

Article

Stage Shifting by Modifying the Determinants of Breast Cancer Stage at Diagnosis: A Simulation Study

Gyanendra Pokharel ¹, Qinggang Wang ², Momtadin Khan ², Paula J. Robson ^{3,4}, Lorraine Shack ⁵
and Karen A. Kopciuk ^{2,6,7,*} 

¹ Department of Mathematics and Statistics, University of Winnipeg, 515 Portage Avenue, Winnipeg, MB R3B 2E9, Canada; g.pokharel@uwinnipeg.ca

² Cancer Epidemiology and Prevention Research, Cancer Care Alberta, Alberta Health Services, 2210 2nd Street SW, Calgary, AB T2S 3C3, Canada; qinggang.wang@albertahealthservices.ca (Q.W.)

³ Department of Agricultural, Food and Nutritional Science and School of Public Health, University of Alberta, 116 Street & 85 Avenue, Edmonton, AB T6G 2P5, Canada; paula.robson@albertahealthservices.ca

⁴ Cancer Care Alberta and Cancer Strategic Clinical Network, Alberta Health Services, 10030-107 Street NW, Edmonton, AB T5J 3H1, Canada

⁵ Cancer Surveillance and Reporting, Cancer Care Alberta, Alberta Health Services, 2210 2nd Street SW, Calgary, AB T2S 3C3, Canada; lorraine.shack@albertahealthservices.ca

⁶ Department of Oncology, University of Calgary, 1331 29 Street NW, Calgary, AB T2N 4N2, Canada

⁷ Departments of Community Health Sciences and Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

* Correspondence: karen.kopciuk@albertahealthservices.ca

Simple Summary: The impact of breast cancer on individuals and in populations can be reduced if the cancer is detected in an early stage of disease as treatments are generally less harsh and more successful. Knowing which factors to modify and by what amount that could shift breast cancer diagnoses to early stages of disease were studied in this project. Factors related to stage at diagnosis in a previous analysis of women diagnosed with breast cancer were assessed using data that mimicked the real data. The proportions of modifiable lifestyle factors in a population were increased or decreased, such as those ever having a mammogram, to see the effect on detecting more cancers in the earliest stages. Similarly, the average amounts of protein eaten per day or calories consumed were also increased and decreased. Increasing the total amount of protein eaten per day by even just 5 grams was the most important factor found in this study to improve early breast cancer diagnoses. The findings from this study could be helpful in breast cancer prevention programs that could target key modifiable factors and by other researchers studying other types of cancer.

Abstract: Background: Breast cancer is the most common cancer in Canadian women; nearly 25% of women diagnosed with cancer have breast cancer. The early detection of breast cancer is a major challenge because tumours often grow without causing symptom. The diagnosis of breast cancer at an early stage (stages I and II) improves survival outcomes because treatments are more effective and better tolerated. To better inform the prevention of and screening for breast cancer, simulations using modifiable rather than non-modifiable risk factors may be helpful in shifting the stage at diagnosis downward. Methods: Breast cancer stages were simulated using the data distributions from Alberta's Tomorrow Project participants who developed breast cancer. Using multivariable partial proportional odds regression models, modifiable lifestyle factors associated with the stage of cancer at diagnosis were evaluated. The proportions or mean levels of these lifestyle factors in the simulated population were systematically changed, then multiplied by their corresponding estimated odds ratios from the real data example. The effects of these changes were evaluated singly as well as cumulatively. Results: Increasing total dietary protein (g/day) intake was the single most important lifestyle factor in shifting the breast cancer stage downwards followed by decreasing total dietary energy intake (kcal/day). Increasing the proportion of women who spend time in the sun between 11 am and 4 pm in the summer months, who have had a mammogram, who have been pregnant or reducing the proportion who are in stressful situations had much smaller effects. The percentage of Stage I diagnoses could be increased by approximately 12% with small modifications of these lifestyle factors.



Citation: Pokharel, G.; Wang, Q.; Khan, M.; Robson, P.J.; Shack, L.; Kopciuk, K.A. Stage Shifting by Modifying the Determinants of Breast Cancer Stage at Diagnosis: A Simulation Study. *Cancers* **2024**, *16*, 1201. <https://doi.org/10.3390/cancers16061201>

Academic Editor: Andrea Nicolini

Received: 4 March 2024

Accepted: 14 March 2024

Published: 19 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Conclusion: Shifting the breast cancer stage at diagnosis of a population may be achieved through changes to lifestyle factors. This proof of principle study that evaluated multiple factors associated with the stage at diagnosis in a population can be expanded to other cancers as well, providing opportunities for cancer prevention programs to target specific factors and identify populations at higher risk.

Keywords: Alberta's Tomorrow Project; breast cancer stage; breast neoplasms; early detection of cancer; lifestyle; neoplasm staging; simulation study; stage shift

1. Introduction

Breast cancer mortality has been declining at a steady rate in women since 1988 [1], primarily due to early detection and enhanced treatment options [1,2]. Still, 13% of all cancer deaths in Canada are breast-cancer-related, many of which might have been averted had they been diagnosed at an earlier stage [1]. Breast cancer stages are based on tumour characteristics, including size (T), lymph node involvement (N) and spread (metastases, M) to other tissues [3]. Breast cancer stages range from 0 to IV, with early-stage disease defined as stages I and II [4]. Early-stage disease at diagnosis confers substantial survival benefits because treatments are more effective and better tolerated [2]. The five-year net survival rates for Canadian females diagnosed at stage I is 99.8% but decreases to 74.0% for those diagnosed at stage III and only 23.2% at stage IV [4]. Since the primary prevention of breast cancer is not always possible, the detection of early-stage disease is critically important for cancer control [2,5].

Previous studies have examined delays in cancer diagnoses at a population or neighbourhood level, including effects of ethnicity and socioeconomic status [6]. A Canadian report on population-level factors showed that the likelihood of early cancer diagnosis differed among income groups, geographical areas, and immigration status [7]. Other Canadian studies also found lower annual mammography screening participation for women with low socioeconomic status despite free access to breast cancer screening programs for age-eligible women [8,9]. Organized breast cancer screening programs have led to more women being diagnosed at stages I and II (82%); however, there is still a significant proportion of breast cancer patients diagnosed in later stages [10]. Other reasons for late-stage diagnosis may involve individual-level factors, including health behaviours, lifestyle factors and health insurance coverage [6,11–13]. A prospective healthy cohort, such as Alberta's Tomorrow Project (ATP), is the ideal study design to investigate many individual-level factors associated with stage at diagnosis [14]. Access to free breast cancer screening for women aged 50–74 (the screening program eligibility age range at time of diagnosis) province-wide in Alberta, Canada reduces the barriers to healthcare access that have been identified in other studies.

