THE UNIVERSITY OF CHICAGO

CANCER ANALYSIS WITHOUT BORDER: RISK, CLASSIFICATION, AND PREVENTION OF ARSENIC-INDUCED NONMELANOMA SKIN CANCER

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PUBLIC HEALTH SCIENCES

BY

XUAN HUI

CHICAGO, ILLINOIS

AUGUST 2021

To my husband, Henan Ma,

who made all this possible,

for his constant support and numerous sacrifices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Dissertation Committee

I would like to thank my advisor, Dr. Habibul Ahsan, for his guidance, insights, and dedication to my training and these research projects. My most valuable training in epidemiology has come through the projects I have done with him. I greatly appreciate him for holding me to a high research standard and always inspiring me to think critically and think big. I would also like to thank my dissertation committee members, Professor Yuan Ji, who provided precious guidance and advice especially for the second part of the project. I wanted to thank Professor James Dignam as I learned so much on the clinical trial design and analysis. I am very grateful for Professor Aly Khan, who guided me through this interdisciplinary analysis with cutting-edge expertise on deep learning aspect. As I struggled on many occasions in understanding the concepts and mechanism of these complex models, I am greatly appreciated with Professor Khan's patience and guidance. I am very grateful for Dr. Christopher Shea's contribution and guidance on both parts of this project. When working with Dr. Shea on annotating the slides, Dr. Shea also taught me to look at the differences of the cancer features on the microscopic images and introduced how he make the judgments for each slide and each patient condition. This experience really inspired me and made me think deeper on the gaps between the clinical practice and the machine learning platform.

Dissertation Contributors

There are other people who are also very important contributors. Professor Maria Argos and Dr. Muhammad Kibriya contributed significantly on everything about BEST. Dr. John

Baron and Dr. Mohammad Yunus who also provided great feedback and insights on my manuscript. I am very thankful for Professor Donald Hedeker who patiently walked through every step of the analysis in my first aim. I am very grateful for Renyu, the PhD student from Department of Computer Science and advisee of Prof Khan, who is the go-to person for any technical questions on the deep learning model. I also want to thank Lin Tong and Sameep Shah, who helped me a lot for the data management and analysis in both projects. I also want to thank Dr. Mohammad Shahriar and Sarwar Golam for great support and help in the process of me working with the Bangladesh team. I am grateful for Hari Thomallari, who helped on scanning the images within very short period. Last but not least, I am in debt for all the great contribution from the entire BEST research team and the 7000 participants.

Administrative Support

I am sincerely grateful to Michele Thompson for administrative support throughout my graduate training in the Department of Public Health Sciences. I am also grateful for Kim Cox for administrative support during the time I work with Dr. Ahsan. Additionally, I am grateful to Colleen Mullarkey of the University of Chicago for the critical review and editorial support on my dissertation draft.

ABSTRACT

In Bangladesh, chronic arsenic exposure through drinking water has affected millions of individuals and remains a public health issue. Due to the carcinogenic effects after long-term or high-dose arsenic exposure to multiple organs and systems in the human body, affected individuals are subject to an increased risk for non-melanoma skin cancer and other cancers. While the effort to eradicate this contamination is inconclusive due to unsustainability, complexity, potentially greater secondary pollution, and other drawbacks, oxidative stress and DNA damage have been suggested and hypothesized to underlie arsenic carcinogenesis.

To develop well-rounded prevention strategies, we divided the goal into two parts: first, through lowering the impact from the exposure to lower the disease risk; and second, through early detection to minimize morbidity and mortality.

The primary prevention strategy for nonmelanoma skin cancer (NMSC) was carried out through intervening in the disease. In cancer prevention, the blood levels of antioxidants, namely vitamin E and selenium, have demonstrated protective effects by combating free radicals in our body. This mechanistic evidence set up a foundation for us to explore therapeutic options for the affected populations and their subgroups. In the first study of this dissertation (Chapter 1 and Chapter 2), I evaluated the overall and differential treatment effects of selenium and vitamin E, alone or in combination on the risk of NMSC and mortality outcomes. There were no statistically significant overall and treatment effects observed on both endpoints of the trial. Additionally, we did not observe differential treatment effects by ten baseline population characteristics, including gender, BMI, smoking status, sun exposure, occupation, skin lesion severity, urinary arsenic level, blood selenium, plasma α-tocopherol, and plasma γ-tocopherol.

The second approach to reduce morbidity and mortality due to NMSC was through early and accurate recognition of skin cancer. In the second part of this dissertation (Chapter 3), I applied an attention-based deep neural network to analyze histopathological images collected from the participants who were suspected of NMSC. Our model showed promising diagnostic accuracy for non-cancer and basal cell carcinoma subtypes. Moreover, we also generated the heatmaps from the model and visualized the key areas within each image that drive the cancer diagnosis. These machine-generated heatmaps were proven accurate in pinpointing the lesions by the expert dermatopathologist.

In summary, we conducted a thorough evaluation to alleviate the impact of NMSC for the susceptible population in Bangladesh through both treating the disease as well as making efforts for early and more accessible cancer detection procedures. These prevention strategies would serve as a model for the similar frameworks of other cancers and complex diseases in a similar setting.

INTRODUCTION

In medicine, physicians use many tools to measure patient's conditions in their wellness and illness. An all-encompassing assessment usually includes patients' demographic characteristics, family history, social history, medical history, as well as their physical examination, laboratory tests and imaging evaluations. The enormous contribution from clinical medicine to human's abilities to treat and cure the sick is beyond contention. However, this form of medical practice is still based on treating symptoms, which may work for most patients by delivering symptom relief, reducing the risk of complications, and improving survival chances. Nevertheless, this solution is not for all [1].

Precision medicine is rapidly evolved in recent five years, serving as a vanguard for leading researchers to transform healthcare, and showing great promise in disease treatments [2]. However, proponents argued that these treatments might benefit only a few individuals for a narrow set of conditions that are primarily genetically determined [1], whereas preventions may apply to all [3–5]. This situation calls for a need to reframe and broaden precision medicine beyond "omics" by integrating clinical, imaging, social, environmental, and demographic profiles altogether [3].

Precision in the context of public health is described as "improving the ability to prevent disease, promote health, and reduce health disparities in populations by applying emerging methods and technologies for measuring disease, pathogens, exposures, behaviors, and susceptibility in populations" [6]. Moreover, population stratification based on social determinants could provide valuable insights on the relationship between disease and treatment [6,7]. This shift

implicitly demands a platform that is capable of bridging multiple fields in a secured environment for diverse population and mixed data formats for analysis and visualization and ultimately inform healthcare decisions [4,6,8].

Over the years, biomedical informatics has evolved immensely. Computers are becoming faster in speed; data is accumulating robustly in both volume and diversity. These advancements are propelling machine learning to solve existing complex problems in medicine. Machine learning as a branch of computer science potentiates a machine to imitate and even enhance human behavior. While on the individual level, the complexities of diseases do not inform clinical decision-making, some existing limitations might be minimized by technology improvements. Researchers have demonstrated the potentials of machine learning models to reduce medical errors by analyzing definitive histopathological images [9,10]. Machine learning model was also applied to sift through unstructured electronic health records and identify medical conditions and diagnoses [11]. In another study, machine learning platform enabled a secure framework for data sharing to identify gene variations and similarities among diverse cohorts [12].

Various machine learning applications have already been used to improve healthcare services in high income countries [13–15]. One analysis showed machine learning applications in clinical health could save approximately $150 billion for US healthcare costs by 2026. These demonstrated positive prospects for its implementation in resource-poor countries, where such applications are most needed. In 2017, the United Nations had held a meeting to discuss the development and deployment of AI applications to reduce poverty and deliver a broad range of critical public services on a globally scale. Later, in another meeting held by the United Nations,

different stakeholders were brought together to evaluate how AI could help in achieving the Sustainable Development Goals.

Little has been documented in the academic literature regarding AI applications for population health in resource-poor settings. However, this should not be taken as a sign of the current activities, nor the trend in this area. Based on the current development of AI in high-income countries, one lesson from experience is that "AI should build intelligence into existing systems and institutions rather than starting from scratch or hoping to replace existing systems, however, broken" [16]. In addition, the pervasive use of computer, mobile phones, and digital cameras in daily life and medical practice in these countries implies that the necessary tools have already been in place to initiate this application. In return, the large volume of data being generated could be used to improve individual and population health in these countries.

In this dissertation project, all the aims were fulfilled with data from the Bangladesh Vitamin E and Selenium Trial (BEST). BEST is a triple blind (participants, care provider, and investigators) two-by-two factorial assignment randomized controlled trial, evaluating vitamin E (100 IU daily) and selenium (200 μg daily) for primary prevention of NMSC and chronic disease-related mortality. The high-risk cohort contains 7,000 participants, recruited between April 2006 and August 2009 from the Narayanganj, Comilla, Noakhali, and Chandpur districts in two regions (Araihazar and Matlab) of Bangladesh. The eligibility criteria to participate in BEST included: 1) adult participants with visible arsenical skin lesions; 2) aged 25 to 65 years; and 3) signed informed consent. Exclusion criteria included: 1) pregnant 2) not a permanent resident of the study areas; 3) not willing to discontinue current vitamin use; 4) history of cancer (including NMSC); 5) too ill to participate; or 6) unwilling to provide biological samples (blood or urine).

All the participants from BEST had long-term exposure to arsenic through the contaminated drinking water. The participants, with skin lesion manifestations, without NMSC at the time of recruitment, were also characterized by their unique demographic, lifestyle, medical, molecular, and genomic factors, and their disease indication and diagnosis were captured through images.

In the hope to develop all-encompassing prevention strategies, we divided the goal in two parts: primarily, through lowering the impact from the exposure to lower the disease risk; and secondarily, through early detection to minimize morbidity and mortality. The primary prevention for NMSC was embodied through intervening the disease. Elaboration on the efficient prevention strategies requires good knowledge of the intervention mechanisms and risk factors. In cancer prevention, blood level of antioxidants, namely vitamin E and selenium, have demonstrated protective effects. The mechanistic evidence set up a foundation for further exploring therapeutic options for the affected populations and their subgroups.

The secondary prevention to reduce morbidity and mortality due to NMSC was mainly through early and accurate recognition of skin cancer. Early initiatives have already been implemented for melanoma towards comparable diagnosis capability by machine learning approaches [17–19], a similar platform for NMSC, especially in the resource-poor setting for the high-risk population would fill the gap in precision population health. Furthermore, the platform of these prevention strategies on NMSC would serve as a model for similar framework of other cancers and complex diseases.

Taking into account the epidemiologic trend of NMSC on a global scale as well as zooming in on the high-risk regions, it becomes obvious that there is a rigorous need to control

the increase of incidence of NMSC and the subsequent socioeconomic burden. Efforts towards this goal were demonstrated in this dissertation project.

1. SELENIUM AND VITAMIN E SUPPLEMENTATION AND PREVENTION OF NON-MELANOMA SKIN CANCER AND CHRONIC DISEASE MORTALITIES: RESULTS FROM A RANDOMIZIED CONTROLLED TRIAL

BACKGROUND

In Bangladesh, chronic arsenic exposure through drinking water has affected 35-77 million people and remains a public health issue [20], resulting in an increased risk for non-melanoma skin cancer and other cancers [21,22]. Due to the carcinogenic effects after long-term or high-dose arsenic exposure to multiple organs and systems in human body, World Health Organization (WHO) has established the value of 10 μg/L as the maximum contaminant level for total arsenic in potable water. In order to meet this requirement, a number of remedial techniques for arsenic removal had been developed and attempted worldwide. However, due to unsustainability, complexity, potentially greater secondary pollution and other drawbacks, these efforts mostly ended inconclusive [23,24].

Although the mechanism remains largely unknown, mounting evidence on arsenical carcinogenesis suggested that long-term arsenic consumption may give rise to various types of cancer [22]. Amongst these mechanisms, arsenic exposure promoted oxidative stress is recognized as an underpinning component, and this activity is proven to be inversely correlated with antioxidant capacity in blood [25,26].

The ability of antioxidants to prohibit cancer cell growth has been revealed in a large body of laboratory studies, attributing to its effect on DNA stability, cell proliferation, necrotic

and apoptotic cell death in both healthy and malignant cells, potentially contributing to regulate oxidative stress and maintain immune system homeostasis [27–30]. Observational epidemiologic studies have also established the evidence of inverse association between selenium and vitamin E exposures with cancer outcomes [31–36], this further strengthened the role of cancer prevention from these antioxidants. Furthermore, an increased blood and urinary selenium and vitamin E were also found associated with higher arsenic metabolism in arsenic-exposed populations [37] , while intake of selenium and vitamin E were associated with lower skin lesion incidence [38–40].

These notable findings have led to a few randomized controlled trials (RCTs) for researchers to seek stronger evidence on the effects of antioxidants for cancer. The Nutritional Prevention of Cancer (NPC) Trial evaluated the effect of selenium on NMSC and other types of cancers in high-risk population and discovered a significant protective effect only on prostate cancer [41] . The Alpha-Tocopherol, Beta-Carotene Cancer prevention trial (ATBC) examined the effect of low-dose vitamin E (50 mg daily) and observed no beneficial effects on the risk of lung cancer [42]. A study conducted in Linxian, China reported two RCTs to study the effects of multiple-vitamin and multiple-mineral supplements on cancer outcomes [43]. However, these trials included interventions with multiple nutrients, individual effect from each supplement wasn't explored. Later, in multiple reports [44,45], the Selenium and Vitamin E Chemoprevention Trial (SELECT) in part evaluated selenium (200 μg/day) and vitamin E (400 IU daily) on prostate and other cancers and suggested no treatment effect. A pilot trial in Bangladesh investigated the impact of a six-month supplementation with vitamin E (400 mg daily) or selenium (200 μg daily) in an arsenic-exposed population and reported slightly improved arsenical skin lesion status; however, the results were not statistically significant [46].

Interestingly, except for the inconsistent findings for overall cancer risk, these studies further implied differential cancer risk by various population characteristics, such as demographics, lifestyle characteristics, clinical manifestations, and biomarker levels. One explanation suggested by observational epidemiologic studies is that antioxidant levels found in human specimen globally vary due to variations in other factors, such as sex, individual metabolism, BMI, and smoking status, especially in at-risk populations [47–52]. Hence, the characteristic-specific nutritional status and antioxidant metabolism may contribute to the observed discrepancies in cancer risk.

To evaluate the impact of selenium and vitamin E as chemopreventive agents in a population with manifest arsenic-toxicity, we conducted the Bangladesh Vitamin E and Selenium Trial (BEST); a 2 × 2 factorial RCT of vitamin E (100 IU daily) and selenium (200 μg daily) for the primary prevention of NMSC and chronic disease-related mortality [53] .

METHODS

Data Source

Trial participants were recruited between April 2006 and August 2009 from the Narayanganj, Comilla, Noakhali and Chandpur districts in two main regions (Araihazar and Matlab) in Bangladesh. Field staff from two centralized offices managed the field work, data/sample collection and processing. The eligibility criteria included: 1) adult participants with existing arsenic-related skin lesions; 2) aged 25 to 65 years at the time of recruitment; and 3) signed informed consent. Exclusion criteria included: 1) pregnancy at the time of enrollment; 2) not a permanent resident of the trial areas; 3) not willing to discontinue current vitamin use; 4)

history of cancer; 5) too ill to participate; or 6) unwilling to provide biological samples (blood or urine). A flowchart in Figure 1 demonstrated the detailed follow-up information for BEST.

The trial stratified randomization by two trial sites aforementioned, and participant's gender. Participants were then randomly assigned vitamin E (α-tocopherol, 100 IU/day), selenium (L-selenomethionine 200 μg/day), both vitamin E and selenium, or placebo. Randomization was implemented using random sequence numbers from a computer generator, and the sequence of treatment allocations was concealed and kept in the data center. All trial personnel and participants were blinded to the treatment assignment.

Power Calculation

Based on our pilot trial and a baseline descriptive study of BEST [46,53], we do not expect interaction between vitamin E and selenium, i.e., the treatment effect of vitamin E does not vary with or without the presence of selenium on NMSC or mortality outcomes. Hence, the full sample size was used to evaluate the treatment effects of vitamin E and selenium.

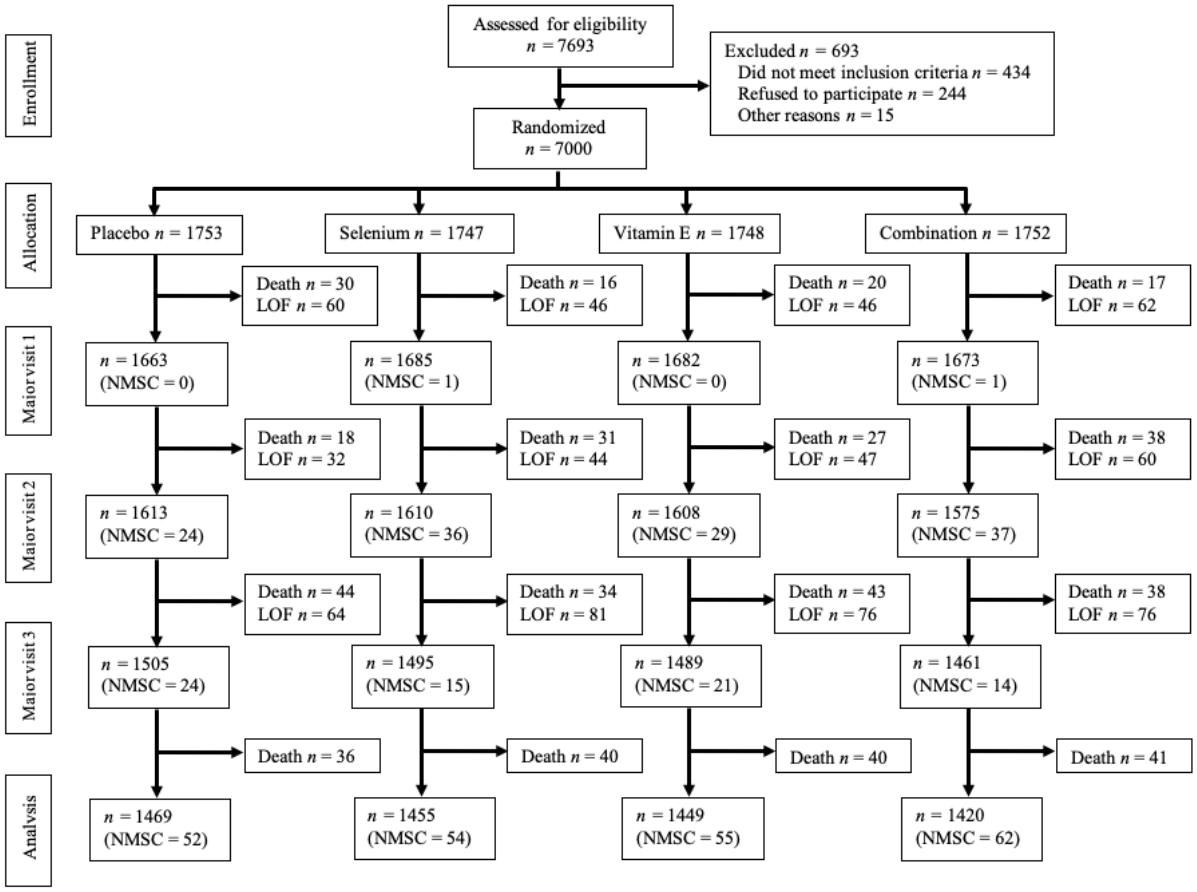
The power and sample size calculations were reported in the baseline descriptive paper [53]. Statistical power for NMSC was presented at a range of assumptions shown in the table below. The power was calculated for time-to-event, taking into account staggered enrollment of the first three years and numbers of death or loss of follow-up [54].

Table 1. Statistical power for detecting the treatment effects on NMSC with 7,000 participants for 2x2 factorial design

NMSC	Annual incidence rate of new NMSC			
	0.008	**0.010**	**0.012**	**0.013**
Relative risk				
0.775	0.67	0.77	0.84	**0.86**
0.75	0.77	0.86	0.91	**0.93**
0.70	0.91	0.96	0.98	**0.99**
0.65	0.97	0.99	1.00	**1.00**

Power estimates in bold are based on currently observed actual event rates.
*1,750 subjects in each of the four arms; 3,500 in each of the two treatment factor groups.

