Anne Burnett Marion School of Medicine at TCU

Skin Cancer Detection Rates in Total Body Skin Examinations during First Half Versus Second Half of Shift; a Retrospective Review

Preliminary Thesis

Adam Stanley Rosencrans
12-16-2022

Abstract

Research Question: In licensed Dermatologists, what is the effect of hours worked prior to examination on Skin Cancer Detection Rate (CDR) during Total Body Skin Examinations (TBSEs), when measured as 1st vs 2nd half of work shift?

Background, Significance, and Rationale for the Question: A pivotal study in the field of gastroenterology suggested that higher rates of colon cancer are detected during colonoscopy when the procedure is performed in the morning vs the afternoon. It was theorized that this difference is due to progressive fatigue among the practicing physicians. The primary focus of my research will be to see if this same principle holds true in the field of dermatology in regard to skin exams and skin cancer detection.

Materials and Methods: Under the direction of my primary SPT mentor Dr. Catherine Harrell, 3,407 medical charts were reviewed for skin exam findings. Cases included all TBSEs and Above Waist exams but did not include "spot" checks or otherwise limited exams, as well as excluding exams performed on minors or other vulnerable populations. Data was collected on the total number of skin checks performed by each physician, the number of hours worked by the physician prior to the exam, and the number and description of biopsy-confirmed skin cancers found by these exams. Using a combination of descriptive statistics, Student's T-tests, Pearson's Chi-square tests, the data was analyzed for significant differences between exams performed in the 1st vs 2nd half of shift, including differences in demographics, cancer pathology, and cancer detection rate (CDR).

Results: 3,709 charts were reviewed, of which 1,718 were included in data analysis, of which 183 had positive skin cancer findings. When split into 1st vs 2nd half of shift, no significant difference was found in subject demographics, cancer descriptors, or CDR (p=0.157). However, a breakdown of 1st and 2nd half of shift by physician suggested variability between physicians, and a potential correlation between CDR and hours worked was subjectively seen, warranting further data collection and analysis.

Conclusions: We anticipated that first-half-of-shift skin exams would demonstrate higher rates of skin cancer detection when compared to second-half-of-shift appointments. Although this is only a preliminary analysis based on incomplete data collection, initial data suggests no difference in skin exam CDR between the first half of the physician's working day when compared to the second half, with the contradictory suggestion that one of the physicians may have an improved CDR in their 2nd half of shift. In addition, a visual trend emerged correlating CDR with number of procedures performed in that 1-hour time bucket, although statistical confirmation is pending access to advanced statistical software. While still preliminary, if these trends remain after final analysis, these results could suggest value in dermatologists analyzing their personal CDR trends for future practice structuring.

Research Question

Among licensed Dermatologists, is there an association between hours worked prior to examination and a reduction in Skin Cancer Detection Rate (CDR) during Total Body Skin Examinations (TBSEs), when measured as 1st vs 2nd half of work shift?

This study was centered around three licensed dermatologists at Dermatology Specialists of Fort Worth, with results generalizable to the larger population of practicing Dermatologists. The primary objective was to analyze CDR during skin examinations, with CDR defined as the percent of skin exams which result in a histologically confirmed diagnosis of skin cancer. The primary exposure was the number of hours worked prior to performing the exam, split into groups of first-half vs second-half of working shift. The desired outcome was determining any significant difference in CDR associated with number of hours worked for the purposes of practice structuring and patient safety.

Hypothesis: We hypothesized that TBSEs performed in the first half of the physician's working day would demonstrate increased skin cancer detection rates in comparison to TBSEs performed in the second half of the physician's working day.

Introduction, Significance and Rationale

Introduction

Skin cancers are among the most common cancers in the United States, with approximately 9,600 non-melanoma skin cancers and 250 melanomas diagnosed in the US every day[9–11]. Mortality varies widely by type of skin cancer, with yearly estimated deaths of 6,850 and 2,000 for melanoma and non-melanoma skin cancers respectively[11]. Some common risk factors for skin cancer include age, gender, and history of skin cancer. Risk of developing skin cancer increases linearly with age and is highest in older adults[6,7]. Cisgender women have been consistently shown to have significantly lower lifetime prevalence of skin cancer than cisgender men, with differences among gender non-conforming individuals a relatively new area of research[8]. A history of past diagnosis of skin cancer is linked with an increased risk of developing future skin cancer, with one study finding up to 17% of those with a past diagnosis eventually developed another cancerous lesion[5].

Total Body Skin Examinations (TBSE) screening for skin cancers are very commonly performed procedures by dermatologists, with 86% of dermatologists performing them either annually or every 2-3 years for their low-risk patients[12], as well as more frequently for their high-risk patients. TBSEs have been found to be both safe and well-tolerated among patients, with one study on the psychosocial impact of TBSE finding that visual skin examination did not worsen patient psychological well-being, and may have actually improved it[1]. TBSEs have also been shown to be effective at detecting skin cancer. Cancer Detection Rate (CDR) is defined as the number of TBSEs with findings positive for cancer on biopsy divided by the total number of TBSEs performed. CDR for TBSEs used for general screening purposes has been shown to be approximately 1% when applied to a population[3]. This rate increases dramatically for those who present to the dermatology clinic, with one study finding that CDR was 13.2% for those with a dermatology referral, and 7.7% for those without[4].

While some studies exist, there is not enough data at this time to recommend TBSE screening for the general population according to the United States Preventive Services Task Force (USPSTF). However, the data that does exist clearly demonstrates links between delayed TBSE and increased stage of cancer, as well as increased stage of cancer and increased morbidity and mortality[2]. For instance, melanoma survival rates are extraordinary if caught early, with 5-year survival rates of 98.4% and 17.9% for stage IA and stage IV disease respectively, and data suggests that regular TBSEs aid in early diagnosis[6].