The majority of research on individual-level factors associated with delayed diagnosis of breast cancer has found that a history of breast cancer screening decreases the risk of late-stage diagnosis, but other factors, such as body mass index (BMI), smoking, physical activity, diet, distance from health centre, pregnancy and comorbidity, have inconsistent associations [13,15,16]. Discrepancies between previous research studies may be explained by different methods of statistical adjustment, measurement error or misclassification, recall bias, and different populations or study designs.

The prevalence of individual-level risk factors that can change (i.e., increase or decrease) over time can be evaluated for their impact individually and collectively on shifting breast cancer stages at diagnosis in a population setting. The comprehensive modelling of individual-level rather than neighbourhood-level factors associated with breast cancer stage at diagnosis requires only limited resources and time, unlike intervention studies. The primary aim of this proof of principle simulation study is to explore the shifting of breast cancer stage at diagnosis downward by modifying individual factors singly or collectively

using simulated data. In addition, we also aim to evaluate factors that could be targeted in policy or prevention programs that collectively maximize the shift to stage I at diagnosis.

2. Methods

This simulation study required input values for individual-level factors that are associated with breast cancer stage at diagnosis, obtained from an appropriate statistical model fit to data on women diagnosed with breast cancer. These input values were acquired from a study of women who were diagnosed with breast cancer after enrolling in ATP between 2000 and 2008 [14]. ATP is a prospective health cohort that began in 2000 in which participants aged 35–69 years of age completed baseline questionnaires on personal and family history of cancer, dietary intake, cancer screening participation, reproductive history, and demographic information, such as household income, education, and employment status.

A total of 492 women who joined ATP and were subsequently diagnosed with an invasive breast cancer up to January 2018 were included (Figure 1). Most women were diagnosed with Stage I disease (51%), followed by 37% of the women with Stage II disease, then 10% with Stage III disease and 2% with Stage IV disease. The American Joint Committee on Cancer (AJCC) version 6 (2004–2017), 7 (2010–2017) and 8 (2018+) staging was used in this study [17–19]. The outcome of interest is stage at breast cancer diagnosis, which is ordered from lowest to highest (I–IV), resulting in ordinal-scale data. Estimates of the association between individual-level factors measured at enrolment and stage at breast cancer diagnosis were obtained in a previous study [20]. In brief, multivariable partial proportional odds (PPOs) models were utilized as well as random forests to identify individual factors associated with breast cancer stage at diagnosis. We categorized the factors from this PPO regression model from Wang et al. [20] as follows: (1) non-modifiable or not easily modifiable factors (age at diagnosis, household income, Elixhauser comorbidity index and breast cancer family history); and (2) modifiable or potentially modifiable factors. The estimated odds ratios from the multivariable PPO model for these individual-level modifiable factors that were statistically significant ($p < 0.08$) were selected for this study (Table 1, Supplementary Table S1). Additional details on the modelling approach and results for the analysis of real data can be found in Wang et al. [20].

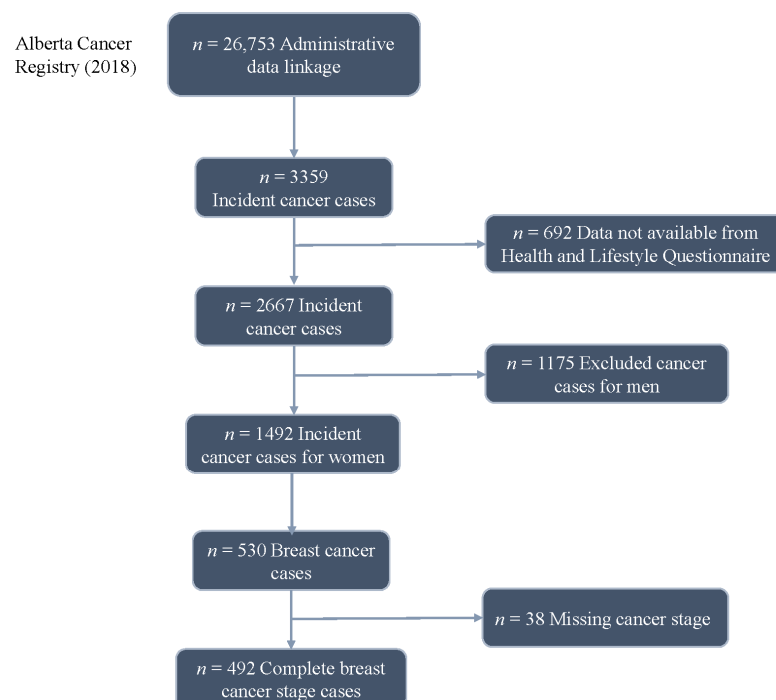


Figure 1. Study population selection from ATP participants who were diagnosed with breast cancer (reproduced from [20]).

Table 1. Properties of the selected modifiable factors from the ATP data set (N = 492).

Lifestyle Factor	Effect *	Mean	Standard Deviation
Total dietary energy intake (kcal/day)	Increased risk	1605	576
Total dietary protein intake (g/day)	Decreased risk	64	25
		N	%
Have had mammogram test (Yes)	Decreased risk	409	83.1
Have been pregnant (Yes)	Decreased risk	444	90.2
Spend time in the sun between 11 am and 4 pm from June to August (≥ 1 h/day)	Decreased risk	248	50.4
Currently experiencing stressful situation (≥ 1)	Increased risk	263	53.5
Use birth control pills (Yes)	Increased risk	427	86.8

* Effect: Odds ratio for increased risk factor > 1 and for decreased risk (protective factor) < 1 for later stage at diagnosis.

2.1. Proportional and Partial Proportional Odds Models

Two common regression models for ordinal response data are the proportional odds (POs) and partial proportional odds (PPOs) models. The proportional odds (POs) model assumes that each factor has the same effect across all levels of the ordered stage outcome variable. For each ordered level of stage, denoted by level j , the PO model is given as follows:

$$Stage_j = \alpha_j + \mathbf{X}_i\boldsymbol{\beta} + \varepsilon \quad (1)$$

where $\boldsymbol{\beta}$ is a vector of unknown regression coefficients for the individual risk factors, $\mathbf{X}_i$ is the matrix of individual risk factors for n subjects, α_j is the intercept for the j^{th} level of cancer stage and $\varepsilon \sim N(0, \sigma^2)$ is the random effect. Now, the probability of the cancer stage being higher than stage j is given as follows:

$$P(Stage > j) = \frac{\exp(\alpha_j + \mathbf{X}_i\boldsymbol{\beta} + \varepsilon)}{1 + \exp(\alpha_j + \mathbf{X}_i\boldsymbol{\beta} + \varepsilon)} \quad (2)$$

