8-6-2009

Temporal Evolution in the Histopathologic Diagnosis of Borderline Melanocytic Lesions

Jason Elihu Frangos
Yale University

TEMPORAL EVOLUTION IN THE HISTOPATHOLOGIC DIAGNOSIS OF BORDERLINE MELANOCYTIC LESIONS

A Thesis Submitted to the

Yale University School of Medicine

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Medicine

by

Jason Elihu Frangos

2009

TEMPORAL EVOLUTION IN THE HISTOPATHOLOGIC DIAGNOSIS OF BORDERLINE MELANOCYTIC LESIONS. Jason E. Frangos, Lyn Duncan and Alexa B. Kimball. Departments of Dermatology and Dermatopathology, Massachusetts General Hospital, Harvard Medical School, Boston MA. (Sponsored by Dr. Robert Tigelaar, Department of Dermatology, Yale University School of Medicine).

While the incidence of cutaneous malignant melanoma has risen steeply over the past half century, increases in the mortality rate have been relatively modest. In an effort to understand this discrepancy, we sought to determine whether a shift toward more malignant diagnoses may have been made by dermatopathologists (DPs) diagnosing severely dysplastic nevi over a time period of 20 years. Forty biopsy slides of dysplastic nevi (28) and thin melanomas (12) from the period 1988-1990 were obtained from the pathology files of the Massachusetts General Hospital (MGH). All DPs that had rendered an original diagnosis for any of the 40 slides as well as the current staff in the MGH Dermatopathology department were invited to re-evaluate the slide-set. Three original DPs and 3 current MGH staff DPs re-read the slide-set. The mean number of melanoma diagnoses by the 6 study participants was 19.7 (median=19.5), an increase of 64% from the original number of melanoma diagnoses in the slide set (12). For lesions originally diagnosed as "Melanoma", study participants had a high level of agreement between each other (kappa=0.74) and between each rater and the original diagnosing DP (kappa=0.86). For lesions originally diagnosed as "Not Melanoma" study participants had a low level of agreement between each other (kappa=0.22) and a low level of agreement between each rater and the original diagnosing DP (mean kappa=0.39). The results of this study indicate that a small set of DPs at a major academic institution tended to read prior non-malignant diagnoses of borderline melanocytic lesions as malignant but not to revise prior diagnoses of malignant melanoma as benign.

ACKNOWLEDGMENTS

Sincerest gratitude to my primary research mentor and collaborator over the past year, Dr. Alexa Kimball. Thank you also to Dr. Lyn Duncan for her help in the design and execution of the project, Dr. Martin Weinstock for his guidance and advice on the epidemiology of melanoma, Dr. Robert Tigelaar for his supervision and guidance with the writing of this thesis and Dr. Peter Heald for his inspiration and mentorship. Deepest love and appreciation to my wife Cindy.

TABLE OF CONTENTS

GUIDE TO FIGURES

INTRODUCTION

Cutaneous malignant melanoma (CMM) is a pigmented neoplasm of the skin that often arises in pre-existing nevi, although half appear on previously normal appearing skin. The cell of origin for malignant melanoma is the melanocyte, a pigmented cell derived from the neural crest that is normally located in the basal layer of the epidermis. (1)

There is broad consensus that the worldwide incidence of CMM has been rising over the last century. (2-7) It has been rising at different rates depending on the time epoch, gender, age cohort or geographic location considered. The rate of rise has been particularly steep during the most recent decades and has recently been characterized by many in the medical community as an "epidemic". (8-11) A number of researchers have questioned whether the apparent rise in the incidence of CMM is real and have raised the idea that it may be, at least in part, due to artifactual causes. (10, 12-18)

In the United States, the rise in incidence has been precipitous; in 2005, the age-adjusted incidence rate of cutaneous melanoma in the U.S. among whites of both sexes was 26.4 per 100,000 per year. This represents an overall increase of approximately 200% since 1975 when the age adjusted incidence was 8.7 per 100,000 per year. (19) [Figure 1] In contrast to the Caucasian population, the age adjusted incidence rate among blacks of both sexes in 2005 was 1.0 per 100,000 per year, a decline of approximately 40% since 1976 when it was 1.7 per 100,000 per year. (19) Notably, the rise in incidence of CMM is mostly accounted for by thin lesions (<1 mm) while the incidence of thick lesions (>1 mm) has remained static. (20)

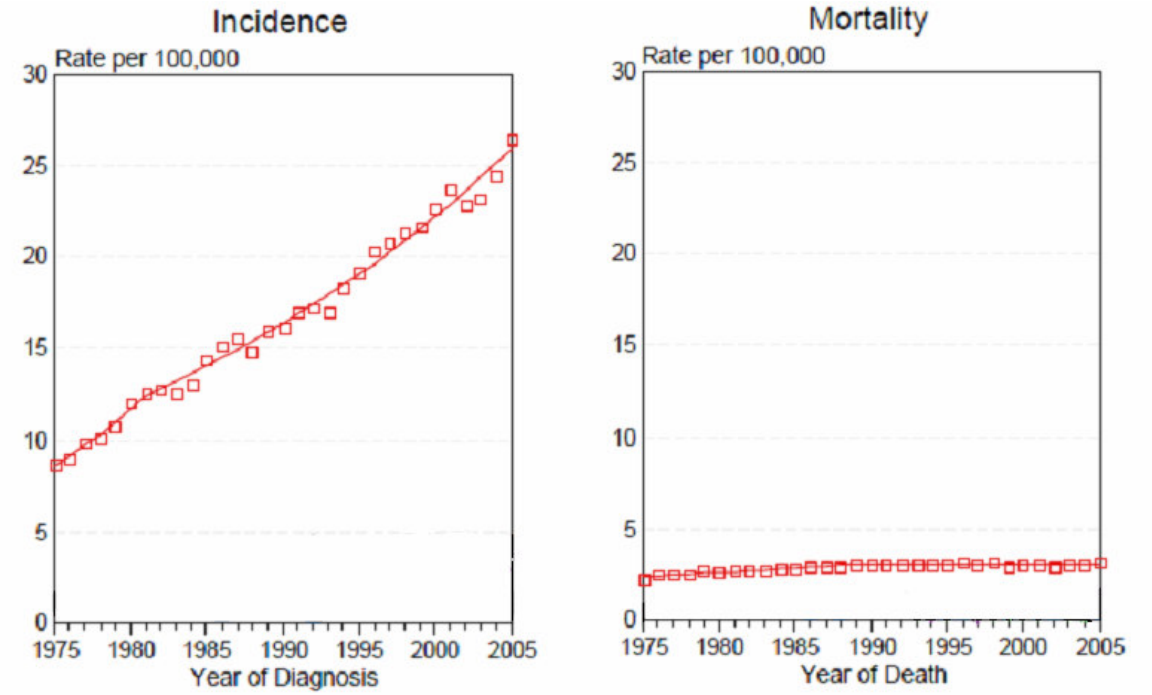

Figure 1. Age Adjusted Incidence and Mortality of Cutaneous Malignant Melanoma among US Whites of Both Sexes

Data and figures are adapted from the US Surveillance, Epidemiology and End Results (SEER) Cancer Statistics Review 2005 (National Cancer Institute)

While the incidence of CMM has been steeply rising over the past half century, the rise in the mortality rate has been much more modest. In 1975 the age-adjusted death rate for CMM for both sexes in the US was 2.1 per 100,000 per year. In 2005, it had risen to 2.7 per 100,000 per year, an increase of only 29%. (19) [Figure 1] Considering selected birth and gender cohorts separately, the death rate has been observed to rise and fall by relatively small amounts from year to year. The difference between the rise in incidence and the mortality rate, approximately an order of magnitude, has until now not been adequately explained.

While over four-fifths of patients initially diagnosed with malignant melanoma present with localized disease at the time of presentation, survival outcomes for all patients with

melanoma have significantly improved over the past few decades. The projected 5 year relative survival rate at the time of diagnosis for US patients with melanoma of any stage during the period between 1995 and 2001 was 92%, a statistically significant increase from the period 1974-1976 when 5 year survival was 80%. (21) For those melanoma patients with evidence of regional spread beyond the primary site, the 5 year relative survival drops to 64%. For those patients presenting with metastatic disease, the 5 year survival is only 16%. CMM accounted for roughly three fourths of all deaths from skin cancer in the United States (US) and accounted for 4.4% of all cancers in the US in 2006. (21)

The physical and psychosocial consequences of a positive diagnosis of CMM are quite serious. Some of the psychosocial ramifications commonly associated with a diagnosis of CMM include anxiety and depression (22, 23) as well as considerable difficulty obtaining life and health insurance or securing a home mortgage or pension plan (24). Depending on the stage of the malignancy, the protocol for re-excision of a lesion varies among clinicians. At many institutions, the margins mandated for re-excision of early stage melanoma are the same as for dysplastic nevi with severe atypia, though that is not universally the case. (25, 26) Though no longer recommended for patients with thin lesions (27), in the case of more advanced disease there is a low but significant risk of complications from sentinel node dissection that includes scarring, pain, sequelae resulting from disrupted lymphatics, and complications due to anesthesia. (28) Since patients who have been diagnosed with malignant melanoma have an increased risk of developing an additional primary melanoma, there is broad consensus that melanoma

survivors should undergo more frequent skin surveillance schedules with annual or semi-annual visits to the dermatologist. Melanoma survivors undergo increased rates of laboratory testing and diagnostic imaging that result in increased healthcare utilization over their lifetime. (29) In some cases, first degree relatives of melanoma survivors may be urged to undergo initial melanoma screening. (26)

EPIDEMIOLOGY AND ETIOLOGY OF MELANOMA

Increases in the incidence of cutaneous malignant melanoma have been attributed to a variety of both intrinsic and environmental factors such as genetics, ultraviolet radiation exposure, latitude, and age. The extent to which each of these forces has contributed to the rise in the incidence of CMM has been the subject of controversy; the relative contribution of each factor toward the development of clinically relevant disease remains uncertain.