Because of the importance of early diagnosis of skin cancer and especially Malignant Melanoma (MM), it is important to optimize sensitivity of TBSEs. A proxy for measurement of TBSE sensitivity is CDR, although this must take into account a number of confounding factors, such as cancer prevalence in the patient population, presence of risk factors among patient demographics, and both inter-rater and intra-rater reliability. A similar conversation is ongoing in multiple medical specialties, with one proposed factor impacting intra-rater reliability for cancer screening being physician fatigue.

Colonoscopies are another commonly performed cancer screening, with an estimated 14 million performed in 2013 by gastroenterologists[13]. A 2009 study by Sanaka et al. found that colonoscopies are

more effective at detecting cancer when performed in the morning versus in the afternoon, which they speculated may be due to physician fatigue or tiredness increasing as the day progresses[14]. This study led to much debate in the field of gastroenterology, with eight retrospective reviews, one prospective trial, and multiple meta-analyses published discussing the topic over the next 11 years, although a consensus has not yet been reached[15].

At this time, there are no such studies that currently exist examining the potential effect of physician fatigue on performance in the field of dermatology. With that in mind, this study was designed to provide a robust introduction to a discussion on physician fatigue and any potential link to skin cancer detection rates. We hypothesized that similar to colonoscopies, skin cancer detection rates during TBSEs performed by dermatologists would be significantly higher in the first half of a physician's shift when compared to the second half.

Significance

The publication of the Sanaka et al. study in 2009[14] had the implication of potentially changing practice flow for gastroenterologists, with the intention of providing better care to patients. Although no consensus has been reached on clinical significance of the results, the Sanaka study began a new discussion on physician fatigue and proxies for measurement. We hypothesize that this current study may have a similar impact for dermatologists. Significant results may influence dermatologists to change their clinic flow, scheduling more TBSEs in the first half of their shift, or shifting patients they expect to have more complex TBSEs to earlier appointment times. There is potential that this change could improve skin cancer detection rates with TBSEs if the data supports that conclusion. If no significant difference is found, we believe that the publication of this study may still lead to a discussion around physician fatigue and dermatologist performance that may produce future significant findings.

Rationale

This is a reasonable study due both to its likelihood of producing significant results and to a cost-benefit analysis. The costs of performing a retrospective chart review are low, and mostly involve IRB costs and the cost of time on the part of the researchers. The potential benefits are significant, as discussed in the above significance section.

Research Materials and Methods

MATERIALS

As this was a retrospective chart review, there was little needed in the way of materials. Access to Electronic Medical Records (EMRs), access to an encrypted and password-secured database, and access to statistical analysis software were the primary requirements for this study.

METHODS

In the interest of replicability, this study's design was largely modeled after the design of the original colonoscopy timing study performed by Sanaka et al. in 2009[14]. The study was reviewed by the Institutional Review Board at Texas Christian University, and it was granted exempt status on 5/3/2021. A report was pulled of 13,467 patient charts who had skin exams performed by either Dr. Harrell, Dr. Roberts, or Dr. Volkman at Dermatology Specialists of Fort Worth between 1/1/2019 and 12/31/2020. This initial report included patient name (Last, First), date of birth (MM/DD/YYYY), gender (M/F), appointment date (MM/DD/YYYY), appointment start time (XX:XX), appointment type, provider, and reason for visit. This report was generated in Microsoft Excel, where it served as the base for future data collection on subjects. Although alternative options for recording patient gender were discussed, no patient records were found in this sample to necessitate additional notation options.

Once this information from all 13,467 charts were generated as line items on the primary data collection sheet, each patient was assigned a random number between 0 and 1 up to 10 decimal places with no duplicates. The subjects were then re-ordered according to their randomized number. Each subject was then assigned a unique subject ID from 00001 to 13467 according to their position on the randomized list. This was done to minimize bias by randomizing which charts were reviewed first in the event that not all charts were completed prior to thesis submission, as is the case for this paper. At this time, only the first 3,709 charts have been reviewed, although this should represent a randomized sample of the overall 13,467 chart pool.

Regarding the 3,709 charts that have been reviewed at this time, the next step was applying Inclusion and Exclusion criteria (Table 1).

Table 1. **Inclusion and Exclusion criteria**

Category	Criterion
Inclusion Criteria	
	Date of exam: Between 1/1/2019 and 12/31/2020
	Location of exam: Dermatology Specialists of Fort Worth
	Exam provider: Dr. Harrell, Dr. Roberts, or Dr. Volkman
	Exam type: TBSE or above-the-waist exam
	Age: 18-100 at the time of the exam
Exclusion Criteria	
	Exam limited to one location or lesion
	Age <18 at time of exam
	Patient record not able to be accessed

After applying Inclusion and Exclusion criteria, 1,718 charts were included in the study. Of the 1,991 charts not included, 1,165 were excluded due to a limited exam (not TBSE or above-the-waist), 790 were excluded due to being minors at the time of exam, and 36 were excluded due to difficulty accessing patient records. Of note, a TBSE was defined as a visual skin exam involving the hands, feet, extremities, trunk, neck, face, scalp, eyes, and mouth, while an Above Waist exam included the trunk, upper extremities, hands, neck, face, scalp, eyes, and mouth. Genital region exam was not considered a necessary component of a complete TBSE or Above Waist exam. We recorded the above information as the variable “Meets inclusion criteria” with a value of yes or no (y, n).

