Etiology of Familial Aggregation in Melanoma and Squamous Cell Carcinoma of the Skin

Linda S. Lindström, Benjamin Yip, Paul Lichtenstein, Yudi Pawitan, and Kamila Czene
Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

Abstract

Background: Melanoma and squamous cell carcinoma of the skin (SCC) have been previously shown to coaggregate in families. To shed light on the etiology, we estimated the relative contributions of genetic and environmental factors on the occurrence of each disease, in addition to their influence on coaggregation of the two diseases. Because the malignancies are dependent on UV radiation, we did separate analyses for sun-covered and sun-exposed sites.

Methods: Our Swedish population-based data included 11 million individuals in 3 million families. We used an extended generalized linear mixed model to estimate the genetic and environmental contribution.

Results: In melanoma, the genetic contribution was 18% [95% confidence interval (95% CI), 13-22%] in the all-sites analysis, whereas the family-shared contribution was slightly higher in the sun-covered compared with sun-exposed sites.

Conclusions: Genetic factors are important in familial aggregation of melanoma and the higher sun-covered compared with sun-exposed site estimate of family-shared environment may convey benefit from cautious sunbathing.

Introduction

Overexposure to the sun and severe burning or blistering in childhood is associated with an increased risk of developing melanoma in later life (1, 2). Other risk factors for melanoma include the frequent use of sun beds and sunshine holidays abroad. A family history of melanoma is one of the most significant risk factors of disease, with a familial risk of 2.4 between first-degree relatives (3). Inherited traits such as density of nevi, skin type and pigmentation, and color of hair and eyes may accommodate a part of the familial risk by modulating the response to solar UV radiation or by other mechanisms (4-9).

For squamous cell carcinoma of the skin (SCC), the genetically defined individual skin type is considered as one of the most important factors in the liability to disease. Individuals with sensitive skin that burn easily are at greatest risk, often including people with fair skin, fair or red hair, and blue eyes (10). Next to the genetic factors (including an inherited trait of skin type), the accumulated exposure of UV radiation during one’s lifetime is essential for this type of skin cancer, and the majority of SCCs appear on sun-exposed areas such as the face, lower lip, neck, ears, and hands (10). Additionally, the presence of chronically injured skin is believed to increase the risk of developing SCC; thus, scars from burning or constantly inflamed skin areas are risk-afflicted conditions (11, 12).

Using the methods of quantitative genetics allows us to disentangle whether familial aggregation of cancers is a result of genetic or environmental similarity among family members because family members naturally share many lifestyle factors that may increase or decrease susceptibility to cancer. The genetic and environmental contribution to SCC has not yet been studied, and little is still known on the exposure pattern of UV radiation and the genetic and environmental liability to melanoma. Melanoma and SCC have previously been shown to be aggregated within families (3, 13). To our knowledge, however, no scientific study has investigated the etiology of the familial coaggregation of SCC and melanoma.

Here, using unique population-based data including 3 million families, we present the genetic and environmental contributions to melanoma and SCC in all sites, sun-exposed sites and sun-covered sites. Moreover, we illuminate the common etiology of melanoma and SCC apportioning the familial coaggregation of the two diseases into genetic and shared environmental factors.

