

A Mathematical Framework Of SMS Reminder Campaigns for Pre- and Post-Diagnosis Check-Ups using Socio-Demographics: An In-Silco Investigation Into Breast Cancer

Elizaveta Savchenko^{1*}, Ariel Rosenfeld², Svetlana Bunimovich-Mendrazitsky¹

¹ Department of Mathematics, Ariel University, Ariel, Israel

² Department of Information Science, Bar Ilan University, Ramat-Gan, Israel

* Corresponding author: svetlanabu@ariel.ac.il

December 21, 2024

Abstract

Timely pre- and post-diagnosis check-ups are critical for **various diseases, in general, and for cancer , in particular**, as these often lead to better outcomes. Several socio-demographic properties have been identified as strongly connected with both clinical dynamics and (indirectly) with different individual check-up behaviors. Unfortunately, existing check-up policies typically consider only the former association explicitly. In this work, we propose a novel **computational** framework, accompanied by a high-resolution computer simulation, to investigate and optimize socio-demographic-based Short Messaging Service (SMS) reminder campaigns for check-ups. **We demonstrate our computational framework using extensive real-world data from the United States (US) population, focusing on breast cancer.** Our results indicate that optimizing an SMS reminder campaign based solely on simple socio-demographic features can bring about a statistically significant reduction in mortality rate compared to alternative campaigns. **These results indicate SMS reminder campaigns for pre- and post-diagnosis check-ups can be instrumental in improving healthcare outcomes. However, additional research is needed to bring about applicative tools.**

Keywords: Cancer; Check-Up Reminders; Socio-clinical dynamics; Healthcare policy management.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest/Competing interests

None.

Data availability

The data that has been used in this study is available by a formal request from the authors.

Author contribution

Conceptualization, formal analysis, investigation, software, visualization, and original draft were performed by Elizaveta Savchenko.

Validation, supervision, and manuscript editing were performed by Ariel Rosenfeld.

Supervision was performed by Svetlana Bunimovich-Mendrazitsky.

1 Introduction

Cancer is a generic name for a wide range of diseases in which cells in the human body grow and reproduce uncontrollably, resulting in a broad spectrum of clinical conditions and complications, commonly resulting in low life quality and early death [Weinberg, 1996]. In addition to its potentially deadly clinical consequences, cancer is also associated with poor quality of life and a substantial economic burden on patients, their families, and the entire healthcare systems [Mariotto et al., 2011, Yabroff et al., 2007]. The exact causes of cancer are yet to be fully understood, but a combination of genetic and environmental factors, including socio-demographic ones, are known to be strongly linked with the onset and progression of the disease [Weinberg, 1996].

There are many different types of cancer and each one may have a different set of risk factors and causes. Nonetheless, it is generally acknowledged that the early detection of the disease (via pre-diagnosis check-ups) and its appropriate monitoring for recurrence (via post-diagnosis check-ups) are pivotal in determining treatment options, reducing treatment costs, improving quality of life, and, arguably most important, lowering mortality rates across all patient groups [Schiffman et al., 2015, Mahvi et al., 2018]. Thus, developing and implementing proper cancer check-up policies, both Pre- and Post-Diagnosis (PPD), is crucial [McCall et al., 1995, Sano et al., 1993]. Unfortunately, determining an optimal PPD check-up policy for a given individual is still an open, yet active, area of research [Crosby et al., 2022, Etzioni et al., 2003, Dunstan et al., 2023, Elitzur et al., 2023]. For example, Yuniar et al. (2020) reviewed multiple policies for breast cancer PPD check-ups and found that the current policies lack clinical or economic evidence as to their cost-effectiveness from a healthcare service provider (HSP) perspective [Yuniar et al., 2020]. In a similar manner, Carles et al. (2011) performed an evaluation of 20 check-up policies for breast cancer by considering the various costs associated with implementing these policies and the potential subsequent medical costs [Carles et al., 2011]. The authors found that policies that are specifically tailored to different age groups result in significantly better outcomes. Accordingly, in order to derive socio-demographic-based PPD check-up policies, researchers commonly use mathematical models, a practice which has proved to be very powerful [Yaniv-Rosenfeld et al., 2023, Shami and Lazebnik, 2023, Lazebnik et al., 2021b, Bai et al., 2021, Piguillem and Shi, 2020, Yaniv-Rosenfeld et al., 2024]. That is, researchers rely on data-driven models, usually trained with machine learning algorithms, and optimization techniques to derive approximated or optimal PPD check-up policies in a fast, secure, and affordable manner [Heesterbeek et al., 2015, Nokes and Anderson, 1993, Deng et al., 2015, Lazebnik et al., 2021a, Veturi et al., 2023]. Unfortunately, an optimized PPD check-up policy need not necessarily be followed by all individuals alike [Evans et al., 2005, Roetzheim et al., 1999]. That is, the real-world effectiveness of a PPD check-up policy strongly depends on individual compliance [Rao et al., 2020] which, in turn, is known to be strongly linked to one's socio-demographic characteristics [Metasari and Nurlina, 2022]. For example, Pornet et al. (2010) showed compliance with colorectal cancer screening is significantly higher in women than in men and changes non-linearly with age [Pornet et al., 2010]. In order to increase individual compliance, particularly in high-risk patient groups, various stakeholders such as HSPs and governmental agencies have been implementing diverse compliance-increasing strategies such as health education programs, taxation, discount offers, and Short Messaging Service (SMS) reminder campaigns [Irmayati et al., 2019, Dube-Pule et al., 2021]. These strategies differ in their effectiveness, costs, and operational overhead. However, SMS reminder campaigns are often considered to be very effective, cheap, flexible, and operationally simple to implement compared to the mentioned alternatives. For example, Manakongtreecheep (2017) reviewed seven research projects concerning SMS reminder campaigns in Africa, concluding that vaccination reminders have led to improvements in vaccination uptakes under various metrics, whether through the increase in vaccination coverage, decrease in dropout rates, increase in completion rate or decrease in delay for vaccination [Manakongtreecheep, 2017]. In particular, Eze and Adeleye (2015) showed that the SMS campaign gain similar clinical benefits to other approaches such as

home visits while being significantly cheaper and much more scalable [Eze and Adeleye, 2015]. Mao et al. (2018) showed that SMS reminders can be used to reduce health and social inequity while providing better clinical outcomes for patients suffering from Human Immunodeficiency Virus [Mao et al., 2018]. Unfortunately, to the best of our knowledge, existing cancer PPD check-up policies are currently accompanied by naïve “one-size-fits-all” SMS reminder campaigns, where all patients are treated the same.

We instantiate our computational framework and provide a simplified *in silico* investigation into Breast Cancer. We choose breast cancer as this is the most common cancer worldwide, with more than 2.2 million new cases during 2020 alone, responsible for eight of all cancer cases this year according to the world cancer research fund¹. A more recent report reveals 680 thousand deaths due to breast cancer during 2018, being the leading cancer in the number of deaths in women with around 15.5% out of the total number of cancer-related deaths in females [Chhikara and Parang, 2023]. Breast cancer is also associated with a relatively high recurrence rate of around 4.5% [Chhikara and Parang, 2023]. As such, breast cancer is a prime candidate for the implementation of compliance-increasing strategies such as SMS reminder campaigns.

In this work, we propose a novel computational framework encompassing the fundamental aspects and decisions inherent to the optimization of socio-demographic-based SMS reminder campaigns for cancer PPD check-ups. Our computational framework, can be instantiated to any type of cancer and PPD policy, for which an optimal socio-demographic-based SMS reminder campaign is approximated through a Monte Carlo optimization technique. Fig. 1 shows a schematic view of the proposed model’s structure, input, and objective.

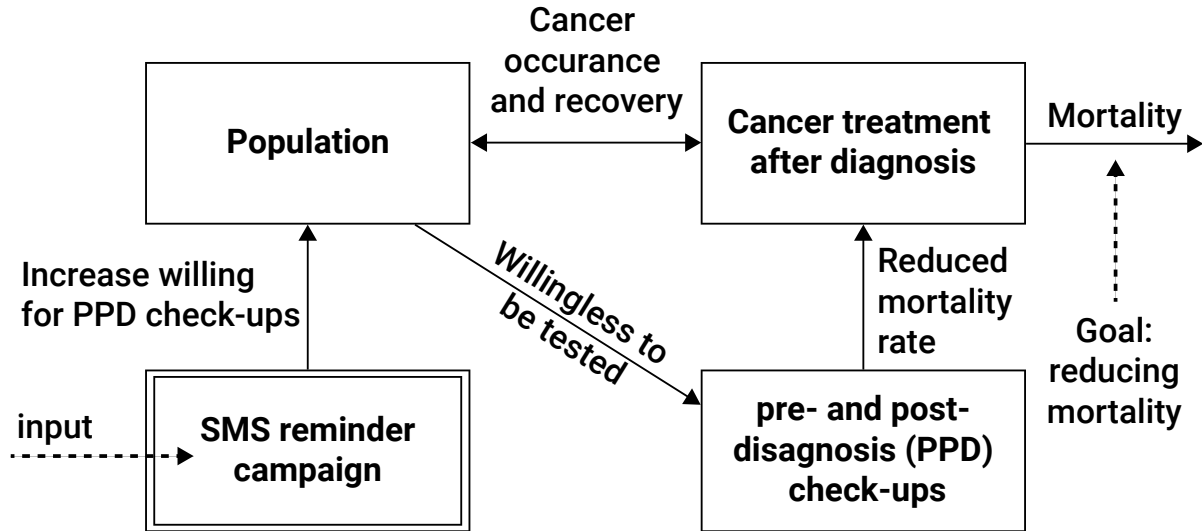


Figure 1: A schematic view of the framework’s input and objective and how it interacts with the current components in the dynamic.