Figure 1. Flowchart of participants during follow-up visits



Measures

Incidence of NMSC

The primary endpoint was the incidence of NMSC. NMSC was screened at each biennial in-person follow-up exam. Specifically, we adopted a three-level evaluation to select participants for a skin biopsy. First, the designated study physician preliminarily triaged potential cases of

invasive NMSC or squamous cell carcinoma (SCC) in situ (Bowen's disease) at the scheduled follow-up visit. Participants identified as having erythema or altered pigmentation of a skin lesion and at least one of the following lesion characteristics—rough, elevated, or interrupted surface or border—were referred for second-level evaluation. The second-level assessment was conducted by senior study physicians specially trained by dermatologists and pathologists to evaluate the lesions for induration, scaly surface, ulceration, blood vessel prominence, and crusting/oozing. When one of these characteristics was present, the participant was referred for a third-level assessment performed by a specialist dermatologist (R.K.) who made the final recommendation for skin biopsy, based upon overall clinical manifestation of invasive skin cancer or SCC in-situ. Lesions < 5 mm in diameter were punch-biopsied, and those ≥ 5 mm in diameter were excised. Among individuals with multiple biopsy-eligible lesions, the most likely malignant lesion was biopsied. If the severity of all lesions was similar, a lesion on or closer to the trunk was preferentially selected over the extremities. Cryosurgery was also provided for the lesions that were not selected for biopsy, but still suspicious for malignancy. Before the biopsy, tetanus immune globulin was administered to those who did not have active immunization to prevent tetanus infection. After the biopsy, the participants were followed up for possible unusual pain, infection, or slowed healing of the biopsy site.

Formalin-fixed biopsy tissues were processed at a specialized surgical pathology laboratory in Dhaka, Bangladesh, and hematoxylin and eosin-stained slides were reviewed by a single pathologist (M.K.) at Bangabandhu Sheikh Mujib Medical University. The slides were transported to The University of Chicago and reviewed blindly and validated by a second pathologist (C.S.). A structured protocol on the scoring system was developed by the two pathologists, based on the histopathological criteria for basal cell carcinoma (BCC) and SCC

(invasive and in-situ forms). All the criteria to define each subtype were based on structured, coded criteria followed to diagnose each condition.

As secondary endpoints, all-cause mortality, as well as deaths attributed to cancer and cardiovascular disease, were evaluated. Vital status was ascertained from semi-weekly village health worker visits. At the report of a participant's death, a trained physician who was blinded to the treatment assignment of the participant conducted a verbal autopsy to ascertain the cause of death. A verbal autopsy questionnaire validated by the icddr,b in a Bangladeshi population was used [55]. A group of expert physicians assigned and coded the cause of death using the WHO's tenth revision of the International Classification of Disease (ICD-10).

Adverse events potentially due to the study treatment were collected, assessed, and treated through a semi-annual assessment conducted by the study physicians, if clinically warranted. Self-report of skin burning, and itching, gastrointestinal symptoms, headache, irritability, and weakness were ascertained.

Statistical Analyses

Baseline characteristics balance for RCT

Data on baseline participant characteristics, including age, sex, smoking status, skin lesion subtypes, education level, baseline urinary arsenic level, weight and height, systolic blood pressure, diastolic blood pressure, baseline blood selenium level, baseline serum vitamin E (α-tocopherol and γ-tocopherol) level, and trial site will be compared across different treatment assignments, for continuous variables with Analysis of Variance (ANOVA) and categorical variables with $\chi 2$ test.

Discrete time hazard model for NMSC

To quantify the conditional probability for NMSC incidence and the 95% confidence interval, I used discrete time hazard models. This model was based on the probabilities of skin lesion incidence at each biennial follow-up visit if the individual was free of outcome in the previous visit. The conditional probability was estimated by a log-linear model, with a different intercept for each study interval, and with common regression coefficients across all intervals. The model was comparable to logistic regression model in that the probability of NMSC at each biennial visit will be estimated. Like the hazard ratio from traditional proportional hazard function with continuous time, the discrete time hazard ratios were estimated.

The formal representation of the conditional probability is defined as:

$$h_{ij} = \Pr\left[T_i = j | T_i \geq j \text{ and } X_{1ij} = x_{1ij}, X_{2ij} = x_{2ij}, \ldots, X_{Pij} = x_{Pij}\right)$$

Where T_{ij} is a discrete random variable indicating the time of the event j for a person i, and the discrete time hazard, h_{ij}, is the probability of that event occurring at time j, conditional on no prior event occurrence.

The discrete time hazard for an individual is modeled as:

$$h_{ij} = \frac{1}{1 + e^{-(A)}}$$

Logit transformation yields the logit hazard of individual i at time j as:

$$\text{logit}\, h(t_{ij}) = \text{logit}\left(\frac{h_{ij}}{1 - h_{ij}}\right)$$

The population discrete-time hazard model with time indicators (the $D's$) and predictors (the $X's$):

$$\text{logit } h(t_{ij}) = [\alpha_1 D_{1ij} + \cdots + \alpha_n D_{nij}] + [\beta_1 X_{1ij} + + \cdots + \beta_m X_{mij}]$$

Where, $\alpha_1, \ldots, \alpha_n$ is each intercept parameter, represents the value of logit hazard (the log odds of event occurrence) in that particular time period for individuals in the "baseline" group; $\beta_1, \ldots, \beta_m$ is each slope parameter, represents the effect of a one-unit difference in that predictor on event occurrence, statistically controlling for the effects of all other predictors in the model.

In the model above, the logit hazard of having NMSC of the individual at time j is a function of each baseline hazard ($\alpha_1 D_{1j} \ldots \alpha_n D_{nj}$), the difference in hazard for each treatment arm relative to the placebo arm (β_1). Therefore, the treatment effects from each arm are estimated as:

$$\text{when treatment} = \text{placebo: logit } h(t_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3]$$

$$\text{when treatment} = \text{selenium: logit } h(t_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3] + \beta_1$$

$$\text{when treatment} = \text{vitamin E: logit } h(t_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3] + 2 * \beta_1$$

$$\text{when treatment} = \text{combination: logit } h(t_j) = [\alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3] + 3 * \beta_1$$

Cox proportional hazard models for deaths

As all-cause, cancer, and CVD deaths were captured continuously throughout the trial, I used Cox proportional hazard models to assess the hazards of the mortality outcomes.

The proportional hazard is:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}) h_0(t)$$

Where, $\beta_1, \ldots, \beta_p$ is each slope parameter, represents the log hazard ratio for one-unit increase in x_j, statistically controlling for the effects of all other predictors in the model.

Hence, the relative hazard (hazard ratio) for all-cause deaths, comparing an individual i (in selenium group) to an individual j (in placebo group), no covariates included:

$$\frac{h_i(t)}{h_j(t)} = \frac{\exp(\beta_1 \text{treatment} = \text{selenium}) h_0(t)}{\exp(\beta_1 \text{treatment} = \text{placebo}) h_0(t)} = \exp(\beta_1)$$

Generalized linear model for adverse events

Each adverse event will be evaluated using relative risk (RR) between treatment arm and placebo, respectively. The general model form is:

$$\text{RR} = \frac{P(Y = 1 \mid \text{treatment} = \text{selenium})}{P(Y = 1 \mid \text{treatment} = \text{placebo})}$$

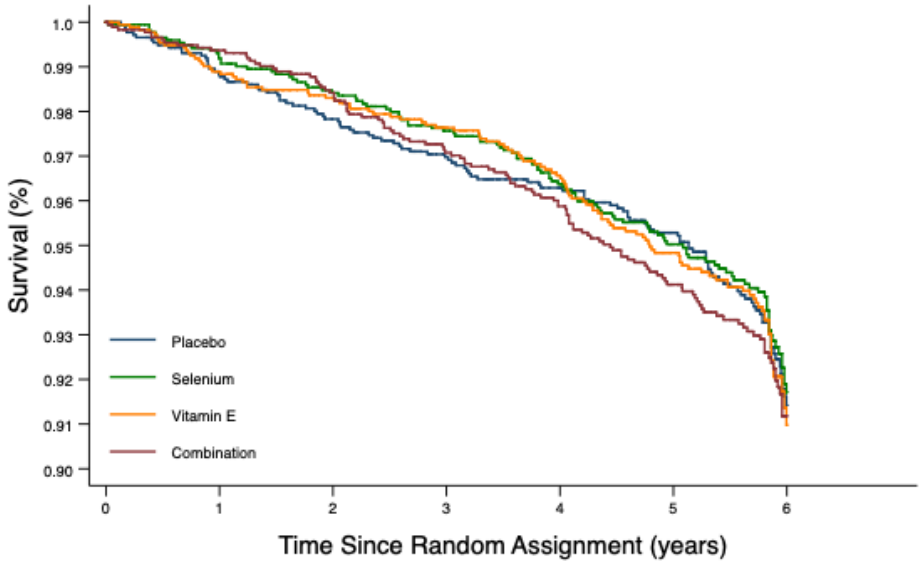
RESULTS

A total of 7,000 participants with arsenical skin lesions were randomized into BEST: 1747 (25%) were assigned to selenium, 1748 to vitamin E (25%), 1752 to both selenium and vitamin E (25%), and 1753 to placebo (25%). Participants were followed for an average of 5.22 years (SD = 1.45). Among the 7,000 participants, 6,963 (99.5%) had complete baseline information.

The mean age of participants was 41.5 years (SD = 10.4) (Table 1). The majority of participants were female (59%). All participants had visible skin lesions whose skin lesion subtypes were classified as pre-cancer (1.4%), keratosis (53.5%), and leucomelanosis & melanosis (45.1%). Among male participants, 66% self-reported ever smoking, while 1% of female participants reported ever smoking. Approximately 40% of study participants reported no formal education. The average body mass index was 19.6 (SD = 3.1), average systolic blood pressure was 116.1 (SD = 16.7), and average diastolic blood pressure was 75.1 (SD = 10.5). Baseline characteristics were balanced across the four treatment groups.

There were 918 (13%) participants with at least one missed biennial follow-up visit, and 5,569 (80%) had complete follow-up (Figure 2-4). As shown in Table 2, the percentage of participants who reported good adherence (taking at least 80% of their assigned tablets) participants was similar across all treatment groups throughout the study (Table 2). On average, more than 92% of participants had good adherence through the first biennial follow-up (year 2), approximately 94% through the second biennial visit (year 4), and about 90% at the third biennial visit (year 6). Self-reported compliance largely concurred with pill counts. Significant increase of blood and plasma concentrations of α-tocopherol and selenium suggest excellent adherence (Table 2).

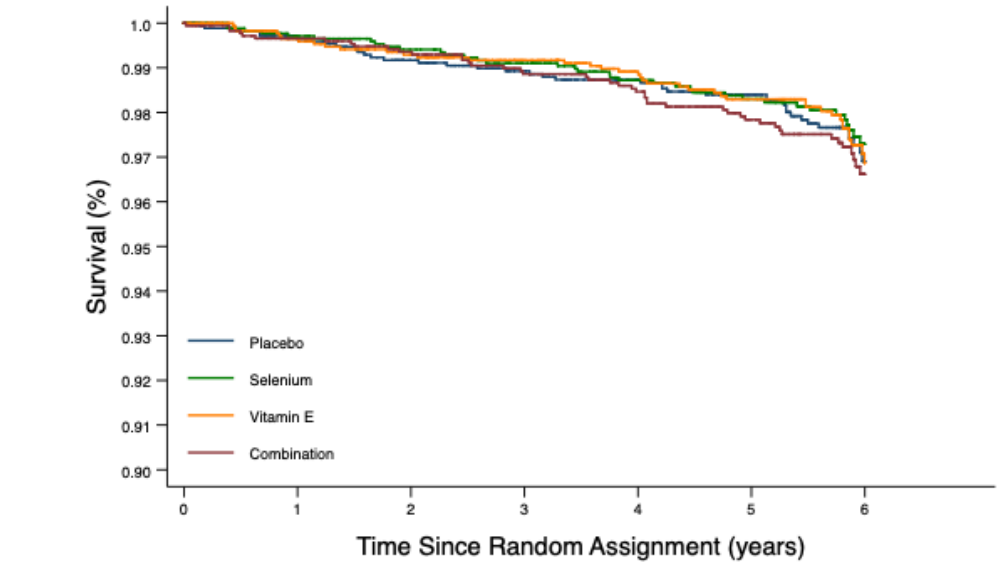
Figure 2. Kaplan-Meier Survival Curves for All-cause Mortality, Censoring at Six Years



Number at risk

trt = Placebo	1753	1679	1630	1564	1496	1351	458
trt = Selenium	1747	1689	1635	1590	1496	1314	462
trt = Vitamin E	1748	1679	1646	1582	1494	1332	455
trt = Combination	1752	1701	1636	1563	1477	1299	477

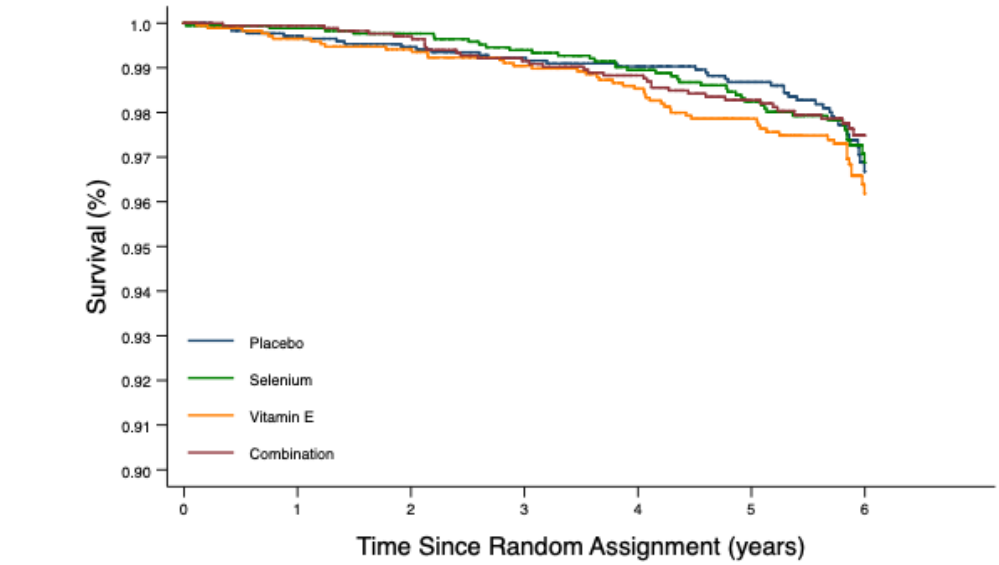
Figure 3. Kaplan-Meier Survival Curves for Cancer Mortality, Censoring at Six Years



Number at risk

trt = Placebo	1753	1679	1630	1564	1496	1351	458
trt = Selenium	1747	1689	1635	1590	1496	1314	462
trt = Vitamin E	1748	1679	1646	1582	1494	1332	455
trt = Combination	1752	1701	1636	1563	1477	1299	477

Figure 4. Kaplan-Meier Survival Curves for Cardiovascular Disease Mortality, Censoring at Six Years



Number at risk

trt = Placebo	1753	1679	1630	1564	1496	1351	458
trt = Selenium	1747	1689	1635	1590	1496	1314	462
trt = Vitamin E	1748	1679	1646	1582	1494	1332	455
trt = Combination	1752	1701	1636	1563	1477	1299	477

Table 2: Baseline Characteristics of The Participants in The BEST Study

Baseline characteristics[1]		Placebo 1753 (25)	Selenium 1747 (25)	Vitamin E 1748 (25)	Combination 1752 (25)
Age (years)	mean (SD)	41.7 (10.3)	41.7 (10.3)	41.6 (10.5)	41.2 (10.5)
Sex	male	714 (41)	704 (40)	712 (41)	710 (41)
	female	1039 (59)	1043 (60)	1036 (59)	1042 (59)
Male smoking (ever)	No	246 (34)	240 (34)	251 (35)	243 (34)
	Yes	468 (66)	464 (66)	461 (65)	467 (66)
Female smoking (ever)	No	1026 (99)	1026 (98)	1027 (99)	1027 (99)
	Yes	13 (1)	17 (2)	9 (1)	15 (1)
Education (years)	0 years	729 (42)	746 (43)	717 (41)	731 (42)
	1-5 years	476 (27)	482 (28)	506 (29)	508 (29)
	6-17 years	548 (31)	519 (30)	525 (30)	513 (29)
Urinary total arsenic level[2] (μg/L)	median (IQR)	15.5 (5.2, 33.1)	14.8 (4.7, 32.2)	15.4 (4.7, 32.5)	16.7 (4.9, 32.6)
Body mass index (Asian) (kg/m2)	< 18.5	717 (41)	711 (41)	737 (42)	729 (42)
	18.5-22.9	780 (44)	789 (45)	756 (43)	793 (45)
	$\geq$ 23	256 (15)	247 (14)	255 (15)	230 (13)
Plasma α-tocopherol (mg/L)	mean (SD)	11.9 (4.2)	10.7 (5.2)	11.2 (5.2)	11.4 (6.6)
Plasma γ-tocopherol (mg/L)	mean (SD)	0.94 (0.86)	0.88 (0.77)	1.01 (1.62)	0.94 (0.72)
Systolic blood pressure (mmHg)	mean (SD)	115.8 (16.4)	116.1 (17.1)	116.0 (16.6)	116.3 (17.0)
Diastolic blood pressure (mmHg)	mean (SD)	75.1 (10.7)	75.0 (10.5)	75.1 (10.1)	75.1 (10.6)

[1] For categorical variables: numbers in parentheses show percentages if not denoted otherwise. and skin lesion subtype information; the remaining 11 (0.2%) participants had only missing urinary arsenic concentration

[2] 37 participants had missing baseline urinary arsenic concentration

Table 3: Physician-report and Self-report Pill Counts of Participant Adherence (in Percentage) Over Six Years and Bioadherence Over Two Years

Pill counts (%)[1]	Placebo	Selenium	Vitamin E	Combination
Year 1	93	93	92	93
Year 2	94	94	94	93
Year 3	95	95	94	95
Year 4	95	95	95	94
Year 5	95	95	95	96
Year 6	90	90	92	92

Self-report (%)	Placebo	Selenium	Vitamin E	Combination
Year 1	94	95	95	95
Year 2	93	93	93	93
Year 3	93	95	94	94
Year 4	95	95	95	94
Year 5	95	96	97	95
Year 6	97	97	97	97

Bioadherence[2] (median (IQR))	Placebo	Selenium	Vitamin E	Combination
Blood selenium, μg/L				
Baseline	129 (115-147)	131 (116-147)	130 (115-147)	130 (115-147)
Two-year visit	123 (109-139)	537 (400-671)	125 (111-141)	500 (354-633)
Plasma α-tocopherol, mg/L				
Baseline	10.5 (8.6-13.1)	9.9 (7.9-12.6)	10.3 (8.2-13.2)	10.4 (8.1-13.2)
Two-year visit	11.3 (9.0-14.5)	10.7 (8.6-13.6)	17.5 (13.0-22.6)	17.2 (13.4-21.9)
Plasma γ-tocopherol, mg/L				
Baseline	0.8 (0.4-1.2)	0.7 (0.4-1.1)	0.7 (0.4-1.2)	0.4 (0.8-1.2)
Two-year visit	0.6 (0.3-1.1)	0.6 (0.3-1.1)	0.4 (0.2-0.7)	0.3 (0.2-0.7)

A total of 425 participants were diagnosed with NMSC and 513 (7%) died. Per the trial design, we conducted comparisons by treatment factor (selenium and vitamin E). For selenium, (alone or in combination with vitamin E) versus no selenium (vitamin E alone or dual placebo groups), there was no difference in the NSMC rate (HR = 1.08, 95% CI [0.89 - 1.30]). Similarly, for vitamin E (alone or in combination with selenium) versus no Vitamin E, there was no reduction in NSMC (HR = 1.09, 95% CI [0.90 - 1.32]). There was no evidence of an interaction between factors; nonetheless, the contrasts of each active treatment arm to placebo-control are

[1] Percentage of participants adherent, defined as taking at least 80% of study supplements.
[2] Bioadherence table: plasma level of the agents for the first 2000 participants were analyzed at two-year visit.

shown (Table 3), again showing no evidence of a reduction in NSMC for any of the treatment groups. Adjustment for the randomization factors (sex and site) had no appreciable effect on the effect estimates. In sensitivity analyses, loss-to-follow-up was not associated with any baseline participant characteristics.