Materials and Methods

Study Population

Linkage of different records of personal information is possible in Sweden due to the fact that each resident has his/her own unique national registration number. Our study is based on a record linkage between several Swedish population-based registers: the Multi-Generation Register, which records familial relationships, the Swedish Cancer Register, the Death Register, and the Migration Register. The Multi-Generation Register includes individuals born in Sweden between 1932 and 2001 and their biological parents. Incident cancers in Sweden since 1958 are recorded in the Swedish Cancer Register [using a four-digit diagnostic code according to the seventh revision of the International Classification of Diseases (ICD-7) for cancer-type identification] including information on histopathologic type. In the 1970s, completeness of cancer registration was assessed to be around 95%. In recent years, the percentage of cytologically or histologically verified cases has been 100% (14).
The structure of the Multi-Generation Register (and thus, in our database) is as follows; every child exists only once, whereas parents are present as many times as they have children; an individual can be in the database both as offspring and parent, and parents are those who admit to parenthood at birth (not only married individuals). Our database comprised 11.3 million individuals, including more than 1 million cancer cases. Restricting our data to families with one partnership, i.e., in case of several partners, we randomly selected one of them, and excluding all families with twins, we retrieved our final database including 10 million individuals in 2.7 million two-generational families. The families consisted of a mother and a father together with their oldest two children (or their only child). Only invasive melanoma cases were considered, whereas both invasive and in situ cancers were included for SCC (ICD-7 190 for melanoma and ICD-7 191 for SCC with histologic code 143 and 144 for in situ and 146 for invasive). Sun-covered sites (ICD-7, 190.5 and 191.5 for melanoma and SCC, respectively) were strictly defined as the trunk reflecting our purpose of capturing the intermittent exposure, and all other sites were regarded as sun exposed for both melanoma and SCC.

**Statistical Methods.** We have previously presented a single-trait generalized linear mixed model for two- and three-generational families estimating the genetic, adult-shared environmental, and childhood-shared environmental contribution to the susceptibility of melanoma (15). This model was extended to a multivariate binary-trait model, enabling us to analyze the coaggregation of two diseases in one family or in one individual and to apportion these effects into genetic and environmental contributions. The coaggregation model can also be restricted to a single-trait model and was used in our independent analysis of melanoma and SCC. Firstly, we will begin to describe the single-trait model and proceed with the more complex coaggregation model.

**Single-Trait Model.** Applying the mixed effects model, we assume that the mean outcome can be explained by a set of fixed effects and a set of random effects. The outcome is assumed to be Bernoulli distributed with parameter \( p_{ij} \), which denotes the probability of member \( j \) in family \( i \) to get diagnosed with cancer. For modeling, the probability is transformed into a standard normal scale with the probit link.

The covariance matrix between the two diseases is defined as the sum of genetic (\( \alpha_m \)), family-shared environmental (\( f \)), and childhood-shared environmental (\( c \)) effects, so that

\[
\begin{align*}
\Phi^{-1}(p_{ij}) &= x_{ij}^T \beta + z_{ij}^T b_i \\
\beta &= \begin{pmatrix} \alpha_m \\ f \end{pmatrix} \\
b_i &= \begin{pmatrix} c \end{pmatrix}
\end{align*}
\]

where \( \Phi(\cdot) \) is the normal distribution function.

In our analyses, the random component is modeled as the sum of genetic (\( \alpha_m \)), family-shared environmental (\( f \)), and childhood-shared environmental (\( c \)) effects, so that

\[
\Phi^{-1}(p_i) = x_i^T \beta + g_i + f_i + c_i
\]

where, for convenience, we use \( p_i \) as a vector to represent family \( i \).

Different family structures are assumed to be independent, but the random effects within the families are correlated according to the usual assumptions in quantitative genetic analysis, where the correlations between relatives for environmental and genetic effects are set to fixed values according to their degree of genetic and environmental relationship. Thus, for the genetic effect, all first-degree relatives (parent offspring and full siblings) are correlated by 0.5. All family members are fully correlated for the family-shared environmental factor, whereas only the siblings are assumed to share a common childhood-shared environmental component. In summary, the model expectations are seen below.

<table>
<thead>
<tr>
<th>Genetic, family, and childhood model</th>
<th>Variance (random effects)</th>
<th>Covariance (spouse-spouse)</th>
<th>Covariance (sibling-sibling)</th>
<th>Covariance (parent-child)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_g^2 ) + ( \sigma_i^2 ) + ( \sigma_c^2 )</td>
<td>( \sigma_i^2 )</td>
<td>( \frac{1}{2} \sigma_g^2 + \sigma_i^2 + \sigma_c^2 )</td>
<td>( \frac{1}{2} \sigma_g^2 + \sigma_i^2 )</td>
<td></td>
</tr>
</tbody>
</table>

Note that nonshared environmental effects seen in within-family differences are implied by the randomness of the probit model and the variance of these nonshared environmental effects is set to 1 for identifiability.