Once the 1,718 charts which met inclusion criteria were identified, the next step was finding further information from the chart report that was generated. In order to determine the patient’s age at the time of the exam, we found the difference between their appointment date and their date of birth. We then rounded this age down to the nearest integer and recorded the variable “age” in years (XX). Next, for each patient, we recorded what time the physician had begun working that day as the variable “work start time” (XX:XX on a 24-hour clock). To determine work start time, we sorted the data sheet by appointment time. Knowing that the three physicians began work at 07:30 for the vast majority of days, we only needed to change chart entries for those days in which an appointment was recorded at 07:00 or 07:15, with 07:00 overriding 07:15 on any day that included appointments at both times. Overall, only 19 days had work start times of 07:15, and only 3 had work start times of 07:00. The next calculated variable was hours worked by dermatologist prior to procedure start, or “hours worked”, recorded as (X.XX to 2 decimal places). As appointments were scheduled on the quarter hour, this variable only changed in quarter hour increments.

To calculate hours worked, we subtracted the appointment start time from the work start time for each subject. While this does not take into account any delay between the start of the appointment and the start of the skin exam, we worked under the assumption that this would not be significantly different with such small variations in time in such a large sample size. Finally, we recorded the variable "half of shift" as either 1st half or 2nd half (1, 2). There was debate on how to best measure this, and this is an area of ongoing discussion between the research team before a final paper is produced. Of note, we do not have access to what time the providers stopped work each day. For the purposes of this preliminary report, we determined first half of shift to be any exam which had a recorded hours worked of less than or equal to 4 hours, with any exam with an hours worked greater than 4 considered 2nd half of shift. This is a simplified method which takes into account the length of the average work day and is unaffected by work start time. Additional methods that will be discussed before a final analysis is performed include a cutoff of noon (similar to the original Sanaka colonoscopy study[14]), a cutoff of the median hours worked (approximately 3.25 which would give 2 roughly equal groups, although the 2nd group would have much more variation in hours worked), or the creation of an excel formula to calculate an approximate "work end" variable for each day so that each subject's half of shift would differ depending on the physician's day. This could theoretically be done by sorting the excel sheet by appointment day and using IFS statements to search for the provider for that subject, search for the latest appointment the provider had scheduled that day, and adding 15 minutes to the start time of their last appointment. The drawback to this method is that 1st and 2nd half of shifts measured in this way would not take into account the effect of a shorter or longer work day, as well as presenting additional work in creating and verifying the necessary programming.

Once all data of interest was calculated from the chart reports, the next step was to individually search each chart for additional information of interest. The first variable recorded was whether a subject had a history of skin cancer, recorded as "Hx of cancer" (S-skin cancer, N-none, O-other). To determine skin cancer history, we looked for any reference to a history of skin cancer present in the patient's chart on the day of the appointment. We made no distinction made between cancer pathology: basal cell carcinoma, squamous cell carcinoma, melanoma, both in situ and invasive were all recorded as a positive skin cancer history. We did not consider pre-cancerous lesions to be positive. Of note, the data sheet was set up to record other cancer history beyond skin cancer, although this was omitted in the data collection. The next variable recorded was whether or not the TBSE or Above Waist exam detected cancer, recorded as "cancer found" with a value of yes or no (y, n). This was determined by histologically confirmed cancerous lesion found from a biopsy that was either taken or ordered on the day of the exam. As above, any cancerous pathology was determined to be a positive finding, and we did not record benign or pre-cancerous findings.

For only those patients with cancer found on exam, we then recorded descriptive variables of their cancer. The first of these variables was number of cancerous lesions, recorded as "lesion #" with an integer value depending on the number of positive biopsies taken due to that exam, according to the previous criteria for cancer found (1, 2, 3, etc.). Next we recorded "lesion location" as recorded in the medical chart (H-head, N-neck, T-trunk, U-upper extremities, L-lower extremities, V-various). For those patients who had multiple positive lesions on different locations, V for various was used. In future analysis, it may be worth splitting these patients up so that their lesion locations are reflected by the various groups above, although care must be taken to not bias any data through this process. The next

recorded variable was lesion pathology, determined by the pathology listed in the biopsy result (B-basal cell carcinoma, S-squamous cell carcinoma, M-melanoma, V-various, O-other). For patients with multiple cancerous lesions, V for various was used again, with the same caveat for future analysis. Of the 1,718 charts included, only 3 findings required the O category for other, these being cutaneous T-cell lymphoma, keratoacanthoma, and sebaceous carcinoma.

The EMR used was Dermatology Specialists of Fort Worth's Modmed Dermatology EMR. The process for gathering the data in the above paragraphs for each subject is outlined below.

1. Navigate to the patient search function by clicking "patients" -> "advanced patient search"
2. Copy the patient name from the chart and paste it into the name section, search
3. If multiple patients with the same name exist, verify using birthday
4. Once in the chart, click on "visit summaries", and find the summary matching the appointment date in the data sheet.
5. Confirm the exam performed on that date meets inclusion criteria. Chart will either say " Exam - Full Skin Detailed - Appearance, Orientation, Mood" or "Exam - Above the Waist Detailed - Appearance, Orientation, Mood", with any other exam not meeting criteria.
6. If meeting criteria, look under diagnoses for that visit date of history of skin cancer. Any variation of "history of basal cell carcinoma", "history of squamous cell carcinoma", or "history of malignant melanoma" will count as a positive history, using clinical judgement for variations
7. Navigate to the "Path/Labs" tab to determine whether cancer was found. To count, a biopsy result must have the "cancerous" flag, and the listed date under "visit" must match the appointment date from the data sheet.
8. If cancer was found, continue to 9, otherwise go back to 1. for the next subject
9. Record lesion # by counting how many cancerous biopsies are linked to the correct visit date
10. Record location by sorting into above buckets using location listed under biopsy result.
11. Record pathology using pathology result, using clinical judgement when necessary to include lesion under correct pathology bucket
12. Go to 1. for the next subject

STATISTICAL METHODS

While the plan for any final paper involves the use of advanced analytical tools such as Stata, at the time of this preliminary paper the research team only had access to Microsoft Excel, which limited access to some statistical tests. The first step of analysis involved descriptive statistics. All 13,467 charts were analyzed for whether they were included in any future data analysis, separated by either "y", "n", or blank under the value "meets inclusion criteria", with blank indicating a chart that had not been reviewed at the time of this paper. Frequencies were also found for each of the reasons a subject could be excluded from the study.