Table 4: Effect of Study Treatments on NMSC Incidence, for Factorial and Non-factorial Analyses

	# of participants	# of cases	HR (95% Cl)
	Two treatments by factorial design[3]		
Vitamin E	3500	219	1.08 (0.89 - 1.30)
Selenium	3499	220	1.09 (0.90 - 1.32)
	Four treatments		
Placebo	1753	100	Ref.
Selenium	1747	106	1.08 (0.82 - 1.43)
Vitamin E	1748	105	1.07 (0.81 - 1.41)
Combination	1752	114	1.17 (0.90 - 1.54)

We also evaluated treatment effects on NMSC subtypes: i) Bowen's disease (SCC in-situ), ii) basal cell carcinoma (BCC), and iii) squamous cell carcinoma (SCC). There were 291 participants with Bowen's disease, 135 with BCC, and 33 with SCC lesions identified. Two participants were identified as having all three types of skin cancers; seven were identified as having Bowen's disease and SCC, five were diagnosed as having both BCC and SCC, and 18 were diagnosed with both Bowen's disease and BCC. As shown in Table 4, no significant treatment effects for selenium or vitamin E were observed with any subtype.

[33] Estimates from factorial analyses indicate Vitamin E + Combination vs. Selenium + Placebo for vitamin E vs. no vitamin E; Selenium + Combination vs. Vitamin E + Placebo for selenium vs. no selenium.

Table 5: Effect of Study Treatments on Incidence of BCC, Bowen's Disease and SCC, for Factorial and Non-factorial Analyses

		Bowen's disease		BCC		SCC	
	# of participants	# of cases	HR (95% Cl)	# of cases	HR (95% Cl)	# of cases	HR (95% Cl)
				Two treatments by factorial design[4]			
Vitamin E	3500	150	1.08 (0.86 - 1.36)	63	0.88 (0.63 - 1.24)	17	1.07 (0.54 - 2.13)
Selenium	3499	151	1.10 (0.87 - 1.38)	63	0.89 (0.63 - 1.25)	21	1.78 (0.87 - 3.61)
				Four treatments			
Placebo	1753	67	Ref.	37	Ref.	8	Ref.
Selenium	1747	74	1.13 (0.81 - 1.58)	35	0.97 (0.61 - 1.54)	8	1.02 (0.38 - 2.72)
Vitamin E	1748	73	1.11 (0.80 - 1.55)	35	0.96 (0.61 - 1.53)	4	0.51 (0.15 - 1.69)
Combination	1752	77	1.18 (0.85 - 1.64)	28	0.78 (0.48 - 1.27)	13	1.67 (0.69 - 4.02)

[4] Estimates from factorial analyses indicate Vitamin E + Combination vs. Selenium + Placebo for vitamin E vs. no vitamin E; Selenium + Combination vs. Vitamin E + Placebo for selenium vs. no selenium.

Overall, for the prespecified secondary endpoint (overall mortality, cancer mortality, and CVD mortality) censoring at six years, there were no statistically significant effects by treatment assignment (Figure 2). Similar results were observed by two treatment factors as well as four treatment factors (Table 5).

Reported adverse events were not associated with treatment assignment (Table 7). No life-threatening adverse events related to the study treatment were reported. No subjects stopped participating in the study due to adverse side effects.

Table 6: Effect of Study Treatments on Incidence of All-cause, Cancer, CVD Mortality and Other Cause of Mortality Outcomes, Censoring at Six Years, for Factorial and Non-factorial Analyses

		All-cause death		Cancer death		CVD death	
	# of participants	# of cases	HR (95% CI)	# of cases	HR (95% CI)	# of cases	HR (95% CI)
				Two treatments by factorial design[5]			
Vitamin E	3500	240	1.07 (0.89 - 1.29)	84	1.11 (0.81 - 1.51)	84	1.08 (0.80 - 1.47)
Selenium	3499	232	1.00 (0.83 - 1.20)	81	1.03 (0.76 - 1.40)	74	0.85 (0.62 - 1.15)
					Four treatments		
Placebo	1753	115	Ref.	40	Ref.	39	Ref.
Selenium	1747	110	0.96 (0.74 - 1.25)	36	0.90 (0.58 - 1.42)	39	1.00 (0.64 - 1.57)
Vitamin E	1748	118	1.03 (0.80 - 1.33)	39	0.98 (0.63 - 1.52)	49	1.26 (0.83 - 1.92)
Combination	1752	122	1.07 (0.83 - 1.38)	45	1.13 (0.74 - 1.74)	35	0.91 (0.58 - 1.43)

[5] Estimates from factorial analyses indicate Vitamin E + Combination vs. Selenium + Placebo for vitamin E vs. no vitamin E; Selenium + Combination vs. Vitamin E + Placebo for selenium vs. no selenium.

Table 7: Adverse Events Known to Be Associated with The Study Supplements

	Placebo (n = 1753)		Selenium (n = 1747)		Vitamin E (n = 1748)		Combination (n = 1752)	
Events	N	RR	N	RR (95% CI)[6]	N	RR (95% CI)	N	RR (95% CI)
Burning and itching	2	Ref.	2	1.00 (0.14, 7.12)	3	1.50 (0.25, 8.99)	0	NA
Diarrhea	11	Ref.	13	1.19 (0.53, 2.64)	11	1.00 (0.44, 2.31)	10	0.91 (0.39, 2.14)
Gastritis	1	Ref.	0	NA	1	1.00 (0.06, 16.02)	1	1.00 (0.06, 15.98)
Gastroenteritis	1	Ref.	1	1.00 (0.06, 16.03)	1	1.00 (0.06, 16.02)	3	3.00 (0.31, 28.83)
Headache	1	Ref.	2	2.01 (0.18, 22.11)	2	2.01 (0.18, 22.10)	3	3.00 (0.31, 28.83)
Intestinal obstruction	0	Ref.	2	NA	1	NA	1	NA
Irritability	1	Ref.	0	NA	0	NA	0	NA
Vomiting	8	Ref.	7	0.88 (0.32, 2.42)	4	0.50 (0.15, 1.66)	6	0.75 (0.26, 2.16)
Weakness	8	Ref.	6	0.75 (0.26, 2.16)	4	0.50 (0.15, 1.66)	9	1.13 (0.44, 2.91)

[6] RR: Relative Risk compared to placebo group.

DISCUSSION

Results from BEST show that neither selenium nor vitamin E, alone or in combination, reduced risks of NMSC in a population with manifest arsenic toxicity during a median treatment period of 5.8 years. Analyses of NMSC subtypes revealed no treatment effects for selenium or vitamin E with incident BCC, SCC, or Bowen's disease.

There are differences in the form and dose of the supplements used in BEST compared with other published trials. In SELECT, investigators reported no beneficial effect of selenium or vitamin E on overall cancer risks [44]. SELECT participants received 400 IU/day (all rac-α-tocopheryl acetate) of vitamin E, which is a higher dose and different form of vitamin E than utilized in the current trial, and the same dose and form of selenium (200 μg daily *L*-selenomethionine). Of note, the administration of α-tocopherol may cause a 50% decrease in median plasma γ-tocopherol, while the latter has been proposed as a promising alternative cancer prevention supplement than α-tocopherol [56,57]. The ATBC Cancer Prevention Study reported that daily supplementation of 50 IU of synthetic dl-α-tocopheryl acetate for 5 to 8 years did not show protective effects on incident lung cancer [42]. The NPC trial [41] evaluated the treatment effects of selenium 200 μg/day as selenized yeast compared to placebo for BCC and SCC and reported no protective effect. However, the achieved levels of selenomethionine, the natural form of selenium, is unknown. In addition, the authors of the NPC trial also acknowledged substantial batch-to-batch variations in the levels of organoselenium compound samples [44]. Furthermore, two interventional trials conducted in Linxian, China [43,58,59] also utilized different forms of selenium from BEST. One trial used inorganic selenium (sodium selenite, 50 μg daily) and observed protective effects on mortality endpoints, while the other adopted selenized yeast (50

μg daily) and found no treatment effect. The differences in dose, form, and duration of vitamin E and selenium supplementation across trials makes it difficult to summarize the existing literature.

Approaches for endpoint ascertainment may explain the difference in findings between BEST and other published trials. The NPC trial suggested a higher risk of incident NMSC in the selenium arm compared to placebo [41]. Investigators from the NPC trial indicated a higher rate of prostate biopsy in the placebo arm compared to the selenium arm (35% vs. 14%) [60], particularly for the lowest baseline selenium concentration stratum, which happened to show the strongest inverse association between selenium and prostate cancer incidence. This potential detection bias may have affected the assessment of the primary endpoint of NMSC in the stratified analysis by baseline selenium concentration and puts the NMSC findings in that study into question. Two small trials of selenium reported NMSC as an adverse event. In one trial of patients at high risk of prostate cancer [61], the authors reported a 1% incidence of NMSC in the placebo group compared to 3% and 1% incidence in the 200 μg/L and 400 μg/L selenium groups, respectively. In the ECOG trial conducted among patients with lung cancer [62], 3.6% of placebo-treated patients developed SCC or BCC, and 2.4% of selenium-treated patients developed SCC or BCC.

The BEST was the first trial using selenium and vitamin E in an arsenic-exposed population. The population characteristics of BEST participants are considerably different from other trial target populations. For example, SELECT [44] enrolled 32,400 multi-ethnic men aged >50 years with a low Prostate-Specific Antigen (PSA) level. The ATBC trial [42] enrolled 29,133 male smokers aged 50 to 69 years from southwestern Finland. The NPC trial [41] enrolled patients aged 18 to 80 years (mean = 63) with a history of BCC and SCC. Two interventional trials conducted in Linxian, China [43,59] evaluated patients with esophageal cancer and the general population, respectively. While in BEST, the participants are residents of rural Bangladesh and

have been chronically exposed to arsenic through drinking water and developed cutaneous manifestation of arsenicosis.

The BEST study had several strengths. It is the largest RCT evaluating oral supplementation of selenium and vitamin E with NMSC incidence, making the conclusions drawn from BEST especially robust. As the study population was ascertained on the population scale, BEST is a population-based chemoprevention trial conducted in two regions in Bangladesh, where the health system relies heavily on the government or the public sector. In this resource-limited setting, implementing a long-term chemoprevention trial faced several unique challenges. Limited public facilities, compromised access to medical care, lack of essential commodities, and lack of public health and management expertise at the district and upazila (regional administration in Bangladesh) levels altogether made it difficult for researchers to collect detailed health information of the participants. BEST addressed these issues early in the study design stage and serves as a model for the conduct of future RCTs in similar resource-limited countries. Furthermore, no prior RCT of selenium or vitamin E was conducted among a South Asian population, nor in a community at high risk of developing NMSC and other cancers due to chronic environmental exposure.

We also acknowledge the limitations of BEST. First, due to financial constraints, we only measured α-tocopherol and γ-tocopherol levels for the first 2,000 randomized participants at baseline and the first biennial follow-up visit. However, based on self-reported adherence and pill counts, we believe that adherence was high for the duration of the trial. The lack of vitamin E or selenium treatment effect on the incidence of NMSC may be explained by an insufficient length of treatment, dose, or form of supplements utilized.

CONCLUSION

Health outcomes from arsenic exposure in drinking water may persist for decades after the exposure is eliminated [63]. While initiatives to reduce exposure have been ongoing and making considerable progress, investigations concerning biological approaches to reducing disease risk are still desirable for affected populations. We conducted a thorough evaluation of long-term supplementation of vitamin E and selenium with NMSC incidence and mortality endpoints and found no beneficial effects on these endpoints. Potential benefits of these agents on other health outcomes need to be investigated.

2. SELENIUM AND VITAMIN E SUPPLEMENTATION AND PREVENTION OF NON-MELANOMA SKIN CANCER AND CHRONIC DISEASE MORTALITIES: CONFOUNDING AND EFFECT MODIFICATION FROM A RANDOMIZED CONTROLLED TRIAL

BACKGROUND

Every year, more than one million nonmelanoma skin cancer (NMSC) cases are diagnosed in the United States, surpassing the annual combined incidence of lung, prostate, colon, bladder, and kidney cancers (PMID: 28515985, 21034989). The American Cancer Society estimates that in 2012, 5.4 million cases of NMSC were diagnosed in 3.3 million people. The increase in NMSC incidence has resulted in both a direct and an indirect cost to our society. In 2004, the total direct cost associated with the treatment for NMSCs was more than $1 billion (PMID: 16908356). In 2013, the cost soared to over $4.5 billion, not including prescription/OTC medications, screening, vaccination, and other related medical services (PMID: 28259441). In the most serious circumstances, some aggressive and neglected NMSC can even lead to deaths. In the same report (PMID: 28259441), NMSC accounts for 19.07% of the total deaths related to skin disease in 2013. Among established mechanisms, these deaths are largely due to preventable factors (PMID: 25651787) However, in cases where the factors yet to be addressed, prophylactic trials are most desired to impede epidemic. In Bangladesh, arsenic exposure through drinking water has affected about 35-77 million people and remains a public health issue (PMID: 11019458, 23226896).

The mechanism of arsenic toxicity has yet to reach consensus. One dominant hypothesis is that oxidative stress induced by arsenic is the key underlying the carcinogenesis (PMID: 26861378, 16430879), which subsequently were extensively investigated in both in vitro studies (PMID: 8902524, 9653147, 12771042) and epidemiologic studies (PMID: 11940449, 15276408, 11675266, 10417614). This proposed mechanism underlies the inter-individual variability in arsenic metabolism capacity (PMID: 19168087, 28796632). Building upon this, there has been significant number of studies examining the effects of antioxidants in arsenic-exposed population (PMID: 21652291, 12505432, 16337848, 21227482, 23590571, 18709164), especially in the Bangladesh population (PMID: 16160703). However, stronger evidence generated through randomized controlled trials (RCT) have been inconsistent with laboratory and observational epidemiology studies. The Nutritional Prevention of Cancer (NPC) study (PMID: 8971064, 12699469) further evaluated the effect of selenium on various types of cancers and discovered significant protective effect only on prostate cancer. The Alpha-Tocopherol, Beta-Carotene Cancer prevention study (ATBC) (PMID: 8127329) examined the effect of vitamin E administered 50mg daily and observed early beneficial effects. Later, in multiple reports (PMID: 15657339, 19066370, 21990298), the Selenium and Vitamin E Cancer Prevention Trial (SELECT) in part evaluated selenium (200 μg/day) and vitamin (400 IU daily) on prostate cancer and the higher dose vitamin E showed no efficacy. Another investigation revealed that higher doses of vitamin E may increase the risk of hemorrhagic stroke in the subjects (PMID: 21051774), which had raised significant safety concerns.

Treatments towards NMSC has always been “one cure for all”, however, this may dilute effects within subgroups. Therefore, in this aim, we further evaluated treatment effects accounting for different population characteristics.

METHODS

Study population

BEST participants who were recruited between April 2006 and August 2009 from two main regions (Araihazar and Matlab) in Bangladesh were included for analyses. Participant eligibility criteria included: 1) adult participants with existing arsenic-related skin lesions; 2) aged 25 to 65 years at the time of recruitment; and 3) signed informed consent. Exclusion criteria included: 1) pregnancy at the time of enrollment; 2) not a permanent resident of the trial areas; 3) not willing to discontinue current vitamin use; 4) history of cancer; 5) too ill to participate; or 6) unwilling to provide biological samples (blood or urine).

Covariates

Study participants provided comprehensive demographic, medical history, clinical testing, and risk factors to the trained trial staff at baseline. Factors include age, gender, BMI, smoking status, occupation, sun exposure, baseline urinary arsenic level, baseline blood level of selenium, baseline plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol. We calculated BMI as weight (kg) divided by the square of height (cm), then categorized using both WHO standard. Smoking status was assessed as ever and never smokers. Continuous variables such as urinary level of arsenic, blood selenium level, plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol were categorized into quartiles and evaluated with both continuous and categorical forms.

Statistical analyses

Confounding effect

To assess the confounding effect from baseline characteristics, we adjugated all the covariates aforementioned to evaluate the treatment effects on the endpoint of NMSC as well as all-cause mortality. Age was included as a continuous variable. Other continuous variables such as urinary level of arsenic, blood selenium level, plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol were adjusted as categorial variables.

Effect modification

To quantify the conditional probability for NMSC incidence and the treatments in each subgroup according to gender, BMI, smoking status, occupation, sun exposure, baseline urinary arsenic level, baseline blood level of selenium, baseline plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol, discrete time hazard models were implemented. Variables in continuous format and without a standardized categorization, such as baseline urinary arsenic level, baseline blood level of selenium, baseline plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol, were evaluated in both continuous fashion and categorial fashion. These factors have first assessed the skewness; when skewed, log transformation was applied. Then, these factors were categorized into four levels, with the lowest level set as the reference group. The conditional probability of NMSC with effect modification was estimated by a log-linear model, with a different intercept for each study interval, and with common regression coefficients across all intervals. The model was comparable to logistic regression model in that the probability of NMSC at each biennial visit will be estimated.

Similarly, effect modification was also evaluated for mortality outcomes. Cox proportional hazards models were used to quantify the association between all-cause mortality with study treatments. Mortalities due to cardiovascular disease, cancer, and other causes were

not assessed due to limited event numbers. We first modeled factors such as baseline urinary arsenic level, baseline blood level of selenium, baseline plasma levels of cholesterol-adjusted α-tocopherol and γ-tocopherol as continuous variables in log-transformation, as appropriate. Then, evaluated these factors as categorial effect modifiers.

To evaluate interaction, models with and without product terms between treatment and hypothesized effect modifiers were tested. Overall interaction was tested based on a likelihood ratio test comparing the models with and without the product terms. A p-value <0.05 (2-tailed) was considered statistically significant. Statistical analysis was performed using Stata 16 (StatCorp LP, College Station, TX).

Multiplicative interaction model form:

H_0: Relative Risk(A_+ B_+) = Relative Risk(A_+) * Relative Risk(B_+)

H_A: Relative Risk(A_+ B_+) ≠ Relative Risk(A_+) * Relative Risk(B_+)

RESULTS

In our study cohort, mean age for participants identified with NMSC was 44.5 years old (SD = 0.45), marginally older than participants without NMSC (mean = 41.3 years old, SD = 0.13) during the study duration ($p < 0.001$). More male than female was identified with NMSC (57.18% vs. 42.82%, $p < 0.001$) and majority of NMSC cases have never smoke (62.12% vs. 37.88%, $p < 0.001$). Among all NMSC cases, majority were unemployed, home maker, or retired and does not work under the sun. Baseline urinary arsenic level, blood selenium level, and plasma α-tocopherol and γ-tocopherol were relatively evenly distributed in NMSC cases (all p-values > 0.05) (Table 8).

Table 8: Baseline Characteristics by The Status of NMSC

Covariates	Categories	non-NMSC	NMSC	p-value
Age, mean (SD)		41.3 (0.13)	44.5 (0.45)	< 0.001
Sex, n (%)	Male	2597 (39.50)	243 (57.18)	< 0.001
	Female	3978 (60.50)	182 (42.82)	
Site	Araihazar	3289 (50.02)	240 (56.47)	0.010
	Matlab	3286 (49.98)	185 (43.53)	
BMI, n (%)	< 18.5	2711 (41.23)	183 (43.06)	
	18.5 - 24.9	3459 (52.61)	218 (51.29)	0.732
	25+	405 (6.16)	24 (5.65)	
Ever smoker, n (%)	No	4822 (73.34)	264 (62.12)	< 0.001
	Yes	1753 (26.66)	161 (37.88)	
Occupation, n (%)	daily laborer	1411 (17.35)	93 (21.88)	
	farmer	631 (9.60)	63 (14.82)	
	other	796 (12.11)	85 (20.00)	< 0.001
	unemployed, home maker, retired	4007 (60.94)	184 (43.29)	
Sun exposure	Mild	1393 (53.66)	147 (60.49)	0.041
	Heavy	1203 (46.34)	96 (39.51)	
Arsenic level in urine, mean (SD)	1st quartile	1644 (25.14)	101 (23.88)	
	2nd quartile	1620 (24.77)	117 (27.66)	0.551
	3rd quartile	1649 (25.21)	99 (23.40)	
	4th quartile	1627 (24.88)	106 (25.06)	
Blood selenium, mean (SD)	1st quartile	1643 (25.12)	101 (23.88)	
	2nd quartile	1667 (25.49)	95 (22.46)	0.293
	3rd quartile	1646 (25.17)	110 (26.00)	
	4th quartile	1584 (24.22)	117 (27.66)	
α-tocopherol, mean (SD)	1st quartile	440 (25.48)	34 (31.48)	
	2nd quartile	432 (25.01)	23 (21.30)	0.247
	3rd quartile	442 (25.59)	21 (19.44)	
	4th quartile	413 (23.91)	30 (27.78)	
γ-tocopherol, mean (SD)	1st quartile	486 (28.14)	30 (27.78)	
	2nd quartile	499 (28.89)	34 (31.48)	0.944
	3rd quartile	345 (19.98)	21 (19.44)	
	4th quartile	397 (22.99)	23 (21.30)	

Table 9 shows the results from multivariate discrete time hazard regression for the association between treatments and NMSC endpoint. In general, none of the confounding factors turned out statistically significant.