In quantitative genetics, the estimated variance of liability attributable to various factors is usually presented. The heritability, in particular, is the proportion of susceptibility due to genetic factors given by

\[
h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_i^2 + \sigma_c^2 + 1} \tag{16}
\]

where 1 in the denominator is derived from the non-shared environment as explained above, and \( s \) stands for single trait.

To reduce the amount of computation and without any impact on the results, we first summarized the data according to all possible configurations in a family with regard to the event. The probability of each family pattern is computed using a fast Monte Carlo integration method as described by Pawitan et al. (16). The total likelihood is retrieved by summing the individual log-likelihood contributions from all families, assuming that the family patterns are independent. Optimization of the total log-likelihood is done by employing optim(), a derivative-free simplex algorithm available in the statistical package R. Although the exact location of the maximum of the likelihood cannot be found because of the Monte Carlo approximation used in the computation of the likelihood, we have confirmed that the estimate is within the statistical uncertainty in the data by smoothing the log-likelihood around each estimated parameter value.

**Familial Coaggregation Model.** For the coaggregation model, the same principles apply as for the single-trait model. However, studying the coaggregation of two diseases in families results in the binary outcome of disease \( y_{ik} \), where \( j \) is member in family \( i \) with disease \( k = 1,2 \). For a family including mother, father, and two children, the vector is of length 8, consisting of four family members with binary outcomes for each disease.

The covariance matrix between the two diseases is defined as a matrix combining the effects of melanoma (\( m \)) and SCC (\( s \)) due to the independent assumption of the random effects. In the coaggregation analysis, each member can either be diagnosed with melanoma and/or SCC, resulting in four submatrices. The covariance matrix contains two variance components (\( \sigma_m^2 \) and \( \sigma_s^2 \)) and one covariance component (\( \sigma_{ms} \)).

We, for clarification of the model, present the covariance component

\[
\sigma_{ms} = \begin{pmatrix}
M_{ms} & M_{ms} & M_{ms} & M_{ms} \\
F_{ms} & F_{ms} & F_{ms} & F_{ms} \\
C_{ms} & C_{ms} & C_{ms} & C_{ms} \\
C_{ms} & C_{ms} & C_{ms} & C_{ms}
\end{pmatrix}
\]

where \( M \) stands for mother, \( F \) stands for father, and \( C \) and \( C \) stand for sibling 1 and sibling 2, respectively.
We model the genetic, family-shared environmental, childhood-shared environmental and unshared environmental factors in the familial coaggregation model inhabiting two malignancies. The unshared effect ($u_i$) is defined by an identity matrix because the unshared effect only describes exposures that affect the coaggregation of two diseases within an individual and, thus, not shared with the rest of the family. The proportion of variance for these effects can be defined employing the quantitative genetic approach in a similar manner as for the single-trait model. To elucidate the model, we define the genetic and childhood-shared environmental contribution to the coaggregation of two diseases given by

$$h_{Co}^2 = \frac{\sigma_{msg}}{\sigma_{msg} + \sigma_{msf} + \sigma_{msc} + \sigma_u^2},$$

$$c_{Co}^2 = \frac{\sigma_{msc}}{\sigma_{msg} + \sigma_{msf} + \sigma_{msc} + \sigma_u^2}$$

(E)

where Co stands for coaggregation.

Results

The numbers of invasive and in situ cases of melanoma and SCC, respectively, are seen in Table 1. In contrast to SCC, the majority of melanoma cases are invasive. In view of that, we decided to restrict our melanoma outcome to only invasive cases, while we included invasive and in situ cases for SCC in subsequent analyses.

Nuclear families consisting of a mother and a father, together with the only or the two oldest children, were used in the analysis resulting in 2.7 million parental pairs with 4.7 million children. The number of individuals and concordances for melanoma and/or SCC are presented in Table 2. We identified 291 families in which both a parent and child were diagnosed with melanoma in the model including all body sites (Table 2). In the sun-exposed and sun-covered sites analyses of melanoma, the number of parent-child concordant families was 94 and 60, respectively (Table 2). Unfortunately, there were no sibling concordances for SCC on sun-covered sites, making it impossible to study the genetic and environmental contribution for SCC on sun-covered sites. Thus, subsequent analyses on SCC are presented only for all sites and sun-exposed sites.