Genetics

The most significant host factor to confer susceptibility to CMM appears to be a family history of melanoma. (30) On a very basic level, it appears that an individual's genetic susceptibility to develop melanoma reflects the innate ability of the individual and their blood relatives to withstand or adapt to ultraviolet light exposure as manifested by their inherited skin phototype; both constitutive pigmentation as well as tanning, the adaptive darkening of skin through up-regulation of melanocytes in response to UV radiation, determines to a large extent an individual's potential for protection from the mutagenic events that lead to skin cancer. (31, 32)

It has been demonstrated that familial ethnic variations in skin type are directly related to the presence or absence of polymorphisms in the melanocortin 1 receptor gene (MC1R) which governs the activity of melanocyte stimulating hormone at its receptor in the skin. (33-35) Fair skinned and/or red-headed individuals have a reduced ability to generate a

tanning response to UV light because they tend to harbor a variant of the MC1R gene and they subsequently suffer the highest incidence of melanoma. (36)

Being primarily a disease of light skinned individuals, CMM affects populations of European origin an order of magnitude greater than dark skinned individuals. (6, 37) During the period 1992-2002, the mean annual age-adjusted incidence of melanoma for American whites per 100,000 was 18.4 while the incidence for African Americans was calculated to be 0.8. (38) Other non-white ethnic groups in the U.S. demonstrated similarly low incidence rates of melanoma per 100,000: Hispanics 2.3, Native Americans 1.6 and Asian Americans 1.0. (38)

Latitude

Among Caucasian populations, the incidence of CMM generally trends higher with decreasing latitude, although this effect is complicated by varying patterns of recreational travel among inhabitants of northern countries. According to 2001 data, the highest age-adjusted incidences of melanoma worldwide were found in Australia (men= 40.5/100,000, women= 31.8/100,000) and New Zealand (men= 36.7/100,000, women= 34.9/100,000). (6) In 2001, North America had the third highest incidence of melanoma (men= 24.3/100,000, women= 16.2/100,000) followed in decreasing order by Scandinavia, the rest of Northern Europe, Israel and Eastern Europe. (6, 19, 39)

Within North America, incidence does not clearly correlate with latitude. The 2004 age adjusted incidence of melanoma per 100,000 in white populations of northern states such

as New Hampshire (men= 30.8, women= 26.3), Vermont (men= 31.7, women= 29.0) and Minnesota (men= 21.5, women= 15.8) are on par or higher than some southern tier states such as North Carolina (men= 26.4, women= 23.2) and Texas (men=20.2, women=10.7). (40) The white population of Hawaii has by far the highest melanoma rates (men= 89.3, women= 56.8).

Ultraviolet Radiation Exposure

Among the environmental factors that may influence the incidence of melanoma, there is broad but incomplete consensus as well as much debate over the idea that different rates and patterns of exposure to solar and solar-type UV radiation drives the major differences in melanoma incidence between historical, geographic, and age cohorts. (41-43)

The incidence of CMM does not appear to be proportional to total sun exposure as are other non-melanoma skin cancers. Instead, intermittent sun exposure appears to confer a greater risk of melanoma, suggesting that chronic sun exposure might be protective due to host tanning responses. The relationship between intermittent sun exposure and an increased incidence of melanoma are suggested by studies showing increased melanoma incidence associated with total number of holidays abroad (44), accessibility to air travel (45), and non-occupational sun exposure vs. occupational exposure (46).

Age

The risk of CMM increases significantly with advanced age. The probability of developing an invasive melanoma for the cohort encompassing subjects from birth to 39

years of age in the United States was 1 in 800 for men and 1 in 470 for women whereas for the cohort 70 years of age and older, the probability for men was 1 in 80 and for women was 1 in 178. (40)

Despite the correlation of incidence and increasing age, melanoma tends to affect the young more than other solid tumor cancers. The mean age of melanoma diagnosis is 58, roughly a decade earlier than other common cancers like those affecting the lung (70 years), colon (73 years), uterus (68 years), and prostate (68 years). (19) For patients aged birth to 39, there is a significantly increased chance of developing melanoma (men= 0.13%, women= 0.21%) than of developing other cancers like lung cancer (men and women= 0.03%) and colon cancer (men= 0.7%, women= 0.6%). Malignancies with comparable rates to melanoma in the age 0 to 39 cohort include non-Hodgkin's lymphoma (men= 0.14%, women= 0.09%) and breast cancer in women (0.48%). (21)

Birth/Period/Cohort Analyses

Since researchers first began to study the rise in melanoma incidence, epidemiological analysis has been brought to bear on the question of whether the increase in the incidence of melanoma is real, artifactual or a combination thereof. A birth-cohort model would postulate that the incidence of a disease varies when one identifies cohorts by the year they were born whereas a period model would detect changes in incidence occurring over specific periods of time.

In the late 1980s, when mortality rates in the U.S. were rising more uniformly across all age cohorts than they are presently, the rise in melanoma incidence was explained in several studies by applying a birth cohort model to the epidemiologic data. (3, 47)

When a team of Yale researchers in the late 1980s compared birth cohort vs. period cohort effects in data from the Connecticut tumor registry, the rise in melanoma incidence was reported to be almost entirely explained by the birth cohort effect, increasing in proportion to more recent birth cohorts. (47) Adding period as a variable was reported not to have changed the outcome as would be expected with an artifactual variable. Subsequently, the results of this study have been used as evidence that the apparent rise in the incidence of melanoma has been primarily due to real increases in the rate of disease.

In contrast, fitting data to a period-cohort analysis would allow for the possibility that artifactual factors are contributing to the apparent increase in disease. Artifacts are commonly introduced into epidemiologic analyses by phenomena such as increases in disease detection due to the introduction of new tests or imaging technology as well as changes in diagnostic criteria or an increase in the reporting of cases.

It should be noted that in the many decades since most of the birth cohort analyses were conducted, the mortality rate of melanoma has slowed or reversed among various age and birth cohorts. (19) It is not clear whether these analyses remain valid given the ever more uniformly static death rate. Furthermore, artifactual changes occurring gradually over a

long period of time may not be detectable with period analysis. Ultimately, the birth cohort explanation for the rise in the incidence of melanoma is not a settled issue. (48) Efforts to determine the relative contribution of a multitude of potential factors toward the increase in melanoma incidence remains a challenging task.

ARTIFACTUAL FACTORS

While the notion has been advanced that artifactual factors could account for no more than a small portion of the rise in incidence of melanoma, (49) careful scrutiny of the issue brings to light certain facts that appear to undermine this position. A summary of these facts are briefly enumerated below and will be explored more deeply further in this section.

First, the mortality rate of malignant melanoma has remained essentially static over decades of steep rises in incidence. (19) Compared to most other aggressively fatal cancers, the apparent dissociation between the incidence and mortality rate of CMM is unusual and has until now not been well explained.

Secondly, biopsy rates have been rising and closely parallel the overall increase in incidence. (15) Whereas the incidence of thin melanomas has been rising and almost completely accounts for the overall rise in disease incidence, (20, 50) the incidence of thick lesions has essentially remained static. (20) This preponderance of thin lesions with no attendant decrease in thick lesions calls into question the success of surveillance and prevention efforts and suggests that clinicians may be removing biologically indolent lesions at an increasing rate.

Third, many studies have suggested that dermatopathologists cannot consistently agree on the diagnosis of borderline melanocytic lesions. (49, 51-56) This raises the possibility that factors other than diagnostic criteria, such as subjective bias, may be influencing the

diagnosis of melanoma. Moreover, there is a significant risk of medico-legal liability in the diagnosis of melanoma; this invites questions about the extent to which the threat of litigation impacts diagnostic trends.

Death Rate, Disease Surveillance and Tumor Thickness

The 2005 age-adjusted death rate in the U.S. was 4.0 per 100,000 per year for males and 1.8 per 100,000 for females. (19) While the age adjusted death rate for both genders has increased significantly since the mid 1970s, changes have occurred within a very narrow range and trends in mortality according to the latest SEER data indicate that overall mortality from melanoma was decreased during the period 1989-2001 as compared to the period 1975-1981. The U.S. age-adjusted death rate rose from 2.1 in 1975 to a high of 2.8 in 1990 but has remained between 2.6 and 2.8 in all subsequent years. (19)

Some have argued that the stable death rate may be accounted for by better disease prevention secondary to surveillance and early screening efforts. (11) However, others have argued that this relationship is implausible (14). In order for screening and prevention efforts to have mitigated mortality to a static rate, changes in surveillance and screening must have exactly matched the rise in melanoma incidence over a long period of time.

There is evidence to suggest that increased surveillance for skin cancer has led to an increase in the reported incidence of melanoma (18, 57); however, the impact of surveillance and prevention on the mortality of CMM is uncertain. The frequency of skin

exams among the US population has not been well documented. According to an annual, cross-sectional in-person household survey conducted by the National Center for Health Statistics, the prevalence of lifetime skin cancer screenings was low with a mere 15% of US workers reporting ever having had a skin exam in their lifetime. (58) While 69% of respondents (26,225/38,124) reported seeing a primary physician over the past year, only 8% reported having had a skin exam in the past 12 months. In light of the fact that most people do not perform self skin examinations and most people never see a dermatologist, (58) it seems unlikely that surveillance and screening can account for the relatively stable death rate.

Early and frequent screening would certainly explain the increase in the number of thin lesions. However, it does not explain the stability in the rate of thick lesions. If advanced melanomas were being prevented by screening measures, the incidence of thick lesions should decrease over time. That is, again, unless the overall rise in incidence of disease exactly matched the number of melanomas prevented -- an implausible scenario. It appears that the increase in the number of thin melanomas being detected has had no mitigating effect on the incidence of advanced disease. (20)

Likewise, if it is assumed that the incidence of melanoma has truly increased across the whole population, then an increase in the number of people with advanced, fatal disease should be observed. It should follow that since the majority of patients with advanced disease never see a physician, the death rate should still rise significantly despite any screening efforts. Nevertheless, this phenomenon has not been observed.