Table 9: Confounding Effects on The Association between Effects of Four Treatment Groups on Non-melanoma Skin Cancer

Characteristics	Categories	HR	p-value	95% Confidence Interval	
Age		1.02	0.084	1	1.05
Sex	Male	Ref.	-	-	-
	Female	omitted	-	-	-
Site	Araihazar	Ref.	-	-	-
	Matlab	1.07	0.836	0.58	1.94
BMI	< 18.5	Ref.	-	-	-
	18.5 - 24.9	0.91	0.731	0.53	1.56
	25+	0.54	0.417	0.12	2.42
Ever smoker	Yes	0.86	0.602	0.48	1.53
Occupation	daily laborer	Ref.	-	-	-
	farmer	1.36	0.412	0.65	2.82
	other	1.61	0.169	0.82	3.15
	unemployed, home maker, retired	0.97	0.960	0.32	2.98
Sun exposure	Mild	Ref.	-	-	-
	Heavy	0.89	0.721	0.46	1.71
Arsenic level in urine	1st quartile	Ref.	-	-	-
	2nd quartile	0.80	0.499	0.42	1.52
	3rd quartile	0.64	0.236	0.30	1.34
	4th quartile	0.91	0.835	0.36	2.26
Blood selenium	1st quartile	Ref.	-	-	-
	2nd quartile	0.93	0.808	0.51	1.70
	3rd quartile	0.62	0.219	0.29	1.33
	4th quartile	0.69	0.401	0.29	1.65
α-tocopherol	1st quartile	Ref.	-	-	-
	2nd quartile	0.70	0.335	0.34	1.44
	3rd quartile	0.61	0.202	0.29	1.30
	4th quartile	0.97	0.935	0.46	2.04
γ-tocopherol	1st quartile	Ref.	-	-	-
	2nd quartile	1.02	0.941	0.53	1.97
	3rd quartile	1.09	0.824	0.50	2.36
	4th quartile	0.92	0.845	0.38	2.22

When evaluating all-cause mortality, however, with one year older in age, there is an increased risk of NMSC for 5% (HR = 1.05, 95% CI [1.03, 1.08]). The hazard of NMSC is also 2.5 times higher in smokers compared to non-smoker (HR = 2.50, 95% CI [1.29, 4.85]). Baseline arsenic levels in the middle ranges presented protective effect relative to the lowest quartile with 45% decreased risk if in the 2nd quartile (urinary arsenic level (4.9 - 15.6), HR = 0.55, 95% CI [0.32, 0.97]), and 3rd quartile (urinary arsenic level (15.7 - 32.5), HR = 0.53, 95% CI [0.29, 0.98]) (Table 10).

Table 10: Confounding Effects on The Association between Effects of Four Treatment Groups on Deaths of All Causes

Characteristics	Categories	HR	p-value	95% Confidence Interval	
Age		1.05	< 0.001	1.03	1.08
Sex	Male	Ref.	-	-	-
	Female	omitted	-	-	-
Site	Araihazar	Ref.	-	-	-
	Matlab	1.87	0.021	1.10	3.18
BMI	< 18.5	Ref.	-	-	-
	18.5 - 24.9	0.99	0.955	0.62	1.58
	25+	0.17	0.086	0.02	1.28
Ever smoker	Yes	2.50	0.007	1.29	4.85
Occupation	daily laborer	Ref.	-	-	-
	farmer	0.9	0.736	0.49	1.65
	other	1.24	0.491	0.67	2.28
	unemployed, home maker, retired	1.04	0.927	0.47	2.31
Sun exposure	Mild	Ref.	-	-	-
	Heavy	0.96	0.891	0.57	1.64
Arsenic level in urine	1st quartile	Ref.	-	-	-
	2nd quartile	0.55	0.038	0.32	0.97
	3rd quartile	0.53	0.044	0.29	0.98
	4th quartile	0.61	0.224	0.27	1.36
Blood selenium	1st quartile	Ref.	-	-	-
	2nd quartile	1.05	0.846	0.63	1.75
	3rd quartile	0.93	0.807	0.51	1.68
	4th quartile	0.45	0.104	0.17	1.18
α-tocopherol	1st quartile	Ref.	-	-	-
	2nd quartile	0.98	0.957	0.54	1.80
	3rd quartile	0.65	0.184	0.34	1.23
	4th quartile	1.12	0.746	0.58	2.16
γ-tocopherol	1st quartile	Ref.	-	-	-
	2nd quartile	0.99	0.980	0.58	1.69
	3rd quartile	0.93	0.829	0.47	1.85
	4th quartile	0.81	0.585	0.38	1.74

For effect modification by the pre-specified characteristics, we performed our analyses for four treatment groups first, when statistically significant, we then repeat the analysis with factorial design two groups. Among all the potential effect modifiers, likelihood only indicated

that there is differential treatment effect on NMSC by plasma α-tocopherol, and the highest quartile seemed the driver of the significant association. Therefore, we combined the first three quartile as the reference group and discovered that in the lower quartiles, there were largely non-significant protective effect from selenium, vitamin E, and the combination groups relative to the placebo. Yet, in the highest quartile of α-tocopherol group, participants in vitamin E group revealed 3.24-time higher risk of getting NMSC compared to placebo (HR(VE) = 3.24, 95% CI [0.65, 2.82]) (Table 11). In further analysis for factorial design (Table 12), the significant association disappeared (HR(VE) = 1.35, 95% CI [1.07, 9.84]). Since the study population is rather homogenous in their age range, we did not assessment age as an effect modifier.

Table 11: Effect Modification of Baseline α-tocopherol for The Association Between Effects of Four Treatment Groups on Non-melanoma Skin Cancer

	Alphatoc_4 (1st-3rd quartile)	Alphatoc_4 (4th quartile)
Placebo	1	1
	-	-
SE HR (p)	0.67 (p = 0.211)	2.27 (p = 0.181)
95% CI	[0.35, 1.26]	[0.68, 7.54]
VE HR (p)	0.78 (p = 0.434)	3.24 (p = 0.038)
95% CI	[0.43, 1.44]	[1.07, 9.84]
Combo HR (p)	0.92 (p = 798)	0.98 (p = 0.973)
95% CI	[0.51, 1.68]	[0.24, 3.91]

Table 12: Effect Modification of Baseline α-tocopherol for The Association Between Treatment Effects by Factorial Design on Non-melanoma Skin Cancer

	Alphatoc_4 (1st-3rd quartile)	Alphatoc_4 (4th quartile)
No SE	1	1
	-	-
SE HR (p)	0.89 (p = 0.595)	0.73 (p = 0.399)
95% CI	[0.57, 1.38]	[0.35, 1.52]
No VE	1	1
	-	-
VE HR (p)	1.02 (p = 0.918)	1.35 (p = 0.420)
95% CI	[0.66, 1.60]	[0.65, 2.82]

DISCUSSION

In this study, we explored confounding effects and effect modification from participant characteristics in demographic, medical, and lifestyle collected at the baseline of this RCT.

In the assessment for confounding effects, all the factors were well-balanced attributable to randomization. However, factors such as age, sex, smoking status, occupation, level of sun exposure was unevenly distributed among NMSC cases and non-cases (Table 8). In the multivariate discrete time hazard model for four treatments and NMSC endpoint, we did not observe statistically significant confounding effects from any of these factors (Table 9).

When evaluating the confounding effects for the association between treatments and all-cause mortality, several factors stood out. Being older and smoking increased the risk of all-cause death compared to the counterparts. However, if participants' urinary arsenic level is in the range of 4.9 and 32.5, there is a protective effect against all-cause death (Table 10).

RCT is widely accepted as the "gold standard" for comparing treatment effects, and when randomization is conducted properly, it usually takes out the potential biases introduced by confounding factors in the allocation phase, thereby, produce unbiased effect estimates. Therefore, in the context of an RCT, adjustments of confounding factor for baseline covariates in the analysis is less common than in observational epidemiologic studies. BEST was designed as an RCT with the stratification factors of female and enrollment site. And based upon baseline characteristics distribution (Table 2), the randomization was conducted properly, and all the characteristics were well balanced. However, in current chapter, we still adjusted potential confounding factors aside from the stratification factors for three main reasons. First, we can effectively correct imbalance in baseline prognostic covariates despite randomization. Second, we can increase the power by modelling the variability in endpoint explained by the relationship with highly prognostic covariates. Third, by adjusting baseline characteristics, the effect estimates obtained would be more closely relevant to treatment effect on individual patients as opposed to average population [64–67].

In the multivariate Cox regression, being older and using cigarettes are two deleterious factors for death, whereas having arsenic level in urine in between 15.7- 32.5 is protective from death. This is largely consistent with published findings for the known risk factors of death.

The second part of this study was to examine the effect modification of baseline factors in both four treatment groups and factorial two groups design. When evaluating four treatment groups, participants in VE assignment with the highest quartile of alpha-tocopherol group presented 3.24-time higher risk of getting NMSC compared to placebo. In further analysis for factorial design, the significant association disappeared. We did not observe any differential treatment effects in terms of mortality endpoint.

The main controversy of having uniform treatment regimen without accounting for demographic, medical, lifestyle and molecular traits revolves around the relationship between antioxidants and NMSC. In this context, our findings that the contribution of high plasma alpha-tocopherol to increase risk of NMSC is clinically important. Although no significant association observed in our study, this in-depth evaluation within this high-risk population warranted future research for individualized treatment accounting for patient characteristics.

CONCLUSION

We conducted thorough evaluations of confounding effects and effect modification for the association between long-term supplementation of vitamin E and selenium with NMSC incidence and all-cause death. Our study revealed no confounding effects from all the factors, and differential treatment effects was observed in VE group for people with elevated level of baseline plasma α-tocopherol. However, this finding disappeared when in the evaluation of factorial analysis. Potential effects generated by these factors on NMSC and mortality outcomes warrant further evaluation.

3. ATTENTION-BASED DEEP NEURAL NETWORKS FOR DETECTION OF CANCEROUS AND PRECANCEROUS TISSUE FOR NON-MELANOMA SKIN CANCER ON HISTOPATHOLOGICAL IMAGES

BACKGROUND

Arsenic contamination of ground water in Bangladesh is considered to be the largest mass poisoning of a population in history by the World Health Organization (WHO), with estimated 35-77 million Bangladeshi people who have been chronically exposed to arsenic through drinking water [20,68,69]. Arsenic toxicity is closely dependent on the amount of ingestion, and once consumed, 40-60% arsenic is retained in human body and passes slowly through skin, resulting in skin malignancies, named nonmelanoma skin cancer (NMSC) [68,70]. Over 95% of NMSC cases consists of basal cell carcinoma (BCC) and cutaneous squamous cell carcinoma (SCC), the former is a slow growing locally invasive epidermal tumor, and the latter arises from dysplastic epidermal keratinocytes. SCC can be either in-situ (Bowen's disease) or invasive. Bowen's disease is generally considered a low-grade form of SCC with the reported risk of progression to invasive SCC at 3-5% [71,72], however, 20% of the tumor developed into invasive SCC eventually become metastatic [71]. Current evidence supports that the delay in detection as the main underlying cause for aggressive tumor behavior and subsequent morbidity in NMSC patients [71–73]. Hence, early detection is critical for controlling disease progression and could lead to substantially higher success rate in treatment. Usually, both types are readily identified by a pathologist in a timely manner, and in this situation, patients would benefit from timely treatment. In resource limited setting, however, such accurate and timely detection becomes

unaccountable, poses affected individuals to poor prognosis. In Bangladesh, as major attributable risk factor yet to be eradicated, people who are chronically exposed to arsenic through consuming arsenic-contaminated water are deemed at high risk of NMSC. Meanwhile, the cancer detection and diagnostic facility infrastructure is still underway, resulting in delayed treatment and poor disease prognosis for the affected population [74].

Over the past two decades, dramatic advancement in computational power and improvement in machine learning algorithms have led to the development of powerful computer-assisted approaches for image analysis. Among applications in various medical fields, Computer-Aided Diagnosis (CADx) in analyzing skin lesions is fast growing owing to the disease nature and common cutaneous manifestation [75]. The most common approach for analyzing histopathology whole-slide images involves first performing tissue or tumor detection, and then generating tiles of a certain size and magnification from the whole-slide image, which are dictated by computational requirements. For classification models, predictions are usually performed on the tiles, which is followed by an aggregation step where tile-level predictions are combined to produce an inference for the whole-slide image [76]. Thus, for this approach, a pathologist must develop a heuristic to aggregate tile predictions into a whole-slide diagnosis.

This approach has several limitations. First, each tile must be assigned a diagnosis by the pathologist, which is expensive in time and resources. Second, during classification, each tile is analyzed independently of their neighbors without considering relationship in between neighboring tiles. Together with the large volume of NMSC cases, the high demand in manual work by the pathologist, and potential inaccurate diagnosis caused by intra- and inter- observer variations, this application becomes suboptimal in resource limited setting [77–79].

We propose an efficient deep learning framework for achieving accurate diagnosis and interpretable detection of NMSC from hematoxylin and eosin-stained (H&E) images. Our approach is inspired by Ilse et al. [80], which utilizes an attention-based multiple instance learning (MIL) method to integrate contextual information across multiple tiles in whole-slide images. Moreover, the attention mechanism can be interrogated to produce visually interpretable predictions. Generally speaking, our model reads each NMSC image as a bag of instances that can be interpreted from small tiles and learns to predict an individual status of NMSC (e.g., BCC, Bowen's disease, SCC, or non-cancer) assigned to the entire H&E image. Further, we obtain crucial instances that can reveal the location of cancer cells, which improves the interpretability and visualization. Our model applies the attention-based MIL pooling strategy that gives insight into every instance's contribution on the bag label. As such, it could better facilitate early and more accessible cancer detection in resource limited settings.

METHODS

Data Source

This study utilized histopathological images collected from the parent study, Bangladesh Vitamin E and Selenium Trial (BEST), which included 7000 participants from two regions (Araihazar and Matlab) in Bangladesh for six years on the status of NMSC. Incidence of NMSC was identified at each biennial in-person follow-up exam. Participants who had undergone all three levels of evaluation were eligible for skin biopsies of their suspicious lesions. Smaller lesions (diameter < 5 mm) were punch-biopsied, and large lesions (diameter ≥ 5 mm in diameter) were excised. Among individuals with multiple biopsy-eligible lesions, the most likely malignant lesion was biopsied. When an individual developed lesions at two different timepoints, two

biopsies were collected. Formalin-fixed biopsy tissues were processed at a specialized surgical pathology laboratory in Dhaka, Bangladesh, and made into H&E slides. Subsequently, these slides were transported to the University of Chicago Medical Center (UCMC), scanned by a Leica Aperio ScanScope XT at 20X magnification scanner and saved as the SVS format for analysis. In total, 2828 whole-slide images were collected. In total, we scanned 2084 histopathological slides from 455 participants as some participants were referred to biopsy procedure multiple times during the entire six-year follow up, and each biopsy generated on average of four slides. We randomly partitioned about three fourths of these images (n = 2084) for initial model training and testing, set aside the remaining one fourth (n = 744) as an independent set for independent model testing.

Data pre-processing:

To apply a deep learning classifier, the images were divided into several thousand non-overlapping tiles; the tile size was set as 1000 x 1000 pixels under 20x magnification, then resized to 224 x 224 pixels prior to feeding into ResNet18. In processing the selection of tiles, tiles too light (intensity > 0.85) or too dark (intensity < 0.15) were excluded (i.e., the tile color is too light when the tile region does not contain any tissue or contains fat tissues; the tile color is too dark when the selected tile region contains too much ink, folded tissues, or other image artifacts).

Owing to the rarity of cancer cells in each slide, we applied several procedures for data augmentation, like other related digital pathology works [81]. First, since there are no canonical orientations for histopathological slides, we rotated the images in eight orientations. Next, we

applied stain perturbation, a technique that is specifically tailored for H&E image augmentation [82]. Finally, we added color jittering on the brightness, contrast, saturation, and hue of each image.

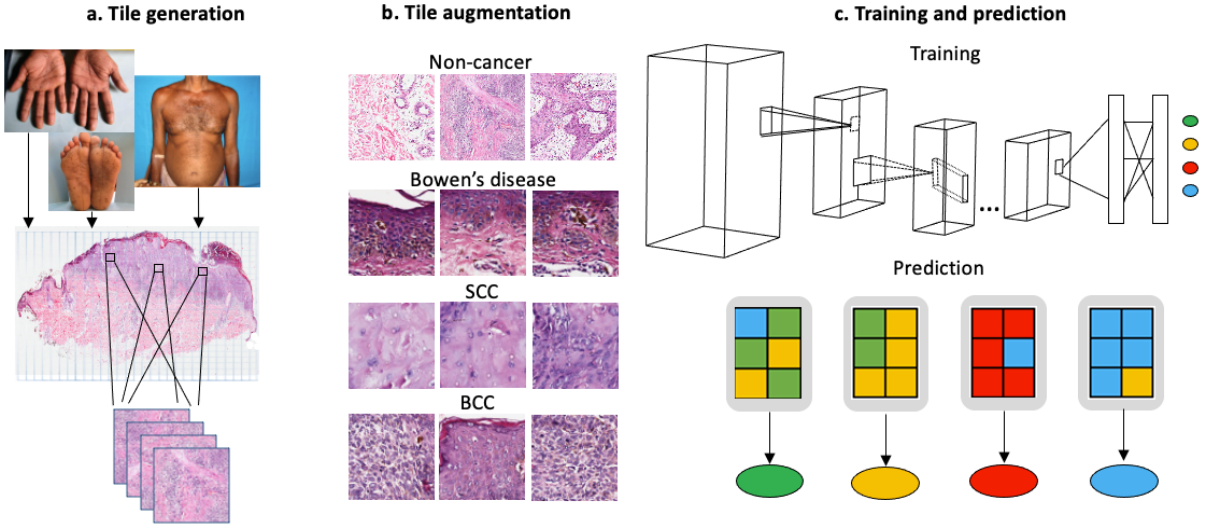
Ascertainment of NMSC

All whole-slide images were given a diagnosis by a dermatopathologist (Dr. C.S.) from the Department of Dermatology at UCMC. The diagnosis included whether there is the indication for NMSC, if there is, what is the histology subtype of NMSC (Figure 2).

Analyses Framework

There were two steps to achieve the proposed attention-based model (Figure 1). First, we used a convolutional neural network (CNN) to extract image-based features. In this step, tiles generated from the whole slide were analyzed to generate a feature vector. Second, we applied the attention mechanism on the extracted feature vectors across a set of tiles and generated a single whole-slide classification. The CNN, serving as the feature extractor, was jointly optimized with the attention module in an end-to-end manner, using a cross-entropy loss function to model the pathologist labeled diagnoses and guide the optimization of all parameters in the network, including both CNN and the attention module.

Figure 5. Overview of The Proposal Attention-based Model



Feature extraction through ResNet18

In our study, we leveraged deep residual network ResNet18 [83], a type of CNN that uses residual blocks to achieve promising performance on image recognition benchmarks. For each biopsy during training or testing, we randomly sampled tiles without replacement from all slides of each biopsy and fed them to the model. Subsequently, train-validation-test split for the entire sample and 5-fold cross-validation was conducted. In this process, slides from multiple biopsies of the same person were put into the same fold. This process guaranteed the model, either in training, validating, or testing, only "sees" data from one person once so that the model does not memorize latent information of this person. Finally, the CNN model extracted image features from each tile, and outputs a one-dimensional feature vector representation on the tile level, which is a high-level feature expression from the original image.