Equation (2) is the PO model which assumes that the effect of factor $\mathbf{X}_i$ does not vary across the levels of stages. This means the odds for a factor modelled as stage I ($j = 1$) vs. stages II + III + IV is the same as those for stage I + stage II vs. stage III + IV, etc. However, this assumption is not always realistic. For example, several studies have found that the probability of diagnosing cancer at a higher stage is greater for older individuals than younger ones, contradicting the POs assumption [21,22]. Wang et al. [20]’s study found that four factors (age at diagnosis, having been pregnant, spending time in the summer sun and currently experiencing stressful situation) did not satisfy the POs assumption. Therefore, they used the partial proportional odds (PPOs) model which has the following form:

$$P(Stage > j) = \frac{\exp(\alpha_j + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_j + \varepsilon)}{1 + \exp(\alpha_j + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_j + \varepsilon)} \quad (3)$$

In Equation (3), $\mathbf{X}_i^T\boldsymbol{\beta}$ is the contribution of the factors with equal slopes ($\boldsymbol{\beta}$) across all four levels of the stage outcome, and $\mathbf{Z}_i^T\boldsymbol{\gamma}_j$ is the contribution of the factors with different slopes ($\boldsymbol{\gamma}_j$) for each level j of the stage outcome. Both POs and PPOs regression models were used in our simulation study.

2.2. Data Simulation

Seven lifestyle factors that are potentially modifiable in a population and associated with breast cancer stage at diagnosis were considered in this study (see Table S1 for factor definitions). Protective factors from late stage at diagnosis included parity, spending time in the sun, having had a mammogram and high daily protein intake; however, increased risk of later stage at diagnosis included factors such as stress levels, high caloric intake and the use of birth control pills [20]. Table 1 describes these selected lifestyle factors, their individual effects on breast cancer stage and their distributional properties of mean and standard deviation (not median and quartiles as in Wang et al. [20]) in the study sample.

The distributions of these selected lifestyle factors from the dataset for the 492 ATP women diagnosed with breast cancer provided the inputs needed for generating the simulated data. For the continuous dietary factors (total dietary energy intake and total dietary protein intake), 1000 samples were simulated using a truncated multivariate normal distribution that used the estimated means, correlations and standard deviation parameters from the real data set. A thousand samples for each binary variable were simulated using a Bernoulli distribution based on the estimated probabilities and pairwise correlations from the real data set. Distributions of the simulated data were checked by comparing summary statistics with the summary statistics of observed data from the ATP study and found to be similar (Supplemental Table S3). The values of the regression coefficients β_j s and γ_j s as well as intercepts α_j for each j were obtained from the estimates found in Wang et al. [20]. Since the proportion of women diagnosed at stage IV was very small, stages III and IV were combined in our simulation study.

2.3. Stage Shifting

Three different scenarios were evaluated in the simulation study. In Scenario I, each of the seven selected lifestyle factors identified by Wang et al. [20] were individually varied over five possible values, including the mean or proportion obtained from their study of real data. The remaining variables not being shifted were fixed at the observed value from the aforementioned study. In Scenario II, the factors that shifted stage at diagnosis to stage I by at least 0.5% in Scenario I were modelled using cumulative not individual effects. Lastly, Scenario III evaluated four factors (total dietary energy intake, total dietary protein intake, having had a mammogram and spending time in the sun from 11 am to 4 pm, June–August) that could be targeted in policies regardless of their impact on the shift downwards, also using cumulative effects.

Changes in the mean levels of the dietary factors from the values observed in the ATP study data were based on the effect of the factor. Higher-fat diets have been associated with weight gain, and previous studies show that excess weight is a significant risk factor in breast cancer aetiology [23]. Therefore, total dietary energy intake levels were modelled using the mean value of 1605 kcal/day, as well shifting it higher and lower by 100 and 200 kcal/day in the first scenario (Design 1, Table 2). Taking the value of 64 g/day of protein as the baseline, four increasing values of average daily total protein consumption by 5 g/day increments were evaluated in Design 2. Only increasing values of average daily total protein consumption were considered as higher levels as recent evidence suggests higher recommended daily allowances can improve one's health [24].

Ever having a mammogram was found to be protective as mammograms can detect breast cancer in early stages [25]. In Ontario, a mammogram cancer screening report reported a decline from 2012 to 2018 in cancer screening participation rates (83% to 66%) [26]. Although participation rates are based on the percentage of eligible women having had a mammogram in a 30-month period and not having ever had one in their lifetime, as measured in this study, our surrogate variable can provide insights into possible participation trends. Therefore, it is reasonable to check the effect of both increasing as well as decreasing the percentage of women having ever had a mammogram by 5–10%; this was evaluated in Design 3. Additionally, women who are pregnant at a younger age and have more children have a reduced long-term risk of breast cancer because of hormone-induced

genetic changes in mammary glands, allowing mature breast cells to protect against breast cancer [27,28]. Women who are pregnant at an older age or women with low parity have a higher risk of breast cancer from increased oestrogen exposure, a hormone which can promote breast cancer [21,22]. Information about the ages of women when they were pregnant was not available in the ATP data. Therefore, increased proportions of 0.05 and decreased proportions of 0.05, 0.10 and 0.15 from the observed value of 0.90 were evaluated in Design 4 as women now tend to have children later in life and fewer children (e.g., declining birth rate) [29,30].

Table 2. Modified means and proportions of the lifestyle factors for Scenario I.

Design	Factor	Modified Distribution	* ORs (95% CI)
1	Total dietary energy intake (numeric, kcal/day)	(1800, 1700, 1605 , 1500, 1400)	1.07 (1.01–1.14)
2	Total dietary protein intake (numeric, g/day)	(64 , 70, 75, 80, 85)	0.98 (0.97–1.00)
3	Ever had a mammogram (proportion)	(0.7, 0.75, 0.83 , 0.90, 0.95)	0.57 (0.33–0.97)
4	Ever been pregnant (proportion)	(0.75, 0.80, 0.85, 0.9 , 0.95)	0.51 (0.27–0.98)
5	Spending time in the sun from 11 am to 4 pm, June–August (proportion)	(0.3, 0.4, 0.5 , 0.6, 0.7)	0.56 (0.32–0.99)
6	Currently experiencing stressful situations (proportion)	(0.7, 0.6, 0.53 , 0.4, 0.3)	2.87 (1.55–5.32)
7	Using birth control pills (proportion)	(0.70, 0.80, 0.87 , 0.92, 0.98)	1.64 (0.94–2.86)

* ORs (95% CI) are considered from Wang et. al. [20] Effects: increased risk if OR > 1 and decreased risk if OR < 1. Observed means and proportions from ATP study data in bold, underlined text.