The rest of the paper is organized as follows: Section 2 formally presents the computational framework, followed by Section 3 which outlines its implementation in the **simplified** case of breast cancer. **Finally, in section 4, we analyze and discuss the results in context, outline the limitations of this study, and highlight future work directions.**

2 Framework

Our proposed framework consists of several interconnected components which are detailed and discussed below. First, we define the clinical dynamics of cancer’s onset and progression in the context of PPD check-ups and

¹<https://www.wcrf.org/cancer-trends/worldwide-cancer-data/>

treatment. Then, we formalize the challenge of determining a PPD check-up policy as a resource-bounded optimization task. Based on these two components, we formulate the SMS reminder campaign optimization task and propose a Monte Carlo optimization technique that is shown to converge to a near-optimal solution given enough computational resources. Then, we propose a fitting procedure to set the parameters of an instance of the framework using historical data and facilitate the fitting of unavailable parameters' values that agree with realistic scenarios. Last, we detail how the different components are assembled into a single framework.

2.1 Clinical dynamics

Individuals are categorized into one of 10 clinical-oncological statuses (denoted by their α parameter): healthy (H), sick at stage j (S_j such that $j \in \{1, 2, 3, 4\}$), Recovered from stage j (R_j such that $j \in \{1, 2, 3, 4\}$), and dead (D) such that $N = H + S_1 + S_2 + S_3 + S_4 + R_1 + R_2 + R_3 + R_4 + D$ where N is the population's size at a given point in time. Individuals in the first (healthy) status were never diagnosed with cancer (H). If the individual never gets sick with cancer, it eventually naturally dies after γ steps in time and transforms to the dead (D) status. Healthy individuals can perform a pre-diagnosis check-up by either following the existing pre-diagnosis policy or due to symptoms. Pre-diagnosis check-ups are part of the proactive medical approach where healthy (or individuals without clear symptoms) obtain medical care to prevent diseases rather than treating them once a disease or symptoms are already preset [Doherty et al., 2019]. We assume that if the individual suffers from symptoms, s/he will choose to perform a pre-diagnosis check-up regardless of the PPD policy. **A policy-based pre-diagnosis check-up will result in one of two outcomes: either indicating that the individual is healthy (H) or s/he has cancer of any stage (S_1, S_2, S_3, S_4) [Maconi et al., 2003, Walter et al., 2015].** In a similar manner, a check-up due to symptoms will either result in a non-cancer diagnosis (i.e., healthy (H) from a cancer perspective – potentially indicating a non-oncological disease), or a cancer diagnosis with either stage 1, 2, 3, or 4 (S_1, S_2, S_3, S_4). The difference between the two cases is the probability of reaching each one of these states. Once an individual is diagnosed, treatment takes place immediately. Each individual has a personal duration and probability to recover, according to their socio-demographic properties and cancer stage. If the individual dies during the treatment, it transforms into the dead (D) status, and mortality due to the disease is recorded. Otherwise, the individual recovers and transforms to the corresponding recovery stage ($S_j \rightarrow R_j$). Similar to healthy individuals, any other individual eventually naturally dies after γ steps in time and transforms into the dead (D) status. Here, for simplicity, we assume that getting sick with cancer does not affect the individual's life expectancy if s/he recovers from it, while clinical evidence shows a correlational relationship between life expectancy and recovering from cancer [Sugita et al., 2022]. Recovered individuals can perform post-diagnosis check-ups following the post-diagnosis policy and/or due to symptoms. Here, both check-up types may result in a healthy outcome (i.e., the individual remains in the same clinical status) or in a cancer outcome (i.e., recurrent cancer) of stage $k \in [1, 2, 3, 4]$ [Beaugerie et al., 2014, Friedlander and Grogan, 2002]. For this model, we assume the recurrent cases are identical to the first cancer cases in any manner. During the illness, the transition from one stage to the consecutive one, until death, follows a socio-demographic-based dynamics indicated by $T_{1 \rightarrow 2}$, $T_{2 \rightarrow 3}$, $T_{3 \rightarrow 4}$, and $T_{4 \rightarrow D}$ [Rosevear et al., 2011]. Fig. 2 provides a schematic view of the clinical statuses and the flow between them as a result of PPD check-ups.

2.2 PPD check-up policy

Each individual in the population $p \in P$ is represented by a timed finite state machine [Alagar and Periyasamy, 2011] as follows: $p := (\alpha, \tau, \mu, \rho, e, g, s, \gamma)$ where $\alpha \in \{H, S_1, S_2, S_3, S_4, R_1, R_2, R_3, R_4, D\}$ is the current clinical status of the individual, $\tau \in \mathbb{N}$ is the time passed from the last change of the clinical status (α), $\mu \in [0, 1]$ is the probability that the individual will naturally comply with the pre-diagnosis check-up (without any compliance-increasing strategy), $\rho \in \mathbb{R}^+$ is the individual's degree of openness or susceptibility to compliance-increasing strategies, $e \in \mathbb{N}$ is the individual's age, $g \in \{\text{male}, \text{female}\}$ is the individual's gender, $s \in [1, 2, \dots, 10]$ is the relative socio-economic tenth percentile of the individual, and $\gamma \in \mathbb{N}$ the time steps until the individual naturally dies. In the scope of this study, we assume an abstract PPD check ups as there are a large number of possibilities that depend on multiple factors. For example, a standard PPD check up can be a blood test in order to obtain an initial evaluation of a patient's state. The blood test's outcomes may raise questions about abnormalities in the

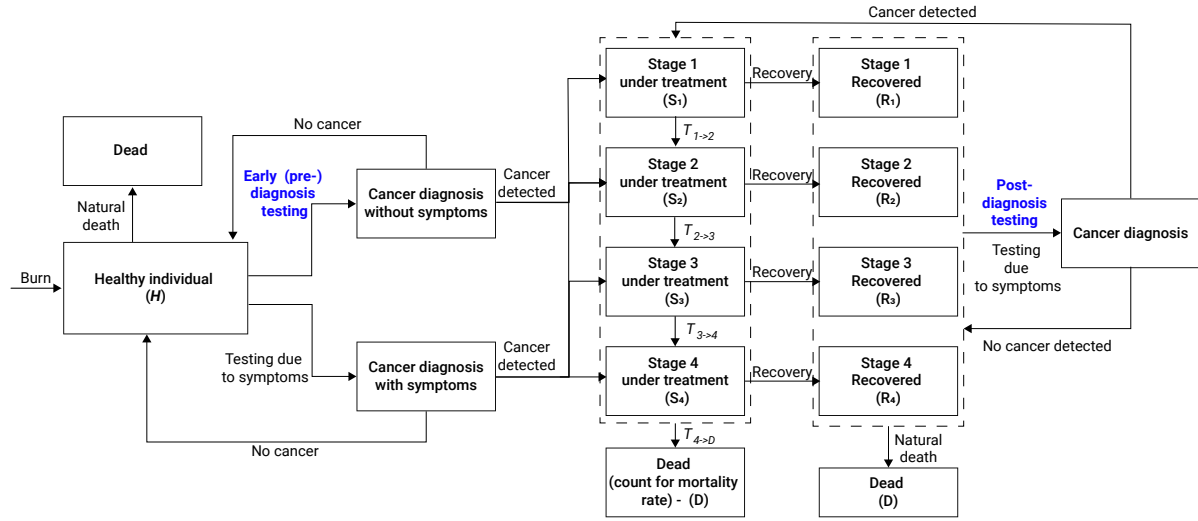


Figure 2: A schematic view of the clinical statuses and the transitions between them due to PPD check-ups and clinical deterioration. The optimization components of the framework are highlighted in blue.

obtained results which may require further, more dedicated, analysis of the patient. In turn, these can be resolved as false alarms, a suspicion that needs to be re-checked at a later time, or a detection of a disease. **Moreover, we assume that individuals decide to come to check ups voluntarily.**

From a socio-demographic perspective, the model uses 12 parameters as indicated by the tuple (e, g, s) : $T_{1 \rightarrow 2}, T_{2 \rightarrow 3}, T_{3 \rightarrow 4}, T_{4 \rightarrow D}, b, \gamma, \psi_i, \psi_r^j, \delta_i, \delta_r$ where $\psi_i \in [0, 1]$ is the probability that an individual would get sick with cancer for the first time, $\psi_r^j \in [0, 1]$ is the probability that the individual would get sick with recurrent cancer after recovering from stage $j \in \{1, 2, 3, 4\}$, δ_i is the duration delta between two consecutive pre-diagnosis check-ups recommended to that individual, and δ_r is the duration delta between two consecutive post-diagnosis check-ups recommended to that individual. We assume that the life expectancy of an individual, γ , is dependent on the socio-economic status (s) and gender (g) alone.