Attention-based classification

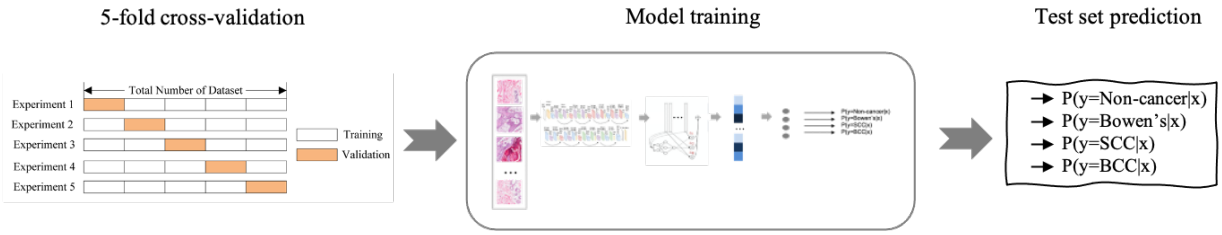
In deep learning, embedding attention mechanisms is analogous to mimicking human brain actions by focusing on the important details among heterogeneous information. After feature extraction, we applied the attention module to the feature map with their weights determining the importance of each tile for diagnosis. The importance of each tile was estimated based on features extracted from the tiles. Next, this information was aggregated in the attention module and an attention vector was created for whole-slide classification through a fully connected layer. The confusion matrix and area under the curve (AUC) were computed to evaluate model discrimination for the presence/absence of NMSC subtypes. In addition, the attention map illustrating each tile's importance in biopsy-level prediction, indicating high and low likelihood of containing cancer cells. The model was implemented in PyTorch [84].

Validation

5-fold cross-validation

We tested our model in 5-fold cross-validation manner (Figure 6). We chose three folds for training, one fold for validation, and one fold for testing. Under each assignment, we trained the models four times with Adam optimization method with learning rate=1e-4, $\beta 1=0.9$, $\beta 2=0.999$ and $\epsilon=1e-8$.

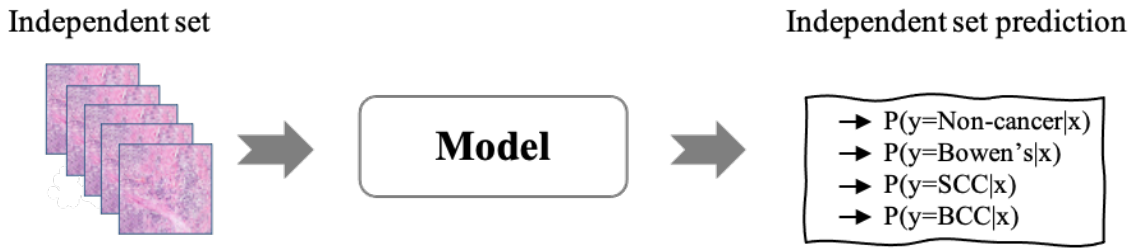
Figure 6: Training Set 5-fold Cross-Validation



Independent set validation

We set aside a dataset of 774 images from an independent group of patients for secondary model validation after model building (Figure 7). We followed the same pipeline to pre-process the dataset into tiles, augmented and balanced the data before testing. For all the 20 checkpoints of each model, we combined all the models using ensemble method, and tested the performance on the test dataset. Average AUCs with 95% confidence intervals (CIs) for each classification were computed, receiver operating curves (ROCs) were generated to illustrate the performance. Confusion matrices were also created to show the correctly and incorrectly classified cases under 4-class and 3-class classifications.

Figure 7: Independent Set Model Validation



Physician validation and model interpretation

A subset of 64 images were annotated by the dermatopathologist (Dr. C.S.) with contouring of each ROI in the image, who was masked from the results generated by the model. We then compared the model-generated ROIs against annotators by the physician.

Batch effect assessment

To evaluate whether there was batch effect between the training set and the independent testing set during random data partition, we output 100 features for each dataset, then applied t-distributed stochastic neighbor embedding (t-SNE) [85] for non-linear dimensionality reduction and visualize the distribution on two dimensions. In the visualization, every sample of the high-dimensional feature map was given a location in a two-dimensional map. The perplexity of the t-SNE algorithm was set to 30 and the maximum number of iterations was set to 1000 before reducing the dimensions of features into two dimensions.

RESULTS

The main data set contained a total of 2,084 H&E images generated from 553 patients, of which 229 specimens (41.4%) were in the non-cancer class, 171 (30.9%) were in the Bowen's

disease class, 37 (6.7%) were in the SCC class, and 116 (21.0%) were in the BCC class. Separately, 744 images from 248 patients were placed in an independent testing set, 95 (38.3%) non-cancer, 110 (44.4%) Bowen's disease, 3 (1.2%) SCC, and 40 (16.1%) BCC to validate the trained model and evaluate performance (Table 10).

Table 13. Class Distribution of Images by Non-melanoma Skin Cancer Subtypes on Specimen Level

	Training set	Independent set
Non-cancer, n (%)	229 (41.4%)	95 (38.3%)
Bowen's, n (%)	171 (30.9%)	110 (44.4%)
SCC, n (%)	37 (6.7%)	3 (1.2%)
BCC, n (%)	116 (21.0%)	40 (16.1%)
Total specimen	553	248
Total images	2084	744
Total individuals	455	243

The 5-fold cross validation model performance on classification of each cancer type was demonstrated in Table 11 and Table 12. For this classification, our model achieved 89% (95% CI [0.84, 0.93]) overall macro AUC, with SCC performing the best (AUC = 0.98, 95% CI [0.94, 1.00]), followed by BCC (AUC = 0.93, 95% CI [0.88, 0.98]), non-cancer (AUC = 0.84, 95% CI [0.72, 0.95]) and Bowen's disease (AUC = 0.80, 95% CI [0.71, 0.88]) (Table 11). In general, SCC and BCC are histopathologically well defined, whereas Bowen's disease is thought to be an intermediate transition between to SCC, and non-cancer also contains pre-cancer stages, such as arsenical and actinic keratoses. These equivocal stages challenged the performance on our classification of non-cancer and Bowen's disease.

Table 14: Area Under the Curve for Model Prediction with 5-fold Cross-Validation for Four Classes Using Training Dataset

	AUC	95% CI
Non-cancer	0.84	[0.72, 0.95]
Bowen's	0.80	[0.71, 0.88]
SCC	0.98	[0.94, 1.00]
BCC	0.93	[0.88, 0.98]
macro AUC	0.89	[0.84, 0.93]

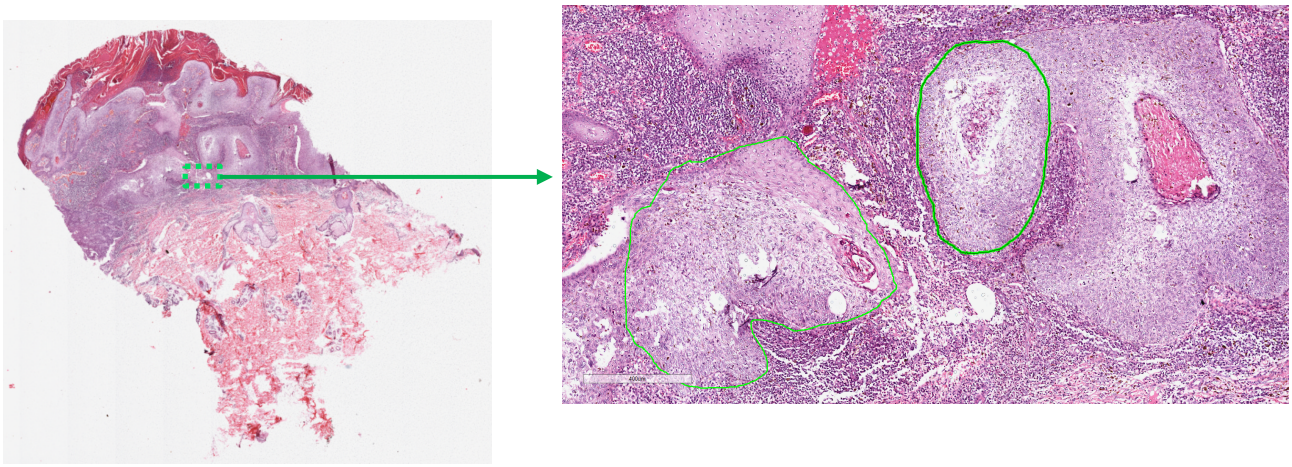
Clinically, since the treatment for Bowen's has a clear distinction to treatment for SCC, we combined Bowen's disease into the non-cancer class to perform 3-class classification. AUCs were improved to 93% (95% CI [0.88, 0.98]) all the entire 3-class classification, and to 90% (95% CI [0.83, 0.97]) for non-cancer & Bowen's disease (Table 12).

Table 15: Area Under the Curve for Model Prediction with 5-fold Cross-Validation for Three Classes Using Training Dataset

	AUC	95% CI
Non-cancer & Bowen's	0.90	[0.83, 0.97]
SCC	0.98	[0.95, 1.00]
BCC	0.91	[0.84, 0.98]
macro AUC	0.93	[0.88, 0.98]

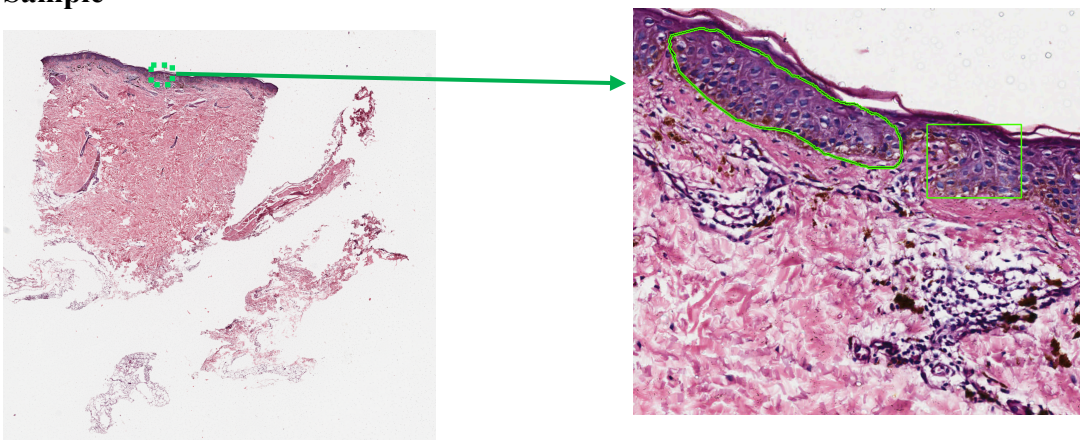
The attention heatmaps were generated for all the testing images to visualize the key instances. Characteristic examples for Bowen's disease, invasive SCC and BCC are presented in Figure 8-10. Different shades of color indicate the distributions of the attention weights – light tiles represent the presence of the cancer features identified by the attention module, in contrast, dark color shows the absence of key information in the tile.

Figure 8: Typical Example of Squamous Cell Carcinoma in Study Sample



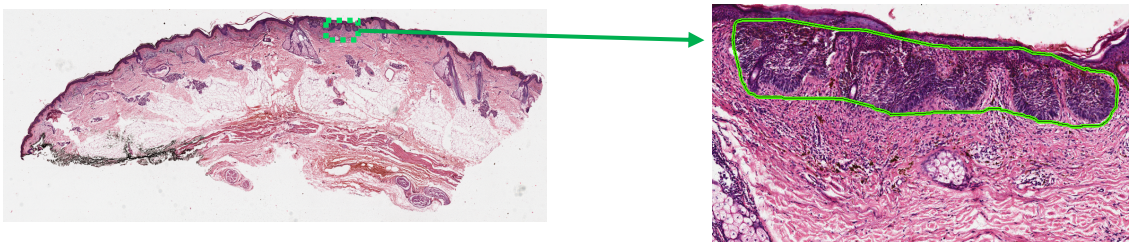
This example shows invasive squamous cell carcinoma (SCC) from an excisional biopsy. These lesions demonstrate early keratinocyte invasion of the dermis. Well-differentiated lesions show prominent keratinization and may form "pearl-like" structures where dermal nests of keratinocytes attempt to mature in a layered fashion (5X).

Figure 9: Typical Example of Bowen's Disease/Squamous Cell Carcinoma in-situ in Study Sample



There is prominent dyskeratosis and aberrant mitoses at all levels of the epidermis, along with marked parakeratosis. The basement membrane remains intact (10X).

Figure 10: Typical Example of Basal Cell Carcinoma in Study Sample



The lesions arise from the epidermis with nests of basaloid cells with scant cytoplasm and elongated hyperchromatic nuclei, basal palisade, and peritumoral retraction clefts (5X).

We then evaluated our model with an independent dataset of 774 images. The distribution on correctly and incorrectly classified cases were shown in Table 13 and Table 14. Each row shows the cases diagnosed by the dermatopathologist, and each column indicates the predicted results by our model. Our results showed that significant amount of Bowen's disease cases was misclassified into non-cancer category, but the classification improved greatly when we

combined Bowen's disease into the non-cancer class. Receiver operating curves (ROCs) were generated, and average AUCs were computed for each class and suggested 70.3% for non-cancer, 69.8% for Bowen's disease, and 96.4% for BCC (Table 13). After combining non-cancer and Bowen's disease, the AUC improved to 94.4%. SCC results were omitted due to the small number of samples (Table 14).

Table 16: Area Under the Curve for Model Prediction of Four Classes Using Independent Dataset

Ground truth	Prediction			
	Non-cancer	Bowen's	SCC	BCC
Non-cancer	87	2	2	5
Bowen's	99	6	0	5
SCC	2	0	1	0
BCC	9	0	0	31

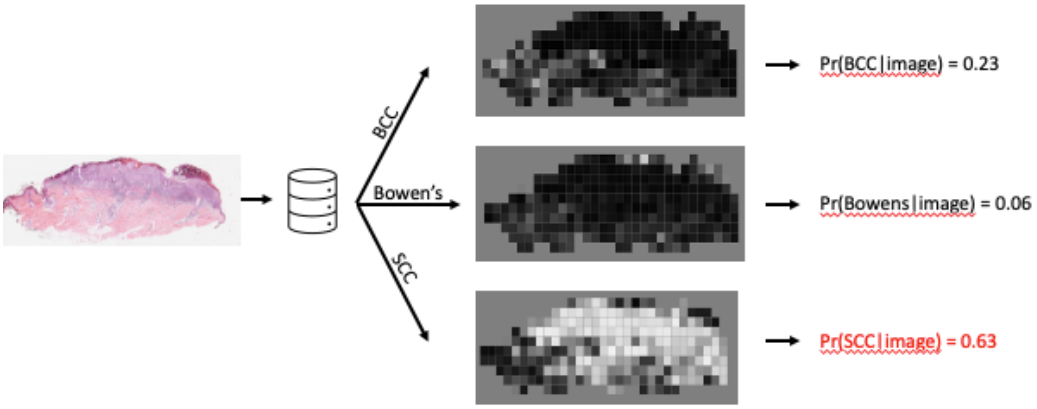
Table 17: Area Under the Curve for Model Prediction of Three Classes Using Independent Dataset

Ground truth	Prediction		
	Non-cancer & Bowen's	SCC	BCC
Non-cancer & Bowen's	182	2	20
SCC	1	2	0
BCC	4	0	36

To validate the results on the heatmaps, we compared the key instances on the attention against the ROI contoured by the dermatopathologist (Figure 11-13). Lighter tiles suggest higher chances of having cancerous features, whereas darker tiles indicate lower chance of having cancerous features. Expert evaluation of the examples demonstrates broadly coherent and interpretable results. In the SCC example (Figure 11), the predicted visualization picked up the presence of infiltrative cells passing through the basement membrane into the dermis. Under 20X, we could also observe the involvement of full thickness epidermal atypia and hair follicles.

In the BCC example (Figure 13), the predicted visualization illustrates the involvement starts from the epidermis with palisade and clefts forming from the adjacent tumor stroma. The nuclei in the center become crowded with scattered mitotic figures. Presence of mucinous stroma can also be seen in many locations. Last, in the Bowen's disease example, the predicted visualization successfully identified hyperkeratosis and parakeratosis. The atypia was observed spread out the full thickness of the epidermis, the keratinocytes showing intense mitotic activity, enlarged nuclei, and pleomorphism.