Several previous studies found that regular sun exposure is associated with substantial decreases in death rates from certain cancers. A systematic review found that while increased sun exposure is detrimental for cutaneous cancers, it may be protective for certain cancers through the production of vitamin D from sun exposure [31]. For breast cancer, there is a potential protective effect on aetiology [32]. Thus, it is reasonable to consider changing the amount of sunlight exposure, so we considered five distributions taking the observed distribution for the middle of the design matrix and changing it by 10% (see Design 5 in Table 2).

Chronic stress can negatively impact health by weakening the immune system which can promote tumorigenesis [33]. Chronic stress is a modifiable factor, which can be mitigated through various stress management methods [34]. Therefore, we considered the effects of decreasing or increasing the proportion of people experiencing at least one stressful situation in Design 6. Lastly, the proportion of women using birth control pills for any reason was also varied above and below the observed proportion of 0.87 who had taken them. A recent study reported that the use of oral contraceptives amongst Canadian women dropped by almost 20% between 2006 and 2016 [35]. Design 7 varied the proportion of women who had ever taken the pill.

Scenario II evaluated the individual factors from Scenario I that shifted stage lower by at least 0.50% (Table 3. Designs 1–4, 7), as this could decrease the prevalence of late-stage disease by about 20%. Here, we only considered three values that shifted the stage at diagnosis proportions downwards towards the lower stages (see Table 4). Lifestyle factors were introduced one at time, building sequentially by adding each new factor to the previous ones in model. The final design, Design E in Table 4, represents the cumulative

effects of all five factors fixed at the shift amounts denoted in Table 3. The final scenario, III, evaluated these four factors regardless of their impact to shift stage at diagnosis: total dietary energy intake, total dietary protein intake, having ever had a mammogram and spending time in the sun from 11 am to 4 pm, June–August. These factors could be potential policy or guideline targets so they were evaluated in this final scenario.

Table 3. Average percentage change for each cancer stage after modifying the lifestyle factors' distributions in Scenario I by the shift amount and the effect of that shift on late-stage risk.

Design	Factor	Shift Amount	Average % Change			Effect
			Stage I	Stage II	Stage III and IV	
1	Total dietary energy intake (numeric, kcal/day)	−100 kcal/day	0.83	−0.33	−0.41	Increased Risk
2	Total dietary protein intake (numeric, g/day)	+5 g/day	2.18	−1.24	−0.91	Decreased Risk
3	Having ever had a mammogram (proportion)	+5%	0.63	−0.46	−0.49	Decreased Risk
4	Having ever been pregnant (proportion)	+5%	0.91	−1.44	0.38	Decreased risk
5	Spending time in the sun between 11 am and 4 pm, June–August (proportion)	+10%	0.48	0.30	−0.60	Decreased Risk
6	Currently experiencing stressful situations (proportion)	−10%	0.16	0.58	−0.68	Increased Risk
7	Using birth control pills (proportion)	+10%	−0.65	0.15	0.38	Increased Risk

Table 4. Modified distributions of the five most significant lifestyle health factors from Scenario II, their shifted values and the effect of these shifts on late-stage risk.

Design	Factor	Shifted Values	Effect
A	Total dietary energy intake (kcal/day)	(1400, 1500, <u>1605</u>)	Increased Risk
B	Total dietary protein intake (g/day)	(<u>64.0</u> , 70.0, 85.0)	Decreased Risk
C	Having ever had a mammogram	(<u>0.83</u> , 0.90, 0.95)	Decreased Risk
D	Having ever been pregnant	(<u>0.90</u> , 0.95, 0.95)	Decreased Risk
E	Using birth control pills	(0.70, 0.80, <u>0.87</u>)	Increased Risk

Observed means and proportions from ATP study data in bold, underlined text.

We simulated 20 realizations of the covariates and response for each combination of the distributions and generated 1000 samples per realization. The proportions of stages I, II and III were calculated, and median and the interval between lower and upper 2.25% of the ranked proportions (95% percentile interval (PI)) of the stages from 20 realizations were obtained. The R code used to generate the lifestyle factors and predict stage at diagnosis in our simulation study can be found in Supplemental Table S4.

3. Results

The percentage of ATP women diagnosed at stage II and higher (49%) is very similar to Alberta population estimates (52%) for women [36]. These women have access to breast cancer screening without a referral from a doctor if they are aged 50 or older, so the barriers experienced by women in other jurisdictions are minimized. About 20% of the ATP women

were less than 50 years of age when they were diagnosed with breast cancer, although most of them reported having had a mammogram previously (60.9%). The percentage of women between 40 and 49 years of age with a positive family history of breast cancer was 18.4%. However, other reasons for having had a mammogram previously are unknown for all participants.

Table 3 shows the average percentage change, either up (+) or down (−), for each cancer stage when one lifestyle factor is changed from its observed value in the ATP data set. The other six factors are fixed at their observed values from the ATP data set in this scenario. For example, in Design 1, if the total dietary energy intake is decreased by 100 kcal/day, it results in 0.83% more women being diagnosed at stage I and fewer women diagnosed at stage II (−0.33%) and stages III and IV (−0.41%). When the total dietary energy intake and other factors are fixed to their observed ATP data set values and the total dietary protein is increased by 5 g/day to 70 g/day, this results in 2.18% more women being diagnosed at stage I (Design 2). Although modifying the proportion of people currently experiencing stressful situations or spending more time in the sun from 11 am to 4 pm June–August did not increase the proportion of women diagnosed at stage I, both did substantially reduce the proportion of women diagnosed at stages III and IV. Generally, increasing protective factors and decreasing risk factors shifted more women being diagnosed with early-stage disease downwards, with the greatest effect being seen by modifying dietary protein intake. Figure 2 illustrates these single-factor effect changes graphically.