Biopsy Rate

A recent study by Welch et al. that focused on patients 65 years of age and older reported that increases in the average biopsy rate have roughly paralleled the increase in melanoma diagnoses, rising 2.5 fold, from 2847 per 100,000 in 1986 to 7222 per 100,000 in 2001. (15) The rise in incidence in this cohort over the same time period was 2.4 fold, from 45 to 108 per 100,000 population. The authors suggested that the proportionality of the rise in biopsy rate to the rise in incidence in the setting of a static mortality rate implied that the rise in incidence was due to "overdiagnosis—the increased incidence being largely the result of increased diagnostic scrutiny and not an increase in the incidence of disease." (15)

Although higher socioeconomic status has been associated with a higher incidence of CMM, the average thickness of lesions in an affluent cohort was found to be less than for lower socioeconomic groups. (59) Subsequently, higher socioeconomic groups were found to enjoy an overall better prognosis than lower socioeconomic groups – a finding that raises the possibility that higher levels of wealth, corresponding to better healthcare access and more frequent skin surveillance, may also result in the increased biopsy rate of otherwise biologically benign lesions. (12)

Histopathologic Diagnosis

In contrast to CMM, dysplastic nevi are thought to be relatively clinically stable benign lesions that possess some of the histopathologic features that characterize "frank" malignant melanoma such as cytologic atypia, disordered proliferation, hyperchromasia

and irregular nuclear contours. Most dermatologists consider dysplastic nevi one step along a continuum of melanocytic lesions with increasingly malignant potential. (60) It is commonly believed that certain dysplastic nevi represent precursor lesions to malignant melanoma. (61-64)

Dysplastic nevi are often characterized as possessing “mild”, “moderate” or “severe” cellular atypia. (65) The diagnostic criteria used by dermatopathologists to classify melanocytic lesions are, however, less definitive for those borderline lesions that share features of both benign and malignant disease. Histologic diagnosis is based on evaluation of a collection of findings with no single element being diagnostic. There is no gold standard for the diagnosis of malignant melanoma and there appears to be a significant measure of subjectivity inherent in the process of pathologic diagnosis.

The reliability and reproducibility of histologic criteria used to denote dysplastic nevi has not been well established. Historically, there has been a marked lack of consensus among dermatopathologists in characterizing “borderline” dysplastic nevi. (66) With few exceptions, most of the studies that have examined the reliability of the histologic diagnosis of melanocytic dysplasia by examining diagnostic concordance between different dermatopathologists have found inconsistent application of diagnostic criteria to characterize the histopathology of melanocytic lesions. (49, 51-56)

One of the first studies to demonstrate a lack of consensus among dermatopathologists evaluating borderline lesions was conducted by a German researcher in the mid 1980s.

Fifteen dermatopathologists from around the world with a special interest in melanocytic lesions were asked to evaluate a set of nine "precursors of malignant melanoma". Single slides were provided to participants without accompanying clinical information. Results revealed that there was no agreement between the dermatopathologists in designating the lesions "benign", "pre-malignant" or "malignant" and there was little agreement in diagnostic nomenclature. The author concluded that the ability of pathologists to render reliable interpretations of biopsies containing atypical melanocytes was "limited". (55)

Van der Esch and colleagues undertook an extensive international study in 1991 to determine if pathologists' diagnostic threshold for malignancy had changed over time. (49) This landmark study has been widely cited since its publication as the strongest evidence thus far that changes in diagnostic criteria used to evaluate melanocytic lesions have not changed as a function of time. In this study, ten pathologists from various international institutions read a total of 2506 slides of melanocytic lesions originally biopsied in the 1930s, the 1950s and the 1980s. The diagnostic material originated from a variety of international medical centers. In choosing slides for inclusion in the study, the authors chose "to give greater emphasis to those lesions – the junctional and compound naevi – where a change of opinion…as to malignancy would be most likely to arise". Original diagnoses of the slides were obtained and classified according to the original diagnosis as "benign", "dubious benign", "dubious malignant", or "malignant".

The authors reported an astounding degree of agreement among participating pathologists in classifying the study lesions; only 2.8% of lesions changed diagnostic categories upon

re-diagnosis. However, despite the authors' stated interest in focusing on "borderline" lesions, it is important to note that only 108/2506 (4.3%) of the slides evaluated in the study had an original diagnosis classified as "dubious benign" or "dubious malignant". With the majority of slides (N=1700/2506, 67%) originally diagnosed as "benign", it is not surprising that the authors found a modest "overall percentage of change" in diagnostic category of 2.8%.

In contrast, although "dubious" diagnoses constituted only a small fraction of the total pool of study lesions, nevertheless a large portion of the lesions originally classified as "dubious benign"/ "dubious malignant" – over 1/3 -- were re-evaluated as frankly "malignant". Taking the "dubious malignant" slides alone from all periods, well over half (23/41, 56%) were re-classified as frankly "malignant" while only 10/41 (24%) were re-classified as "benign" or "dubious benign". Conversely, of the 692 lesions originally diagnosed as "malignant", participating pathologists re-confirmed the diagnosis in 665 (96%) of cases. (49) Taken in this light, the study by Van der Esch et al. appears to be less than conclusive about the effect of time trends in the diagnosis of borderline melanocytic lesions.

A similar study to that of Van der Esch, et al. was conducted around the same time by Philipp, et al. in the United Kingdom. (54) Seventy lesions from each of three time periods: the 1940s, 1950s and 1980s were chosen that included roughly 1/4 malignant lesions, roughly 2/3 "junctional or compound naevi for which a change of opinion over time was thought most likely to arise", and about 5% "intradermal lesions which were

considered less likely to be confused with malignant melanoma." The published study reported the results of re-diagnosis of the diagnostic material by only a single pathologist. The authors reported that 206/210 slides (98%) were not reclassified with a different diagnostic category than the originally issued diagnosis.

A study conducted at Yale in 1992 by Duray et al. compared the responses of five observers reading 50 slides of "nevomelanocytic tumors" in a blinded fashion. (53) The study demonstrated only "moderate" inter-observer agreement with regard to the characterization of the histologic components of dysplastic nevi.

In a 1993 study, Duncan et al. found a high concordance rate (77%) among five dermatopathologists asked to grade a set of previously diagnosed melanocytic lesions in distinguishing between benign nevi, various grades of dysplastic nevi and melanoma. (67)

A study undertaken by a Danish team of dermatopathologists, Hastrup et al., examined the inter-observer reproducibility of the various histological criteria used to distinguish nevi as "dysplastic". (52) After analyzing the responses of four observers asked to re-diagnose a set of previously diagnosed melanocytic lesions, they found "slight" to "fair" inter-observer reproducibility of histological features, particularly cytological features.

In 1995, Farmer et al. found "moderate" concordance (kappa=0.50) between eight "expert pathologists" rating a set of biopsies of 37 melanocytic lesions as "benign",

"malignant" or "indeterminate". (68) Thirty five percent of the cases had complete agreement, 27% had one discordant observer and 38% had two or more discordant observers.

Corona and colleagues measured agreement among four dermatopathologists evaluating a large set of mixed melanomas and benign pigmented lesions. (69) They reported an overall kappa value of 0.61 for melanoma vs. benign lesion but a poorer level of agreement for presence of absence of specific histologic features.

A 1997 study by Weinstock et al. compared agreement among five dermatopathologists and two "melanoma experts" grading a heterogeneous collection of 112 biopsy slides of melanocytic tumors with the help of a photomicrographic atlas. (51) Graders' observations were assigned points according to a 5-point scale that grouped responses into different categories according to pre-determined diagnostic terms, i.e. "no melanocytic dysplasia", "melanocytic dysplasia with slight, moderate or severe cellular atypia" or "melanoma". When raters' responses were grouped in a dichotomous fashion (benign or malignant) and compared against each other, the intra-class coefficient was 0.58, suggesting a significant level of discordance.

Considering the available literature on the subject as a whole, a substantial majority of the prior studies examining concordance rates of dermatopathologists reading melanocytic lesions have not reported robust levels of agreement, particularly in the case of borderline lesions. Of the nine studies mentioned above, none except Duncan et al. and

Phillipp et al., demonstrated high levels of agreement in the case of borderline lesions. The evidence so far seems to suggest that dermatopathologists, whether or not they agree on normative diagnostic criteria, have not historically achieved consistently high levels of concordance in the diagnosis of borderline melanocytic lesions.

Medicolegal Liability

There have been few studies to date examining the influence of malpractice claims and legal liability on the decision making behavior of pathologists. In an examination of the records at a professional liability insurer in California, researchers found that 8.6% of all malpractice claims generated against pathologists involved the words “skin cancer” and/or melanoma. (70)

In an analysis of published verdicts and jury settlements on a popular legal database, 26 out of the 171 cases examined involved the underdiagnosis of melanoma on skin biopsies by dermatopathologists. (71) The results of this study corroborated the work of Troxel et al. in revealing that “false-negative” diagnoses of melanoma constitute the most common claim against surgical pathologists.

HYPOTHESIS

It is hypothesized that the diagnostic threshold of malignancy for the histopathologic diagnosis of melanoma has decreased over the past 20 years; that dermatopathologists are diagnosing more melanoma now than in the period 1988-1990, in lesions that would previously have been regarded as dysplastic nevi. This study has sought to determine if diagnostic decision-making in a cohort of dermatopathologists evaluating severely atypical melanocytic lesions and thin superficial spreading melanomas has changed as a function of time over the two decades that elapsed between the late 1980s and 2008.