It is important to note that we consider PPD check-up policies to be mere *recommendations* that cannot be enforced on any individual.

2.3 SMS reminder campaign

SMS (Short Messaging Service) reminder campaigns are a popular compliance-increasing strategy in healthcare [Manakongtreecheep, 2017, Mao et al., 2018]. The implementing agency (e.g., HSP) can send any number of reminders to any subset of patients in order to encourage them to comply with the PPD policy [Bamba and Barnes, 2007]. Formally, an SMS reminder campaign is a function, Φ , that accepts the population, represented by a set of finite state machines, over time and returns when and how many SMSs each socio-demographic group (defined by the e, g , and s parameters of each individual) should get. For simplicity, we ignore the specific text used in the SMS campaign and the influence it may have with different configurations. Each SMS increases an individual's likelihood to follow the PPD policy as proposed by [Manakongtreecheep, 2017]:

$$\mu \leftarrow \mu + \rho \left(c_1 + c_2 (\log_{10}(n) - \log_{10}(n-1)) \right),$$

where c_1 and c_2 are the SMS effectiveness coefficients and n is the number of SMSs the individual received thus far. In addition, each SMS has a fixed cost $b \in \mathbb{R}^+$. Since the implementing agency is limited by some budget $B \in \mathbb{R}^+$ for a fixed duration $[t_0, t_f]$, deriving an SMS reminder campaign can be formulated as the following resource-bounded optimization task:

$$\min_{\Phi} MR_{[t_0, t_f]}(\Phi) \text{ s.t. } cost(\Phi) \leq B, \quad (1)$$

where $MR_{[t_0, t_f]}$ is a function that returns the average mortality rate during $[t_0, t_f]$ and $cost(\Phi)$ is a function that returns the total cost of an SMS reminder campaign Φ . Note that Φ makes decisions in a discrete manner. Since Φ is not able to pick a specific individual from each socio-demographic group as the μ and ρ parameters of each individual are not available to Φ in realistic cases, an SMS is sent in random inside each socio-demographic group in an equally distributed manner.

In order to solve the SMS reminder campaign optimization task (Eq. (1)), we used a Monte Carlo approach [Zio, 2013]. Namely, we sample the SMS reminder campaign parameter space. It takes the form of a four-dimensional tensor with one temporal dimension and three dimensions representing the age, gender, and socio-economic status of each socio-demographic group. Each value in the resulting matrix represents the relative part of the entire budget (B) allocated to SMS distribution among each socio-demographic group at each step in time. After the parameter values are set, we run the model for $t_f - t_0$ rounds and calculate the average mortality rate. If a configuration resulted in an average mortality rate smaller than any previous parameter configuration, we declare this configuration to be the best one so far. Since the parameter configuration space is finite, this computational procedure is guaranteed to converge to the optimal solution as the number of samples goes to infinity [Peres et al., 2006]. Overall, for the case of stochastic processes with random functions that are piecewise convex and a discrete state space, such as the case here, it has been proven elsewhere that this optimization process has a probability of reaching the optimal solution that approaches one exponentially fast with an increase in the number of simulations [Shapiro and Homem-de Mello, 2000].

2.4 Fitting procedure

In order to obtain the parameters that best fit historical records, we use the gradient descent (GD) method for the parameters' space following [Lazebnik et al., 2022b]. Formally, given the model's initial condition, the parameter space, historical data, and a loss function d we use the GD method [Curry, 1944] to find the parameters that minimize d on a fixed and finite duration in time $[t_0, t_f]$ such that $t_0 < t_f$. Formally, let us denote the parameter space by $\mathbb{P} \in \mathbb{R}^\epsilon$ where $\epsilon \in \mathbb{N}$ is the number of parameters in the implemented framework. In addition, a specific parameter configuration is denoted by $P \in \mathbb{P}$. We also denote the parameter configuration of the i_{th} iteration of the GD by P_i . Since the GD is computed on the parameter space with respect to a loss function d , the gradient is numerically obtained by following the five-point stencil numerical scheme [Zingg, 2000]:

$$\nabla P_i := \forall j \in [1, \dots, \epsilon] : \frac{-d(P_i(j, -2)) + 8d(P_i(j, 1)) - 8d(P_i(j, -1)) + d(P_i(j, 2))}{12h}$$

where $P_i(j, k) := P_i(x_1, \dots, x_j + kh, \dots, x_\epsilon)$ and $h \in \mathbb{R}^+$ is the step's size.

$$P_{i+1} \leftarrow P_i + \nabla P_i.$$

For our case, let us assume an instance of the framework, M_P , with the parameter configuration P . In order to fit historical data, we define a metric d between the prediction of the mortality rate and the historically recorded mortality rate:

$$d(M_P, H) := \sum_{t=t_0}^{t_f} |MR(M_P, t) - H(t)|, \quad (2)$$

where H is the historical recorded mortality rate, MR is a function that gets the simulation's state as defined by the distribution of the individuals' clinical state M_P , and $t \in \mathbb{N}$ is a point in time and returns the average predicted mortality rate of the framework at this point in time. We define d to be the mean absolute error of the predicted mortality rate since this is the metric that one, presumably, wishes to minimize using a designated SMS reminder campaign.

2.5 Assembling the components into a single framework

The framework has a global and discrete clock that all the individuals in the population follow. Namely, let us define each step in time as a round $t \in [1, \dots, T]$, where $T < \infty$. In the first round ($t = 1$), the population is created to satisfy a pre-define co-distribution of socio-demographic and clinical distributions. The PPD check-up policy is defined and fixed at this stage as well. Then, at each round $t \geq 1$, each individual is following the

PPD policy with its personal probability μ , according to its clinical status α . Afterward, the clinical dynamics are executed for each individual in the population, in a random order. Right after, the SMS reminder campaign is activated based on the population's state, sending, if any, SMSs to each socio-demographic group according to Ψ . In addition, the population naturally grows at a rate, $r \in \mathbb{R}$ which corresponds to its current size. All born individuals assume to keep the population's socio-demographic co-distribution as identical as possible and set to be healthy ($\alpha \leftarrow H$). Fig. 3 shows a schematic view of the framework's components and the interactions between them.

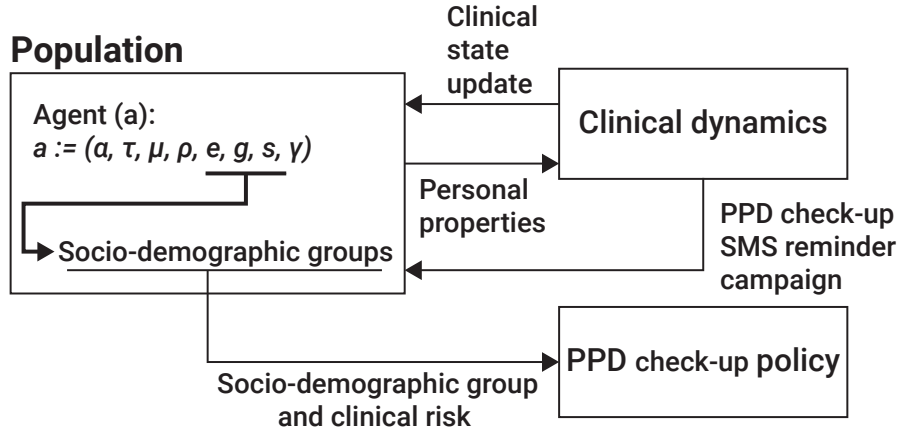


Figure 3: A schematic view of the framework's components and the interactions between them. Recall that the socio-demographic groups are defined by the individual's age (e), gender (g), and socio-economic status (s). These properties, along with clinical risk, determine the PPD check-up policy and the accompanying SMS reminder campaign. All personal features together are reflected in the clinical dynamics.

3 Investigation Into The United States with Breast Cancer

The following analysis consists of three parts: First, we outline the implementation of the above generic framework for the case of the United States with breast cancer. Then, we propose five candidate SMS reminder campaigns targeted at minimizing the expected mortality rate. Last, we explore the characteristics of the best-performing SMS reminder campaign and its statistical relationship to its underlying socio-demographic characteristics.