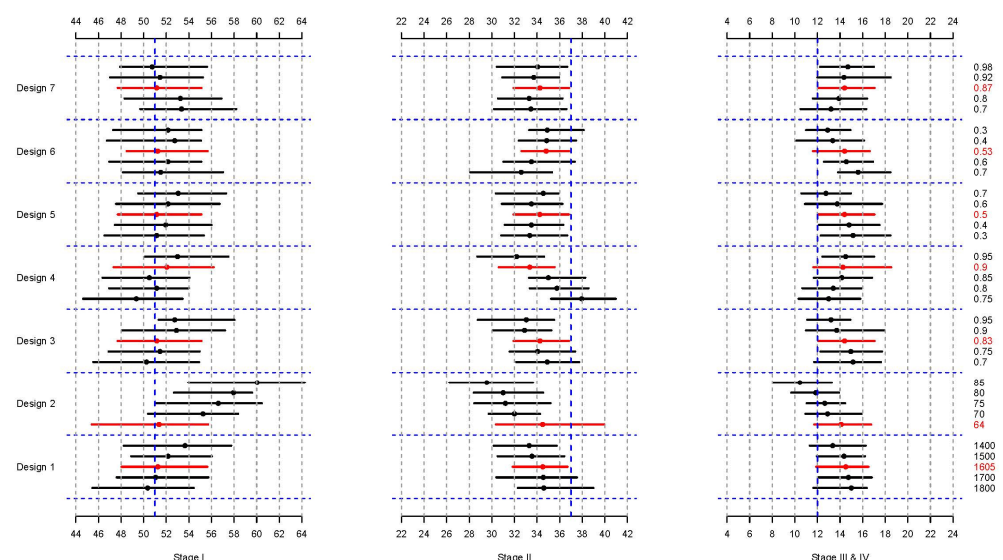


Figure 2. Impact of modifying the distributions for each lifestyle factor while fixing the other six factors at their observed values in Scenario I. The medians (dots) and 95% percentile intervals (PIs, horizontal lines) of the observed (red) and predicted (black) breast cancer stages after the modifications are shown. The blocks between the blue horizontal dotted lines denote the effects of each factor. The vertical black dotted lines represent 2% while the vertical blue dashed lines represent the percentage of women diagnosed in that stage in the real data set.

The factors that singly increased stage at diagnosis by at least 0.5% were then evaluated by modifying their values (means or proportions) to increase their protective effect or decrease their risk effect (Table 4 and Figure 3). Beginning with total dietary energy intake, three lower shifted values were evaluated with the remaining four factors fixed at their observed ATP values (Design A). It was then fixed at 1400 kcal/day in the subsequent designs because that value resulted in the greatest increase at stage I diagnoses. Next, the total dietary protein was varied with the remaining three factors fixed at their observed ATP values in Design B. In Design C, the proportion of women having ever had a mammogram was varied from 0.83 to 0.95 with the total dietary protein now fixed at 85 g/day (optimal value) and the remaining two factors fixed at their observed values. With the total dietary

intake fixed at 1400 kcal/day, the total dietary protein fixed at 85 g/day and the proportion of women having ever had a mammogram fixed at 0.95, the proportion of women who had ever been pregnant was increased while the proportion of women using birth control was fixed at 0.87 in Design D. Lastly, the first four factors were fixed at their optimal values and the proportion of women who had ever used birth control pills was decreased (Design E). Figure 3 illustrates the cumulative effects of increasing the proportion of breast cancer cases at stage I by optimizing these five factors and the corresponding decrease in the proportion of women diagnosed at breast cancer stages II, III and IV. Total dietary protein intake had the greatest single effect while the cumulative effect of additional changes in risk factors did not substantially increase the proportion of women diagnosed with early disease.

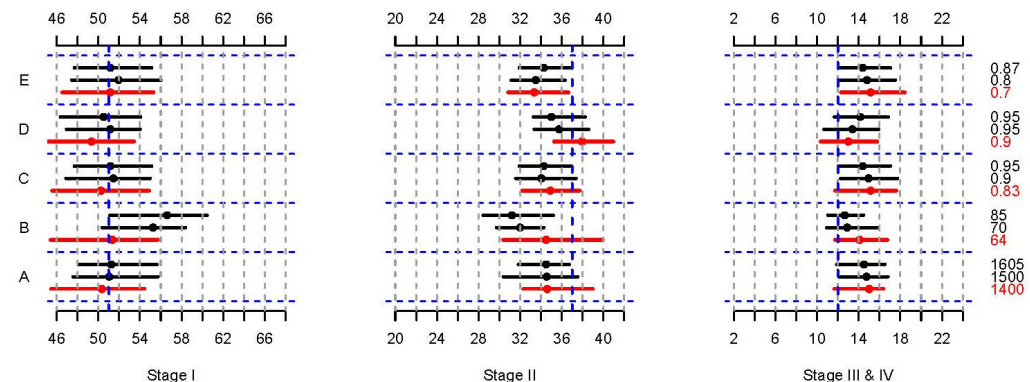


Figure 3. Cumulative effects of modifying the distributions of the factors from Designs A to E and setting them to their optimal values for most significant lifestyle factors in Scenario II. The medians (dots) and 95% percentile intervals (PIs, horizontal lines) of the observed (red) and predicted (black) breast cancer stages after the modifications. The blocks between the blue horizontal dotted lines denote the effects of each factor. The vertical black dotted lines represent 2% while the vertical blue dashed lines represent the percentage of women diagnosed in that stage in the real data set.

The results for Scenario III, concerning shifting lifestyle factors that could be potential policy targets, are given in the additional materials section (Additional File 2: Supplementary Table S2 and Figure S1). Using the same modelling strategy as that in the second scenario, the cumulative effects from adding each factor to the model at its most impactful level while fixing the others also showed that increasing the total dietary protein intake to 85 g/day had the biggest impact. Figure S1 shows these results graphically.

4. Discussion

This simulation study focused on modifying selected lifestyle factors associated with the stage of breast cancer at diagnosis to shift the cancer stage at diagnosis downward. Data were simulated based on lifestyle factors and their corresponding odds ratio estimates were obtained by Wang et. al. [20] using partial proportional models. Factors that increased the risk of early-stage diagnosis or those that decreased the risk of late-stage diagnosis were evaluated, the former by increasing its prevalence in the population and the latter by decreasing them. Both individual and cumulative effects were considered, including those that could be targeted in cancer prevention programs or policies. Increasing the total protein intake by 5 g/day on average in a population had the greatest effect on both increasing the proportion of breast cancer diagnoses at stage I and on decreasing the proportion diagnosed at stage III and IV. The next most impactful factor was total dietary energy intake for which a reduction of 100 kcal/day also led to a shift in the cancer stage downwards.