The single-trait results on all, sun-covered and sun-exposed sites are shown in Table 3. In melanoma, the genetic contribution was significant in all three analyses, ranging from 12% [95% confidence interval (95% CI), 6-18%] to 13% (95% CI, 5-21%) in the sun-covered and sun-exposed analyses to 18% (95% CI, 13-22%) in the all-sites analysis. Interestingly, for melanoma, it could be seen that the family-shared contribution of 9% (95% CI, 6-11%) in the analysis of sun-covered sites was higher than the sun-exposed sites estimate of 6% (95% CI, 3-8%). In addition, we found a significant childhood-shared environmental factor of 8% to 9%. We also analyzed the genetic, family-shared environmental, and childhood-shared environmental contribution, including combined invasive and in situ cases of melanoma. The estimates were almost identical to the estimates only including invasive melanoma.

SCC revealed quite similar estimates, contrasting the sun-exposed and all-sites analyses for both the genetic and environmental factors. A family-shared environmental effect of 17% to 18% (95% CI, 16-19%) was found in the all-sites and sun-exposed sites analyses of SCC, together with a significant genetic and childhood-shared environmental contribution of 8% and 7%, respectively.

The common etiology for melanoma and SCC was disentangled using a coaggregation model apportioning the variance to genetic and environmental contributions (Table 4). A considerable genetic factor of 47% (95% CI, 43-51%) was found, along with a high family-shared environmental component of 36% (95% CI, 33-39%), a childhood-shared environmental component at 8% (95% CI, 4-13%), and an unshared environmental component at 8% (95% CI, 3-14%; see Table 4).

We initially fitted three models: the family model (only including familial effect); the genetic and family-shared environmental model; and the genetic, family-shared environmental and childhood-shared environmental model. However, the model including genetic, family-shared environment and childhood-shared environment had the best fit in addition to being a preferable model in a biological perspective. Finally, the log-likelihood curves for the single-trait and coaggregation analyses of melanoma and SCC are very well behaved.

Discussion

In melanoma and SCC, the corresponding genetic and environmental susceptibility due to the exposure pattern of UV radiation, in addition to the etiology of the familial coaggregation between the two diseases, is mainly unexplored. Using Swedish population-based data, we analyzed melanoma and SCC, focusing on sunlight exposure to sun-covered and sun-exposed sites with the belief that exposure to sun-covered sites is generally of a more intermittent nature (10). In addition, we shed light on the common etiology of melanoma and SCC by disentangling the familial coaggregation of the two diseases into environmental and genetic contributions.

Reflecting on the incidence of melanoma and SCC, the pattern of sunlight to sun-covered sites differs from that of sun-exposed sites in that the pattern of exposure to sun-covered sites is of a more intermittent nature (10). However,
the interpretation of the amount of exposure is complicated by the need to understand the pattern of sunbathing. One interpretation is that at low levels of sun exposure, a majority of the exposure is of intermittent type, and as exposure increases, the exposure becomes more continuous. Continuous exposure has been found to increase the risk of SCC, whereas unoccupational exposure, believed to reflect an intermittent pattern of exposure to sunlight and sunburn, is highly associated with melanoma (10).

Our results, contrasting sun-covered (12%) to sun-exposed sites (13%), indicate that the pattern of sun exposure does not interact with the genetic susceptibility to melanoma. The difference in genetic effects for sun-exposed, sun-covered, and all-sites might be due to random variation; on the other hand, the etiology of sun-covered and sun-exposed sites might also be different from the all-sites because these subanalyses exclude discordant sites. Furthermore, today, only a small proportion of the estimated genetic factor (18%) can in fact be explained by identified mutations. Susceptibility to melanoma, particularly in families with many affected individuals, is linked to mutations in the cell cycle regulator CDKN2A (p16) gene and less frequently to the p16 gene and less frequently to the CDK4 gene (4, 6, 17, 18). However, mutations are rare in kindreds with a fewer number of affected individuals, and only 7.8% of the Swedish melanoma and dysplastic nevus syndrome families show CDKN2A mutations (19). Genetic susceptibility to melanoma may also include low-penetrance genes, such as the melanocortin-1 receptor as well as defects in various DNA repair mechanisms (6). To our knowledge, there are no confirmed associations for single nucleotide polymorphism variants related to melanoma; however, a few potential candidates have been published, such as the vitamin D receptor BsmI (20).

