify those who regularly engage in the behavior (ie, “always” or “most of the time”) compared with those who do not (ie, “sometimes,” “rarely,” and “never”). Participants who selected “don’t go out in the sun” were recoded as engaging in the behavior.

Statistical Analysis Data analysis was conducted in SPSS, version 25.0 (IBM Corp). Using listwise deletion, 28 558 cases with complete data were identified. Sun sensitivity and the 6 sun-protection variables were used as predictors, and incidence of sunburn was used as the outcome. The Chi Square Automatic Interaction Detection (CHAID) approach11 was used to grow the tree. CHAID is designed to work with categorical or discretized variables; the analysis determines which variables by relative importance (ie, highest χ² value). The aim of pruning the tree is to create the simplest structure while maintaining the predictive value. Pruning can be performed using a top-down or bottom-up approach. A top-down approach involves modifying the stopping criteria to be more stringent (ie, more cases required for parent and child nodes). A bottom-up approach involves growing the tree to its maximum depth and sequentially removing child nodes and reexamining the model. Although by default the CHAID method is designed to minimize the size of the tree (ie, a top-down approach), the large sample size used in this analysis makes it prone to overfitting. Therefore, a bottom-up approach was also used. Predictors that did not result in a statistically significant node split, those in which the child node represented a very small proportion of the sample, and those that resulted in multiple splits (eg, the same predictor as a parent and child node within the same branch) were removed successively to allow for model comparison. The tree that resulted in the simplest structure with the lowest prediction and generalization error was chosen as the final model.

Results

Participant Characteristics Among the 28 558 respondents with complete data, 13 104 (45.9%) were men and 15 454 (54.1%) were women. Participants ranged in age from 18 to 85 years, with a mean (SD) age of 49.0 (18.0). Most of the sample (n = 22 285 [78.0%]) was white; 3614 participants (12.7%) were black; and 2575 (9.0%) indicated their race as Asian, Native American Indian or Alaskan Native, or more than 1 race. The sample included 4704 participants (16.5%) who indicated their race as Hispanic. Most of the sample (15 992 participants [56.0%]) had sun-sensitive skin; 12 566 participants (44.0%) did not. Most people (22 110 participants [77.4%]) engaged in at least 1 sun-protective behavior; shade seeking was the most common behavior (11 369 instances [39.8%]), and wearing long sleeves was the least common behavior (n = 4502 instances [15.8%]).

Decision Tree The final model consisted of sun sensitivity and 4 protective behaviors (using sunscreen, seeking shade, wearing long sleeves, and wearing a hat) as predictors. Wearing caps and long pants were dropped as predictors during pruning. The tree had a depth of 5 branches, with 39 parent nodes and 20 child nodes. In the training sample (n = 20 185), the risk estimate (proportion of cases incorrectly classified after adjustment for prior probabilities and misclassification costs) was 0.305 (SE, 0.003). The model correctly classified 14 023 cases (69.5%) overall (of the 13 362 cases without sunburn, 11 052 [82.7%] were correctly classified; of the 6823 cases with sunburn, 2971 [43.5%] were correctly classified). In the validation sample (n = 8373), the risk estimate was 0.296 (SE, 0.005). The model correctly classified 5898 cases (70.4%) overall (of the 5565 cases without sunburn, 4655 [83.6%] were correctly classified; of the 2808 cases with sunburn, 1243 [44.3%] were correctly classified).
The Figure presents the partial decision tree from the validation sample (some nodes are not displayed for ease of interpretation; eFigures 1 and 2 in the Supplement show the full model, which also includes percentage values at each branch). The first split was made based on sun sensitivity.

Among sun-sensitive individuals, those who engaged in all 4 protective behaviors (seeking shade, wearing long sleeves, wearing a hat, and using sunscreen) had a 26.2% probability of sunburn. Participants who only reported shade seeking had a 37.3% probability of sunburn; similarly, those who reported wearing long sleeves and wearing a hat had a 37.4% probability of sunburn. Although participants who did not use sunscreen, seek shade, or wear protective clothing had a higher probability of sunburn (54.8%), the group with highest likelihood of sunburn consisted of those who used only sunscreen (62.4%). The group with the lowest probability of sunburn did not report using sunscreen but reported engaging in the other 3 protective behaviors (24.3%).

Among non-sun-sensitive individuals, decisional paths were less complex; however, similar patterns emerged. Those who engaged in all 4 protective behaviors had the lowest probability of sunburn (6.6%). Those who did not engage in any protective behavior had a 17.3% probability of sunburn; however, the highest probability of sunburn was among those who reported only sunscreen use (26.2%). Those who report seeking shade but not using sunscreen had an 8.7% probability of sunburn.