Dietary factors have been previously identified with the risk of developing breast cancer or with impacts on survival [37]. A recent meta-analysis suggests that higher-fat diets are associated with an increased risk of breast cancer [23]. There is strong evidence that oestrogen level is associated with breast cancer risk and that increased dietary fat may increase the production of endogenous oestrogen [15,23]. Our results found that

increasing total dietary energy (caloric) intake can increase the stage at breast cancer diagnosis, suggesting that low-fat diets could be beneficial in reducing one's breast cancer risk and stage of cancer diagnoses. Results from the Nurses' Health Study cohort of 6348 women diagnosed with stage I to III breast cancer concluded that there was a modest survival advantage with higher intakes of protein [38]. However, overall, biomarker-calibrated total protein intake was not associated with breast cancer incidence or mortality in another recent study [39]. Some studies suggest that higher vegetable protein intake was associated with lower breast cancer risk while higher animal protein intake was associated with an increased risk [39]. Our results also suggest that increased dietary protein intake is associated with early-stage breast cancer diagnoses; however, the model is limited to total intake and cannot evaluate vegetable protein and animal protein intake individually. There are potential mechanisms linking body fat and oestrogen levels with breast cancer risk but how these factors are associated with diet and physical activity requires substantially more research [15,23]. A recent systematic review and meta-analysis suggests that spending greater than one hour a day in the sun during the summer months could reduce the risk of breast cancer development [32]. However, few studies in the review accounted for skin type and sunscreen use [32]; studies which did account for these effects did not determine a significant effect [32]. Our results did not find a significant effect of outdoor sunlight exposure on lowering the stages of breast cancer at diagnosis, but skin type and sunscreen use were not incorporated as these were not available.

Breast density and BMI are also factors which may affect breast cancer stage at diagnosis. A recent review found that denser breasts can mask cancers in mammograms due to only a 50% sensitivity for the highest breast density class [40]. As a result, there is a lower mortality reduction in women with dense breast tissues, and they have potentially increased chances of breast cancer detection at later stages [40]. However, this is not a modifiable risk factor nor was it available in the cohort data; thus, it was not included. In a recent Korean study, increased BMI was associated with advanced stage breast cancer diagnosis: 41% of people with obesity had breast cancer stages T2 to T4, compared to 28% and 23% in patients with a BMI of 25–29 or <25, respectively [41]. The Korean study also determined that obesity was potentially associated with a reduced likelihood of having palpable tumours [41]. While BMI was not statistically significant in the analysis of real data and so was not evaluated in this study, total dietary caloric intake may represent an individual's potential BMI, based on average caloric requirements by age but not physical activity levels. Therefore, future studies should consider breast density, BMI and other stage-related factors.

While other early-stage breast cancer models exist, they differ in the factors included in their models [42,43]. A recent early breast cancer staging model utilized mammography and biopsy data to predict the stage at diagnosis; however, while that model used data from cancer screening or diagnostic procedures, it did not suggest how to reduce the risk of breast cancer being diagnosed at late stages [42]. Similarly, Cancer Intervention and Surveillance Modeling Network (CISNET) models also do not utilize lifestyle factors in addition to screening and screening-related factors to determine breast cancer risk [43]. Furthermore, the models determine risk reductions in incidence and mortality but do not focus on breast cancer stage [43]. The approach taken in our proof of principle study incorporated modifiable lifestyle factors, which enabled the effects of specific factors associated with breast cancer stage to be systematically evaluated.

An important advantage of our modelling approach is its flexibility—it can be adapted for various cancer screening and prevention programs that focus on targeted screening and prevention messages. Another strength of this study is that the combined effects of two or more factors were modelled, as in real-life scenarios in which individuals may be affected by multiple factors. The single or cumulative effects of multiple factors could also be evaluated. Furthermore, the outcome, stage at diagnosis, was modelled on an ordinal scale rather than combining stage categories into early versus late, thereby potentially providing greater insights into factors that vary in their association across stage levels. A

final strength of this simulation study is that our results could be obtained quickly and at a very low cost, compared to large population interventions that typically focus on a single factor.

However, this study also has some limitations. The estimated ORs for the significant lifestyle factors were based on a single study utilizing data from one province, so they may not be similar to those of other populations, reducing the simulation's generalizability. Although this simulation study approach could utilize various factors for various cancers, the focus was only on breast cancer in women. Some factors cannot be modified after the fact, such as having ever been pregnant, although changes in a population over time can occur. In addition, total dietary protein intake was only evaluated with increasing consumption as that would lead to better health, so the effects of decreasing consumption were not considered. The simulation model met most of the components of the STRESS guidelines, although additional validation and verification would be needed prior to implementing new cancer prevention programs [44]. Additional validation approaches and a decision-making framework could be used to verify and validate this modelling approach with specific validation metrics considered beyond the ones utilized here [45]. Lastly, some factors could be measured differently in other studies or with more granularity, such as the source of dietary protein.

5. Conclusions

Reducing the diagnosis of late-stage breast cancer, that is, shifting the stage at diagnosis downwards, will naturally lead to reductions in mortality rates. This is an important benefit to diagnosing breast cancer—or most cancers—at an early stage. Our proof of principle modelling approach enables cancer prevention programs to evaluate the potential impact of changes in modifiable lifestyle factors in a population and then develop interventions that target those factors. If lifestyle factors that increase the risk of late-stage diagnoses are becoming more prevalent in a population, then cancer control programs can prepare for the potential increase in late-stage diagnoses and devise strategies to mitigate them. Policies could also be created or modified too. Other researchers can generalize this approach to factors beyond lifestyle and to different types of cancer. Identifying the most influential factors that contribute to breast cancer stage at diagnosis, especially those that are modifiable, continues to be relevant for cancer risk reduction.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers16061201/s1>, Figure S1: Impact from modifying the distributions for policy target factors from Scenario I; Table S1: Data Dictionary; Table S2: Modified distributions of lifestyle and health factors for policy targets; Table S3: Comparison of summary statistics between the observed data and simulated data for all of the variables included in the final multivariable partial proportional odds regression model for the real data analysis; Table S4: Simulation study R code.

Author Contributions: Conceptualization, G.P. and K.A.K.; methodology, G.P. and K.A.K.; software, G.P.; formal analysis, G.P. and Q.W.; resources, K.A.K.; data curation, G.P. and Q.W.; writing—original draft preparation, G.P. and M.K.; writing—review and editing, all authors; visualization, G.P.; supervision, K.A.K.; project administration, K.A.K.; funding acquisition, P.J.R., L.S. and K.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by an operating grant received from the Canadian Institutes for Health Research (Primary Investigators Kopciuk, Robson, and Shack), grant number PJT-148774.

Institutional Review Board Statement: Ethics approvals from the Health Research Ethics Board of Alberta—Cancer Committee were obtained for this study (HREBA.CC—16-0814) and for ATP (HREBA.CC—17-0461). ATP participants provided written informed consent for the use of their de-identified data, including linkage with administrative databases, by researchers whose projects have been scientifically reviewed and who have obtained ethics approval. All methods were carried out in accordance with relevant guidelines and regulations.

Informed Consent Statement: Not applicable.

Data Availability Statement: Access to individual-level data is available in accordance with the Health Information Act of Alberta and Alberta's Tomorrow Project (ATP) Access Guidelines at <https://myatpresearch.ca/DataAccess> (accessed on 14 October 2022).