Once that was completed, all future data analysis focused solely on those subjects which were included in the study. Of those 1,718 charts, frequencies were determined for all categorical variables. Those variables included "Hx of cancer" (y/n), "Cancer found" (y/n), "Lesion #" (0/1/2/3/4), "Lesion location" (h/n/t/u/l/v), "Lesion Pathology" (b/s/m/o/v), "Gender" (m/f), "Work start time" (hrs), "Shift half" (1/2), "Dermatologist" (Harrell/Roberts/Volkman), and "Age" (yrs). In this study, age was analyzed primarily as a categorical variable, as ages were rounded down to the nearest integer and then sorted

into buckets. The buckets chosen were 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+. These specific buckets were chosen as they reflected those used in previous related studies examining CDR in TBSEs, allowing for greater comparability and replicability[15]. Additionally, mean, median, mode, and quartiles were determined for continuous variables. Both Age and Hours worked were analyzed in this way, although there is an interesting discussion on whether they should solely be treated as categorical variables. While time is traditionally seen as a continuous variable, Time worked was based off appointment times, which were scheduled on the quarter hour. Due to this, Time worked increased in increments of 0.25 hours. While this alone would not make time a categorical variable, an argument could be made that the skin exam procedure itself did not start exactly at the appointment start time, and that by rounding down the time of the procedure start to the nearest quarter hour in this way, we are treating time as a categorical variable. This same argument may also be applied to Age, as discussed above.

The next statistic calculated was Cancer Detection Rate (CDR) for a number of different independent variables. The overall CDR for all included subjects was calculated using the formula CDRov=[All skin exams with positive findings/(All cancer positive exams+All cancer negative exams)]*100. This same formula was then used for calculating CDR and different variables by limiting "all exams" to "only those exams which meet the criteria". Independent variables for which CDR was calculated was CDR for those with and without a history of skin cancer, CDR for male vs female subjects, CDR breakdown by the above-listed 10-year age buckets, CDR for each physician, and CDR for both first and second half of shift, both overall and per physician. Because CDR is calculated using the binary categorical variable of Cancer found, it is tempting to perform a Pearson's Chi-square test on CDR per half-of-shift, as this was our primary endpoint. However, it would be inappropriate to report a significance at this time without further breakdown of the demographics of both groups, as well as using a multivariable regression analysis to determine the effects the other variables may have on CDR.

To that end, the next analysis performed was on demographic differences between subjects in the 1st and 2nd half of shift groups. To do this, all 1,718 charts were separated into 2 groups depending on the variable Half of shift. For each independent variable, a count was preformed for frequency, standard deviations were reported for continuous variables, percentage of total qualifying subjects was reported for categorical variables, and p-values were reported using the appropriate statistical test. When testing for differences in age between 1st and 2nd half of shift subjects, age was treated as a continuous variable as opposed to using age buckets as a categorical variable. Because of this, a Student's T-test was used. Because it was checking for differences in either direction, it was chosen to be a 2-tailed test. It was further chosen to be unpaired due to no pairing between subjects, and with the assumption of equal variance between groups (confirmed through a variance analysis). Testing for all other demographic variables involved Pearson's Chi-square tests for independence, as we were examining 2 categorical variables for significant differences between the groups. This involved the creation of actual and expected value tables in excel. To determine expected values, the row total was multiplied by the column total, which was divided by the overall total. A Chisq.test analysis was then run in excel, which automatically takes into account appropriate degrees of freedom. In this way, p-values were determined for differences between 1st and 2nd half of shift groups in terms of gender (1 degree of freedom), history of skin cancer (1 degree of freedom), and CDR% (1 degree of freedom).

The next analysis was performed only on those 183 subjects which had cancer positive exams, further separated into 2 groups by Half of shift. Of these subjects with cancerous findings, a similar analysis was performed to look for differences between 1st and 2nd half of shift in regards to lesion location, pathology, and multiplicity (1 lesion vs 2+ lesions). For each of these independent variables, frequency, percentage of total, and a p-value was determined. While each of these are categorical variables, special attention had to be paid to counts of subgroups. While the Pearson's Chi Square test for Independence is the preferred test in this setting, it is inappropriate to use when greater than 20% of counts have a value of less than 5. Because the breakdown of lesion pathology by Half of shift had 40% of its values less than 5, it would be more appropriate to use a Fisher's Exact Test to analyze this variable. Both lesion location and multiplicity were able to be analyzed using a Pearson's Chi Square test for Independence.

As of the writing of this preliminary paper, the research team did not have access to advanced statistical analysis software. As such, all tests were performed in Microsoft Excel, and some were unable to be performed. Planned analysis for the final version of this paper includes using a Fisher's Exact test on lesion pathology by half of shift, using a multivariable logistic regression on variables impacting CDR, and using a Cochran-Armitage test for trend on CDR by hours worked.