Figure 11: Examples of Visualized Heatmaps for Squamous Cell Carcinoma



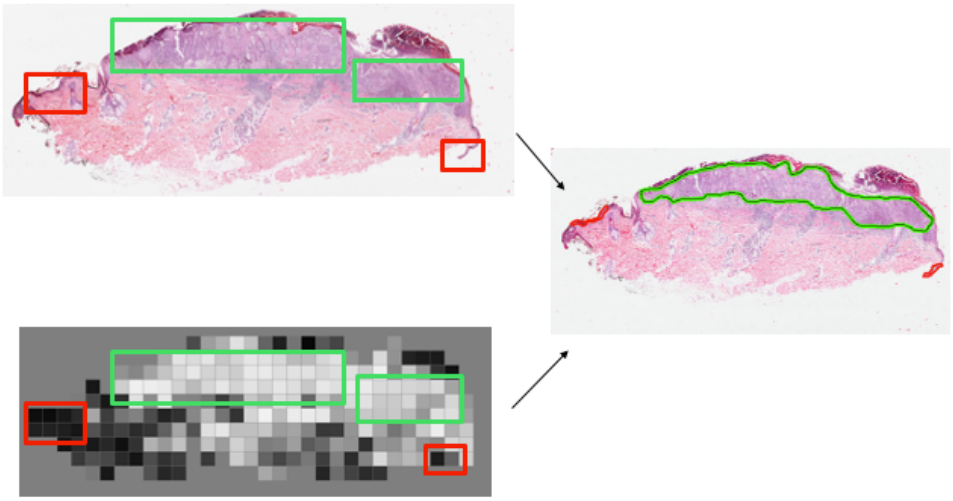


Figure 12: Examples of Visualized Heatmaps for Bowen's Disease

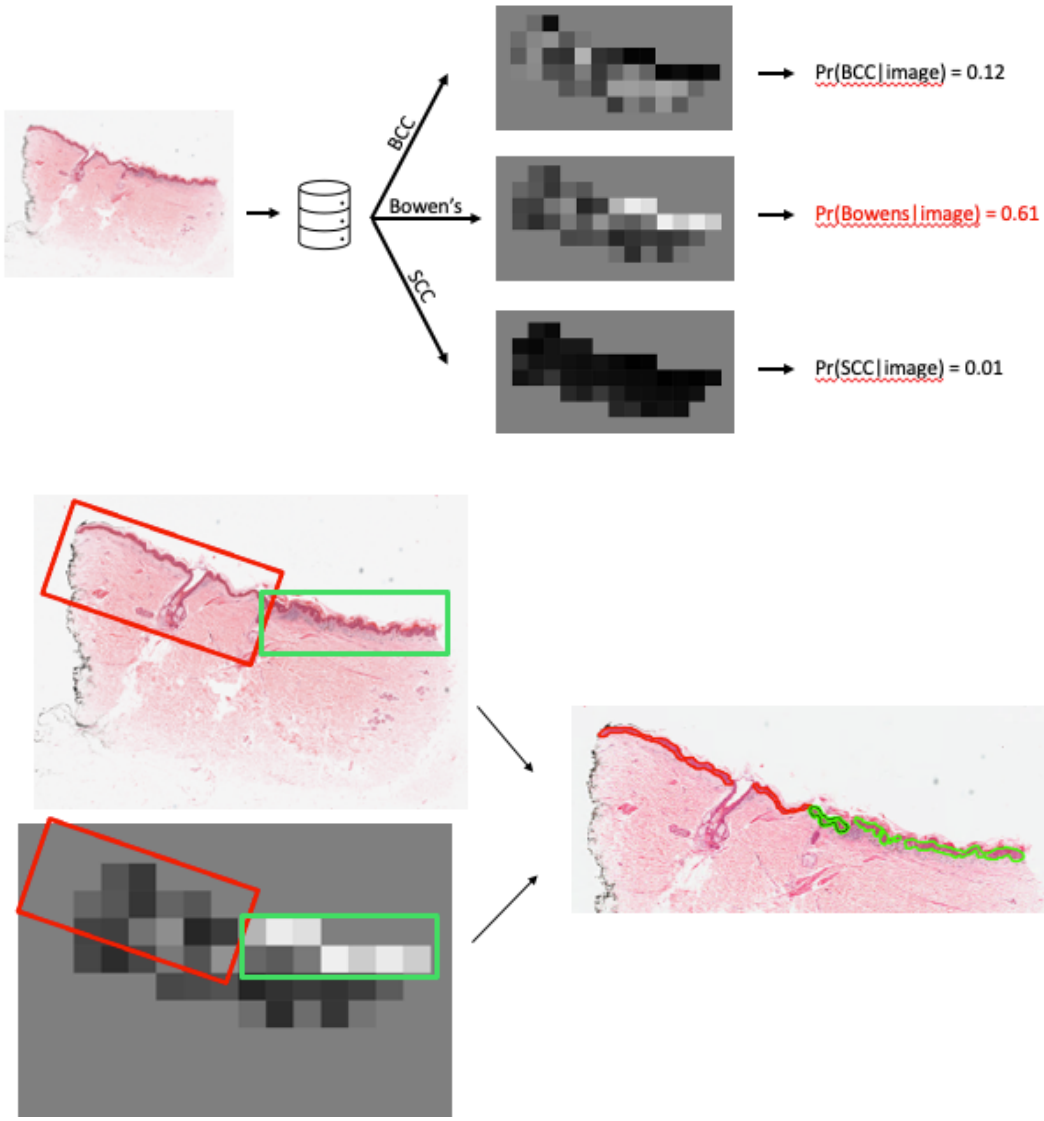
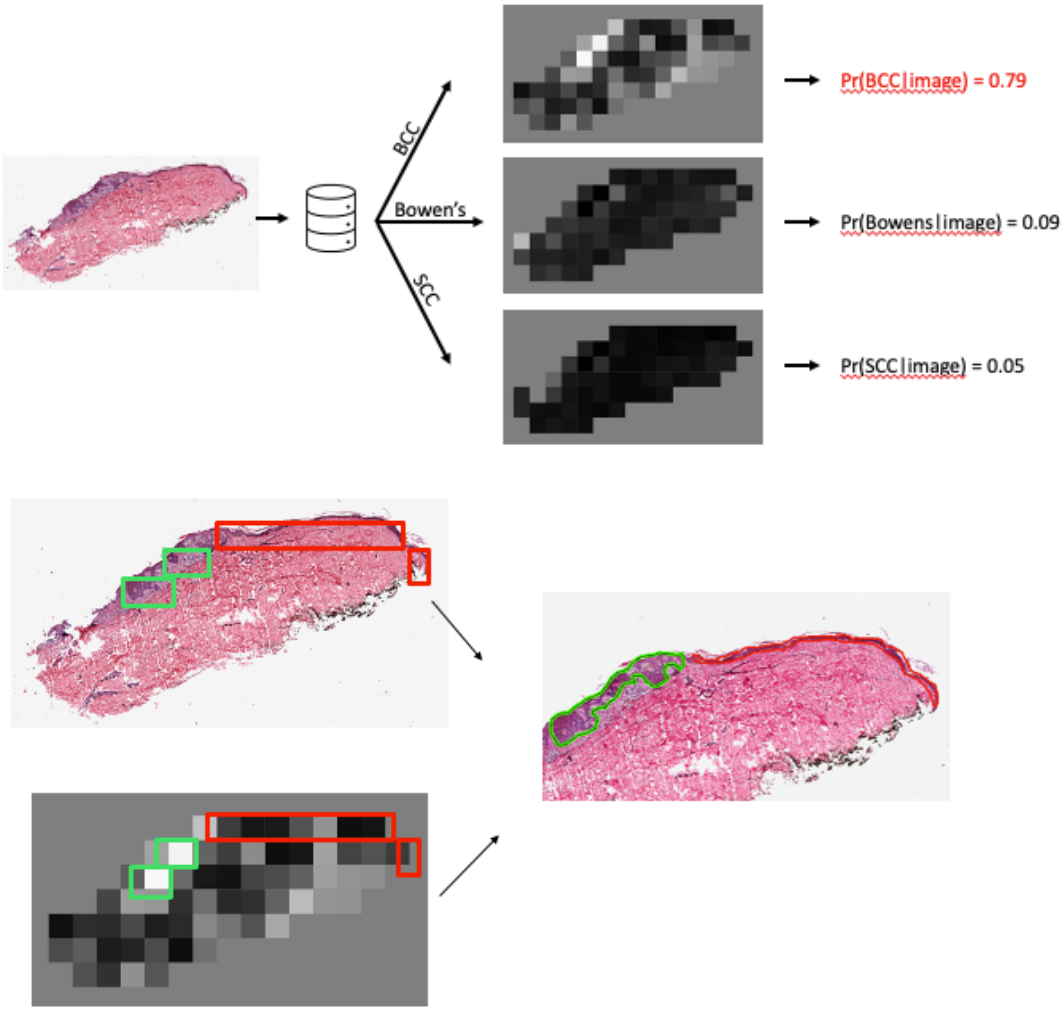


Figure 13: Examples of Visualized Heatmaps for Basal Cell Carcinoma



Our results also indicated a non-negligible misclassification rate between non-cancer and Bowen's (Table 11-12). In an effort to investigate the misclassified Bowen's disease cases, 11 images were selected for the dermatopathologist to scrutinize. This investigation yielded insights that Bowen's cases which were mostly identified as non-cancer contain atypical keratinocytes from the basal cell layer extending into granular and cornified layers. In some cases, the dying granular layer as well as the involvement of the full thickness of epidermis were not identified in the heatmap as signs for cancer.

Table 18: Area Under the Curve for Model Prediction with 5-fold Cross-Validation for Four Classes Using Training Dataset

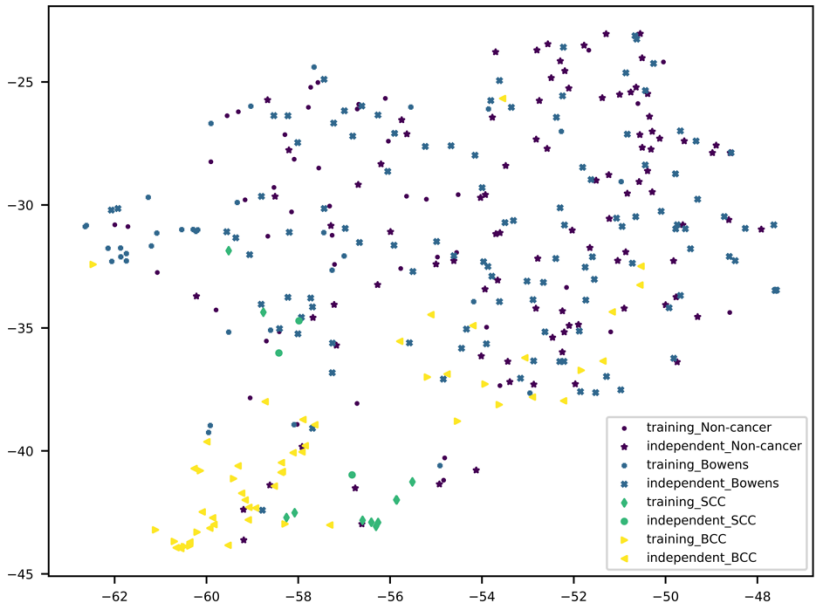
Ground truth	Prediction			
	Non-cancer	Bowen's	SCC	BCC
Non-cancer	87	2	2	5
Bowen's	99	6	0	5
SCC	2	0	1	0
BCC	9	0	0	31

Table 19: Area Under the Curve for Model Prediction with 5-fold Cross-Validation for Three Classes Using Training Dataset

Ground truth	Prediction		
	Non-cancer & Bowen's	SCC	BCC
Non-cancer & Bowen's	182	2	20
SCC	1	2	0
BCC	4	0	36

Last, to evaluate the potential of batch effects for training and independent datasets was evaluated (Figure 14). One hundred features extracted from CNN for both datasets were outputted and visualized by t-SNE with four color-coded labels. In the 2D embedding figure, every point represents a sample image, and the axis represents the t-SNE dimension [85], the overlapping phenomena for each label between training and independent sets indicates similar image features extracted, therefore, confirmed that the random partition of the dataset did not exhibit unexpected batch effects.

Figure 14: t-SNE to Visualize Batch Effect



DISCUSSION

Our study successfully demonstrated comparable accuracy in differentiating NMSC subtypes from non-cancer tissue with an attention-based CNN. In the 4-class classification, our model achieved over 80% AUC for all classes, with SCC and BCC over 90%; when we combined non-cancer and Bowen's disease into one class, our model achieved over 90% for all classes. The testing results on the independent dataset showed consistent AUC for non-cancer, Bowen's disease and BCC (Figure 15-16). The comparison of heatmap with key areas that drive the cancer diagnosis and physician annotators also indicate promising consistency.

Figure 15: Area Under the Curve for The Independent Set Using Ensemble Method for Four Classes

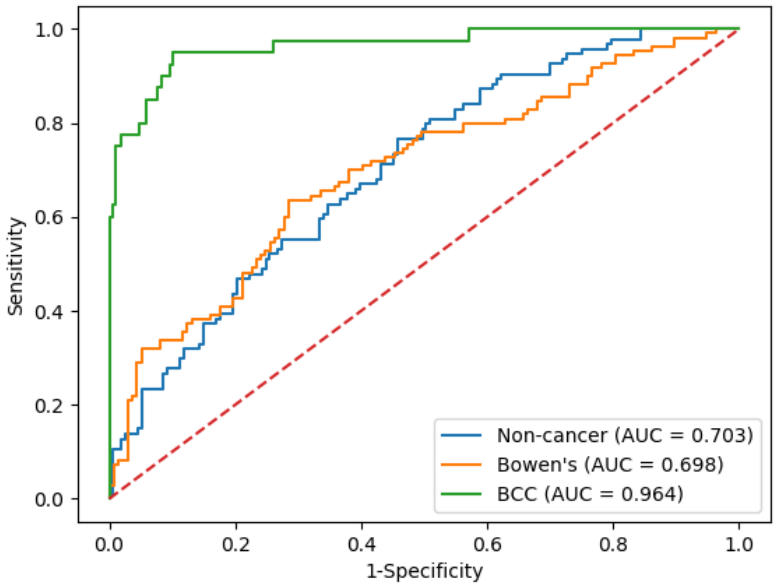
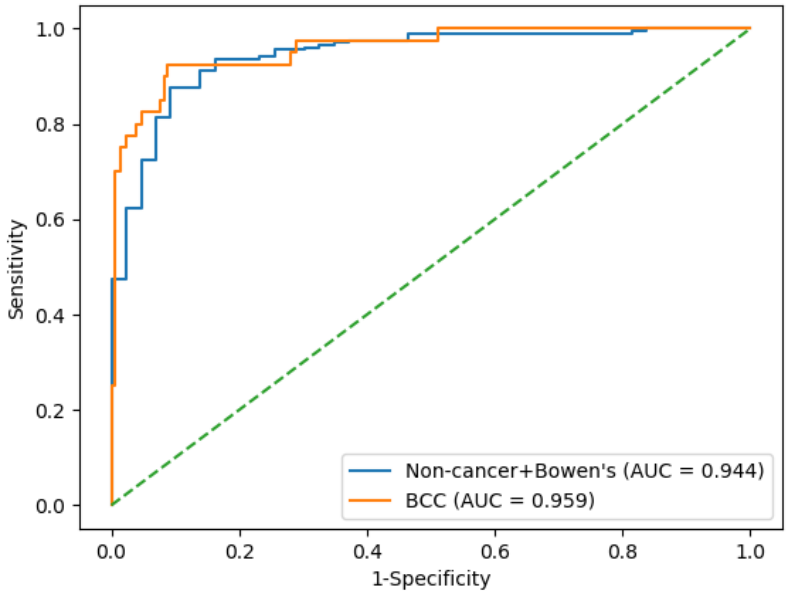


Figure 16: Area Under the Curve for The Independent Set Using Ensemble Method for Three Classes



In the validation with the independent set, we observed pronounced misclassification between non-cancer and Bowen's disease. Among 110 Bowen's disease cases, only six were correctly classified, while 99 were misclassified as non-cancer (Table 3). Histopathologically, Bowen's disease often presents hyperkeratosis and parakeratosis. These are usually accompanied by marked acanthosis with elongation and thickening of the rete ridges. The keratinocytic cells might be saturated and often highly atypical. Unlike arsenical keratosis and actinic keratosis that we lumped into non-cancer category, the atypia spread the full thickness of the epidermis, and the keratinocytes demonstrating intense mitotic activity, pleomorphism, and greatly enlarged nuclei. We may also observe a loss of maturity and polarity, giving the epidermis a disordered or "windblown" appearance. Occasionally, cells of the upper epidermis will undergo vacuolization, demonstrating an abundant and strongly eosinophilic cytoplasm [86].

In a typical case, the whole epidermis is involved and sometimes epithelium of the pilosebaceous glands is affected. The non-cancer cells are usually replaced by modified keratinocytes with loss of orientation and hyper-chromatic bigger nuclei with disorderly aspect of the epidermis [87]. In arsenic-induced Bowen's disease, modified cells with large number of vacuoles are also present [88]. However, these features are not unique to Bowen's disease only, which potentially explained the misclassification in our study. Also, the patterns that occur in small amounts made it more challenging to differentiate even for the skilled pathologist. Hence, these patterns can be interpreted differently or easily overlooked, leading to higher chances of misclassification.

Our attention-based model can effectively localize the ROIs that contain key information in the image, which means the model can potentially pick out the tiles that contain cancerous cells and assign more weights to those tiles. Further, this information can be utilized to visualize

the location of lesion in the image. Thus, analogous to how pathologists scrutinize slides under the microscope, the model uses weighted features from the entire slide to classify images. Last, the prediction from our model can be tested for potential relationship with patient outcomes. When only small number of microscopic patterns are present, they can be easily concealed on the whole-slide image, resulting in the overlook of the associative poor prognosis of the patient. Further evaluation of our model's prediction for those who had unexpected worse prognosis on NMSC could shed light on elusive histopathological patterns that are easily missed by the pathologists.

Although our model is rooted in strong deep learning methodology and achieved pathologist level performance on testing sets, there are limitations to be acknowledged and addressed. First, all the slides were collected from a single medical setting in Bangladesh, so our data may not be representative of all NMSC histopathological patterns. Second, all the images were scanned with the same equipment at UCMC, therefore, our results may be subject to errors in samples generated with different scanners. Third, in our data collection, when a patient came to the doctor with multiple suspicious lesions, only the most likely malignant lesion was biopsied. Given that it is not uncommon for a patient to have multiple NMSC lesions in a single visit, the case estimates in our study might be under-reported. Third, images used in this study was from the parent interventional trial, suggesting that the histopathological changes might have been more severe had it been from a non-interventional setting. However, based on the in-depth evaluation of the parent trial, there is no statistically significant treatment effect in all treatment arms. Hence, it's reasonable to generalize current finding to affected individuals in Bangladesh. Last, although our whole-slide images are of high resolution, and image augmentation procedure was applied to generate and balance training samples, our dataset is still relatively small in size

compared to classical deep learning datasets with more than ten thousand samples for each class. To evaluate the robustness and generalizability of current work, further validation and large datasets are needed.

An automated system for detection and visualizing histopathological patterns of NMSC has a wide variety of applications in clinical settings. Given the quick turnaround of our model, it could potentially be integrated into existing clinical management systems to automatically pre-populate diagnosis for histopathological patterns on slides and provide complementary opinion on challenging cases for the pathologists. Also, visualization of the entire slide after being examined by our model at tile-level could highlight elusive areas of suspected patterns and the identified cancer cells. In addition, our model could significantly expedite the cancer diagnosis process for the pathologist, free up time for them to focus on more cases and make cancer detection more accessible when medical resources are in shortage.

In Bangladesh, the problem of arsenic exposure in drinking water has started since the 1970s and a considerable amount of time has gone by before an effective remedial solution to be developed, as a result, many people have become affected and suffer the debilitating symptoms. Yet, without a timely diagnosis, these affected individuals mostly are unaware of the need to seek medical attention, needless to say proper treatments [89]. The participants included in our study only represent a small portion of the total affected population, the rest who developed cancerous indication may have never presented themselves to medical professionals. The potential big number of cases is very alarming and could be the tip of the iceberg of the real problem. Future studies should be targeted at developing nation-wide cancer detection protocol in different tertiary medical facilities.

CONCLUSION

In this diagnostic study, we developed an attention-based model for high-resolution histopathological image analysis. Our results showed promising diagnostic accuracy for two clinically meaningful classes (i.e., non-cancer and BCC) for nonmelanoma skin cancer. This fast, scalable method can be incorporated into the hospital clinical management systems and holds the potential for substantial clinical impact, including expediting healthcare delivery, broadening the scope of primary care practice, and augmenting clinical decision-making for dermato-oncology specialists particularly in areas with limited resources. Further research is necessary to evaluate performance in a real-world clinical setting, in order to validate this approach across the full spectrum of NMSC lesions encountered in typical practice. While we acknowledge that a physician's clinical diagnosis is also based on contextual factors beyond microscopic inspection of a lesion, the ability to classify histopathological images with the physician comparable accuracy has the potential to profoundly expand access to medical care.

SUMMARY AND FUTURE DIRECTIONS

Machine learning for population health and healthcare delivery

As machine learning tools continue to demonstrate better performance in handling large volume and mixed data, they enabled us to pursue answers in the population health problems previously deemed impossible. Being part of the effort for Precision Public Health (PPH), first introduced by Khoury in academic literature, this modernization of surveillance, epidemiology, information systems, targeted interventions would potentially benefit population health in a broader extent [4].

One powerful dimension under this effort is to pool various formats of the data together. This integration of datasets is merely trivial. When implemented well, the linkage of previously siloed datasets embraces new possibilities for discovering genetic, biological, and clinical associations that might explain disease pathogenesis and progression. It also enables the evaluation of the treatment effects on disease outcomes through stratified patient characteristics.

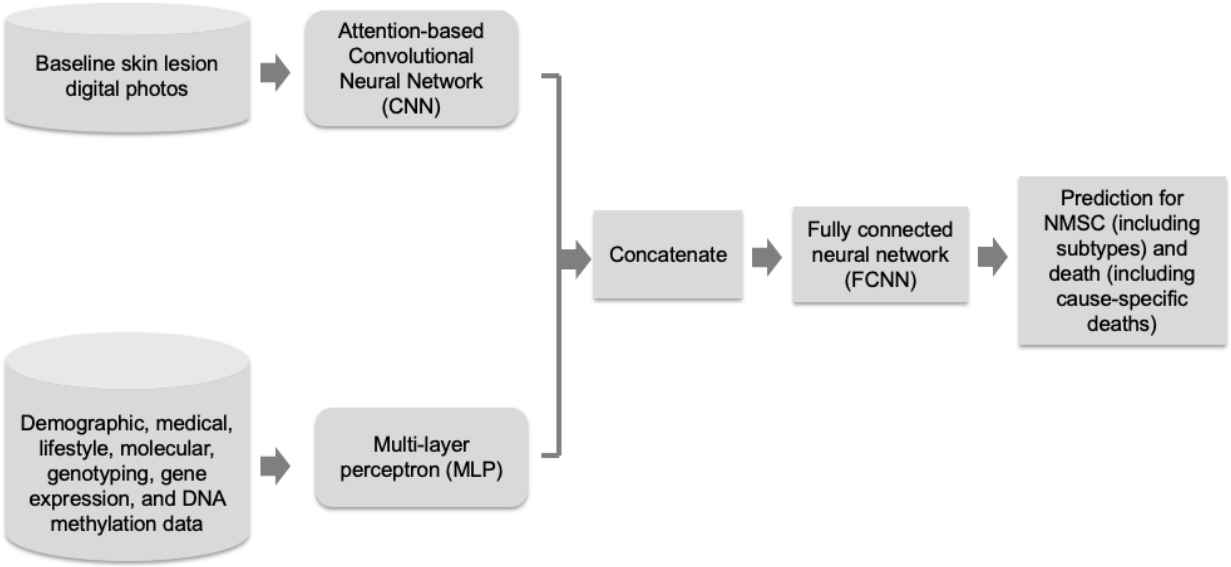
These implementations are now possible owing to the unique features in machine learning. For instance, the flexibility of the neural network made it possible to combine demographic, lifestyle, medical, molecular, and imaging data together, resulting in complex and powerful models to reveal early indication in the medical image and the disease incidence. This machine learning infrastructure is also renowned for its efficiency - enabling the same algorithm to be applied with minimal modifications to address other clinical problems using different datasets [90,91].