Acknowledgments: Alberta's Tomorrow Project is only possible because of the commitment of its research participants, its staff and its funders: Alberta Health, Alberta Cancer Foundation, Canadian Partnership Against Cancer and Health Canada and substantial in-kind funding from Alberta Health Services. The cancer registry data were obtained through linkage with Surveillance & Reporting, Cancer Research & Analytics, Cancer Care Alberta. The views expressed herein represent the views of the author(s) and not of Alberta's Tomorrow Project or any of its funders. This study is based in part on data provided by Alberta Health. The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of the Government of Alberta. Neither the Government nor Alberta Health express any opinion in relation to this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

AJCC	American Joint Committee on Cancer
ATP	Alberta's Tomorrow Project
BMI	Body mass index
CISNET	Cancer Intervention and Surveillance Modeling Network
M	Metastases
N	Lymph node involvement
PO	Proportional odds
PPO	Partial proportional odds
T	Tumor size

References

1. *Canadian Cancer Statistics 2017*; Canadian Cancer Society: Toronto, ON, Canada, 2017.
2. World Health Organization. Breast Cancer. Available online: [https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Treatments%20for%20breast%20cancer%20are,the%20whole%20breast%20\(mastectomy\)](https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Treatments%20for%20breast%20cancer%20are,the%20whole%20breast%20(mastectomy)) (accessed on 23 February 2024).
3. Breast Cancer Stages. Available online: <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/stages-of-breast-cancer.html> (accessed on 1 August 2023).
4. Ellison, L.F.; Saint-Jaques, N. Five-year cancer survival by stage at diagnosis in Canada. *Health Rep.* **2023**, *34*, 3–15. [CrossRef]
5. World Health Organization. Promoting Cancer Early Diagnosis. Available online: <https://www.who.int/activities/promoting-cancer-early-diagnosis> (accessed on 22 February 2024).
6. Wang, F.; Luo, L.; McLafferty, S. Healthcare access, socioeconomic factors and late-stage cancer diagnosis: An exploratory spatial analysis and public policy implication. *Int. J. Public Pol.* **2010**, *5*, 237–258. [CrossRef]
7. *Examining Disparities in Cancer Control: A System Performance Special Focus Report 2014*; Canadian Partnership Against Cancer: Toronto, ON, Canada, 2014.
8. Shields, M.; Wilkins, K. An update on mammography use in Canada. *Health Rep.* **2009**, *20*, 7–19.
9. Chan, W.; Yun, L.; Austin, P.C.; Jaakkimainen, R.L.; Booth, G.L.; Hux, J.; Rochon, P.A.; Lipscombe, L.L. Impact of socio-economic status on breast cancer screening in women with diabetes: A population-based study. *Diabet. Med.* **2014**, *31*, 806–812. [CrossRef]
10. Breast Cancer in Canada. Available online: <https://www.canada.ca/en/public-health/services/publications/diseases-conditions/breast-cancer.html> (accessed on 23 February 2024).
11. Keegan, T.H.M.; Parsons, H.M.; Chen, Y.; Maguire, F.B.; Morris, C.R.; Parikh-Patel, A.; Kizer, K.W.; Wun, T. Impact of Health Insurance on Stage at Cancer Diagnosis Among Adolescents and Young Adults. *J. Natl. Cancer Inst.* **2019**, *111*, 1152–1160. [CrossRef] [PubMed]
12. Pinto, B.M.; Trunzo, J.J. Health behaviors during and after a cancer diagnosis. *Cancer* **2005**, *104*, 2614–2623. [CrossRef]
13. Taplin, S.H.; Ichikawa, L.; Yood, M.U.; Manos, M.M.; Geiger, A.M.; Weinmann, S.; Gilbert, J.; Mouchawar, J.; Leyden, W.A.; Altaras, R.; et al. Reason for Late-Stage Breast Cancer: Absence of Screening or Detection, or Breakdown in Follow-up? *JNCI J. Natl. Cancer Inst.* **2004**, *96*, 1518–1527. [CrossRef]
14. Robson, P.J.; Solbak, N.M.; Haig, T.R.; Whelan, H.K.; Vena, J.E.; Akawung, A.K.; Rosner, W.K.; Brenner, D.R.; Cook, L.S.; Csizmad, I.; et al. Design, methods and demographics from phase I of Alberta's Tomorrow Project cohort: A prospective cohort profile. *CMAJ Open* **2016**, *4*, E515. [CrossRef] [PubMed]