STATEMENT OF PURPOSE

The dramatic increase in the incidence of malignant melanoma over the past few decades has been attributed to an array of possible factors, both artifactual and real. One under-explored factor that could help to explain the apparent paradox posed by the dissonance of incidence and mortality rates, is the possibility that dermatopathologists' threshold for rendering a diagnosis of melanoma may have changed over time. Given the dramatic increase in biopsy rates over the past 20 years, it is plausible that shifts in diagnostic decision-making by dermatopathologists could have occurred over the same period. Such shifts, potentially due to increased vigilance in the face of heightened legal liability, may have resulted in an increase in the diagnosis of melanoma in lesions that would have been otherwise diagnosed as benign in the past- a fact that may explain some of the discrepancy between incidence and mortality. In light of the significant morbidity and cost associated with a diagnosis of CMM, it is important that any potential source of overdiagnosis be identified and mitigated.

METHODS

A search was conducted within the surgical pathology computer database of the Massachusetts General Hospital (MGH) for the years 1988 to 1990. The entire database was searched with the terms “dysplastic nevus”, “severe atypia” and “superficial spreading malignant melanoma”. All pathology reports that contained any of the aforementioned terms were collected and were subsequently reviewed in order to determine suitability for inclusion in the study. Slides of dysplastic nevi were deemed to meet inclusion criteria if the diagnosis mentioned “severe atypia” or “moderate to focally severe atypia”. Particular preference was given to slides of dysplastic nevi with severe or moderate to focally severe atypia of the “intraepidermal component”. Slides of malignant melanoma were considered suitable for inclusion if they contained the term “superficial spreading melanoma” and if they contained a designation of Clark level II, III or II/III. Slides of biopsies originally processed at outside institutions were excluded. Slides that contained the terms “spindle cell”, “blue nevus”, or “Spitz nevus” were also excluded.

A total of 1207 pathology reports were generated by the computer search. According to the pre-stated criteria, the total number of suitable cases of dysplastic nevi obtained was 28. Seventy nine suitable cases of superficial spreading melanoma were obtained.

In light of the possibility that the re-classification of any of the non-melanoma study slides as invasive melanoma could have lead to a medico-legal dilemma, consideration was given to the fact that a 15 year interval had elapsed between the study and the original diagnosis. It was concluded that the study could proceed with reasonable

assurance that study slides would not be re-diagnosed as advanced melanoma since the biologic aggressiveness of this type of tumor would have in all likelihood declared itself in a much shorter time interval and the consequences would already have been known and addressed. Furthermore, the standard of care at the MGH at the time that these biopsies were originally performed mandated conservative re-excision of 1 cm for severely dysplastic nevi – the same level of care performed for patients diagnosed with early stage superficial spreading melanoma.

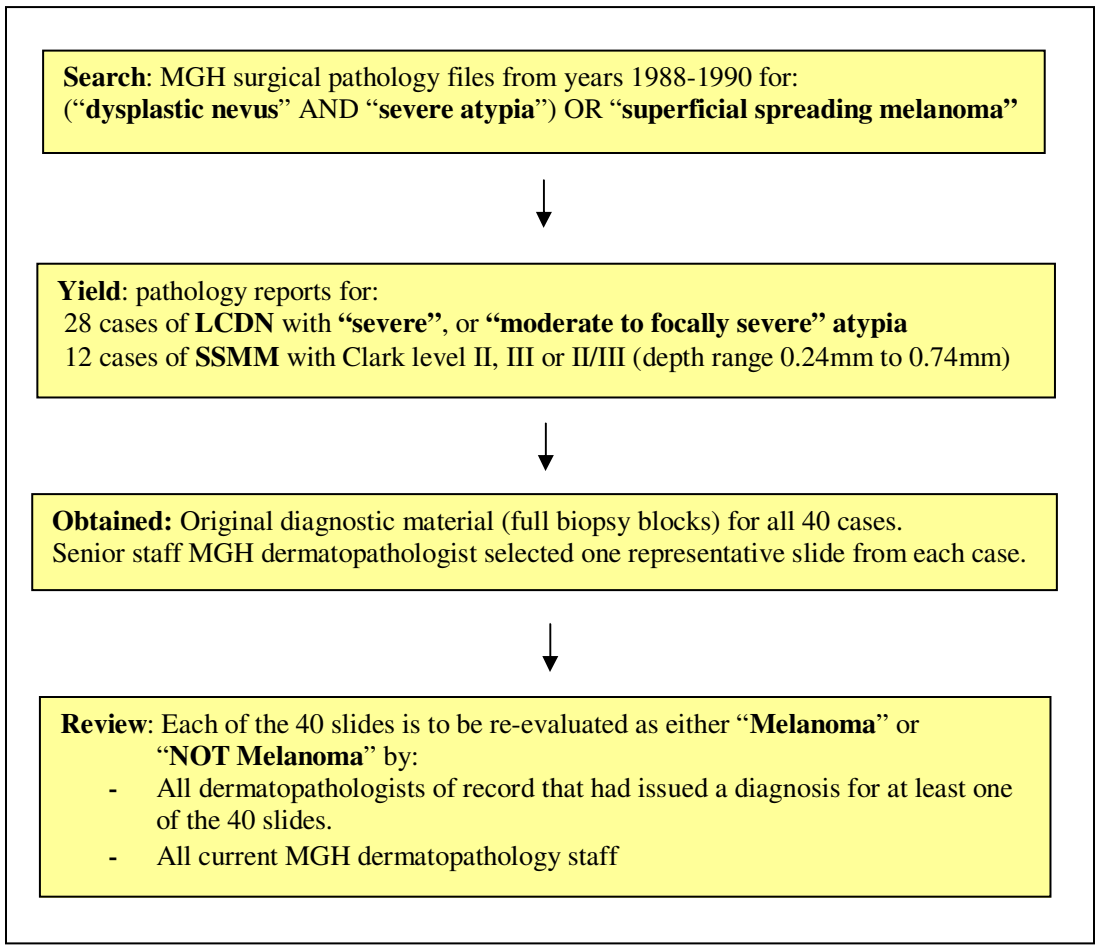

Figure 2. Schematic Diagram of the Experiment

Twenty eight cases of dysplastic nevi and twelve cases of superficial spreading melanoma (depth range 0.24mm to 0.74mm) were chosen for inclusion in the study. The entire set of slides corresponding to the 40 cases selected for the study were reviewed by a senior staff dermatopathologist at MGH in order to select one representative slide from each case as free of artifacts as possible. The dermatopathologist that selected representative slides did not have access to the pathology reports corresponding to the cases reviewed. The total number of study slides was fixed at 40 in order to make the re-evaluation of the study slides a manageable enterprise and thus increase the chances that participating dermatopathologists would re-evaluate the slides in a timely manner. All slides were anonymized by concealing the accession numbers with white tape. All 40 slides were shuffled randomly, numbered 1 to 40, and placed in a slide folder. A schematic of the study protocol has been provided below. [Figure 2]

The identities of the original dermatopathologists that had signed out the diagnoses for each of the 40 study cases were noted and recorded. A total of 9 dermatopathologists on staff at MGH during the years 1988-1990, singly or in tandem, issued the original diagnoses of the 40 study slides.

The identities of the 9 original dermatopathologists were anonymized and designated by letters of the alphabet. The number of slides read by each dermatopathologist ranged from 1 to 12 with a mean of 7.2 slides per reader. Thirteen of the study slides (32.5%) had originally been diagnosed jointly by 2 dermatopathologists. Three of the jointly

diagnosed slides were co-diagnosed by a dermatopathologist (L) that was deceased at the time the study was conducted. [Figure 3]

A= 12 Total	7 Jointly
B= 8 Total	3 Jointly
C= 12 Total	3 Jointly
D= 4 Total	0 Jointly
E= 6 Total	6 Jointly
F= 1 Total	1 Jointly
J= 3 Total	2 Jointly
K= 1 Total	1 Jointly
L= 2 Total	3 Jointly

Figure 3. Number of Study Slides Diagnosed Singly or Jointly by each Dermatopathologist, as Designated by a Letter of the Alphabet.

Dermatopathologists were recruited to participate in the study. Participants were selected if they fulfilled either of two criteria:

1. They had rendered a diagnosis on one of the slides selected for the study or,
2. They were currently serving on the staff of the dermatopathology department at MGH at the time the study was conducted.

Therefore, all participating dermatopathologists were currently serving or had formerly served on the staff in the Department of Dermatopathology at the Massachusetts General Hospital. Attempts were made to recruit all of the original diagnosing dermatopathologists as well as all faculty members in the MGH Department of Dermatopathology.

Participating dermatopathologists were asked to re-evaluate all the slides and decide whether the biopsy represented “melanoma” or “not melanoma”. To facilitate comparison of diagnoses, all 40 original official diagnoses of the study slides were subjected to a dichotomous categorization whereby they were designated either “malignant melanoma” or “not malignant melanoma”. In this way, all slides of dysplastic nevi were considered “not melanoma”. A chart has been provided that illustrates all potential responses by study participants [Figure 4]. Slides with an original diagnosis of “not melanoma” were coded “0” and slides with an original diagnosis of “melanoma” were coded with a “1”.

Participating dermatopathologists received a cover letter explaining the procedures for grading study slides and recording diagnoses. In order to achieve a degree of participant blinding to the primary aims of the study, and in order to approximate as much as possible, a non-biased diagnostic setting, the cover letter contained a description of the study limited to a simplified and general explanation of the study’s aims. All participants were told that at the conclusion of the study they would be apprised of the specific aims of the study, and given a copy of the protocol and the working manuscript.

After the cover letter had been sent to selected dermatopathologists and consent had been obtained, the slide set was sent sequentially to responding dermatopathologists along with instructions for reviewing the slide set and a grading sheet for recording diagnoses. The grading sheet allowed participating dermatopathologists to check a box for each numbered slide designating the slide “melanoma” or “not melanoma” as well as a space

Slide #	Original Reader	Original Diagnosis	DP A	DP B	DP C	DP D	DP E	DP F	DP G	DP H	DP I	DP J	DP K	DP L
1	L	0												
2	A,E	1												
3	C	0												
4	D	0												
5	A	0												
6	A,B	0												
7	C	1												
8	D	0												
9	D	0												
10	B	1												
11	B	0												
12	B,F	1												
13	A	0												
14	J	1												
15	A	0												
16	D	1												
17	C	1												
18	B	1												
19	C	0												
20	U/K	0												
21	L	0												
22	A	0												
23	E	0												
24	A,B	0												
25	A,E	0												
26	C	0												
27	C	0												
28	C,E	0												
29	A	0												
30	A	0												
31	B	0												
32	C,E	0												
33	C	0												
34	B	1												
35	C	1												
36	C	0												
37	A	0												
38	C	1												
39	E	0												
40	A	1												
Total Melanoma		12												

Figure 4. Study Response Tally Outline

The diagnoses of forty study slides were dichotomized as "melanoma" (white= 0) or "not melanoma (black= 1). Potential study participants included nine dermatopathologists (A-F, J-L) that had rendered an original diagnosis on at least one of the study slides as well as current MGH staff dermatopathologists (A, F-I).