Public health researchers collect data from various sources, perform analysis to estimate the incidence and prevalence of different health conditions, and the associations with related risk factors. To present a concerted effort of population health, machine learning methods extract

data in a whole different level of granularity, integrate them together, and harmonize the evidence of populations, communities, and environment into the promise of population health advancements.

For future direction, a platform may be developed through two paralleled pathways (Figure 17). First, baseline skin lesion digital photos will be processed through a CNN for feature extraction and then through an attention module for weights calculation. The ability that CNN models to learn predictive features from raw images is a paradigm shift that presents opportunities for medical imaging. Through this process, the images and patient diagnosis/outcome are presented to a network that contains interconnected layers aiming to highlight important patterns from the images. Second, all other patient information will go through another deep network, through which important information will be extracted and aggregated to represent non-image aspects of patient characteristics. Both pathways will then join together to represent a complete profile of each participant for endpoint prediction.

Figure 17. Platform development workflow



Predicting the patients' outcome is a critical step in treatment. Advancement in tabular data analytics and imaging technologies provide physicians with new possibilities, yet patient prognosis remains largely subjective, leading to suboptimal clinical management. In this aim, we plan to develop a computational platform based on deep learning models to predict the incidence of NMSC and the mortalities from digitized images of participants' cutaneous manifestation, demographic, lifestyle, molecular, and genomic features. This method simultaneously learns the image patterns and participants' characteristics in a tabular format that are associated with the outcome.

Machine learning models and neural networks allow the input of structured data (in the form of numeric and categorical) and achieved decent performance. CNNs handle the input of unstructured data, for example, image data or sentences, and is renowned for its superior performance. Mechanistically, we will bridge the advantages of these two approaches and integrate both structured and unstructured data together for the prediction of NMSC types and mortality outcomes.

In spite of the excitement for the use of machine learning models, the promise of this self-learning, continuously advancing tools needs to be tempered against the challenges when implementing them in the routine clinical practice. Broadly speaking, there are three aspects we need to consider – the relationship between healthcare data and machine learning, the relationship between machine learning and healthcare provider, and the governance platform to enable equitable use of data in clinical practice [15].

First, machine learning is essentially the interplay between datasets and one or a group of algorithms, the most popular and perhaps the current state-of-the-art is deep learning algorithms.

The advent of deep learning can be traced back in decades, but they gained significant popularity in 2012, when highly accurate training was brought the hub of attention [92]. Healthcare dataset, on the other hand, is notoriously in large volume, messy and complex forms. To make the training of a model accurate, data must be pre-processed and curated. For example, if one medication abbreviation was misread by the model and mis-interpreted as another one, and the data is used to train the model to suggest treatment for future patients, the model might erroneously give out recommendations leading to disastrous outcome. Hence, machine learning is like any tool, to make it effective, the problems exist in the entire clinical workflow, involving patient journey, healthcare delivery, payer systems and any possibilities in each step of the intricate flow, should be both the driver and the reference point for the implementation of machine learning [93,94].

Another important consideration is the relationship between machine learning implementation and healthcare practitioners. In agreement with current mainstream publications, we believe that the role of machine learning is to augment the capability and capacity of human doctors rather than to replace them [95,96]. In healthcare delivery, the emotional virtues serve as one core element, through this process, patients receive attention, compassion, empathy, and care. Machines are capable of taking over tasks that are more routine and standardized, therefore, freeing up the time for the human doctors to focus on tasks demand intuition, judgement, and emotions [97–99]. In this process, other issues may also emerge, such as ethical issues, and data privacy and security.

The demand for the concerted efforts calls out the governance platform. This platform provides central control over functions such as data curation, pre-processing, regulation, and the interaction with current healthcare delivery system. Guideline developments has been underway across the globe, such as Singapore's Model Artificial Intelligence Governance Framework,

guiding private sectors to use artificial intelligence ethically. Others developed proprietary platforms that integrated electronic health records together with data governance structures that take advantage of cloud-based machine learning platforms.

Machine learning for Oncology

Medicine and biology have rapidly become data-intensive [100]. When the format of the data is no longer limited to numeric format, algorithms which are capable of sifting through massive amount of data and extracting meaningful patterns can provide a way in which the prevention and treatment developed.

Among all disease types, cancer is considered the most common cause of death in developed countries, and the number of cases is only worsening along with the aging populations [101,102]. In 2021, data shows that 1.9 million people are estimated being diagnosed with cancer, and over 608 thousand deaths would occur in the US. Thus, application of artificial intelligence in the context of cancer continues to be the top priority to prevent cancer and save lives.

With the rapidly declined cost in genomic sequencing as well as the evident associations between cancer occurrence and human genomics, a deeper understanding in the implicit relationships may reveal new insights. However, such studies usually involve the identification between one thousand and one hundred thousand mutations for each tumor sample [103]. And the clinical interpretation is heavily depending on published literature to link the mutation and the disease states and prognosis. In 2019 alone, there were over 200 thousand new cancer related papers published, making manual curation impossible. Hence, the implementation of machine learning becomes a reasonable solution. Relative to the conventional way of transforming genomic data into a binary table, machine performs same task within minutes or even seconds. In

the meantime, by sharing parts of the model, machine performs multiple tasks at the same time and search for implicit patterns, which were not feasible in the past. Moreover, machine allows multi-model learning, meaning that now we can integrate different types of data. Given that cancer is a group of complex diseases, integrating various types of data together could better inform clinical decision.

Another application makes machine learning implementation very attractive is the image analysis. Majority of the cancer types progresses rapidly or have an insidious early stage. By the time that physicians have enough information to diagnose, patients have already compromised their health to a great extent. Therefore, early detection is crucial to saving lives of the affected individuals. Deep learning is one branch of machine learning which revolutionized image analysis in recent years [10,92]. One example demonstrated great impact was the classification of skin cancer based on dermoscopy images where AI showed the capability to achieve accuracy comparable to certified doctors and annotated skin lesions on the image precisely as were expert dermatologists. In another example, AI also achieved promising accuracy when interpreting mammograms during breast cancer screening [104].

By far, there has been countless published studies reporting reliable accuracy on application of machine learning in cancer detection and diagnosis. Collectively, these achievements demonstrated that this technique is matured enough to perform versatile tasks at a level of accuracy to provide trustworthy insights to healthcare providers and healthcare system [105].

Machine learning for NMSC

Nonmelanoma Skin Cancer (NMSC) is considered the most common malignancy, with estimated over 3 million annual cases in the United States alone [106]. Among all NMSC cases,

basal cell carcinoma (BCC) and squamous cell carcinoma represent 95% of the NMSC cases [107], which are readily identified through visual inspection by expert dermatologist. Yet, as NMSC progresses continuously, several benign lesions may mimic the appearance of these subtypes, leading to further evaluation by biopsy. It is not uncommon that the affected individual needs to go through multiple invasive biopsy procedure in order to get one definitive diagnosis [108].

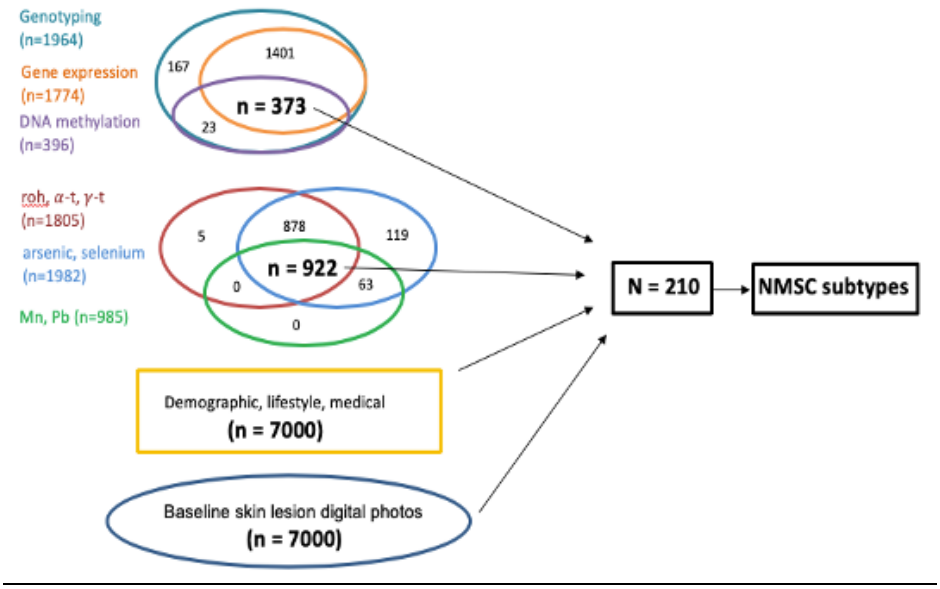
The application of machine learning as a diagnostic assistance has been rapidly growing in dermatology. The implementation involves training the model with images curated by the doctor, then feed the new images for diagnostic insights. This is particularly helpful when the suspected case volumes are large while medical resources are scarce. With the aid from machine learning model, regions of interests are automated for physicians to reference, saving significant amount of time for them to focus on tricky cases.

Numerous studies [10,75,109–116] have piloted on using machine learning approach to aid diagnosis in clinic. While almost all these studies reported superior diagnostic accuracy, many of them are subject to methodological limitations, such as including same images for both training data and testing data [109,110], or including consecutive samples in test sets [111,112,115,116]. Nevertheless, these studies did provide high level insights for machine learning classifiers capable of accurately discriminating NMSC from benign lesions. Further research is warranted to evaluate the viability of these results before incorporating them in a clinical setting.

Data source

In addition to the data used in previous aims, genotyping, gene expression, DNA methylation and blood metal data from a subset of 210 participants will be incorporated. Details are provided in Figure 6.

Figure 18. Data integration for a total of 210 participants



Genotyping

DNA were extracted from the whole blood using the QIAamp 96 DNA Blood Kit from Qiagen (Valencia, USA). DNA sample with a concentration <40 ng/μL, and/or 260/280 ratio outside the range of <1.6 to ≥2.1 (measured by Nanodrop 1000), and/or fragmented DNA <2 Kb (assessed by smearing in Agilent BioAnalyzer) was excluded. Genotyping was performed using the Illumina HumanCytoSNP-12 BeadChip utilizing 250 ng DNA according to the manufacturer's protocol. Samples with very poor call rates (<97%) were excluded, individuals with gender mismatches, duplicated samples were also excluded.

Quality control (QC) exclusion criteria was applied among 299,140 genotyped SNPs using PLINK [117]: (i) SNPs without rs numbers; (ii) SNP call rate <95%; (iii) monomorphic SNPs; (iv) Hardy-Weinberg P<1 × 10−10. This procedure yielded 257,768 SNPs. Then, imputation was performed using MaCH on the basis of the HapMap 3 Gujarati Indians in

Houston (GIH) population (Build 36). Subsequently, additional QC exclusion criteria were applied for SNPs post-imputation: (i) MAF <0.01 and (ii) SNP imputation score <0.3. All QC procedures resulted in 1,211,988 million SNPs for both genotyped and imputed SNPs, which will be included in the analysis. Genotyping data was available in 1,964 individuals (female: 902, male: 1062).

Gene expression data

RNA was extracted from mononuclear cells preserved in RLT buffer, stored at −80°C, using RNeasy Micro Kit from Qiagen (Valencia, USA). The concentration and quality of RNA was checked on Nanodrop 1000. cRNA synthesis was done from 250 ng of RNA using Illumina TotalPrep 96 RNA Amplification kit. Gene expression was measured using the Illumina HumanHT-12-v4 BeadChip utilizing 750 ng of cRNA according to the manufacturer's protocol. The chip contains a total of 47,231 probes covering 31,335 genes. We restricted our analyses to specific probes for expression quantitative trait loci (eQTL) analyses, which yielded 31,583 probes. Quantile normalized expression values were log2 transformed and adjusted for batch variability using ComBat software [118]. Gene expression data was available for 1,774 individuals (female: 795, male: 979) included in these analyses.

DNA methylation

DNA were extracted from whole blood using DNeasy Blood kits (Qiagen). Bisulfite conversion was performed using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA). For each sample, 500 ng of bisulfite-converted DNA was applied to the Illumina HumanMethylation 450K BeadChip kit (Illumina, San Diego, CA, USA) according to the manufacturer's protocol, enabling interrogation of 482 421 CpG sites and 3091 non-CpG sites

per sample. This array contains an average of 17 CpG sites per gene, distributed across the promoter, 5' untranslated region (UTR), first exon, gene body and 3' UTR, covering 99% of RefSeq genes.

Methylation status at each CpG is expressed as a β value that can range from 0 to 1 (unmethylated to completely methylated). Data were quantile normalized. Among the 409 participants, six samples reported sex of the participant did not match the predicted sex based on methylation patterns of the X and Y chromosomes, and seven samples with more than 5% of CpGs either containing missing values or having p over 0.05 for detection. This procedure resulted in 396 samples (female: 188, male: 208) with quality methylation data. In addition, we also removed probes mapping to multiple locations (41937), probes with SNPs (20 869). Individual β values with a p for detection > 0.05 were set to missing, and >10% missing of β values were excluded. Probes on the X (11 232) and Y (416) chromosomes, probes with missing chromosome data (mostly control probes, 65), and probes with >10% missing data across samples (1932) were excluded. In the end, this yielded a total of 423 604 probes available. β values were logit-transformed and adjusted for batch variability, the average inter-assay Spearman correlation coefficient was 0.987 (range, 0.974-0.993) [119,120].

Blood metal data

The blood level of manganese (Mn), lead (Pb), arsenic (As), Selenium (Se), α-tocopherol, γ-tocopherol, and retinol (Rho) were measured at baseline and 2-yeawr visit. Among these measurements, baseline arsenic and selenium levels were measured with all the participants, other measurements were conducted in subsets of participants (Table 13).

Table 20. Blood level of minerals measured at baseline and first major visit

	Baseline (n)	1st major visit (n)
Manganese (Mn)	986	986
Lead (Pb)	986	986
Arsenic (As)	6963	1983
Selenium (Se)	6963	1983
α-tocopherol	1835	1847
γ-tocopherol	1835	1847
Retinol (rho)	1835	1847

Measures

The clinical diagnosis made by the pathologist based on clinical manifestations and histopathological evaluations will be used as the "ground truth" in current analysis. The description is detailed in previous aims.

Analyses

Data and image curation

Baseline skin lesion digital photos will be reviewed and curated. This process will identify digital photos with poor quality arising from blurred images, heaving inky marks on the images where the regions of interest are likely to be targeted. Tabular data, including demographic, lifestyle, medical, molecular and genomics will also be cleaned – genomics data will go through quality control processes, described in the above data source section, before putting into the model for analysis. Other data sources will also be cleaned with missing data and outliers handled, dependent on the proportion with the issue.

Network architecture for each analysis branch

Digital photos will apply the same data preprocessing technique detailed in aim 2 and convert into image patches. Image feature extraction from these patches will be achieved through an attention-based CNN. ResNet 18 again will be adopted to extract features for digital photos, as skin lesions on the participant's feet, hands and other parts of the skins are localized. A feature vector will be the output from ResNet 18 and input into the attention-module to calculate the importance of each patch. The output data from the attention module will be a same-length vector representing feature information for each photo. To handle the tabular dataset, multi-layer network will be applied to extract features on the participant level, integrating all the non-image data into this model and output a feature map.

Integrative modeling

Two-branch neural network will be built to process and integrate all the data together for analyses. First, a multi-layer perceptron (MLP) will be built to handle the categorical/numerical and without doing any prediction, save the output dataset. Then, an attention-based CNN will be used to handle image data and save the output as same-length vector without doing the prediction. Subsequently, these two output datasets will be concatenated together as the input dataset for a fully connected neural network. In this step, we will perform the train-validate-test split and do the outcome predictions. The analysis workflow is demonstrated below in Figure 7.

Future Directions

Through the implementation of current aim, we will be among the first to build the machine learning prediction model for NMSC vs. non-NMSC and survival outcome using mixed data. As part of the effort to realize the goal for precision population health, this model will

bridge the gap between unstructured and structured data and provide a comprehensive view for patient characteristics and disease outcomes.

There are a few foreseeable challenges and limitations. First, machine learning platforms are not widely applied in health care practice, even in the high-income countries. Therefore, the implementation will likely meet resistance, due to its complexity. Second, data sparsity will also hurdle this initiative. Although promising to combine mixed data together for a comprehensive evaluation, the availability of all the data elements for each individual is concerning. In this aim, the subtypes of NMSC as well as cause-specific mortalities are exploratory due to this reason. Alternatively, these outcomes will be explored without incorporating all other co-variates. Third, as more data will be needed to build integrative models, legal and ethical concerns related to machine learning driven platform demands better regulations to be in place.

Decreasing the impact from disease etiology plays a central role in disease prevention and health promotion. The mechanistic evidence of disease causes sets up a foundation for further exploring therapeutic options for any affected populations. However, when exposure to an agent is homogeneous within a population, the case-control or cohort studies will fail to detect the cause of the disease. The corresponding strategy is to identify and therefore, protect the "high-risk" individuals. The "high-risk" strategy evaluates interventions from RCTs that are appropriate to the susceptible individuals [121].

However, there are growing concerns that the results of RCTs might not directly apply to individual patients, even for those within the same trial and the same treatment arm. Although in theory, randomization ensures the comparability of treatment groups, the differences between individuals in each treatment arm can still dramatically affect the likelihood of benefiting from or being harmed by a therapy. The average treatment effects across different patients can present

misleading results to physicians who care for the individual patient, not the average patients. Subgroup analysis allows exploring differences in treatment effects based on patient characteristics, which put a step forward towards personalized population health [122].

Integration of disparate data sources plays a crucial role in the implementation of personalized and population health. The all-encompassing information made it possible to generate a more thorough profile of individual patient, which spurs pattern discovery in patient characteristics and disease progression. As in the high-income countries, the emerging technologies and abundant choices of analytical tools demonstrated transformational power in the provision of healthcare services, this implementation shows even greater potential in resource-limiting countries. However, moving from pilot to large scale-will still require addressing several challenges and concerns.

First, determine which algorithmic approach to use for which research question is a challenge. As numerous machine learning models are widely available, researchers should be cautious as inappropriate algorithm would generate misleading insights. One solution is to select approaches based on available predictor variables.

Second, constructing and training an effective model from scratch would require a good amount of data and time, especially for deep learning neural networks. This might not be feasible in the initial implementation in resource-poor settings. One way to overcome this hurdle is to adopt highly refined features from existing pre-trained models as a starting point in training a new model for a different task. The features selected from the pre-trained models work as basic building blocks, such as lines, edges, and curves for image analysis, and have been shown to be applicable to many different image-recognition tasks [123]. This technique not only reduces

enormous number of computations needed for training but delivers substantial performance benefits [124].

Third, this implementation in clinical practice also raises concerns because this platform does not provide a straightforward explanation to healthcare providers as to how machine learning tools work. Although they might not need to know the nitty-gritty mathematical calculations in an algorithm, knowing the types of data used in making the predictions and the relative weights assigned to each predictor would be beneficial. However, analogous to other clinical tests, metrics such as sensitivity and specificity could also be used as a starting point to inform the performance of the tool in predicting a particular clinical outcome.

Fourth, in spite of the comparable accuracy achieved by machine learning approaches, very few implementations had successfully penetrated the clinical practice. One main reason is that these early applications are lack of external validation. The sparsity of the data affects not only the model training but also the external validation, which is indispensable before we can adopt such platforms.

With proper implementation, machine learning application could lead to a great leap in healthcare in the resource-poor settings. Further effort is needed to accelerate its deployment in such settings.

REFERENCES:

1. Bayer, R. & Galea, S. Public Health in the Precision-Medicine Era. *New Engl J Medicine* **373**, 499–501 (2015).

2. Collins, F. S. & Varmus, H. A New Initiative on Precision Medicine. *New Engl J Medicine* **372**, 793–795 (2015).

3. Lyles, C. R., Lunn, M. R., Obedin-Maliver, J. & Bibbins-Domingo, K. The new era of precision population health: insights for the All of Us Research Program and beyond. *J Transl Med* **16**, 211 (2018).

4. Khoury, M. J., Iademarco, M. F. & Riley, W. T. Precision Public Health for the Era of Precision Medicine. *Am J Prev Med* **50**, 398–401 (2016).