3.1 Setup

To realize the proposed framework, several parameters have to be set. Since most of the relevant data is available from the United States, it is the focus of our analysis. We rely on the following sources [Shariat et al., 2010, Madeb and Messing, 2004, Gupta et al., 2009, Densmore et al., 2019, MahdaviFar et al., 2016, Klapheke et al., 2018, Datta et al., 2006, Yip et al., 2020]. Specifically, in order to obtain the population's growth rate (b), we use the data of the United States population's growth between 1950 and 2020², fitting a standard exponential smoothing time-series forecasting model [Billah et al., 2006]. The window size for this forecasting model is obtained using the grid search method, ranging between 2 and 25 years [Liu et al., 2006] and aiming to minimize the mean absolute error. For the life expectancy, γ , we use the United States average life expectancy as reported by the United Nations³, ranging between 1950 and 2020. In order to get the life expectancy divided into gender and socioeconomic status, we used life expectancy gender differences reported by [Wang et al., 2013] and the socioeconomic status differences reported by [Chetty et al., 2016]. When the socioeconomic status is divided by tenths, this division

²<https://ourworldindata.org/grapher/population-growth-rate-with-and-without-migration>

³<https://www.macrotrends.net/countries/USA/united-states/life-expectancy>

results in 20 time series with 12 constraints - 10 for the socioeconomic status, one for the gender differences, and one to agree with the average reported life expectancy. Since there are more relationships than constraints, there is an infinite number of possible solutions to this computational task. We find a feasible solution using the least mean square method [Bjorck, 1996], obtaining a time-series function for the life expectancy for each gender and socioeconomic status separately in the same way the population’s growth rate was obtained.

As the underlying PPD check-up policy, we are inspired by the recommendation of the American Cancer that the rate of check-ups based on age. Specifically, we assume that individuals are encouraged to perform a pre-diagnosis check-up once a year following the age of 45 (none before that) and post-diagnosis check-ups once a year for those recovered from stage 1 or 2, or twice a year for those recovered from stage 3 or 4 *regardless of age*. These rates agree with several cancer types and are used as a reasonable “average” case. That is, people of all ages may be encouraged to perform check-ups and thus get SMS reminders. In order to find the SMS effectiveness coefficients (c_1, c_2), we used the data reported by [Manakongtreecheep, 2017] and fitted it using the least mean square method [Bjorck, 1996]. Importantly, as the data from [Manakongtreecheep, 2017] does not take into account socio-demographic groups, the SMS effectiveness coefficients are estimated once for the entire population. In addition, we averaged the SMS sending cost of five leading SMS providers in the US, as manually sampled in 2023, obtaining an average SMS cost of 0.049 US dollars.

In order to take the unique characteristics of breast cancer into account, one should find an estimate of the values for the proposed clinical dynamics of the model (see Fig. 2). To this end, we adopted the reported number from the American Cancer Society⁴ with 5-years survival rates of 99% for localized (commonly associated with stages I and II), 86% for regional (commonly associated with stages III), and 31% for distant (commonly associated with stages IV) resulting in $T_{1 \rightarrow 2} = T_{2 \rightarrow 3} = 5.48 \cdot 10^{-6}$, $T_{3 \rightarrow 4} = 3.91 \cdot 10^{-4}$, and $T_{4 \rightarrow D} = 7.94 \cdot 10^{-4}$. Moreover, according to the recent statistics from the Center for Disease Control and Prevention⁵ 12.5% of females and 0.15% of man will have breast cancer during their lifetime. We assume that recurrent probability is similar due to the complexity and diversity of changes depending on the clinical-specific course of treatments and the specific conditions of breast cancer.

For initialization, we used the age, gender, and socioeconomic data from the US in 2022 as reported by the US Census Bureau⁶. Since the data is not provided as the cross of all three properties (i.e., the number of individuals for each combination of age, gender, and socioeconomic status), we assume the age and gender distributions for each socioeconomic group are identical. This assumption is known to be false but it is commonly adopted due to a lack of finer-grained data at a publicly available level [MacIntyre and Hunt, 1997]. Overall, we include 333 million individuals in the initial condition, divided into 140 socio-demographic groups (i.e., two gender groups, seven age groups, and ten socioeconomic groups). In addition, we assume a budget of ten million dollars. We set $t_0 = 0$ and $t_f = 365$ and each round, t_i , to be a single day.

3.2 Parameter fitting

Given the framework’s initial condition and available parameter values (see Section 3.1), we used the proposed fitting procedure (see Section 2.4) in order to obtain the remaining parameter values that best align with historical data. Notably, as part of this process, we find the values of μ and ρ across the population. This is important as these values are not readily available and can only be approximated by fitting a model that describes the dynamics of the observed historical data. Fig 4 shows the model’s mean absolute error (MAE) from the historical data (red line), as defined by Eq. (2), and the coefficient of determination (R^2) (green line). Due to the stochastic nature of the framework, the results are shown as the mean of $n = 50$ simulations. We fitted both signals using the SciMed symbolic regression model [Keren et al., 2023], obtaining: $MAE(i) = 348i^{-0.357}$ and $R^2(i) = 0.498 + 0.110\ln(i)$ with $R^2 = 0.953$ and $R^2 = 0.959$, respectively, where $i \in [10, 20, \dots, 290, 300]$ is the optimization step’s index. One can notice that the optimization process coverage around an MAE of $9.0 \cdot 10^6$ with an R^2 of 0.79.

⁴<https://www.cancer.org/cancer/types/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>

⁵https://www.cdc.gov/cancer/breast/basic_info/index.htm

⁶<https://www.census.gov/topics/population/age-and-sex/data/tables.html>

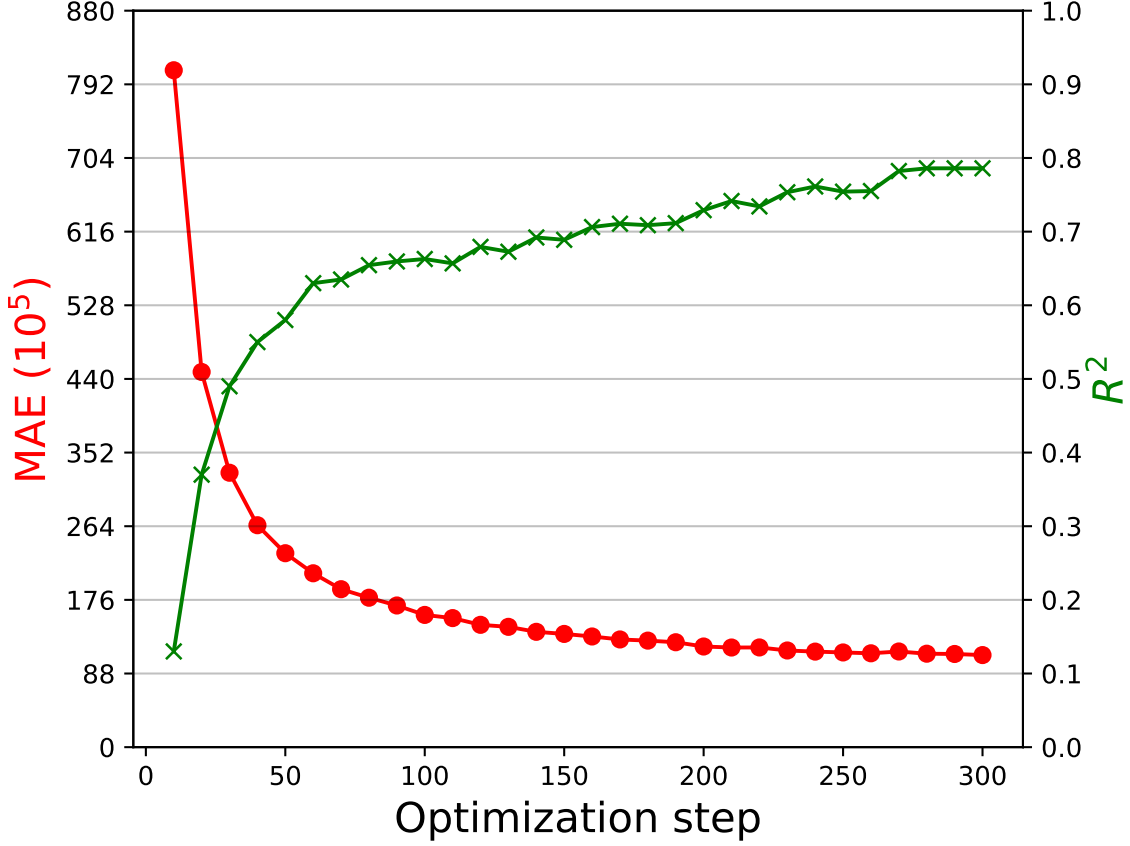


Figure 4: The fitting process. The MAE (red line) decreases as a function of the simulation steps while the R^2 (green line) is increasing. The results are shown as the mean of $n = 50$ simulations.