Interestingly, in melanoma, the higher family-shared environmental factor in the analysis of sun-covered compared with sun-exposed sites (9% and 6%, respectively) may convey benefit from cautious sunbathing, especially on sensitive winter skin unaccustomed to UV radiation. In support for our findings, a previous study on melanoma showed higher familial risks at sun-covered compared with sun-exposed sites (21); however, we can now ascertain that the difference is not due to inherited but shared environmental effects. Finally, in agreement with previous studies, the childhood-shared environmental effect was fairly high and stable (8-9%) in all analyses, revealing the importance of sensible sun habits to avoid risk-inflicted conditions such as sunburns during infancy and youth (1, 15, 22, 23).

For SCC, to our knowledge, the inherited and environmental contribution to the disease has not yet been studied. Intriguingly, the familial shared proportion of the liability to SCC was found to be the highest yet in cancer research (22), accounting for 17% and 18% of the total susceptibility in the sun-exposed and all-sites analyses, respectively. Familial habits of accumulated sun exposure might explain most of the familial shared effect seen. Furthermore, the genetic contribution was 8% for both all-sites and sun-exposed sites analyses. Consequently, genetic variability in individuals enhancing sensitivity to accumulated sun exposure and accounting for an additional susceptibility beyond family-shared environment seems to be involved in the etiology of SCC. The common sensitivity to UV radiation might be explained by inherited traits such as skin type and pigmentation or by other as-yet unknown mechanisms. In agreement with our results, a reasonably high familial risk of 2.7 in SCC has been shown previously (13). Moreover, childhood-shared environment (7%) is important in individual liability to SCC, and supporting our findings, sunburn at a young age has been found to increase the risk of SCC (12). Unfortunately, we did not have enough cases in sibling pairs to study SCC on sun-covered sites, so we instead compared sun-exposed sites with all-sites analysis in this study, resulting in almost identical estimates of both genetic and environmental factors.

Virtually nothing is known about the familial coaggregation of melanoma and SCC. In our study, we disentangled the familial coaggregation and apportioned it to specific genetic and environmental factors. Interestingly, inherited components (47%) seem to have a considerable importance in the common etiology of melanoma and SCC, indicating that in families with melanoma and SCC, a high percentage of coaggregation is due to heritable causes. Previously, a shared sensitivity to a shared risk factor such as UV radiation has been thought to be the reason for the coaggregation of melanoma and SCC (10). However, the explanation has been suggested to be simplified. In a recent study, high relative risks were found when both invasive SCC and melanoma were diagnosed in the same family (3, 13). A few syndromes such as Werner’s syndrome and xeroderma pigmentosum are included in both melanoma and SCC, but still, the main part of the common

Table 3. Effects of genetic and environmental factors in melanoma and SCC

<table>
<thead>
<tr>
<th>Single-trait model</th>
<th>Genetic</th>
<th>Family-shared environment</th>
<th>Childhood-shared environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma, all sites</td>
<td>0.25 (0.20-0.32)</td>
<td>0.09 (0.06-0.11)</td>
<td>0.11 (0.05-0.18)</td>
</tr>
<tr>
<td>Exposed sites</td>
<td>0.16 (0.08-0.25)</td>
<td>0.08 (0.05-0.11)</td>
<td>0.11 (0.02-0.21)</td>
</tr>
<tr>
<td>Covered sites</td>
<td>0.12 (0.07-0.16)</td>
<td>0.26 (0.24-0.28)</td>
<td>0.10 (0.04-0.17)</td>
</tr>
<tr>
<td>SCC, all sites</td>
<td>0.11 (0.05-0.17)</td>
<td>0.12 (0.06-0.19)</td>
<td>0.11 (0.03-0.17)</td>
</tr>
</tbody>
</table>