5. Rubin, R. Precision Medicine: The Future or Simply Politics? *Jama* **313**, 1089–1091 (2015).

6. Khoury, M. J. & Galea, S. Will Precision Medicine Improve Population Health? *Jama* **316**, 1357 (2016).

7. Maas, P. *et al.* Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *Jama Oncol* **2**, 1295 (2016).

8. Khoury, M. J. & Evans, J. P. A Public Health Perspective on a National Precision Medicine Cohort: Balancing Long-term Knowledge Generation With Early Health Benefit. *Jama* **313**, 2117–2118 (2015).

9. Fogel, A. L. & Kvedar, J. C. Artificial intelligence powers digital medicine. *Npj Digital Medicine* **1**, 5 (2018).

10. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).

11. Stein, J. D. *et al.* Evaluation of an Algorithm for Identifying Ocular Conditions in Electronic Health Record Data. *Jama Ophthalmol* **137**, 491–497 (2019).

12. Azencott, C.-A. Machine learning and genomics: precision medicine versus patient privacy. *Philosophical Transactions Royal Soc Math Phys Eng Sci* **376**, 20170350 (2018).

13. CRECRC *et al.* The CRISP colorectal cancer risk prediction tool: an exploratory study using simulated consultations in Australian primary care. *Bmc Med Inform Decis* **17**, 13 (2017).

14. Coudray, N. *et al.* Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med* **24**, 1559–1567 (2018).

15. Ngiam, K. Y. & Khor, I. W. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* **20**, e262–e273 (2019).

16. Remembering the Past for Meaningful AI-D. *AAAI Spring Symposium: Artificial Intelligence for Development* (2010).

17. Rajpara, S. M., Botello, A. P., Townend, J. & Ormerod, A. D. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. *Brit J Dermatol* **161**, 591–604 (2009).

18. Korotkov, K. & Garcia, R. Computerized analysis of pigmented skin lesions: A review. *Artif Intell Med* **56**, 69–90 (2012).

19. Dreiseitl, S. *et al.* A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions. *J Biomed Inform* **34**, 28–36 (2001).

20. Flanagan, S., Johnston, R. & Zheng, Y. Arsenic in tube well water in Bangladesh: health and economic impacts and implications for arsenic mitigation. *B World Health Organ* **90**, 839–846 (2012).

21. Miller, W. H., Schipper, H. M., Lee, J. S., Singer, J. & Waxman, S. Mechanisms of action of arsenic trioxide. *Cancer Res* **62**, 3893–903 (2002).

22. Jomova, K. *et al.* Arsenic: toxicity, oxidative stress and human disease. *J Appl Toxicol* **31**, 95–107 (2011).

23. Shakoor, M. B. *et al.* Human health implications, risk assessment and remediation of As-contaminated water: A critical review. *Sci Total Environ* **601**, 756–769 (2017).

24. Duarte, A. A. L. S., Cardoso, S. J. A. & Alçada, A. J. Emerging and Innovative Techniques for Arsenic Removal Applied to a Small Water Supply System. *Sustainability-basel* **1**, 1288–1304 (2009).

25. Wu, M. M. *et al.* Association of blood arsenic levels with increased reactive oxidants and decreased antioxidant capacity in a human population of northeastern Taiwan. *Environ Health Persp* **109**, 1011–1017 (2001).

26. Valko, M., Morris, H. & Cronin, M. Metals, Toxicity and Oxidative Stress. *Curr Med Chem* **12**, 1161–1208 (2005).

27. Fernandes, A. P. & Gandin, V. Selenium compounds as therapeutic agents in cancer. *Biochimica Et Biophysica Acta Bba - Gen Subj* **1850**, 1642–1660 (2015).

28. Misra, S., Boylan, M., Selvam, A., Spallholz, J. E. & Björnstedt, M. Redox-Active Selenium Compounds—From Toxicity and Cell Death to Cancer Treatment. *Nutrients* **7**, 3536–3556 (2015).

29. Labunskyy, V. M., Hatfield, D. L. & Gladyshev, V. N. Selenoproteins: Molecular Pathways and Physiological Roles. *Physiol Rev* **94**, 739–777 (2014).

30. Hatfield, D. L., Tsuji, P. A., Carlson, B. A. & Gladyshev, V. N. Selenium and selenocysteine: roles in cancer, health, and development. *Trends Biochem Sci* **39**, 112–120 (2014).

31. Barrington, W. E. *et al.* Difference in Association of Obesity With Prostate Cancer Risk Between US African American and Non-Hispanic White Men in the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *Jama Oncol* **1**, 342–349 (2015).

32. Vinceti, M., Filippini, T., Cilloni, S. & Crespi, C. M. Chapter One The Epidemiology of Selenium and Human Cancer. *Adv Cancer Res* **136**, 1–48 (2017).

33. Schrauzer, G. N., White, D. A. & Schneider, C. J. Cancer mortality correlation studies-III: Statistical associations with dietary selenium intakes. *Bioinorg Chem* **7**, 23–34 (1977).

34. Shamberger, R. J. & Frost, D. V. Possible protective effect of selenium against human cancer. *Can Med Assoc J* **100**, 682 (1969).

35. Weinstein, S. J. *et al.* Serum and Dietary Vitamin E in Relation to Prostate Cancer Risk. *Cancer Epidemiology Prev Biomarkers* **16**, 1253–1259 (2007).

36. M.P.H., L. C. C., Cantor, K. P. & Allaway, W. H. Selenium in Forage Crops and Cancer Mortality in U.S. Counties. *Archives Environ Heal Int J* **46**, 37–42 (1991).

37. Chung, C.-J. *et al.* Protective effects of plasma alpha-tocopherols on the risk of inorganic arsenic-related urothelial carcinoma. *Sci Total Environ* **409**, 1039–1045 (2011).

38. Melkonian, S. *et al.* Intakes of Several Nutrients Are Associated with Incidence of Arsenic-Related Keratotic Skin Lesions in Bangladesh. *J Nutrition* **142**, 2128–2134 (2012).

39. Yang, L., Wang, W., Hou, S., Peterson, P. J. & Williams, W. P. Effects of Selenium Supplementation on Arsenism: An Intervention Trial in Inner Mongolia. *Environ Geochem Hlth* **24**, 359–374 (2002).

40. Zablotska, L. B. *et al.* Protective Effects of B Vitamins and Antioxidants on the Risk of Arsenic-Related Skin Lesions in Bangladesh. *Environ Health Persp* **116**, 1056–1062 (2008).

41. Clark, L. C. *et al.* Effects of Selenium Supplementation for Cancer Prevention in Patients With Carcinoma of the Skin: A Randomized Controlled Trial. *Jama* **276**, 1957–1963 (1996).

42. The Effect of Vitamin E and Beta Carotene on the Incidence of Lung Cancer and Other Cancers in Male Smokers. *New Engl J Medicine* **330**, 1029–1035 (1994).

43. Li, J.-Y. *et al.* Nutrition Intervention Trials in Linxian, China: Multiple Vitamin/Mineral Supplementation, Cancer Incidence, and Disease-Specific Mortality Among Adults With Esophageal Dysplasia. *Jnci J National Cancer Inst* **85**, 1492–1498 (1993).

44. Lippman, S. M. *et al.* Effect of Selenium and Vitamin E on Risk of Prostate Cancer and Other Cancers: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). *Jama* **301**, 39–51 (2009).

45. Klein, E. A. *et al.* Vitamin E and the Risk of Prostate Cancer: The Selenium and Vitamin E Cancer Prevention Trial (SELECT). *Jama* **306**, 1549–1556 (2011).

46. Verret, W. J. *et al.* A Randomized, Double-Blind Placebo-Controlled Trial Evaluating the Effects of Vitamin E and Selenium on Arsenic-Induced Skin Lesions in Bangladesh. *J Occup Environ Med* **47**, 1026–1035 (2005).

47. Kuras, R. *et al.* Biomarkers of selenium status and antioxidant effect in workers occupationally exposed to mercury. *J Trace Elem Med Bio* **49**, 43–50 (2018).

48. Gascón-Vila, P. *et al.* Determinants of the nutritional status of vitamin E in a non-smoking Mediterranean population. Analysis of the effect of vitamin E intake, alcohol consumption and body mass index on the serum alpha-tocopherol concentration. *Eur J Clin Nutr* **51**, 723–728 (1997).

49. Karademirci, M., Kutlu, R. & Kilinc, I. Relationship between smoking and total antioxidant status, total oxidant status, oxidative stress index, vit C, vit E. *Clin Respir J* **12**, 2006–2012 (2018).

50. Fairweather-Tait, S. J. *et al.* Selenium in Human Health and Disease. *Antioxid Redox Sign* **14**, 1337–1383 (2011).

51. Goodstein, R. K. & Swift, K. Psychotherapy with Phobic Patients: The Marriage Relationship as the Source of Symptoms and Focus of Treatment. *Am J Psychother* **31**, 284–293 (1977).

52. Weber, D. *et al.* Plasma carotenoids, tocopherols and retinol - Association with age in the Berlin Aging Study II. *Redox Biol* **32**, 101461 (2020).

53. Argos, M. *et al.* Baseline comorbidities in a skin cancer prevention trial in Bangladesh. *Eur J Clin Invest* **43**, 579–588 (2013).

54. Rubinstein, L. V., Gail, M. H. & Santner, T. J. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. *J Chron Dis* **34**, 469–479 (1981).

55. Ronsmans, C., Vanneste, A. M., Chakraborty, J. & Ginneken, J. V. A comparison of three verbal autopsy methods to ascertain levels and causes of maternal deaths in Matlab, Bangladesh. *Int J Epidemiol* **27**, 660–666 (1998).

56. Ju, J. *et al.* Cancer-preventive activities of tocopherols and tocotrienols. *Carcinogenesis* **31**, 533–542 (2010).

57. Helzlsouer, K. J. *et al.* Association Between α-Tocopherol, γ-Tocopherol, Selenium, and Subsequent Prostate Cancer. *Jnci J National Cancer Inst* **92**, 2018–2023 (2000).

58. Li, B. *et al.* Linxian nutrition intervention trials design, methods, participant characteristics, and compliance. *Ann Epidemiol* **3**, 577–585 (1993).

59. Blot, W. J. *et al.* Nutrition Intervention Trials in Linxian, China: Supplementation With Specific Vitamin/Mineral Combinations, Cancer Incidence, and Disease-Specific Mortality in the General Population. *Jnci J National Cancer Inst* **85**, 1483–1491 (1993).

60. Duffield-Lillico, A. J. *et al.* Selenium supplementation, baseline plasma selenium status and incidence of prostate cancer: an analysis of the complete treatment period of the Nutritional Prevention of Cancer Trial. *Bju Int* **91**, 608–612 (2003).

61. Algotar, A. M. *et al.* Phase 3 clinical trial investigating the effect of selenium supplementation in men at high-risk for prostate cancer. *Prostate* **73**, 328–335 (2013).

62. Karp, D. D. *et al.* Randomized, Double-Blind, Placebo-Controlled, Phase III Chemoprevention Trial of Selenium Supplementation in Patients With Resected Stage I Non–Small-Cell Lung Cancer: ECOG 5597. *J Clin Oncol* **31**, 4179–4187 (2013).

63. Fernández, M. I., López, J. F., Vivaldi, B. & Coz, F. Long-Term Impact of Arsenic in Drinking Water on Bladder Cancer Health Care and Mortality Rates 20 Years After End of Exposure. *J Urology* **187**, 856–861 (2012).

64. Assmann, S. F., Pocock, S. J., Enos, L. E. & Kasten, L. E. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* **355**, 1064–1069 (2000).

65. Altman, D. G. Comparability of Randomised Groups. *J Royal Statistical Soc Ser D Statistician* **34**, 125–136 (1985).

66. Hauck, W. W., Anderson, S. & Marcus, S. M. Should We Adjust for Covariates in Nonlinear Regression Analyses of Randomized Trials? *Control Clin Trials* **19**, 249–256 (1998).

67. Senn, S. J. Covariate imbalance and random allocation in clinical trials. *Stat Med* **8**, 467–475 (1989).

68. Smith, A. H., Lingas, E. O. & Rahman, M. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *B World Health Organ* **78**, 1093–103 (2000).

69. DG, K. & PL, S. *Arsenic contamination of groundwater Bangladesh*. http://nora.nerc.ac.uk/id/eprint/11986/1/WC00019.pdf (2001).

70. Seow, W. J. *et al.* Arsenic Reduction in Drinking Water and Improvement in Skin Lesions: A Follow-Up Study in Bangladesh. *Environ Health Persp* **120**, 1733–1738 (2012).

71. PETERKA, E. S., LYNCH, F. W. & GOLTZ, R. W. An Association Between Bowen's Disease and Internal Cancer. *Arch Dermatol* **84**, 623–629 (1961).

72. Kao, G. F. Carcinoma arising in Bowen's disease. *Arch Dermatol* **122**, 1124–1126 (1986).

73. Hoorens, I., Vossaert, K., Ongenae, K. & Brochez, L. Is early detection of basal cell carcinoma worthwhile? Systematic review based on the WHO criteria for screening. *Brit J Dermatol* **174**, 1258–1265 (2016).

74. Hussain, S. A. & Sullivan, R. Cancer Control in Bangladesh. *Jpn J Clin Oncol* **43**, 1159–1169 (2013).

75. Chang, W.-Y. *et al.* Computer-Aided Diagnosis of Skin Lesions Using Conventional Digital Photography: A Reliability and Feasibility Study. *Plos One* **8**, e76212 (2013).

76. Wei, J. W. *et al.* Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep-uk* **9**, 3358 (2019).

77. Peck, M., Moffat, D., Latham, B. & Badrick, T. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. *J Clin Pathol* **71**, 995 (2018).

78. Sirota, R. L. Defining Error in Anatomic Pathology. *Arch Pathol Lab Med* **130**, 604–606 (2006).

79. Gurcan, M. N. *et al.* Histopathological Image Analysis: A Review. *Ieee Rev Biomed Eng* **2**, 147–171 (2009).

80. Ilse, M., Tomczak, J. M. & Welling, M. Attention-based Deep Multiple Instance Learning. *Arxiv* (2018).

81. Liu, Y. *et al.* Detecting Cancer Metastases on Gigapixel Pathology Images. *Arxiv* (2017).

82. Tellez, D. *et al.* Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks. *Ieee T Med Imaging* **37**, 2126–2136 (2018).

83. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *2016 Ieee Conf Comput Vis Pattern Recognit Cvpr* 770–778 (2016) doi:10.1109/cvpr.2016.90.

84. PyTorch. https://pytorch.org (n.d.).

85. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 2579–605 (n.d.).

86. Yanofsky, V. R., Mercer, S. E. & Phelps, R. G. Histopathological Variants of Cutaneous Squamous Cell Carcinoma: A Review. *J Ski Cancer* **2011**, 210813 (2011).

87. Neagu, T. P. *et al.* Clinical, histological and therapeutic features of Bowen's disease. *Romanian J Morphol Embryology Revue Roumaine De Morphol Et Embryologie* **58**, 33–40 (2017).

88. Weedon's Skin Pathology. (2010) doi:10.1016/b978-0-7020-3485-5.x0001-0.

89. Choudhury, Md. I. M. *et al.* Cutaneous Malignancy due to Arsenicosis in Bangladesh: 12-Year Study in Tertiary Level Hospital. *Biomed Res Int* **2018**, 1–9 (2018).

90. Shaban-Nejad, A., Michalowski, M. & Buckeridge, D. L. Health intelligence: how artificial intelligence transforms population and personalized health. *Npj Digital Medicine* **1**, 53 (2018).

91. Kim, D. H. & MacKinnon, T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. *Clin Radiol* **73**, 439–445 (2018).

92. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

93. Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S. & Acharya, U. R. Deep learning for healthcare applications based on physiological signals: A review. *Comput Meth Prog Bio* **161**, 1–13 (2018).

94. Mazzanti, M., Shirka, E., Gjergo, H. & Hasimi, E. Imaging, Health Record, and Artificial Intelligence: Hype or Hope? *Curr Cardiol Rep* **20**, 48 (2018).

95. Hainc, N. *et al.* The Bright, Artificial Intelligence-Augmented Future of Neuroimaging Reading. *Front Neurol* **8**, 489 (2017).

96. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* **25**, 44–56 (2019).

97. Jha, S. & Topol, E. J. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. *Jama* **316**, 2353 (2016).

98. Gelhaus, P. Robot decisions: on the importance of virtuous judgment in clinical decision making. *J Eval Clin Pract* **17**, 883–887 (2011).

99. Luxton, D. D. Recommendations for the ethical use and design of artificial intelligent care providers. *Artif Intell Med* **62**, 1–10 (2014).

100. Shimizu, H. & Nakayama, K. I. Artificial intelligence in oncology. *Cancer Sci* **111**, 1452–1460 (2020).

101. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *Ca Cancer J Clin* **69**, 7–34 (2019).

102. DeSantis, C. E. *et al.* Cancer statistics for adults aged 85 years and older, 2019. *Ca Cancer J Clin* **69**, 452–467 (2019).

103. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703–713 (2017).

104. Rodriguez-Ruiz, A. *et al.* Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *Jnci J National Cancer Inst* **111**, 916–922 (2019).

105. Litjens, G. *et al.* Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep-uk* **6**, 26286 (2016).

106. Rogers, H. W., Weinstock, M. A., Feldman, S. R. & Coldiron, B. M. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the US Population, 2012. *Jama Dermatol* **151**, 1081–1086 (2015).

107. Samarasinghe, V. & Madan, V. Nonmelanoma skin cancer. *J Cutan Aesthetic Surg* **5**, 3–10 (2012).

108. Breitbart, E. W. *et al.* Systematic skin cancer screening in Northern Germany. *J Am Acad Dermatol* **66**, 201–211 (2012).

109. Shakya, N. M. *et al.* Discrimination of squamous cell carcinoma in situ from seborrheic keratosis by color analysis techniques requires information from scale, scale-crust and surrounding areas in dermoscopy images. *Comput Biol Med* **42**, 1165–1169 (2012).

110. Wahba, M. A., Ashour, A. S., Guo, Y., Napoleon, S. A. & Elnaby, M. M. A. A novel cumulative level difference mean based GLDM and modified ABCD features ranked using eigenvector centrality approach for four skin lesion types classification. *Comput Meth Prog Bio* **165**, 163–174 (2018).

111. Ferris, L. K. *et al.* Computer-aided classification of melanocytic lesions using dermoscopic images. *J Am Acad Dermatol* **73**, 769–776 (2015).

112. Møllersen, K. *et al.* Computer-Aided Decision Support for Melanoma Detection Applied on Melanocytic and Nonmelanocytic Skin Lesions: A Comparison of Two Systems Based on Automatic Analysis of Dermoscopic Images. *Biomed Res Int* **2015**, 1–8 (2015).

113. Fujisawa, Y. *et al.* Deep-learning-based, computer-aided classifier developed with a small dataset of clinical images surpasses board-certified dermatologists in skin tumour diagnosis. *Brit J Dermatol* **180**, 373–381 (2019).

114. Han, S. S. *et al.* Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *J Invest Dermatol* **138**, 1529–1538 (2018).

115. Cheng, B. *et al.* Automatic detection of basal cell carcinoma using telangiectasia analysis in dermoscopy skin lesion images. *Skin Res Technol* **17**, 278–287 (2011).

116. Kefel, S. *et al.* Adaptable texture-based segmentation by variance and intensity for automatic detection of semitranslucent and pink blush areas in basal cell carcinoma. *Skin Res Technol* **22**, 412–422 (2016).

117. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genetics* **81**, 559–575 (2007).

118. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

119. Pierce, B. L. *et al.* Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat Commun* **9**, 804 (2018).

120. Price, E. M. *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenet Chromatin* **6**, 4 (2013).

121. Rose, G. Sick individuals and sick populations. *Int J Epidemiol* **30**, 427–432 (2001).

122. Kent, D. M. & Hayward, R. A. Limitations of Applying Summary Results of Clinical Trials to Individual Patients: The Need for Risk Stratification. *Jama* **298**, 1209–1212 (2007).

123. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? *Arxiv* (2014).

124. Razavian, A. S., Azizpour, H., Sullivan, J. & Carlsson, S. CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. *2014 Ieee Conf Comput Vis Pattern Recognit Work* 512–519 (2014) doi:10.1109/cvprw.2014.131.