3.3 SMS reminder campaigns

We compare five SMS reminder campaigns with increasing levels of sophistication: “None”, Naïve, Greedy, Naïve Monte Carlo, and Socio-demographic Monte Carlo. The first, the “None” campaign, indicates that there are no SMS reminders sent at all. The Naïve campaign treats all individuals as if they belong to the same (single) socio-demographic group. As such, the Naïve campaign treats individuals differently only based on the PPD check-up policy and their clinical state (α). The Greedy campaign takes into consideration the mortality rate associated with each socio-demographic group such that the allocated budget to each group is proportional to its relative contribution to the overall mortality rate in the entire population. Namely, the Greedy campaign optimizes the SMS reminder campaign in every single step in time, ignoring the need to optimize for the entire duration $[t_0, t_f]$. Finally, we evaluate two variants of the Monte Carlo optimized SMS reminder campaigns - a Naïve one, which does not consider the socio-demographic characteristics of each individual and thus it only optimizes for the timing and volume of reminders sent to each individual based on their clinical statuses (like the Naïve campaign); and a socio-demographic one which applies the socio-demographic-based Monte Carlo optimization technique (as formally described in Section 2.3). Both variants were trained for 10,000 instances before being applied in the following analysis.

Fig. 5 presents the comparison of these five SMS reminder campaigns in terms of the relative mortality rate. The results are shown as the mean \pm standard deviation of $n = 100$ simulations, normalized to the “None” SMS campaign case. One can notice that the Monte Carlo socio-demographic campaign (the right-most column)

results in the lowest average mortality rate. In order to validate this outcome statistically, we computed ANOVA test [Girden, 1992] with a one-sided T-test post-hoc correction [Kalpić et al., 2011], obtaining that, indeed, this campaign is statistically better compared to the alternative campaigns with $p < 0.01$.

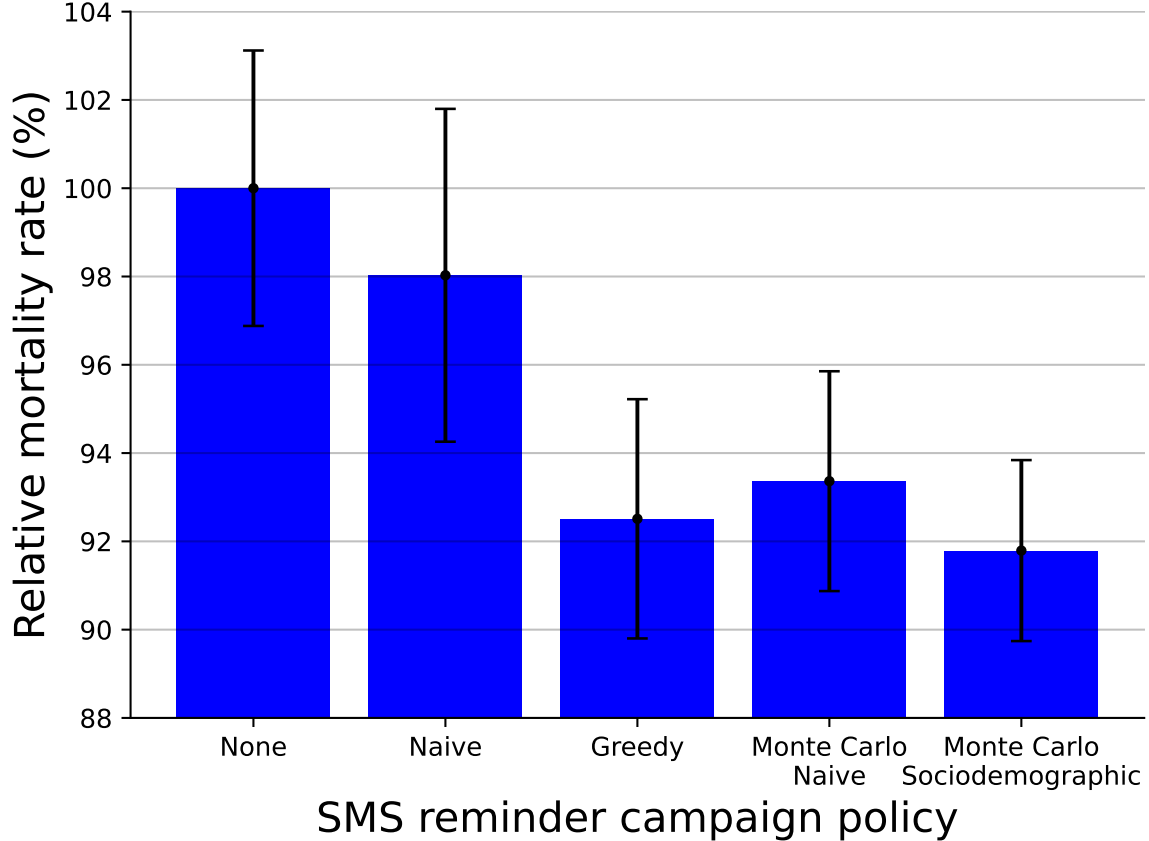


Figure 5: Average relative mortality rate, in percentages, across the five examined SMS reminder campaigns. The results are shown as the mean \pm standard deviation of $n = 100$ simulations, normalized to the “None” SMS campaign case.

3.4 Campaign analysis

We further explore the best-performing SMS reminder campaign and analyze the relative amount of resources invested in each socio-demographic group. Fig. 6 shows the average yearly number of SMSs an individual in each socio-demographic group would get based on the Monte Carlo socio-demographic SMS reminder campaign, divided into male and female heatmaps. The figure shows that, generally speaking, older individuals with higher socio-economic status require more resources. In particular, a sharp shift in resource allocation is observed around age 45 and the 40% percentile of socio-economic statuses. In addition, for most of the age and socio-economic statuses, females get more reminders. Similar results were obtained when testing for two, five, fifteen, and eighteen million dollar budgets with less than a five percent difference in the relative number of SMS an individual in each socio-demographic group gets between the two and eighteen million dollar budget. Notably, these SMS are for the entire PPD checkout campaign. Therefore, SMS reminders for both pre- and post-check ups for each socio-demographic group are allocated. **Importantly, as the model does not differ between pre- and post-check-up SMS, the budget is arbitrarily divided between the two with respect to the cancer survival portion in each**

socio-economic group.

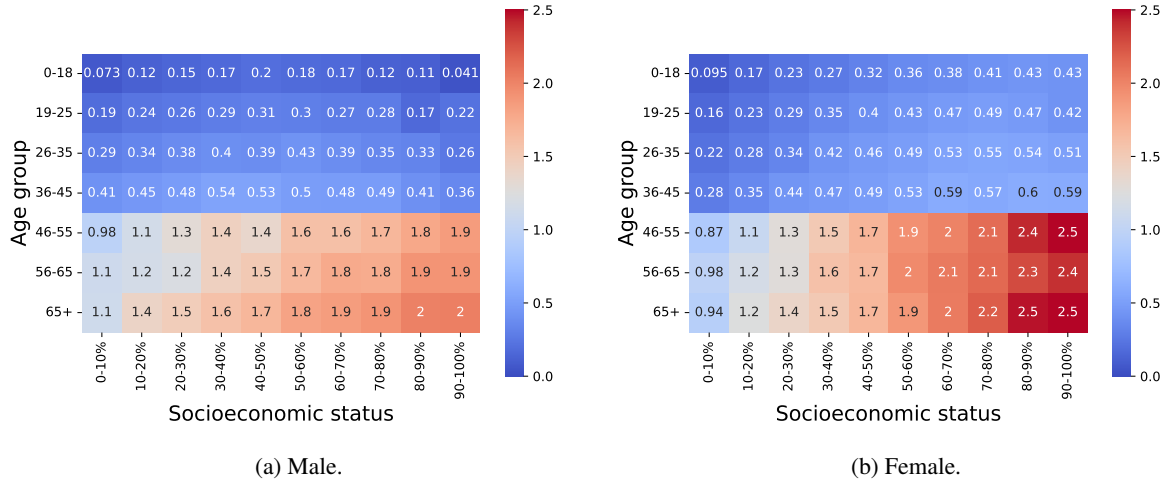


Figure 6: The percentage of resources allocated to each socio-demographic group based on the Monte Carlo socio-demographic SMS reminder campaign.

As Fig. 6 shows the resource allocation matrix (in percentage), an SMS campaign that sends discrete SMS to send to specific phone numbers can use this matrix, M , as follows. Given a budget, ψ , the (x, y, z) socio-demographic in the matrix, M , would receive $\text{floor}(\psi * M(x, y, z)/c)$ SMS messages such that c is the cost of a single SMS. Moreover, as the individuals in each socio-demographic group are identical as far as the model is aware, it would send the SMS in a random manner. **For example, let us assume a budget of one million dollars and an SMS cost of one dollar. Focusing on males, aged 65 or older with the top 10% socioeconomic status, let us assume this population has 10000 individuals. According to the figure, 2% of the resources should be allocated to this group, namely 20000 dollars. Since this socio-demographic group includes 10000 individuals and the cost of a single SMS is one dollar, each individual in the population would obtain exactly two SMS.**

4 Discussion

In this study, we proposed a novel computational framework that allows for the optimization and investigation of SMS PPD reminder campaigns and instantiated it for a simplified case of cancer check-ups. Specifically, considering the US as a representative example, we implemented the framework based on real-world historical data, and derived a Monte Carlo optimized SMS reminder campaign based on individuals' socio-demographic characteristics. The resulting campaign is shown to favorably compare with sensible alternatives.