Key

	Biopsy originally diagnosed as **melanoma**	1
	Biopsy originally diagnosed as **NOT melanoma**	0

Current MGH DPs 2008
A
F
G
H
I

Original Diagnosing MGH DPs (1988-1990)
A
B
C
D
E
F
J
K
L

DP = Dermatopathologist

in which they were instructed to add additional comments about the slides if they so desired.

RESULTS

Five dermatopathologists currently on staff at MGH at the time of the study evaluated the slide set; two of the current staff members belonged to the original set of 9 dermatopathologists that had rendered an original diagnosis for one or more of the study slides.

Of the original 9 dermatopathologists that diagnosed one or more of the study slides, 3 successfully completed re-evaluation of the slide set. One additional member of the original set of 9 diagnosing dermatopathologists was in the process of re-evaluating the slide set at the time of this writing. One of the original dermatopathologists had since passed away. At the time of this writing, five of the original diagnosing dermatopathologists had not responded to study recruitment attempts.

All study participants diagnosed significantly more melanomas than the 12 melanomas that were originally included in the slide set. [Figure 5] The mean number of melanoma diagnoses for all 6 graders was 19.7 (median= 19.5, range 16-23), an increase of 64%.

There were 6 instances where a dermatopathologist disagreed with his or her own diagnosis 20 years ago. These changes of within-rater diagnosis went from benign to malignant in 5 of the 6 new diagnoses. [Figure 5]

The Cohen kappa statistic was used to determine the degree of agreement between diagnoses. The free-marginal kappa as a measure of inter-rater agreement was

Slide #	Original Reader	Original Diagnosis	DP A	DP B	DP F	DP G	DP H	DP I
1	L	0	0	0	0	0	1	0
2	A,E	1	1	1	1	1	1	0
3	C	0	0	0	0	0	0	0
4	D	0	0	0	0	0	0	0
5	A	0	0	1	0	1	1	1
6	A,B	0	0	1	0	0	0	0
7	C	1	1	1	1	1	1	1
8	D	0	0	0	0	0	1	0
9	D	0	0	0	1	1	0	0
10	B	1	1	1	1	1	1	1
11	B	0	1	0	0	0	0	0
12	B,F	1	1	1	0	0	1	1
13	A	0	0	1	1	0	1	1
14	J	1	1	1	1	1	1	0
15	A	0	1	1	0	1	0	1
16	D	1	1	1	1	1	1	1
17	C	1	1	1	1	1	1	0
18	B	1	1	1	1	1	1	1
19	C	0	1	1	1	1	0	0
20	U/K	0	0	1	0	1	0	1
21	L	0	0	0	0	0	0	1
22	A	0	1	0	0	1	0	1
23	E	0	1	0	0	1	0	1
24	A,B	0	0	1	0	1	0	1
25	A,E	0	0	1	0	0	0	1
26	C	0	0	1	0	0	1	1
27	C	0	0	0	1	0	0	0
28	C,E	0	1	0	0	0	0	1
29	A	0	0	0	0	0	0	0
30	A	0	1	0	1	0	1	0
31	B	0	0	0	0	0	0	0
32	C,E	0	0	0	0	0	1	1
33	C	0	0	0	0	1	0	0
34	B	1	1	1	1	1	1	1
35	C	1	1	1	1	1	1	1
36	C	0	0	0	0	0	0	1
37	A	0	0	0	0	0	0	0
38	C	1	1	1	1	1	1	1
39	E	0	0	0	0	0	0	1
40	A	1	1	1	1	1	1	1
Total Melanoma		12	19	21	16	20	19	23

Original Diagnosing MGH DPs (1988-1990)
A
B
C
D
E
F
J
K
L

Current MGH DPs 2008
A
F
G
H
I

→ Mean revised melanoma yield: **19.7**

Figure 5. New Diagnoses of Study Slide Set by Participating Dermatopathologists (DPs) compared to the original diagnoses.

There were 40 slides in the study slide set: 28 dysplastic nevi and 12 superficial spreading melanomas. Six DPs reviewed the study slides. Diagnoses were given a dichotomous categorization: 0 = "not melanoma and 1 = "melanoma". Individual DPs are designated by letters of the alphabet. Three of the six participating DPs (A,B, and F) had rendered an original diagnosis on at least one of the study slides and are represented by a yellow background. DPs who did not render one of the original diagnoses (G, H, and I) have a white background. There were a striking number of instances (21%) where the original diagnosis of "not melanoma" was overturned.

Key

	Slide originally diagnosed "**NOT melanoma**" was re-evaluated as "**melanoma**"
	Slide originally diagnosed "**melanoma**" was re-evaluated as "**NOT melanoma**"
	Biopsy originally diagnosed as **melanoma**
	Biopsy originally diagnosed as **NOT melanoma**
	Dermatopathologist disagreed with their original diagnosis

ascertained by comparing the diagnoses of all 6 dermatopathologists that re-evaluated the slides. The free-marginal kappa was also used as a measure of temporal concordance – the relative difference between raters' diagnoses in 2008 against the diagnoses originally rendered for the study slides in 1988-1990. Temporal concordance was calculated retrospectively between each participating dermatopathologist and the original diagnosing dermatopathologists, whether they were one in the same or not.

The Cohen kappa coefficient is a value that may be generated by comparing sets of paired observations. (72-74) The kappa statistic describes two observers' level of agreement as compared to chance. The kappa statistic ranges between 0 and 1 with increasing value proportional to level of agreement. The "free-marginal" kappa statistic is used in place of the "fixed-marginal" kappa statistic when raters are not obliged to rate a particular number of items in one way or another.

There is considerable disagreement among statisticians about what value of kappa constitutes a sufficient level of agreement. Landis and Koch devised a widely accepted interpretive scale that identified kappa values of 0.61-0.8 as indicating "substantial agreement". (74) [Figure 6] This scale was later corroborated by Rietveld and Van Hout. (75) More recently, Shrout proposed a revision of Landis and Koch's original scale, suggesting that kappa levels of 0.81 to 1.0 should indicate "substantial agreement". (76) A more conservative interpretation of kappa was proposed by Kripendorff, who declared that "definite conclusions" about the kappa statistic can only be drawn for values of 0.8 or greater. (77) Values below 0.67, according to Krippendorff, should be "discounted". In

contrast, some prominent psychiatric researchers have cited kappa values of 0.5 or 0.6 as "adequate" (78).

K	Landis and Koch's Interpretation	K	Statistical interpretation
0.81 — 1.0	Almost perfect agreement	1	Perfect agreement above chance
0.61 — 0.8	Substantial agreement		
0.41 — 0.6	Moderate agreement		
0.21— 0.4	Fair agreement		
0.0 — 0.2	Slight agreement	0	Agreement equal to chance
< 0	No agreement	-1	Perfect disagreement below chance

Figure 6. Interpretation of the Cohen Kappa Coefficient

The Cohen Kappa coefficient ranges from -1.0 to 1.0. There is considerable controversy over how to interpret Kappa values. One common method of interpretation relies on an arbitrary scale as defined by Landis and Koch. Another common approach proceeds from the statistical definition of the Kappa coefficient with -1.0 representing perfect disagreement below chance, 0.0 representing agreement equal to chance and 1.0 representing perfect agreement above chance.

For the purposes of this study, only kappa values generated from the data that fall well below or above the "grey area" of 0.61-0.8 are discussed in a qualitative manner. Despite the lack of a clear threshold for an acceptable level of agreement, the magnitude of the mean kappa statistic for the group as a whole, representing temporal discordance between the diagnosing dermatopathologist in 1988-1990 and those dermatopathologists re-reading the slides in 2008, still reveals in an imperfect way, to what degree diagnostic behavior employed by this group of dermatopathologists differs from those reading the slides 20 years prior.

The mean free-marginal kappa values describing temporal concordance between each of

the 6 participating dermatopathologists and the original diagnosing dermatopathologists for all study slides was 0.53 (range 0.15- 0.7) [Figure 7]. When comparing temporal concordance for lesions originally diagnosed "Melanoma", the mean marginal free kappa was 0.86 (range 0.5- 1.0). [Figure 8] For lesions originally diagnosed "not melanoma", the mean marginal free kappa was 0.39 (range 0.0- 0.64). [Figure 9]

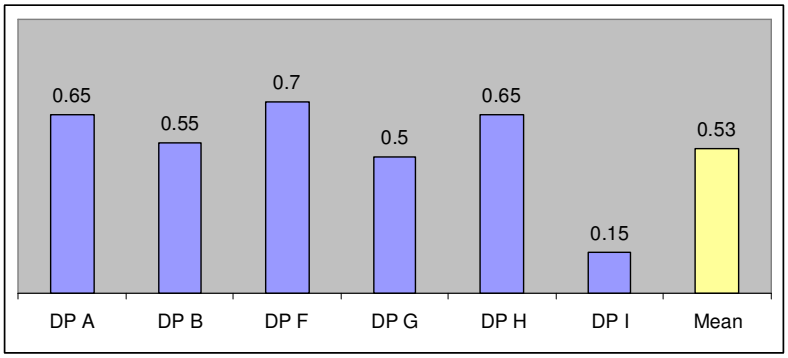

Figure 7. Free-marginal Kappa Values Representing Temporal Concordance between Each Participating Dermatopathologist and Original Diagnosing Dermatopathologists for **All Study Slides**.