Resources

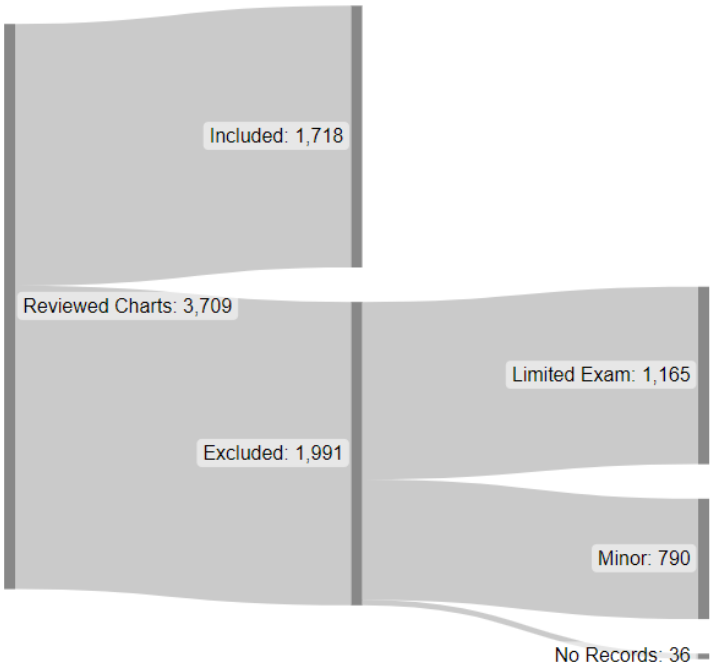
This research study was conducted under the supervision and guidance of my mentor, Dr. Catherine Harrell, who is a dermatologist with Dermatology Specialists of Fort Worth. The personal laptops of Dr. Harrell and myself were used to access TCU Box for data collection, storage, and analysis. All off-site visualization, entry, and analysis of data occurred privately within my residence at 701 E Bluff St, Fort Worth, TX 76102.

Results

All Charts

13,467 charts were pulled from the EMR for first-pass analysis. Of these 13,467 charts, 3,709 were reviewed for this preliminary analysis. 1,991 of these charts were determined not to meet inclusion criteria. 1,165 were excluded due to a limited physical examination not meeting the definition of either TBSE or Above Waist exam, 790 were excluded due to the subject being a minor at the time of the exam, and 36 were excluded due to inadequate access to patient records (see Figure 1). 9,758 charts have not yet been reviewed for inclusion, with the plan to review and analyze them prior to any publication.

Figure 1. Inclusion and Exclusion Sankey Diagram



https://sankeymatic.com/build/

Sample-wide data

For those 1,718 individual subjects and exams which were included in the study, 654 were male (38.1%), and 734 had a history of skin cancer (42.7%). For each of the subjects, 1,682 saw their dermatologist on a day they had started work at 7:30AM (97.9%), with only 30 at 7:15AM (1.7%) and 6 at 7:00AM (0.3%). For 1,059, their dermatologist had worked less than or equal to 4 hours by the time of their exam (61.6%), with 659 in the second half-of-shift group. Physician 1 saw 713 of the patients (41.5%), physician 2 saw 385 (22.4%), and physician 3 saw 620 (36.1%). Cancer was found after 183 exams (10.7%). For those who had cancer positive exam findings, 155 had only a single lesion (84.7%), while 22 had two lesions, 6 had three lesions, and 1 had four lesions. 56 lesions were located on the head (30.6%), 10 on the neck (5.5%), 39 on the trunk (21.3%), 46 on the upper extremities (25.1%), 14 on the lower extremities (7.7%), and 18 were located in multiple locations (9.8%). 107 lesions were

determined to be basal cell carcinoma (58.5%), 53 squamous cell carcinoma (29.0%), 12 melanoma (6.6%), 8 had multiple pathologies (4.4%), and 3 were categorized as other (1.6%), those being cutaneous T-cell lymphoma, keratoacanthoma, and sebaceous carcinoma. The average age was 58.7, 61, and 66 for mean, median and mode respectively. The minimum age was 18 and the maximum 95, with a 25th percentile of 48 and a 75th percentile of 70 years.

Cancer Detection Rate

Using the formulae described in the Methods section, the overall CDR for skin exams was calculated to be 10.7%. CDR in those subjects with a history of skin cancer was 16.6% compared with 6.2% for those without. CDR was 14.2% for male subjects compared to 8.5% for females, and it increased incrementally with age, with a CDR of 1.0%, 1.6%, 5.3%, 10.6%, 12.8%, 14.9%, and 18.9% for the age buckets of 18-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+ respectively. The CDR breakdown by physician was 10.0%, 14.3%, and 9.2% for physicians 1, 2, and 3 respectively. Overall CDR in the first half of shift was 9.8% compared to 12.0% in the second half of shift. When CDR is broken down both by half of shift and individual physician, a Pearson's Chi-Square analysis reveals significant data, although this will not be reported on in this preliminary paper due to a more limited sample size and the fact that a multivariable logistic regression analysis would be more appropriate in this setting than a chi-square test. For that reason, further reporting on the p-value of CDR in relation to different variables is pending final analysis.

Comparison of First vs Second half of shift

A summary of the demographic comparison between the 1st and 2nd half of shift groups can be seen in Table 2. Of note, no significant differences were found between either group, including similar age distributions (p=0.385), genders (p=0.365), histories of skin cancer (p=0.578), and CDRs (p=0.157). Analysis was further performed to compare skin cancer characteristics between the 1st and 2nd half groups (Table 3). Again, no values were found to differ significantly between the groups, including lesion location (p=0.173), pathology (p=0.562*), and multiplicity (p=0.428). Of note, the p-value for lesion pathology was determined using an online Fisher's Exact calculator and should be confirmed with advanced statistics software (chi-square analysis gave a p-value of 0.497, although this is an inappropriate use case). Finally, CDR by hours worked is plotted in Figure 2, with number of exams performed as a secondary y axis. While the figure is suggestive of a trend (other than the clear outlier of exams past 8 hours of work), a p-value will not be able to be determined until a Cochran-Armitage test for trend is able to be run using advanced analytical software.