Our results first demonstrate the presumed adequacy of our proposed computational framework, at least in the case of the US. As can be observed in Fig. 4, after fitting on historical data, the implemented framework can explain up to 80% of the breast cancer-related mortality variance. Using the currently practiced PPD check-up policy, a near-optimal socio-demographic SMS reminder campaign is found to be superior to several alternatives as presented in Fig. 5. This result generally agrees with the previous research on this subject [Evans et al., 2005, Roetzheim et al., 1999]. Namely, taking socio-demographic data into consideration leads to better performance of an SMS reminder campaign. Moreover, campaigns that operate for a long duration and consider the population's clinical distribution seem to operate more favorably compared to greedy approaches which utilize the SMS reminder campaign's budget quickly, raising much awareness yet in a short amount of time. This outcome is not surprising as it aligns with a larger trend of resource allocation strategies in the healthcare

services that show personalized data-driven, in general, and socio-demographic-driven resource allocation results in superior outcomes compared to the “one policy to rule them all” approach [Lazebnik, 2023, Xi et al., 2023, Lazebnik et al., 2022a, Vaingankar et al., 2017].

The derived campaign was also analyzed to identify the more targeted sub-populations. Fig. 6 suggests that the resource distribution is biased towards females and older individuals with a higher socioeconomic status. First, a sharp increase in the average yearly number of SMSs an individual gets occurred crossing the age of 45 years old. This shift can be easily explained by the practiced PPD check-up policy which indicates that pre-diagnosis check-ups are recommended only for individuals older than 45 years old. As such, up to this age group, the SMS reminder campaign is only required for post-diagnosis testing reminders. In addition, one can notice that the number of SMSs increases with age, as the number of cancer-recovered individuals is also increasing with age following more new cases as well as more individuals that recovered and stay healthy since [DePinho, 2000]. The slight increase towards higher socio-economic statuses can be associated with the μ and ρ parameters of these subpopulations. Namely, individuals of higher socio-economic status may be associated with higher alternative costs to come for a check-up and therefore require more SMSs on average, as also suggested by [Subramanian et al., 2009]. This outcome does not align with the overall modern sociological objective of healthcare equality [Rajkomar et al., 2018] as it allocates more resources towards individuals from higher socio-economic groups. This happens as the proposed model aims to reduce the overall mortality in the population (i.e., Eq. (2)), and healthcare equality is not explicitly considered in the model. **As such, the model’s results should be taken with a grain of salt.**

This study has important limitations which offer fruitful avenues for future research. First, it is assumed that the PPD check-ups are perfect and produce 100% accurate and reliable results as to an individual’s clinical status. Unfortunately, this is not true for most clinical tests in general, and oncological tests in particular [Lee et al., 2015]. Thus, our presented results should be treated as slightly over-optimistic of the expected realistic outcomes. In future work, we intend to tackle this shortcoming by integrating PPD check-up accuracy data. Second, since healthcare equality is a desirable goal [Braveman, 2006], extensions of the proposed model can take it into account as part of the model’s goal. **Such extension is not trivial as it raises multiple moral and ethical questions regarding the model’s objective and operations.** For instance, Let us consider the following simplistic scenarios to demonstrate such moral dilemmas - a population with 2 individuals but a budget to send a single SMS. According to the model, by sending an SMS to either of the individuals in the population, one reduces with some arbitrary and non-neglectable probability the chance this individual will die. On one hand, by sending an SMS to one of the individuals, society benefits overall as the mortality rate reduces. On the other hand, picking one out of the two individuals, either by chance (i.e. randomly) or by some informed decision, results in an unfair situation for the second individual. Formally, this is an instance of the Trolley problem [Thomson, 1976]. Future work should adopt a philosophical approach to morals in this context, such as Utilitarianism, Consequentialism, or Deontology, and integrate it into the model. Third, we assumed that the recurrent cases are identical to the first-time cancer patients in any way, such as the survival rates. This assumption does not necessarily hold for many types of cancer [Giordano et al., 2004] and should be refined in future work. **Fourth, further socio-demographic properties such as ethnicity should be included to refine the socio-demographic groups, obtaining more accurate results.** Furthermore, the effect of such socio-demographics on the individuals’ properties should be taken into consideration. For example, the impact of ethnicity on life expectancy. **Fifth, as the content of the SMS is very important [Rowley, 2008], future studies should explore different content and their influence on the individuals’ willingness to do check ups.** Sixth, alternative optimization techniques could be implemented to derive optimal SMS campaigns or potentially reduce the computational burden of obtaining a near-optimal one. Similarly, the lack of self-explainability of the resulting campaign should be tackled in order to promote its acceptance by stakeholders [Blkadek and Krawiec, 2022, Virgolin et al., 2018, Savchenko and Lazebnik, 2022, Kronberger et al., 2022, Liu et al., 2008, Best et al., 2009]. Finally, an individual’s occupation, ethnicity, smoking habits, and other contextual characteristics are known to be strong indicators of the risk of developing several types of cancer. Thus, integrating these and similar features to the age, gender, and socioeconomic status which are currently integrated into the framework, might help to obtain even better outcomes.

Overall, this study introduces a novel computational framework for optimizing socio-demographic-based SMS reminder campaigns. We demonstrated the framework’s effectiveness in reducing cancer-

related mortality rates for the case of the US in the case of breast cancer for a simplified case. Our results emphasize the significance of considering socio-demographic factors to enhance compliance and target specific sub-populations. This is the first work to present SMS reminder camping for clinical check up and future research avenues should address the limitations presented above in order to transform our framework into an applicative tool, offering a cost-effective and tailored approach to cancer check-up reminders with potential benefits for patient outcomes.

Acknowledgement

Elizaveta Savchenko wishes to thank Ariel University's financial support during this research.

References

- [Alagar and Periyasamy, 2011] Alagar, V. S. and Periyasamy, K. (2011). *Extended Finite State Machine*, pages 105–128. Springer London.
- [Bai et al., 2021] Bai, M., Cui, Y., Kong, G., and Zhang, A. Z. (2021). No panic in pandemic: The impact of individual choice on public health policy and vaccine priority. *University of Connecticut School of Business Research Paper*, (21-02).
- [Bamba and Barnes, 2007] Bamba, F. and Barnes, S. J. (2007). Sms advertising, permission and the consumer: a study. *Business Process Management Journal*, 13(6):815–829.
- [Beaugerie et al., 2014] Beaugerie, L., Carrat, F., Colombel, J.-F., Bouvier, A.-M., Sokol, H., Babouri, A., Car-bonnel, F., Laharie, D., Faucheron, J.-L., Simon, T., de Gramont, A., and Peyrin-Biroulet, L. (2014). Risk of new or recurrent cancer under immunosuppressive therapy in patients with ibd and previous cancer. *Gut*, 63(9):1416–1423.
- [Best et al., 2009] Best, A., Terpstra, J. L., Moor, G., Riley, B., Norman, C. D., and Glasgow, R. E. (2009). Building knowledge integration systems for evidence-informed decisions. *Journal of Health Organization and Management*, 23(6):627–641.
- [Billah et al., 2006] Billah, B., King, M. L., Snyder, R. D., and Koehler, A. B. (2006). Exponential smoothing model selection for forecasting. *International Journal of Forecasting*, 22(2):239–247.
- [Bjorck, 1996] Bjorck, A. (1996). Numerical methods for least squares problems. *Society for Industrial and Applied Mathematics*, 5:497–513.
- [Blkadek and Krawiec, 2022] Blkadek, I. and Krawiec, K. (2022). Counterexample-driven genetic programming for symbolic regression with formal constraints. *IEEE Transactions on Evolutionary Computation*.
- [Braveman, 2006] Braveman, P. (2006). Health disparities and health equity: Concepts and measurement. *Annual Review of Public Health*, 27(1):167–194.
- [Carles et al., 2011] Carles, M., Vilapriño, E., Cots, F., Gregori, A., Pla, R., Roman, R., Sala, M., Macia, F., Castells, X., and Rue, M. (2011). Cost-effectiveness of early detection of breast cancer in catalonia (spain). *BMC Cancer*, 11:192.
- [Chetty et al., 2016] Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., and Cutler, D. (2016). The association between income and life expectancy in the united states, 2001-2014. *JAMA*, 315(16):1750–1766.
- [Chhikara and Parang, 2023] Chhikara, B. S. and Parang, K. (2023). Global cancer statistics 2022: the trends projection analysis. *Chemical Biology Letters*, 10(1):451.