Comparison With Traditional Composite Assessments of Sun Protection

Next, we compared the predictive accuracy of the decision tree model with a logistic regression model with a composite
Discussion

The goal of this study was to use an innovative application of a decision tree analytic technique to gain insight into the association between sun-protective behaviors and the incidence of sunburn. In contrast to more typical measurement approaches in which behaviors are thought to function in an additive, linear fashion, the decision tree method provided a test of the nonlinear and interactive patterns between behaviors. Although the overall classification rates were similar between the 2 approaches, the decision tree analysis was superior in its ability to predict the likelihood of having had a sunburn—the outcome of most interest in its relevance to skin cancer prevention. In addition, the convergence between the training and validation samples underscores the strength of this approach. The logistic model only explained a small portion of the variance (18.1%) in sunburn, and the statistical significance of this model may be, in part, an artifact of the large sample size. Finally, the decision tree may offer a more useful clinical interpretation by providing insight into careful, patterned choices of behaviors.

Examination of the specific structure of the tree demonstrates how typical assessments of sun protection may fail to capture important nuances. Although use of multiple forms of sun protection reduced the likelihood of sunburn (for individuals with and without sun-sensitive skin), the use of sunscreen functioned in a counterintuitive manner. Among those with sun-sensitive skin, those who regularly used sunscreen but no other protective behavior had the highest proportion of sunburn. In contrast, participants who regularly engaged in all the protective behaviors except use of sunscreen had the lowest proportion of sunburn. Similarly, sunscreen use was associated with the highest probability of sunburn among participants with non–sun-sensitive skin. Certainly, sunscreen—if used properly—can offer protection from the sun and may be particularly useful for certain groups. However, these findings are consistent with the recommendation that sunscreen not be used in isolation and suggest that, when other behaviors are used regularly, sunscreen might not provide a significant increase in protection. As Lazovich and colleagues noted, inconsistent messaging about skin cancer prevention may stem from uncertainty of the effectiveness of different protection strategies, and the idea that sunscreen should be used in combination with other behaviors is largely missing from this messaging. These findings underscore the importance of being attuned to the differential effects of behavior on relevant outcomes when designing intervention strategies and targets.

Although the NHIS assesses 6 different sun-protective behaviors and all 6 were initially included in the decision tree, wearing a baseball cap or visor and wearing long pants were removed from the tree during pruning. Although these behaviors were engaged in more often (wearing a cap, 39.4%; wearing pants, 37.6%) than wearing wide-brimmed hats (21.5%) or long sleeves (20.3%), they did not provide predictive utility when sunburn was considered as the relevant outcome. This outcome could suggest that the behaviors simply are not associated with UV exposure and subsequent sunburn (indeed, recommendations indicate that wide-brimmed hats that shade the entire head—and not baseball caps or visors—be used to provide protection from the sun). It is also possible that the high percentage of respondents who report engaging in these behaviors is a measurement artifact. That is, people may be indicating that they wear long pants on a regular basis rather than wearing long pants specifically as a means of sun protection. The relevance of the behavior to the outcome of interest and the manner in which that behavior is measured are important considerations going forward.

Limitations

These data and interpretation of results have several limitations. Specifically, this study uses a cross-sectional sample of self-reported outcomes; thus, the potential for recall and/or self-report bias in responses exists, and the temporal association between protective behavior and sunburn is unknown. However, these conditions equally exist for decision tree and composite scoring approaches, and the decision tree approach was superior in predicting sunburn. In addition, this analysis focused on sun sensitivity and behavior as predictors of sunburn, but other important factors may exist. These factors may include level of outdoor exposure, geographical location, family history of cancer, or motivational differences. A primary goal for future work is to address these limitations by obtaining longitudinal measurements and objective assessments of sunburn and UV exposure.

Conclusions

This study provides an innovative approach (decision tree modeling) to the measurement of sun protection and demonstrates...
the need to understand the interactive effects of multiple behaviors and the importance of linking behavior with a clinically relevant outcome. Given the scope of those affected and enormous associated health care costs, improving efforts in skin cancer prevention has the potential to make a tremendous difference.

ARTICLE INFORMATION
Accepted for Publication: April 19, 2018.
Published Online: June 27, 2018.

Author Contributions: Drs Morris and Perna had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Both authors.
Acquisition, analysis, or interpretation of data: Both authors.
Drafting of the manuscript: Both authors.
Critical revision of the manuscript for important intellectual content: Both authors.
Statistical analysis: Morris.
Administrative, technical, or material support: Perna.
Study supervision: Perna.

Conflict of Interest Disclosures: None reported.

Additional Contributions: We are indebted to Erin Morris, MS, Principal Data Scientist at Jabil for her guidance in data analysis and interpretation. She was not compensated for this work.

REFERENCES