Table 2. **Demographics of 1st vs 2nd half of shift**

Factor	1st Half (*n*=1,059)	2nd Half (*n*=659)	*P* value
Age, mean (yrs) (s.d.)	58.4 (16.1)	59.1 (16.9)	0.385
Gender			0.365
Male	412 (38.9%)	242 (36.7%)	
Female	647 (61.1%)	417 (63.3%)	
Hx of Skin Cancer			0.578
Yes	458 (43.2%)	276 (41.9%)	
No	601 (56.8%)	383 (58.1%)	
CDR %	104 (9.8%)	79 (12.0%)	0.157

CDR, cancer detection rate
P value calculated by 2 tailed, unpaired, equal variance t-test for age,
chi-square test of independence otherwise

Figure 2. **CDR by Hours Worked**

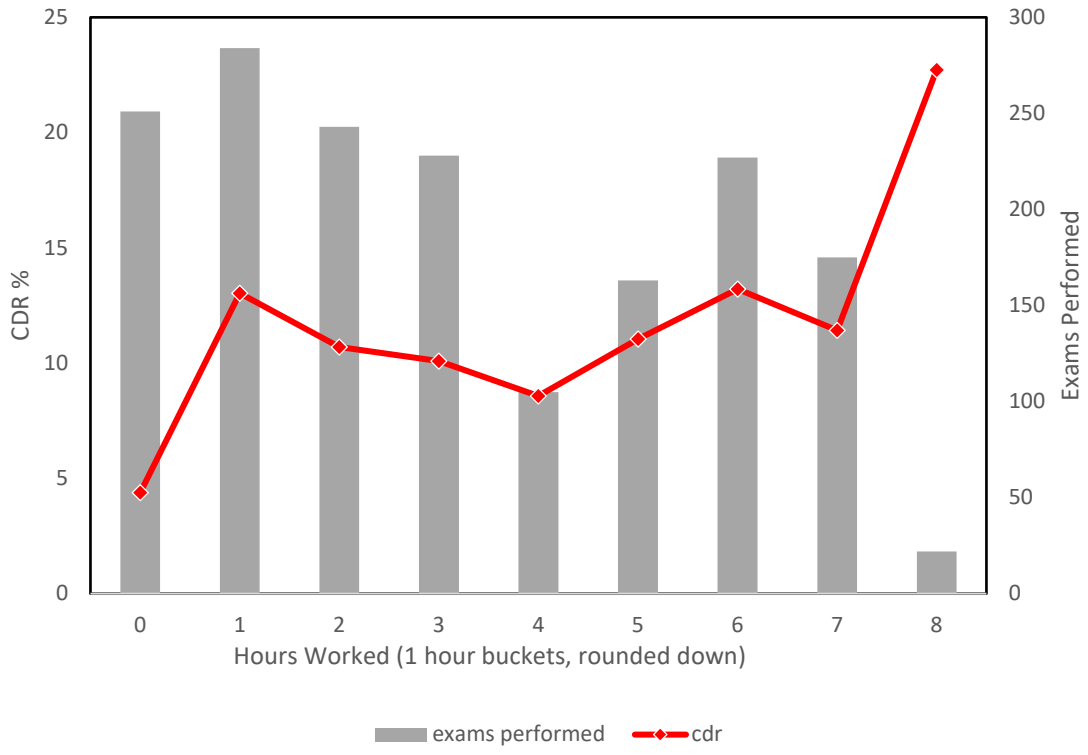


Table 3. **Lesion Characteristics of 1st vs 2nd half of shift**

Factor	1st Half (*n*=1,059)	2nd Half (*n*=659)	*P* value
Location			0.173
Head	34 (32.7%)	22 (27.8%)	
Neck	7 (6.7%)	3 (3.8%)	
Trunk	23 (22.1%)	16 (20.3%)	
Upper Ext.	22 (21.2%)	24 (30.4%)	
Lower Ext.	11 (10.6%)	3 (3.8%)	
Multiple	7 (6.7%)	11 (13.9%)	
Pathology			0.562
BCC	63 (60.6%)	44 (55.7%)	
SCC	27 (26.0%)	26 (32.9%)	
Melanoma	7 (6.7%)	5 (6.3%)	
Other	3 (2.9%)	0 (0.0%)	
Multiple	4 (3.8%)	4 (5.1%)	
Multiplicity			0.428
1	90 (86.5%)	65 (82.3%)	
2+	14 (13.5%)	14 (17.7%)	

BCC, Basal Cell Carcinoma; SCC, Squamous Cell Carcinoma; Ext, Extremity
P value calculated by Fisher's Exact Test for pathology, chi-square test of independence otherwise

Discussion and Innovation

Discussion

Discussion and interpretation of preliminary results is limited by the plan for future data collection and analysis, as well as the plan for advanced testing through statistical software. For instance, we would expect to see the presence of skin cancer risk factors increase CDR, as CDR could be elevated simply through having a higher prevalence of skin cancer in your patient population or sample. Indeed, our data appears to reflect this, as male sex, increased age, and a history of skin cancer all appear to be associated with a higher CDR in our dataset. However, with a *p*-value of <0.05, running multiple independent tests for association on a single dataset and variable is increasingly likely to produce a false-positive result due to chance. Instead, final discussion on the variables that impact CDR should be held until data collection is complete and a multivariable logistic regression is run. Other examples of such suggested but unconfirmed data include a general trend toward Physician 1 having a higher CDR in 1st half of shift while Physician 2 and 3 have a higher CDR in 2nd half, as well as the visually apparent potential positive association between CDR and number of exams performed during a specific timeframe.