- [Crosby et al., 2022] Crosby, D., Bhatia, S., Brindle, K. V., Coussens, L. M., Dive, C., Emberton, M., Esener, S., Fitzgerald, R. C., Gambhir, S. S., Kuhn, P., Rebbeck, T. R., and Balasubramanian, S. (2022). Early detection of cancer. *Science*, 375(6586):eaay9040.
- [Curry, 1944] Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2(3):258–261.
- [Datta et al., 2006] Datta, G. D., Neville, B., Datta, N. S., and Earle, C. (2006). Gender disparities in bladder cancer survival: An assessment of socio-demographic factors. *Cancer Epidemiology, Biomarkers & Prevention*, 15(12):B38–B38.
- [Deng et al., 2015] Deng, R., Yang, Z., Chow, M.-Y., and Chen, J. (2015). A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, 11(3):570–582.
- [Densmore et al., 2019] Densmore, R., Hajizadeh, M., and Hu, M. (2019). Trends in socio-economic inequalities in bladder cancer incidence in canada: 1992–2010. *Can J Public Health*, 110:722–731.
- [DePinho, 2000] DePinho, R. (2000). The age of cancer. *Nature*, 408:248–254.
- [Doherty et al., 2019] Doherty, T., Di Pasquale, A., Michel, J.-P., and Del Giudice, G. (2019). Precision Medicine and Vaccination of Older Adults: From Reactive to Proactive (A Mini-Review). *Gerontology*, 66(3):238–248.
- [Dube-Pule et al., 2021] Dube-Pule, A., Zanoni, B. C., Conolly, C., Shabangu, M., and Archary, M. (2021). Evaluation of an sms-based mhealth intervention to enhance early infant diagnosis follow-up testing and assessment of postnatal prophylaxis. *Southern African Journal of HIV Medicine*, 22(1).
- [Dunstan et al., 2023] Dunstan, J., Villena, F., Hoyos, J. P., Riguelme, V., Royer, M., Ramirez, H., and Peypouquet, J. (2023). Predicting no-show appointments in a pediatric hospital in chile using machine learning. *Health Care Management Science*.
- [Elitzur et al., 2023] Elitzur, R., Krass, D., and Zimlichman, E. (2023). Machine learning for optimal test admission in the presence of resource constraints. *Health Care Management Science*.
- [Etzioni et al., 2003] Etzioni, R., Urban, N., Ramsey, S., McIntosh, M., Schwartz, S., Reid, B., Radich, J., Anderson, G., and Hartwell, L. (2003). The case for early detection. *Nature Reviews Cancer*, 3:243–252.
- [Evans et al., 2005] Evans, R. E., Brotherstone, H., Miles, A., and Wardle, J. (2005). Gender differences in early detection of cancer. *The Journal of Men’s Health and Gender*, 2(2):209–217.
- [Eze and Adeleye, 2015] Eze, G. U. and Adeleye, A. O. (2015). Enhancing routine immunization performance using innovative technology in an urban area of nigeria. *West African Journal of Medicine*, 34(1).
- [Friedlander and Grogan, 2002] Friedlander, M. and Grogan, M. (2002). Guidelines for the Treatment of Recurrent and Metastatic Cervical Cancer. *The Oncologist*, 7(4):342–347.
- [Giordano et al., 2004] Giordano, S. H., Buzdar, A. U., Smith, T. L., Kau, S.-W., Yang, Y., and Hortobagyi, G. N. (2004). Is breast cancer survival improving? *Cancer*, 100(1):44–52.
- [Girden, 1992] Girden, E. R. (1992). *ANOVA: Repeated measures*. Number 84. Sage.
- [Gupta et al., 2009] Gupta, P., Jain, M., Kapoor, R., Muruganandham, K., Srivastava, A., and Mandhani, A. (2009). Impact of age and gender on the clinicopathological characteristics of bladder cancer. *Indian J Urol*, 25(2):207–210.
- [Heesterbeek et al., 2015] Heesterbeek, H., Anderson, R. M., Andreasen, V., Bansal, S., De Angelis, D., Dye, C., Eames, K. T. D., Edmunds, W. J., Frost, S. D. W., Funk, S., Hollingsworth, T. D., House, T., Isham, V., Klepac, P., Lessler, J., Lloyd-Smith, J. O., Metcalf, C. J. E., Mollison, D., Pellis, L., Pulliam, J. R. C., Roberts, M. G., and Viboud, C. (2015). Modeling infectious disease dynamics in the complex landscape of global health. *Science*, 347(6227):aaa4339.

- [Irmayati et al., 2019] Irmayati, N., Yona, S., and Waluyo, A. (2019). Hiv-related stigma, knowledge about hiv, hiv risk behavior and hiv testing motivation among women in lampung, indonesia. *Enfermería Clínica*, 29:546–550.
- [Kalpić et al., 2011] Kalpić, D., Hlupić, N., and Lovrić, M. (2011). *Student's t-Tests*, pages 1559–1563.
- [Keren et al., 2023] Keren, L. S., Liberzon, A., and Lazebnik, T. (2023). A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge. *Scientific Reports*, 13:1249.
- [Klapheke et al., 2018] Klapheke, A., Yap, S. A., Pan, K., and Cress, R. D. (2018). Sociodemographic disparities in chemotherapy treatment and impact on survival among patients with metastatic bladder cancer. *Urologic Oncology: Seminars and Original Investigations*, 36(6):308.e19–308.e25.
- [Kronberger et al., 2022] Kronberger, G., de França, F. O., Burlacu, B., Haider, C., and Kommenda, M. (2022). Shape-constrained symbolic regression—improving extrapolation with prior knowledge. *Evolutionary Computation*, 30(1):75–98.
- [Lazebnik, 2023] Lazebnik, T. (2023). Data-driven hospitals staff and resources allocation using agent-based simulation and deep reinforcement learning. *Engineering Applications of Artificial Intelligence*, 126:106783.
- [Lazebnik et al., 2021a] Lazebnik, T., Aaroni, N., and Bunimovich-Mendrazitsky, S. (2021a). Pde based geometry model for bcg immunotherapy of bladder cancer. *Biosystems*, 200:104319.
- [Lazebnik et al., 2022a] Lazebnik, T., Batouth, Z., Bunimovich-Mendrazitsky, S., and Halachmi, S. (2022a). Predicting acute kidney injury following open partial nephrectomy treatment using sat-pruned explainable machine learning model. *BMC Medical Informatics and Decision Making*, 22:133.
- [Lazebnik et al., 2021b] Lazebnik, T., Bunimovich-Mendrazitsky, S., and Shami, L. (2021b). Pandemic management by a spatio-temporal mathematical model. *International Journal of Nonlinear Sciences and Numerical Simulation*.
- [Lazebnik et al., 2022b] Lazebnik, T., Shami, L., and Bunimovich-Mendrazitsky, S. (2022b). Spatio-temporal influence of non-pharmaceutical interventions policies on pandemic dynamics and the economy: the case of covid-19. *Economic Research-Ekonomska Istraživanja*, 35(1):1833–1861.
- [Lee et al., 2015] Lee, J., Won, K. K., Hyun, C. S., Jimi, H., and Ho, P. S. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: A practical review for clinical researchers-part ii. statistical methods of meta-analysis. *kjr*, 16(6):1188–1196.
- [Liu et al., 2008] Liu, O. L., Lee, H.-S., Hofstetter, C., and Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment*, 13(1):33–55.
- [Liu et al., 2006] Liu, R., Liu, E., Yang, J., Li, M., and Wang, F. (2006). Optimizing the hyper-parameters for svm by combining evolution strategies with a grid search. *Intelligent Control and Automation*, 344.
- [MacIntyre and Hunt, 1997] MacIntyre, S. and Hunt, K. (1997). Socio-economic position, gender and health: How do they interact? *Journal of Health Psychology*, 2(3):315–334.
- [Maconi et al., 2003] Maconi, G., Kurihara, H., Panizzo, V., Russo, A., Cristaldi, M., Marrelli, D., Roviello, F., de Manzoni, G., Di Leo, A., Morgagni, P., Bechi, P., Bianchi Porro, G., and M., T. A. (2003). Gastric cancer in young patients with no alarm symptoms: focus on delay in diagnosis, stage of neoplasm and survival. *Scandinavian Journal of Gastroenterology*, 38(12):1249–1255.
- [Madeb and Messing, 2004] Madeb, R. and Messing, E. M. (2004). Gender, racial and age differences in bladder cancer incidence and mortality. *Urologic Oncology: Seminars and Original Investigations*, 22(2):86–92.