Table 4. Contribution of genetic and environmental factors in the coaggregation of melanoma and SCC

<table>
<thead>
<tr>
<th>Coaggregation model</th>
<th>Genetic</th>
<th>Family-shared environment</th>
<th>Childhood-shared environment</th>
<th>Unshared environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma and SCC</td>
<td>0.16 (0.15-0.18)</td>
<td>0.13 (0.12-0.14)</td>
<td>0.03 (0.01-0.04)</td>
<td>0.03 (0.01-0.05)</td>
</tr>
<tr>
<td>Percentage of variance (95% CI)</td>
<td>36 (33-39)</td>
<td>8 (4-13)</td>
<td>8 (3-14)</td>
<td></td>
</tr>
</tbody>
</table>
genetic susceptibility between melanoma and SCC revealed here remains unexplained and is probably due to genetic variability in individuals.

The high family-shared environmental factor (36%) is of much interest in the common etiology of melanoma and SCC. Furthermore, some attention should also be given to the unshared environment that contributes 8% of the liability in our coaggregation analysis. The meaning of the unshared environment is different in that it describes an exposure that affects both diseases within the same person and is not shared within the family, essentially inhibiting any exposure believed to affect both melanoma and SCC. Contrasting the unshared and the family-shared environmental component reveals that the familial environment is more important than the individual environment. Thus, familial environments such as sunshine holidays abroad and familial sun-tanning habits seem to have more impact on the susceptibility than individual activities such as the use of sun-tanning beds. Finally, the childhood-shared environmental effect accounted for 8% of the coaggregation of SCC and melanoma, enlightening the importance of sun habits to avoid sunburns during infancy and youth.

In this study, we use the unique Swedish population-based data, including 2.7 million families in our analyses. Nevertheless, in SCC, it was still necessary to combine in situ cases with invasive cases due to the low numbers of cases available. Sun-covered and sun-exposed site definitions vary in literature (21, 24, 25), and in this study, we focused on assessing the intermittent exposure, leading to a strict definition of the sun-covered sites. A potential difficulty was that the upper back and chest belong to covered sites, whereas the upper leg was defined as exposed site. This limitation is, however, not easily corrected given the ICD coding system for melanoma and SCC. In the Swedish climate, most solar exposure occurs during summer or holidays abroad in sunny resorts. Consequently, the exposure to covered sites will be of intentional nature, including sunbathing and sun-tanning activities.

An advantage of our generalized linear mixed model with a likelihood-based methodology for parameter inference is that we treat full family data in our large database instead of analyzing pairwise data. Therefore, we consider the model to be superior to the standard Mx methodology previously used for the analysis of familial data. Still, these two models resulted in very similar estimates with respect to melanoma (22). In our current analyses, we disregarded the different follow-up in the two generations; however, previously, the correction for differential length of follow-up resulted in estimates of similar magnitude (15). At present, our model cannot accommodate multiple covariates such as period-corrected prevalence levels or age of onset explicitly.

In conclusion, genetic factors have a considerable effect on melanoma susceptibility, whereas in SCC, the familial environment is of vital importance. Interestingly, the high genetic contribution in the coaggregation of the two diseases elucidates the possibility of genetic variability, creating the need for future genetic studies unraveling the nature of the coaggregation. Finally, the family-shared environment seems more important than the unshared environment, enlightening the impact of familial sun-tanning habits in the coaggregation of the two diseases.

References

Etiology of Familial Aggregation in Melanoma and Squamous Cell Carcinoma of the Skin

Linda S. Lindström, Benjamin Yip, Paul Lichtenstein, et al.


Updated version
Access the most recent version of this article at:
http://ceb.p.aacrjournals.org/content/16/8/1639

Cited articles
This article cites 24 articles, 7 of which you can access for free at:
http://ceb.p.aacrjournals.org/content/16/8/1639.full.html#ref-list-1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.