15. Łukasiewicz, S.; Czezelewski, M.; Forma, A.; Baj, J.; Sitarz, R.; Stanisławek, A. Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. *Cancers* **2021**, *13*, 4287. [\[CrossRef\]](#)
16. Allahqoli, L.; Mazidimoradi, A.; Momenimovahed, Z.; Rahmani, A.; Hakimi, S.; Tiznobaik, A.; Gharacheh, M.; Salehiniya, H.; Babaey, F.; Alkatout, I. The Global Incidence, Mortality, and Burden of Breast Cancer in 2019: Correlation With Smoking, Drinking, and Drug Use. *Front. Oncol.* **2022**, *12*, 921015. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Amin, M.B.; Greene, F.L.; Edge, S.B.; Compton, C.C.; Gershenwald, J.E.; Brookland, R.K.; Meyer, L.; Gress, D.M.; Byrd, D.R.; Winchester, D.P. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA Cancer J. Clin.* **2017**, *67*, 93–99. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Edge, S.B.; Compton, C.C. The American Joint Committee on Cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann. Surg. Oncol.* **2010**, *17*, 1471–1474. [\[CrossRef\]](#)
19. Schmoll, H.J. *AJCC Cancer Staging Manual*, 6th ed.; Lippincott Raven Publishers: Philadelphia, PA, USA, 2002; Part VII.
20. Wang, Q.; Aktary, M.L.; Spinelli, J.J.; Shack, L.; Robson, P.J.; Kopciuk, K.A. Pre-diagnosis lifestyle, health history and psychosocial factors associated with stage at breast cancer diagnosis—Potential targets to shift stage earlier. *Cancer Epidemiol.* **2022**, *78*, 102152. [\[CrossRef\]](#)
21. Van Herck, Y.; Feyaerts, A.; Alibhai, S.; Papamichael, D.; Decoster, L.; Lambrechts, Y.; Pinchuk, M.; Bechter, O.; Herrera-Caceres, J.; Bibeau, F.; et al. Is cancer biology different in older patients? *Lancet Healthy Longev.* **2021**, *2*, e663–e677. [\[CrossRef\]](#)
22. Richie, R.C.; Swanson, J.O. Breast cancer: A review of the literature. *J. Insur. Med.* **2003**, *35*, 85–101. [\[PubMed\]](#)
23. Uhomoibhi, T.O.; Okobi, T.J.; Okobi, O.E.; Koko, J.O.; Uhomoibhi, O.; Igbinosun, O.E.; Ehibor, U.D.; Boms, M.G.; Abdulgaffar, R.A.; Hamed, B.L.; et al. High-Fat Diet as a Risk Factor for Breast Cancer: A Meta-Analysis. *Cureus* **2022**, *14*, e32309. [\[CrossRef\]](#)
24. Lonnie, M.; Hooker, E.; Brunstrom, J.M.; Corfe, B.M.; Green, M.A.; Watson, A.W.; Williams, E.A.; Stevenson, E.J.; Penson, S.; Johnstone, A.M. Protein for Life: Review of Optimal Protein Intake, Sustainable Dietary Sources and the Effect on Appetite in Ageing Adults. *Nutrients* **2018**, *10*, 360. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Miglioretti, D.L.; Lange, J.; van den Broek, J.J.; Lee, C.I.; van Ravesteyn, N.T.; Ritley, D.; Kerlikowske, K.; Fenton, J.J.; Melnikow, J.; de Koning, H.J.; et al. Radiation-Induced Breast Cancer Incidence and Mortality From Digital Mammography Screening: A Modeling Study. *Ann. Intern. Med.* **2016**, *164*, 205–214. [\[CrossRef\]](#)
26. *The Ontario Cancer Screening Performance Report 2020*; Cancer Care Ontario: Toronto, ON, Canada, 2021.
27. Kelsey, J.L.; Gammon, M.D.; John, E.M. Reproductive factors and breast cancer. *Epidemiol. Rev.* **1993**, *15*, 36–47. [\[CrossRef\]](#)
28. Lambe, M.; Hsieh, C.C.; Chan, H.W.; Ekblom, A.; Trichopoulos, D.; Adami, H.O. Parity, age at first and last birth, and risk of breast cancer: A population-based study in Sweden. *Breast Cancer Res. Treat.* **1996**, *38*, 305–311. [\[CrossRef\]](#)
29. *Epilogue Family-Centred Maternity and Newborn Care: National Guidelines*; Public Health Agency of Canada: Ottawa, ON, Canada, 2022.
30. Statistics Canada. Births. 2022. Available online: <https://www150.statcan.gc.ca/n1/daily-quotidien/230926/dq230926a-eng.htm> (accessed on 23 February 2024).
31. Van der Rhee, H.; Coebergh, J.W.; de Vries, E. Is prevention of cancer by sun exposure more than just the effect of vitamin D? A systematic review of epidemiological studies. *Eur. J. Cancer* **2013**, *49*, 1422–1436. [\[CrossRef\]](#)
32. Hiller, T.W.R.; O’Sullivan, D.E.; Brenner, D.R.; Peters, C.E.; King, W.D. Solar Ultraviolet Radiation and Breast Cancer Risk: A Systematic Review and Meta-Analysis. *Environ. Health Perspect.* **2020**, *128*, 16002. [\[CrossRef\]](#)
33. Dai, S.; Mo, Y.; Wang, Y.; Xiang, B.; Liao, Q.; Zhou, M.; Li, X.; Li, Y.; Xiong, W.; Li, G.; et al. Chronic Stress Promotes Cancer Development. *Front. Oncol.* **2020**, *10*, 1492. [\[CrossRef\]](#)
34. Can, Y.S.; Iles-Smith, H.; Chalabianloo, N.; Ekiz, D.; Fernández-Álvarez, J.; Repetto, C.; Riva, G.; Ersoy, C. How to Relax in Stressful Situations: A Smart Stress Reduction System. *Healthcare* **2020**, *8*, 100. [\[CrossRef\]](#)
35. Vogel, L. Canadian women opting for less effective birth control. *CMAJ* **2017**, *189*, E921–E922. [\[CrossRef\]](#) [\[PubMed\]](#)
36. *The 2021 Report on Cancer Statistics in Alberta*; Alberta Health Services: Edmonton, AB, Canada, 2021.
37. World Cancer Research Fund; American Institute for Cancer Research. *Diet, Nutrition, Physical Activity and Cancer: A Global Perspective*; Continuous Update Project Expert Report; 2018. Available online: <https://www.wcrf.org/diet-activity-and-cancer/> (accessed on 23 February 2024).
38. Holmes, M.D.; Wang, J.; Hankinson, S.E.; Tamimi, R.M.; Chen, W.Y. Protein Intake and Breast Cancer Survival in the Nurses’ Health Study. *J. Clin. Oncol.* **2017**, *35*, 325–333. [\[CrossRef\]](#)
39. Pan, K.; Larson, J.C.; Prentice, R.L.; Mortimer, J.E.; Neuhauser, M.L.; Manson, J.E.; Van Horn, L.; Rohan, T.E.; Lane, D.; Chlebowski, R.T. Protein Intake by Source and Breast Cancer Incidence and Mortality: The Women’s Health Initiative. *JNCI Cancer Spectr.* **2020**, *4*, pkaa101. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Gordon, P.B. The Impact of Dense Breasts on the Stage of Breast Cancer at Diagnosis: A Review and Options for Supplemental Screening. *Curr. Oncol.* **2022**, *29*, 3595–3636. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Lee, Y.; Kang, E.; Shin, H.-C.; Lee, H.; Yoon, K.; Kang, B.; Kim, E.-K. The Effect of Body Mass Index on Initial Breast Cancer Stage Among Korean Women. *Clin. Breast Cancer* **2021**, *21*, e631–e637. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Yampaka, T.; Noolek, D. Data Driven for Early Breast Cancer Staging using Integrated Mammography and Biopsy. *Asian Pac. J. Cancer Prev.* **2021**, *22*, 4069–4074. [\[CrossRef\]](#) [\[PubMed\]](#)

43. Van den Broek, J.J.; van Ravesteyn, N.T.; Mandelblatt, J.S.; Cevik, M.; Schechter, C.B.; Lee, S.J.; Huang, H.; Li, Y.; Munoz, D.F.; Plevritis, S.K.; et al. Comparing CISNET Breast Cancer Models Using the Maximum Clinical Incidence Reduction Methodology. *Med. Decis. Mak.* **2018**, *38*, 112s–125s. [[CrossRef](#)] [[PubMed](#)]
44. Monks, T.; Currie, C.S.M.; Onggo, B.S.; Robinson, S.; Kunc, M.; Taylor, S.J.E. Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines. *J. Simul.* **2019**, *13*, 55–67. [[CrossRef](#)]
45. Sargent, R.G. Verification and validation of simulation models. *J. Simul.* **2013**, *7*, 12–24. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.