- [Mahdavifar et al., 2016] Mahdavifar, N., Ghoncheh, M., Pakzad, R., Momenimovahed, Z., and Salehiniya, H. (2016). Epidemiology, incidence and mortality of bladder cancer and their relationship with the development index in the world. *Asian Pacific Journal of Cancer Prevention*, 17:381–386.
- [Mahvi et al., 2018] Mahvi, D. A., Liu, R., Grinstaff, M. W., Colson, Y. L., and Raut, C. P. (2018). Local cancer recurrence: The realities, challenges, and opportunities for new therapies. *CA: A Cancer Journal for Clinicians*, 68(6):488–505.
- [Manakongtreecheep, 2017] Manakongtreecheep, K. (2017). Sms-reminder for vaccination in africa: research from published, unpublished and grey literature. *Pan Afr Med J*.
- [Mao et al., 2018] Mao, L., Buchanan, A., Wong, H. T. H., and Persson, A. (2018). Beyond mere pill taking: Sms reminders for hiv treatment adherence delivered to mobile phones of clients in a community support network in australia. *Health & Social Care in the Community*, 26(4):486–494.
- [Mariotto et al., 2011] Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J., and Brown, M. L. (2011). Projections of the cost of cancer care in the united states: 2010–2020. *JNCI: Journal of the National Cancer Institute*, 103(2):117–128.
- [McCall et al., 1995] McCall, J. L., Cox, M. R., and Wattchow, D. A. (1995). Analysis of local recurrence rates after surgery alone for rectal cancer. *Int J Colorect Dis*, 10:126–132.
- [Metasari and Nurlina, 2022] Metasari, A. R. and Nurlina, A. (2022). Effect of health education about check breast self (sadari) on adolescent motivation in early detection of breast cancer. *International Journal of Public Health Excellence*, 2(1):244–247.
- [Nokes and Anderson, 1993] Nokes, D. J. and Anderson, R. M. (1993). Application of mathematical models to the design of immunization strategies. *Reviews in Medical Microbiology*, 4(1):1–7.
- [Peres et al., 2006] Peres, Y., Schramm, O., Sheffield, S., and Wilson, D. B. (2006). Random-turn hex and other selection games. *arXiv*.
- [Piguillem and Shi, 2020] Piguillem, F. and Shi, L. (2020). Optimal COVID-19 quarantine and testing policies. *Covid Economics*, 27:123–169.
- [Pornet et al., 2010] Pornet, C., Dejardin, O., Morlais, F., Bouvier, V., and Launoy, G. (2010). Socioeconomic determinants for compliance to colorectal cancer screening. a multilevel analysis. *Journal of Epidemiology & Community Health*, 64(4):318–324.
- [Rajkomar et al., 2018] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., and Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12):866–872.
- [Rao et al., 2020] Rao, I., Shaham, A., Yavneh, A., Kahana, D., Ashlagi, I., Brandeau, M. L., and Yamin, D. (2020). Predicting and improving patient-level antibiotic adherence. *Health Care Management Science*, 23:507–519.
- [Roetzheim et al., 1999] Roetzheim, R. G., Pal, N., Tennant, C., Voti, L., Ayanian, J. Z., Schwabe, A., and Krischer, J. P. (1999). Effects of Health Insurance and Race on Early Detection of Cancer. *JNCI: Journal of the National Cancer Institute*, 91(16):1409–1415.
- [Rosevear et al., 2011] Rosevear, H. M., , Lightfoot, A. J., Birusingh, K. K., Maymi, J. L., Nepple, K. G., and O'Donnell, M. A. (2011). Factors affecting response to bacillus calmette-guérin plus interferon for urothelial carcinoma in situ. *The Journal of Urology*, 186(3):817–823.
- [Rowley, 2008] Rowley, J. (2008). Understanding digital content marketing. *Journal of Marketing Management*, 24(5–6):517–540.

- [Sano et al., 1993] Sano, T., Sasako, M., Kinoshita, T., and Maruyama, K. (1993). Recurrence of early gastric cancer. follow-up of 1475 patients and review of the japanese literature. *Cancer*, 72:3174–3178.
- [Savchenko and Lazebnik, 2022] Savchenko, E. and Lazebnik, T. (2022). Computer aided functional style identification and correction in modern russian texts. *Journal of Data, Information and Management*, 4:25–32.
- [Schiffman et al., 2015] Schiffman, J. D., Fisher, P. G., and Gibbs, P. (2015). Early detection of cancer: Past, present, and future. *American Society of Clinical Oncology Educational Book*, (35):57–65.
- [Shami and Lazebnik, 2023] Shami, L. and Lazebnik, T. (2023). Financing and managing epidemiological-economic crises: Are we ready for another outbreak? *Journal of Policy Modeling*, 45(1):74–89.
- [Shapiro and Homem-de Mello, 2000] Shapiro, A. and Homem-de Mello, T. (2000). On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. *SIAM Journal on Optimization*, 11(1):70–86.
- [Shariat et al., 2010] Shariat, S. F., Sfakianos, J. P., Droller, M. J., Karakiewicz, P. I., Meryn, S., and Bochner, B. H. (2010). The effect of age and gender on bladder cancer: a critical review of the literature. *BJU International*, 105(3):300–308.
- [Subramanian et al., 2009] Subramanian, S., Bobashev, G., and Morris, R. J. (2009). Modeling the Cost-Effectiveness of Colorectal Cancer Screening: Policy Guidance Based on Patient Preferences and Compliance. *Cancer Epidemiology, Biomarkers & Prevention*, 18(7):1971–1978.
- [Sugita et al., 2022] Sugita, S., Morita, E., Fujiwara, M., Okuma, T., and Hozumi, T. (2022). Correlation between gait and life expectancy in patients with spinal metastases. *World Neurosurgery*, 163:e156–e161.
- [Thomson, 1976] Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59(2):204–217.
- [Vaingankar et al., 2017] Vaingankar, J. A., Chong, S. A., Abidin, E., Picco, L., Chua, B. Y., Shafie, S., Ong, H. L., Chang, S., Seow, E., Heng, D., Chiam, P. C., and Subramaniam, M. (2017). Prevalence of frailty and its association with sociodemographic and clinical characteristics, and resource utilization in a population of singaporean older adults. *Geriatrics & Gerontology International*, 17(10):1444–1454.
- [Veturi et al., 2023] Veturi, Y. A., Woof, W., Lazebnik, T., Moghul, I., Woodward-Court, P., Wagner, S. K., de Guimaraes, T. A. C., Varela, M. D., Liefers, B., Patel, P. J., Beck, S., Webster, A. R., Mahroo, O., Keane, P. A., Michaelides, M., Balaskas, K., and Pontikos, N. (2023). Syntheye: Investigating the impact of synthetic data on artificial intelligence-assisted gene diagnosis of inherited retinal disease. *Ophthalmology Science*, 3(2):100258.
- [Virgolin et al., 2018] Virgolin, M., Alderliesten, T., Bel, A., Witteveen, C., and Bosman, P. A. N. (2018). Symbolic regression and feature construction with gp-gomea applied to radiotherapy dose reconstruction of childhood cancer survivors. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1395–1402.
- [Walter et al., 2015] Walter, F. M., Rubin, G., Bankhead, C., Morris, H. C., Hall, N., Mills, K., Dobson, C., Rintoul, R. C., Hamilton, W., and Emery, J. (2015). Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *British Journal of Cancer*, 112:S6–S13.
- [Wang et al., 2013] Wang, H., Schumacher, A., Levitz, C., Mokdad, A. H., and Murray, C. J. L. (2013). Left behind: widening disparities for males and females in us county life expectancy, 1985–2010. *Population Health Metrics*, 11:8.
- [Weinberg, 1996] Weinberg, R. A. (1996). How cancer arises. *Scientific American*, 275(3):62–70.
- [Xi et al., 2023] Xi, Y., Ding, Y., Cheng, Y., Zhao, J., Zhou, M., and Qin, S. (2023). Evaluation of the medical resource allocation: Evidence from china. *Healthcare*, 11(6).

- [Yabroff et al., 2007] Yabroff, K., Warren, J., and Brown, M. (2007). Costs of cancer care in the usa: a descriptive review. *Nat Rev Clin Oncol*, 4:643–656.
- [Yaniv-Rosenfeld et al., 2024] Yaniv-Rosenfeld, A., Savchenko, E., Netzer, M., Elalouf, A., and Nitzan, U. (2024). Socio-demographic predictors of hospitalization duration among patients with borderline personality disorder. *Administration and Policy in Mental Health and Mental Health Services Research*.
- [Yaniv-Rosenfeld et al., 2023] Yaniv-Rosenfeld, A., Savchenko, E., Rosenfeld, A., and Lazebnik, T. (2023). Scheduling bcg and il-2 injections for bladder cancer immunotherapy treatment. *Mathematics*, 11(5).
- [Yip et al., 2020] Yip, W., Cacciamani, G., and Bhanvadia, S. (2020). Disparities in bladder cancer outcomes based on key sociodemographic characteristics. *Curr Urol Rep*, 21:24.
- [Yuniar et al., 2020] Yuniar, P., Robinson, S., Moorin, R., and Norman, R. (2020). Economic evaluation of breast cancer early detection strategies in asia: A systematic review. *Value in Health Regional Issues*, 21:252–263.
- [Zingg, 2000] Zingg, D. W. (2000). Comparison of high-accuracy finite-difference methods for linear wave propagation. *SIAM Journal on Scientific Computing*, 22(2):476–502.
- [Zio, 2013] Zio, E. (2013). *Monte Carlo Simulation: The Method*, pages 19–58. Springer London, London.