Kappa values representing concordance of participating dermatopathologists with the original diagnosis for all study slides did not surpass 0.7 for all but one dermatopathologist. The overall mean agreement between the 6 participating dermatopathologists and the diagnosing dermatopathologist for all study slides was represented by a mean kappa value of 0.53.

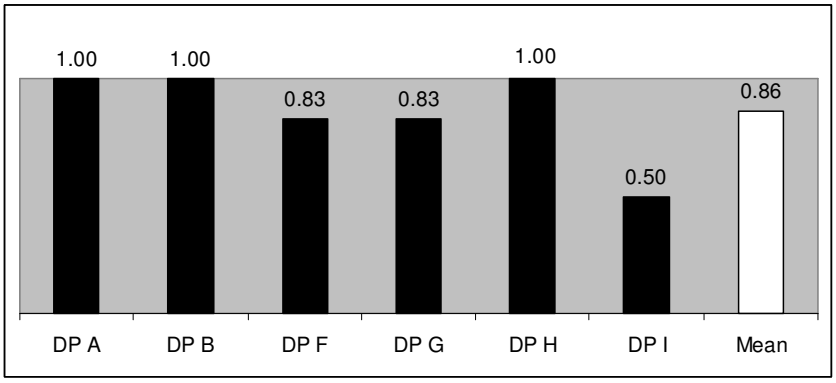

Figure 8. Free Marginal Kappa Values Representing Temporal Concordance between Each Participating Dermatopathologist and Original Diagnosing Dermatopathologists for Slides Originally diagnosed "**Melanoma**".

In evaluating slides originally diagnosed as "Melanoma", participating dermatopathologists demonstrated a high level of concordance with the original diagnosis. Half the participating dermatopathologists had perfect agreement with the original diagnosis of melanoma.

Overall mean agreement between the 6 participating dermatopathologists and the diagnosing dermatopathologist for lesions diagnosed "melanoma" was represented by a mean kappa value of 0.86.

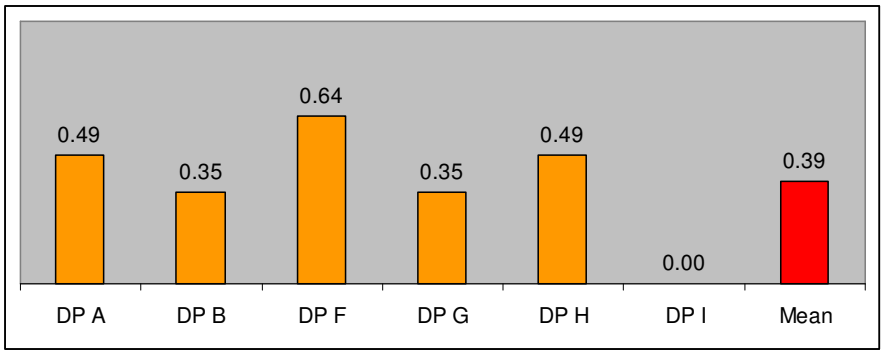

Figure 9. Free Marginal Kappa Values Representing Temporal Concordance between Each Participating Dermatopathologist and Original Diagnosing Dermatopathologists for Slides Originally Diagnosed "**NOT Melanoma**".

In evaluating slides originally diagnosed "NOT Melanoma", participating dermatopathologists demonstrated a low level of concordance with the original diagnosis. Overall mean agreement between the 6 participating dermatopathologists and the original diagnosing dermatopathologist for lesions diagnosed "not melanoma" was represented by a mean kappa value of 0.39.

In considering inter-rater agreement between all 6 participating dermatopathologists, there was a large difference in degree of agreement depending on the original diagnosis. For all study slides, participating dermatopathologists had a low inter-rater agreement rate represented by a mean free-marginal kappa value of 0.38. Only considering lesions originally diagnosed "not melanoma", participating dermatopathologists had a very low agreement rate represented by a mean free-marginal kappa of 0.22. For lesions originally diagnosed "melanoma", the rate of inter-rater agreement was represented by a relatively high mean free-marginal kappa value of .74.

Considering temporal concordance in terms of what proportion of raters agreed with the original diagnosis, unanimous agreement with the original diagnosis was achieved in 66.7% of slides originally diagnosed "Melanoma". In 25% of slides originally diagnosed "Melanoma" 5/6 raters agreed with the original diagnosis. In 8% of slides, 4/6 raters agreed with the original diagnosis. [Figure 10]

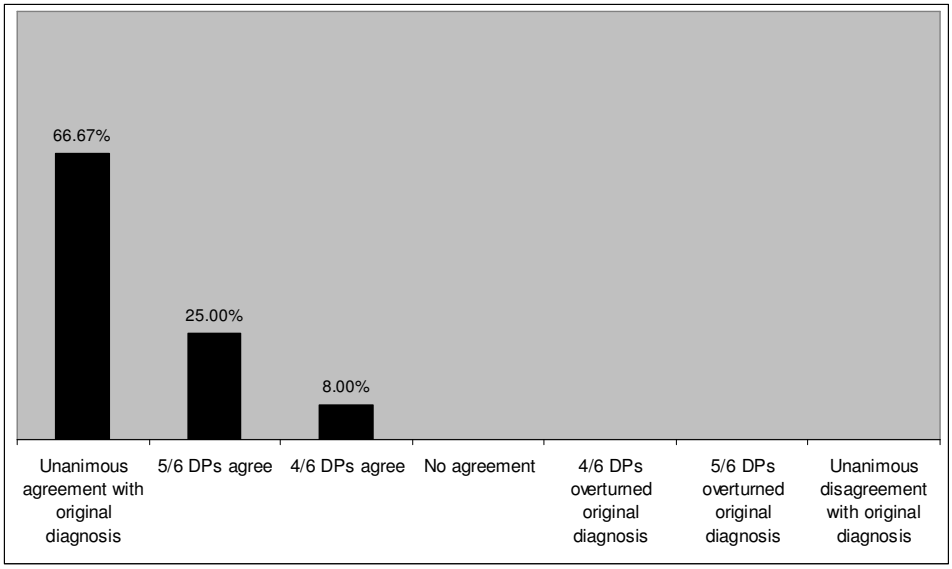

Figure 10. Concordance with original diagnosis among participating DPs: Melanoma Slides Only

Of the 12 slides originally diagnosed "Melanoma" participating DPs had varying levels of concordance with the original diagnosis. There was unanimous agreement (6/6) among participating DPs with the original diagnosis in approximately 2/3 of slides originally diagnosed as "Melanoma.

For slides originally diagnosed "Not Melanoma", unanimous agreement with the original diagnosis was achieved in only 17.9% of slides originally diagnosed "Not Melanoma". In 32.1% of slides originally diagnosed "Not Melanoma" 5/6 raters agreed with the original diagnosis. 14.3% of cases had 4/6 raters agree with the original diagnosis. In 21.4% of slides originally diagnosed "Not Melanoma", there was no agreement one way or the other among the 6 raters. In 14.3% of cases, a majority of raters (4/6) overturned the original diagnosis of "Not Melanoma". [Figure 11]

Although the original diagnoses of the study slides do not represent a "gold standard" for diagnosis of malignant melanoma, it may be instructive to momentarily regard them as

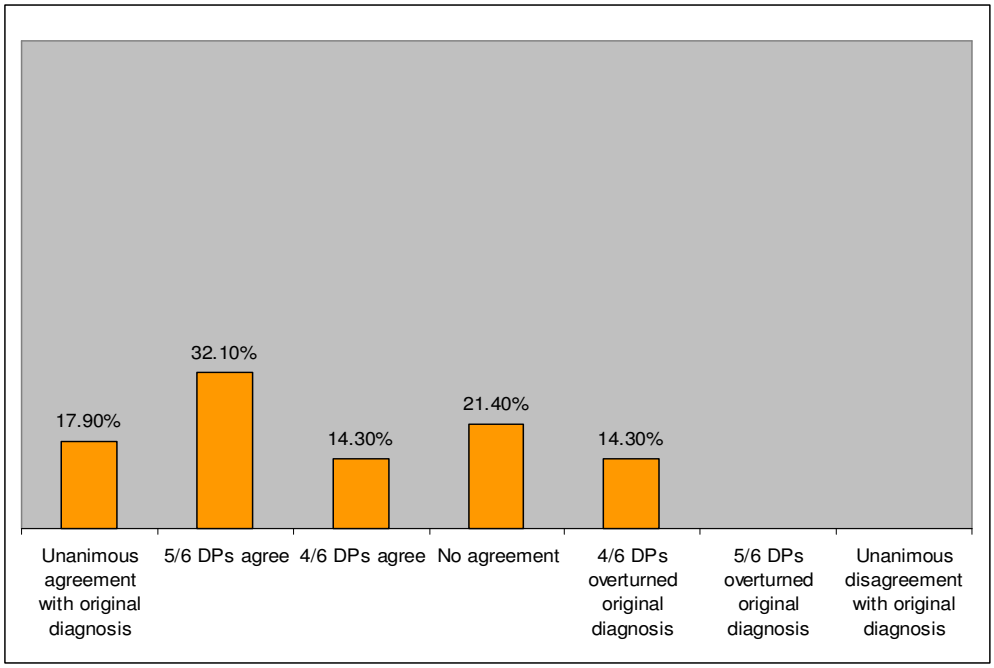

Figure 11. Concordance with original diagnosis among participating DPs: NOT Melanoma Slides Only

Of the 28 slides originally diagnosed "NOT Melanoma" participating DPs had varying levels of concordance with the original diagnosis. Only 1/6th of slides originally diagnosed as "NOT Melanoma" had unanimous (6 out of 6) agreement among participating DPs with the original diagnosis. 1/3 of slides originally diagnosed "NOT Melanoma" had one dissenting DP. 1/5 of slides originally diagnosed "NOT Melanoma" had no agreement whatsoever. In 14.3% of slides originally diagnosed "NOT melanoma" a majority of DPs (4/6) overturned the original diagnosis.

such in order to examine participants' responses as if their diagnoses were a "test" for detecting malignant melanoma. Taking the original diagnoses from 1988-1990 as the "true" diagnoses, participants' responses had a mean positive predictive value for detection of malignant melanoma of 0.58 (range 0.39- 0.69), a mean sensitivity of 0.93 (range 0.75- 1.0) and a mean specificity of 0.7 (range 0.5- 0.82).