However, there are a few statements that can be made given the current state of the data. First, there appears to be no significant difference between the 1st and 2nd half of shift groups, either demographically or in the quality and characteristics of cancer findings. Second, our overall CDR was calculated at 10.7%, which is similar to previous reported averages in the dermatology clinic setting[4]. Our pathology findings were also similar to epidemiological estimates, with 6.6% being melanoma compared to estimates of approximately 5% of all skin cancers[16].

Our preliminary dataset remains promising however, with many intriguing trends between CDR and hours worked awaiting further data collection and analysis. To-date, while many studies exist which explore a potential link between physician fatigue and performance in specialties such as gastroenterology, anesthesia, and surgery, no such study exists in the field of dermatology. While no study exploring a link between fatigue and performance has been found to be definitive and the debate is ongoing, the prospect of being the first such study for skin cancer detection presents an exciting opportunity to introduce a long-running medical debate to a new field.

This study has a number of strengths in its design. First, it has the benefit of similar previous highly regarded studies in different specialties, which the study design was intentionally modeled after to preserve replicability. Second, we have the benefit of a massive dataset, as 13,467 charts represents a larger sample size than other similar studies, increasing the study's power. On a related note, in addition to a large sample size, our samples came from only 3 physicians, greatly increasing intra-rater reliability.

It is important also to acknowledge several limitations with our study. First, with data only coming from one location, there is a greater chance that local patient demographics are not representative of the population, and generalizability is reduced. Second, as data comes from only 3 physicians, there is diminished inter-rater reliability. Third, no record was made of patients' ethnicities, socioeconomic status, benign biopsy results, length of procedure times, or whether the exam performed was a TBSE or Above-Waist exam. While all of these could potentially skew the data, the assumption we were working under was that such a large data set would diminish the effect of any individual differences between patients. Next, much conversation could be had about the method of deciding a cutoff for 1st and 2nd half of shift (currently less than or greater than 4 hours worked), as this is somewhat arbitrary; this cutoff will either remain as-is or be decided before the final paper by an unbiased party who has not seen the preliminary analysis present in this paper. As it is, we unfortunately don't currently have access to the number of hours worked by each physician on each day, leaving the terminology of "first half" and "second half" as somewhat inaccurate. Finally, as with many studies conducted during this time, this study's timeframe includes both before and after the onset of the SARS-CoV-2 coronavirus pandemic. While no current analysis is planned for differences in the dataset before and after March of 2020, this may be an interesting avenue for further research and may present a significant confounding variable to the generalizability of the study.

Innovation

As far as innovations, a common limitation noted by several previous studies was the cutoff of "morning" vs "afternoon" as a proxy for physician fatigue. Notably, the morning vs afternoon cutoff does not account for when a physician begins their day, as a physician who begins work at 5 am would be presumed to be more fatigued by 10am than one who begins their workday at 9am. Furthermore, this cutoff does not account for half-days, as presumably a physician who works only a half-day in the afternoon is less fatigued during the afternoon than one who begins their workday in the morning. One innovation present in our study is the shift of the conversation from a morning and afternoon binary to a first and second half of shift binary. This has the benefit of being agnostic to when a physician begins their workday; if the working hypothesis is that additional hours of work increase physician fatigue, it should not matter the time of day that these work hours occur. However, this new method introduces new limitations. First, splitting shifts directly in half would run into the same problem as the previous method; a 4-hour workday is split into two shifts of two hours, presumably skewing fatigue levels in both halves. Additionally, it introduces the need for additional data collection, potentially increasing time demands or decreasing power of future studies. Finally, both methods of differentiation run into the problem of accounting for a midday lunch break. Whether separated by noon or by first and second half of shift, a lunchbreak falls near the end of one grouping and the start of the next grouping, introducing a new set of confounding variables (distraction or low-blood sugar before lunch vs post-prandial fatigue). While discussing physician fatigue with time as a continuous variable solves many of these problems, doing so does pose the question of whether any significant data would have any clinical relevance for updated best-practices.

Future Directions

As this is the first study of its kind in the field of dermatology, further studies are needed to support or repudiate the claims made in this paper. Ideally, future studies would sample a larger number of dermatologists with greater geographical diversity in patient population, as well as collect data on additional variables such as patient ethnicity, economic or educational background, tanning bed use, or other known risk factors for skin cancer. A well-designed prospective study examining not just CDR but also self-reported physician fatigue values prior to each exam would be ideal for reducing assumptions being made around physician fatigue, although it would be difficult to blind such a study and it would be at risk for introducing observer bias.

There are many additional avenues for future research. By collecting information on benign biopsy results, a researcher could look into not just skin exam CDR but also biopsy "accuracy", or the rate of biopsies leading to positive results. Additionally, an intervention-based study could prove a useful direction for future research. An individual or group of providers could structure their practice to some variation of all TBSEs occurring in the morning or afternoon and record the effect the intervention had on their CDRs.