DISCUSSION

The chief aim of this research was to determine whether diagnostic behavior by dermatopathologists in the evaluation of borderline melanocytic lesions has significantly changed over a 20 year period. The results of this experiment suggest that it has. This study has demonstrated three principal findings concerning a select group of dermatopathologists practicing at a major medical center: One, there is ample disagreement about the malignancy status of lesions originally diagnosed as benign in 1988-1990. Two, diagnosis of borderline lesions trended toward "malignant" for all study participants; in re-evaluation of a slide set that contained 12 original diagnoses of "malignant melanoma", the mean number of revised melanoma diagnoses by the 6 study participants was 19.7, an increase in 64% from the original number of melanoma diagnoses. Three, there are adequate levels of agreement about the malignancy status of lesions originally diagnosed as "malignant" in 1988-1990.

If dermatopathologists have changed their diagnostic habits, they may not have done so in a way that affects all diagnostic categories. It is reasonable to wonder whether changes in diagnostic habits, if they have truly occurred, have affected the diagnosis of "borderline" lesions while leaving melanocytic lesions with a lesser degree of atypia relatively unaffected. Moreover, unlike the transient rise in incidence of prostate cancer in the 1980s which was attributable to the rapid implementation of prostate specific antigen testing, artifactual changes in melanoma incidence such as would be seen in the case of changing subjective diagnostic habits would likely be more gradual and might not be observed in time-cohort model analysis.

Many authors have advanced the argument that some part of the rise in the incidence of melanoma can be attributed to artifactual causes such as increase in biopsy rate (15), changes in histopathologic criteria (8, 13, 14, 39), and the existence of a non-metastasizing form of thin melanoma. (9, 16, 17, 20, 79) As with these theories, it remains difficult to quantify the impact that changes in diagnostic behavior by dermatopathologists may have had on the apparent incidence rates of melanoma.

This study corroborates previous research showing that dermatopathologists achieve little to no consensus on the diagnosis of borderline melanocytic lesions. (49, 51-53, 55, 56, 68, 69) For the lesions whose diagnosis of "not melanoma" was overturned by the study participants, there was poor agreement in most cases, suggesting that if dermatopathologists have lowered their diagnostic threshold, they do not appear to have done so in a uniform manner.

While it could be argued that the trend toward malignant diagnosis revealed by this study reflects the fact that dermatopathologists have better learned to identify malignant melanoma, epidemiologic evidence does not support this; although the number of lesions being biopsied has increased over the past 20 years, there is little evidence that better histopathologic detection has resulted in better outcomes.

With regard to borderline melanocytic lesions, diagnostic criteria appear to go only so far toward enabling a dermatopathologist to render a final diagnosis. It seems plausible that two dermatopathologists may ascribe similar histopathologic descriptors to a given lesion

yet each render a different ultimate diagnosis; one may call the lesion "not malignant" while the other may call it "malignant", yet they both describe the lesion using similar technical language. As this study demonstrates, individual dermatopathologists do not necessarily agree with themselves, let alone with each other, when the malignant status of a lesion is concerned. Specific histopathologic diagnostic criteria for diagnosing CMM need not have changed for this to be the case. What has possibly changed is that the decision-making involved in the subjective final determination of the ultimate diagnosis of a melanocytic lesion, as motivated perhaps by fears of medico-legal liability (70), has slowly pushed lower the "borderline" that is the threshold for diagnosing malignancy.

Limitations

A generalization of the results of this study is limited by a number of factors:

First, the number of slides originally diagnosed as "malignant melanoma" comprised only 30% of the total slide set. The likelihood of detecting a change in both directions is therefore biased toward the set of dysplastic nevi trending to malignancy.

We purposefully selected "borderline" dysplastic nevi -- biopsies that represented the severest grade of atypia. Pathology reports for the dysplastic nevi used in this study often contained language that tempered the diagnosis with caveats and admonitions to perform "conservative re-excision". In some cases, the diagnosing pathologist, though rendering an ultimate diagnosis of dysplastic nevus, added notes that expressed equivocation about the diagnosis and sometimes wrote "melanoma cannot be ruled out". In contrast, none of

the pathology reports for the superficial spreading melanomas expressed any equivocation about the diagnosis. In trying to select "borderline" melanomas, no selection criterion other than "thinness" could be reasonably applied to the melanoma cases. Despite the broad debate about the "borderline" nature of severely dysplastic nevi, there is little discussion at large about "borderline" melanomas. This study was designed according to the assumption that invasive but thin melanomas would be the best examples of "borderline" lesions on the malignant side of the borderline.

Participants in this study received no clinical or demographic information about the patients whose biopsies were selected for review. Without the aid of information regarding the age of the patient, the body part involved and the clinical history of the patient, one could argue that participants were at a significant disadvantage compared to the original diagnosing dermatopathologist. It is true that pathologic diagnosis is often informed by clinical context and demographic data. However, many of the original pathology reports for the biopsies used in this study detailed personal or family history of malignant melanoma which, if disclosed, certainly could have skewed results in the opposite direction.

It could be argued that study participants might have taken a different attitude toward diagnosis of slides for a research study than they would toward diagnosis of slides in the course of their daily work. It is plausible that diagnoses rendered to direct a real patient's ultimate disposition might be generated with an eye toward legal and peer scrutiny – after all, a human life is on the line and the potential consequences of "false negative" errors

pose a formidable disincentive for less conservative behavior. Since the participants read the study slides with no medico-legal or clinical consequences, they could potentially afford to be less conservative than they might be in daily practice. In light of this, the true extent of melanoma re-diagnoses could be larger than the results disclose.

While the limitation placed on participating dermatopathologists in this study to render a dichotomous diagnosis made comparison between study participants relatively easy, it made comparison with the original diagnosis problematic. The fact that the same categories available for diagnosis at the time that the biopsy was originally read were not available during re-evaluation of the slides during the study makes the comparison of diagnoses one step removed from a direct comparison. The latitude normally allowed in wording diagnoses and the fine gradations of language used to characterize melanocytic lesions allow highly nuanced diagnoses that guide treatment more than the simple categories of “benign” and “malignant”. However, as an ultimate designation of “malignant” or “not malignant” is surely an important decision for any dermatopathologist, it is reasonable to assume that two diagnoses by the same dermatopathologist would not change much depending on how nuanced his/her answer was allowed to be.

This study examined the responses of a relatively small number of dermatopathologists. The results reported here are not necessarily representative of dermatopathologists in general.

Lastly, dermatopathologists normally have a larger sample of diagnostic material to review when evaluating cases. Whereas cases in the real world are usually represented by whole blocks of a dozen or more slides, this study limited each case to one representative slide.

Conclusion

The results of this study support the hypothesis that dermatopathologists are more likely to diagnose melanoma in biopsies of lesions that were diagnosed as borderline dysplastic nevi 20 years ago. It is far from clear whether this phenomenon may be extended to dermatopathologists in general and whether it could account for any of the apparent rise in the incidence of melanoma.

These findings also underscore previous research demonstrating that dermatopathologists are not consistently able to achieve consensus in diagnosis of borderline melanocytic lesions.

An expanded version of this study would help to further characterize and refine these conclusions. It would seem worthwhile to extend this study to dermatopathologists at different academic medical centers as well as to dermatopathologists in private practice.

Furthermore, the effect of time as a factor on the change in diagnostic behavior of dermatopathologists reading borderline lesions could perhaps be better understood if the study were repeated with the same participants but this time reading a second set of study

slides that were originally diagnosed within the past year. This would allow a more true comparison of general histopathologic diagnostic habits at two separate points in time.

REFERENCES

1. Balch CM. Cutaneous melanoma. Philadelphia: J.B. Lippincott, 1992.

2. Peterson NC, Bodenham DC, Lloyd OC. Malignant melanomas of the skin. A study of the origin, development, aetiology, spread, treatment, and prognosis. Br J Plast Surg 1962;15:97-116.

3. Lee JA, Carter AP. Secular trends in mortality from malignant melanoma. J Natl Cancer Inst 1970;45:91-7.

4. Rigel DS, Friedman RJ, Kopf AW. The incidence of malignant melanoma in the United States: issues as we approach the 21st century. J Am Acad Dermatol 1996;34:839-47.

5. Hall HI, Miller DR, Rogers JD, Bewerse B. Update on the incidence and mortality from melanoma in the United States. J Am Acad Dermatol 1999;40:35-42.

6. Cancer incidence in five continents. Volume VII. IARC Sci Publ 1997;(143):i,xxxiv, 1-1240.

7. Purdue MP, Freeman LE, Anderson WF, Tucker MA. Recent trends in incidence of cutaneous melanoma among US Caucasian young adults. J Invest Dermatol 2008;128:2905-8.

8. Beddingfield FC,3rd. The melanoma epidemic: res ipsa loquitur. Oncologist 2003;8:459-65.

9. Burton RC, Coates MS, Hersey P, et al. An analysis of a melanoma epidemic. Int J Cancer 1993;55:765-70.

10. Lamberg L. "Epidemic" of malignant melanoma: true increase or better detection? JAMA 2002;287:2201.

11. Rigel DS. Malignant melanoma: perspectives on incidence and its effects on awareness, diagnosis, and treatment. CA Cancer J Clin 1996;46:195-8.

12. Shuster S. Diagnoses of melanoma need further investigation. BMJ 1996;313:627-8.

13. Swerlick RA, Chen S. The melanoma epidemic: more apparent than real? Mayo Clin Proc 1997;72:559-64.

14. Swerlick RA, Chen S. The melanoma epidemic. Is increased surveillance the solution or the problem? Arch Dermatol 1996;132:881-4.

15. Welch HG, Woloshin S, Schwartz LM. Skin biopsy rates and incidence of melanoma: population based ecological study. BMJ 2005;331:481.

16. Burton RC, Armstrong BK. Non-metastasizing melanoma? J Surg Oncol 1998;67:73-6.

17. Dennis LK. Analysis of the melanoma epidemic, both apparent and real: data from the 1973 through 1994 surveillance, epidemiology, and end results program registry. Arch Dermatol 1999;135:275-80.

18. Hiatt RA, Fireman B. The possible effect of increased surveillance on the incidence of malignant melanoma. Prev Med 1986;15:652-60.

19. Seer Cancer Statistics Review 1975-2005. 2006;2008:25.

20. Lipsker D, Engel F, Cribier B, Velten M, Hedelin G. Trends in melanoma epidemiology suggest three different types of melanoma. Br J Dermatol 2007;157:338-43.

21. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2006. CA Cancer J Clin 2006;56:106-30.

22. Spiegel D. Cancer and depression. Br J Psychiatry Suppl 1996;(30):109-16.

23. Schofield PE, Butow PN, Thompson JF, Tattersall MH, Beeney LJ, Dunn SM. Psychological responses of patients receiving a diagnosis of cancer. Ann Oncol 2003;14:48-56.

24. Wilson D, Evans J. Non-medical implications of malignant melanoma. Br J Plast Surg 1993;46:158-9.

25. Del Mar CB, Green AC, Battistutta D. Patterns of excision and referral from primary care of melanocytic lesions. Melanoma Res 1997;7:496-9.

26. Diagnosis and treatment of early melanoma. NIH Consensus Development Conference. January 27-29, 1992. Consens Statement 1992;10:1-25.

27. Warycha MA, Zakrzewski J, Ni Q, et al. Meta-analysis of sentinel lymph node positivity in thin melanoma (</=1 mm). Cancer 2008.

28. Kretschmer L, Thoms KM, Peeters S, Haenssle H, Bertsch HP, Emmert S. Postoperative morbidity of lymph node excision for cutaneous melanoma-sentinel lymphonodectomy versus complete regional lymph node dissection. Melanoma Res 2008;18:16-21.

29. Barzilai DA, Cooper KD, Neuhauser D, Rimm AA, Cooper GS. Geographic and patient variation in receipt of surveillance procedures after local excision of cutaneous melanoma. J Invest Dermatol 2004;122:246-55.

30. Miller AJ, Mihm MC,Jr. Melanoma. N Engl J Med 2006;355:51-65.

31. Bykov VJ, Marcusson JA, Hemminki K. Protective effects of tanning on cutaneous DNA damage in situ. Dermatology 2001;202:22-6.

32. McGovern VJ. Epidemiological aspects of melanoma: a review. Pathology 1977;9:233-41.

33. Sturm RA. Skin colour and skin cancer - MC1R, the genetic link. Melanoma Res 2002;12:405-16.

34. Sturm RA, Duffy DL, Box NF, et al. The role of melanocortin-1 receptor polymorphism in skin cancer risk phenotypes. Pigment Cell Res 2003;16:266-72.

35. Naysmith L, Waterston K, Ha T, et al. Quantitative measures of the effect of the melanocortin 1 receptor on human pigmentary status. J Invest Dermatol 2004;122:423-8.

36. Valverde P, Healy E, Jackson I, Rees JL, Thody AJ. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 1995;11:328-30.

37. Reintgen DS, McCarty KM,Jr, Cox E, Seigler HF. Malignant melanoma in black American and white American populations. A comparative review. JAMA 1982;248:1856-9.

38. Cormier JN, Xing Y, Ding M, et al. Ethnic differences among patients with cutaneous melanoma. Arch Intern Med 2006;166:1907-14.

39. Berwick M, Wiggins C. The current epidemiology of cutaneous malignant melanoma. Front Biosci 2006;11:1244-54.

40. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 2004 Incidence and Mortality. 2007.

41. Tran TT, Schulman J, Fisher DE. UV and pigmentation: molecular mechanisms and social controversies. Pigment Cell Melanoma Res 2008;21:509-16.

42. Menzies SW. Is sun exposure a major cause of melanoma? Yes. BMJ 2008;337:a763.

43. Shuster S. Is sun exposure a major cause of melanoma? No. BMJ 2008;337:a764.

44. Bentham G, Aase A. Incidence of malignant melanoma of the skin in Norway, 1955-1989: associations with solar ultraviolet radiation, income and holidays abroad. Int J Epidemiol 1996;25:1132-8.

45. Agredano YZ, Chan JL, Kimball RC, Kimball AB. Accessibility to air travel correlates strongly with increasing melanoma incidence. Melanoma Res 2006;16:77-81.

46. Elwood JM, Jopson J. Melanoma and sun exposure: an overview of published studies. Int J Cancer 1997;73:198-203.

47. Roush GC, Schymura MJ, Holford TR. Patterns of invasive melanoma in the Connecticut Tumor Registry. Is the long-term increase real? Cancer 1988;61:2586-95.

48. Weinstock, M.A. (Personal Communication). 2008.

49. van der Esch EP, Muir CS, Nectoux J, et al. Temporal change in diagnostic criteria as a cause of the increase of malignant melanoma over time is unlikely. Int J Cancer 1991;47:483-9.

50. Garbe C, Leiter U. Melanoma epidemiology and trends. Clin Dermatol 2009;27:3-9.

51. Weinstock MA, Barnhill RL, Rhodes AR, Brodsky GL. Reliability of the histopathologic diagnosis of melanocytic dysplasia. The Dysplastic Nevus Panel. Arch Dermatol 1997;133:953-8.

52. Hastrup N, Clemmensen OJ, Spaun E, Sondergaard K. Dysplastic naevus: histological criteria and their inter-observer reproducibility. Histopathology 1994;24:503-9.

53. Duray PH, DerSimonian R, Barnhill R, et al. An analysis of interobserver recognition of the histopathologic features of dysplastic nevi from a mixed group of nevomelanocytic lesions. J Am Acad Dermatol 1992;27:741-9.

54. Philipp R, Hastings A, Briggs J, Sizer J. Are malignant melanoma time trends explained by changes in histopathological criteria for classifying pigmented skin lesions? J Epidemiol Community Health 1988;42:14-6.

55. Schmoeckel C. How consistent are dermatopathologists in reading early malignant melanomas and lesions "precursor" to them? An international survey. Am J Dermatopathol 1984;6 Suppl:13-24.

56. Boiko PE, Piepkorn MW. Reliability of skin biopsy pathology. J Am Board Fam Pract 1994;7:371-4.

57. Gong G, Whittemore AS, West D, Moore DH,2nd. Cutaneous melanoma at Lawrence Livermore National Laboratory: comparison with rates in two San Francisco bay area counties. Cancer Causes Control 1992;3:191-7.

58. LeBlanc WG, Vidal L, Kirsner RS, et al. Reported skin cancer screening of US adult workers. J Am Acad Dermatol 2008;59:55-63.

59. MacKie RM, Hole DJ. Incidence and thickness of primary tumours and survival of patients with cutaneous malignant melanoma in relation to socioeconomic status. BMJ 1996;312:1125-8.

60. Tripp JM, Kopf AW, Marghoob AA, Bart RS. Management of dysplastic nevi: a survey of fellows of the American Academy of Dermatology. J Am Acad Dermatol 2002;46:674-82.

61. Mihm MC,Jr, Barnhill RL, Sober AJ, Hernandez MH. Precursor lesions of melanoma: do they exist? Semin Surg Oncol 1992;8:358-65.

62. Sober AJ, Burstein JM. Precursors to skin cancer. Cancer 1995;75:645-50.

63. Langley RG, Sober AJ. Clinical recognition of melanoma and its precursors. Hematol Oncol Clin North Am 1998;12:699-715.

64. Rhodes AR, Sober AJ, Day CL, et al. The malignant potential of small congenital nevocellular nevi. An estimate of association based on a histologic study of 234 primary cutaneous melanomas. J Am Acad Dermatol 1982;6:230-41.

65. Barnhill RL, Piepkorn M, Busam KJ. Pathology of malignant melanoma. New York: Springer, 2004.

66. Larsen TE, Little JH, Orell SR, Prade M. International pathologists congruence survey on quantitation of malignant melanoma. Pathology 1980;12:245-53.

67. Duncan LM, Berwick M, Bruijn JA, Byers HR, Mihm MC, Barnhill RL. Histopathologic recognition and grading of dysplastic melanocytic nevi: an interobserver agreement study. J Invest Dermatol 1993;100:318S-21S.

68. Farmer ER, Gonin R, Hanna MP. Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists. Hum Pathol 1996;27:528-31.

69. Corona R, Mele A, Amini M, et al. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. J Clin Oncol 1996;14:1218-23.

70. Troxel DB. Pitfalls in the diagnosis of malignant melanoma: findings of a risk management panel study. Am J Surg Pathol 2003;27:1278-83.

71. Kornstein MJ, Byrne SP. The medicolegal aspect of error in pathology: a search of jury verdicts and settlements. Arch Pathol Lab Med 2007;131:615-8.

72. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960;XX:37-46.

73. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33:363-74.

74. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

75. Rietveld T, van Hout R. Statistical techniques for the study of language and language behaviour. Berlin ; New York: Mouton de Gruyter, 1993.

76. Shrout PE. Measurement reliability and agreement in psychiatry. Stat Methods Med Res 1998;7:301-17.

77. Krippendorff K. Content analysis : an introduction to its methodology. Thousand Oaks, Calif.: Sage, 2004.

78. Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis. Theory and practice. Arch Gen Psychiatry 1981;38:408-13.

79. Burton RC, Armstrong BK. Recent incidence trends imply a nonmetastasizing form of invasive melanoma. Melanoma Res 1994;4:107-13.