Conclusions

No major conclusions are being drawn from this current paper, as it is a preliminary analysis of the dataset and there are future statistical analyses which must be run. Conclusions which can be reached at this time include that the overall calculated CDR is in line with national averages, that there are no significant differences in our 1st and 2nd half of shift groups in terms of demographics or cancer characteristics, and that there are a number of promising areas for future analysis of the data. If current trends are found to be significant on final analysis (namely that of a correlation between CDR and hours worked, especially in association with number of exams performed), this could have several real-world implications. This correlation would suggest that physician fatigue or some other variable is having a measurable impact on patient care, with implications for increased morbidity and mortality for those patients scheduled for exams at suboptimal appointment times. Knowing this, a practice, hospital, or even national association could create new guidelines around practice structure and scheduling to ensure high-quality patient care.

Whether or not these findings turn out to be significant, this paper plays an important role in introducing the ongoing debate surrounding physician fatigue to the specialty of dermatology. Within 11 years of the Sanaka et al. colonoscopy study, 9 similar studies were published debating the link between time of day and colonoscopy adenoma detection rate, not including reviews and meta-analyses on the same topic[15]. This study may serve as a similar catalyst for future discussion around fatigue and clinic structure in the dermatology setting.

Compliance

On 3/1/21, the project was submitted to the Texas Christian University IRB. On 3/10/21, the IRB requested clarifications which were promptly provided. On 5/3/21, the TCU IRB notified the team that the project was determined to be considered minimal risk, and qualified for exemption from further IRB oversight under 45 CFR 46.104(d)(4)(ii) of the TCU IRB protocol. A waiver of HIPAA authorization and informed consent were also provided. No research activities were completed until after this confirmation was received from the IRB. The steps for data safety and management detailed in the above methods section were still taken according to protocol. No compliance issues were noted throughout the project. All researchers maintained active research certifications through CITI.

References

1. Risica PM, Matthews NH, Dionne L, et al. Psychosocial consequences of skin cancer screening. *Prev Med Rep*. 2018;10:310-316. doi:10.1016/j.pmedr.2018.04.011

2. Wernli KJ, Henrikson NB, Morrison CC, Nguyen M, Pocobelli G, Whitlock EP. *Screening for Skin Cancer in Adults: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force*. Agency for Healthcare Research and Quality (US); 2016. Accessed December 16, 2022. http://www.ncbi.nlm.nih.gov/books/NBK379854/

3. Waldmann A, Nolte S, Geller AC, et al. Frequency of Excisions and Yields of Malignant Skin Tumors in a Population-Based Screening Intervention of 360 288 Whole-Body Examinations. *Arch Dermatol*. 2012;148(8):903-910. doi:10.1001/archdermatol.2012.893

4. Mylle S, Verhaeghe E, Van Coile L, Van de Maele B, Hoorens I, Brochez L. Lesion-directed screening to optimize skin cancer detection in dermatology practice: an observational study. *J Eur Acad Dermatol Venereol JEADV*. 2021;35(6):1309-1314. doi:10.1111/jdv.17129

5. Duarte AF, Sousa-Pinto B, Haneke E, Correia O. Risk factors for development of new skin neoplasms in patients with past history of skin cancer: A survival analysis. *Sci Rep*. 2018;8:15744. doi:10.1038/s41598-018-33763-7

6. Johnson MM, Leachman SA, Aspinwall LG, et al. Skin cancer screening: recommendations for data-driven screening guidelines and a review of the US Preventive Services Task Force controversy. *Melanoma Manag*. 2017;4(1):13-37. doi:10.2217/mmt-2016-0022

7. Sherban A, Waseh S, Hugo A, Bui M, Daskalakis C, Jones E. Skin cancer biopsy and detection rates with total body skin examination: A cross-sectional retrospective analysis. *J Am Acad Dermatol*. 2022;86(4):929-931. doi:10.1016/j.jaad.2021.03.061

8. Singer S, Tkachenko E, Hartman RI, Mostaghimi A. Gender Identity and Lifetime Prevalence of Skin Cancer in the United States. *JAMA Dermatol*. 2020;156(4):458-460. doi:10.1001/jamadermatol.2019.4197

9. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the U.S. Population, 2012. *JAMA Dermatol*. 2015;151(10):1081-1086. doi:10.1001/jamadermatol.2015.1187

10. Rogers HW, Weinstock MA, Harris AR, et al. Incidence Estimate of Nonmelanoma Skin Cancer in the United States, 2006. *Arch Dermatol*. 2010;146(3):283-287. doi:10.1001/archdermatol.2010.19

11. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69(1):7-34. doi:10.3322/caac.21551

12. Stevenson ML, Glazer AM, Cohen DE, Rigel DS, Rieder EA. Frequency of total body skin examinations among US dermatologists. *J Am Acad Dermatol*. 2017;76(2):343-344. doi:10.1016/j.jaad.2016.09.017

13. Joseph DA, Meester RGS, Zauber AG, et al. Colorectal Cancer Screening: Estimated Future Colonoscopy Need and Current Volume and Capacity. *Cancer*. 2016;122(16):2479-2486. doi:10.1002/cncr.30070

14. Sanaka MR, Deepinder F, Thota PN, Lopez R, Burke CA. Adenomas are detected more often in morning than in afternoon colonoscopy. *Am J Gastroenterol*. 2009;104(7):1659-1664; quiz 1665. doi:10.1038/ajg.2009.249

15. Kim SH, Kim JH. When should we perform colonoscopy to increase the adenoma detection rate? *World J Gastrointest Endosc*. 2021;13(12):619-627. doi:10.4253/wjge.v13.i12.619

16. Dimitriou F, Krattinger R, Ramelyte E, et al. The World of Melanoma: Epidemiologic, Genetic, and Anatomic Differences of Melanoma Across the Globe. *Curr Oncol Rep*. 2018;20(11):87. doi:10.1007/s11912-018-0732-8