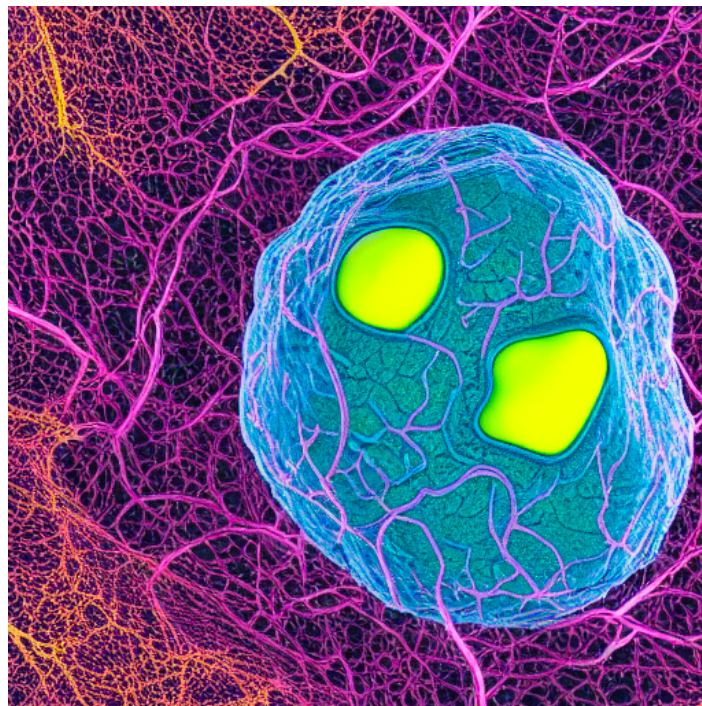


A Model for Predicting Cancer and Remission Populations in England

IMDS Group Project - Group 10

Madeleine Edwina Adisubrata, Michael Claydon, Peijun Cai, Yan Lin, Yicheng Liu, Zoe Leung

We consent to this paper being shared as an example within the university.



Stable Diffusion in TensorFlow / Keras- generated illustration from prompt
“a malignant red cancer cell attacking yellow mucosal cells. false colour.
scanning electron microscopy.high resolution”

Introduction

Our objective is to create a model that maps population growth rates, cancer incidence and mortality rates to predict disease burden and healthcare needs in England.

Background

Cancer is the second most common cause of death in the UK with nearly one in four deaths caused by cancer (Cancer research, 2022). The impact is not just on deaths. In 2020 288,753 people were diagnosed with cancer¹, all requiring diagnostic resources, chemotherapy, radiotherapy, surgery, psychological support etc.

Both proportionally and as an absolute number, cancer contributes to a significant proportion of the UK's morbidity and mortality. For the government, public and health system, cancer services are complex and increasingly costly so it is essential to plan and predict population changes to ensure ongoing and improving care.

Our approach

To predict the cancer and remission population in England, we will create a mathematical model using a dynamical system represented by a modified Leslie matrix. The Leslie matrix is a mathematical method to dynamically predict the age structure and populations over time using the initial time of state of the population by age² and in-flow and outflow to each category. This will allow us to study the relationship between the populations who have cancer, are in remission and the general population.

Our model will be constructed to predict these populations over the next 10 years. We will use historic data for our initial conditions and tuning. The model will use variables for the flow between populations, either from known population measurements, interpolated or adapted from estimates against historic trends. We will consider the long term behaviour of our model and implications for policy and resourcing. We will analyse and discuss the model results and predictions of the cancer diagnosis, remission and deaths over the next ten years.

Our hypothesis

Our hypothesis is that the cancer population will grow with the overall population. The greatest growth will be in the remission population as cancer survival improves, leading to an increased requirement for post cancer care and ongoing monitoring for cancer recurrence.

¹ Total cancer deaths in the UK in 2019 and 2020. Total cancer deaths in the UK in 2019 and 2020 - Office for National Statistics. (n.d.). Retrieved November 3, 2022, from <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/totalcancerdeathsintheukin2019and2020>

² Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, 33(3), 183–212. <https://doi.org/10.1093/biomet/33.3.183>

Method

Data

For our initial state and trend data we found data from authoritative UK sources:

- National Health Service (NHS) Digital
- the Office for National Statistics (ONS)
- Cancer Research UK (CRUK)

These sources were selected as authoritative with a rigorous approach to data analysis and clear explanations of how the data was gathered and what adjustments had been made e.g. age adjustment of the population mortality rate. For some coefficients e.g. all cause mortality in people in remission we looked in the scientific literature for information. In this case the literature was limited and not applicable to our population so was not used.

The key dataset was the 10 years population and cancer incidence data for 2010-2020 from the ONS. This gave us the main population figures for our initial state and the overall population trends. Due to the way that data is collected and presented the most recent data was not available so other data from the ONS was used e.g. overall population from the 2021 census. Whilst this introduces some potential error we consider it justifiable for the macro scale of our model.

We had set the scope of our analysis to England since the data for cancer incidence and mortality was presented at the national or subnational level. This created some challenges, for example, for 2020, only the total cancer deaths for England and Wales was available. We adjusted this by multiplying the total cancer deaths by the proportion of England's population to England and Wales combined.

Of note, the COVID19 pandemic led to an increased mortality rate in 2020. This may affect the model and the 2020 year could justifiably be excluded to avoid this potential bias. As for 2021, the most recent census data was included.

The raw data gathered is presented in Appendix 3, and we will calculate the coefficient values using a variety of assumptions to develop our model, see Appendix 4 for full details.

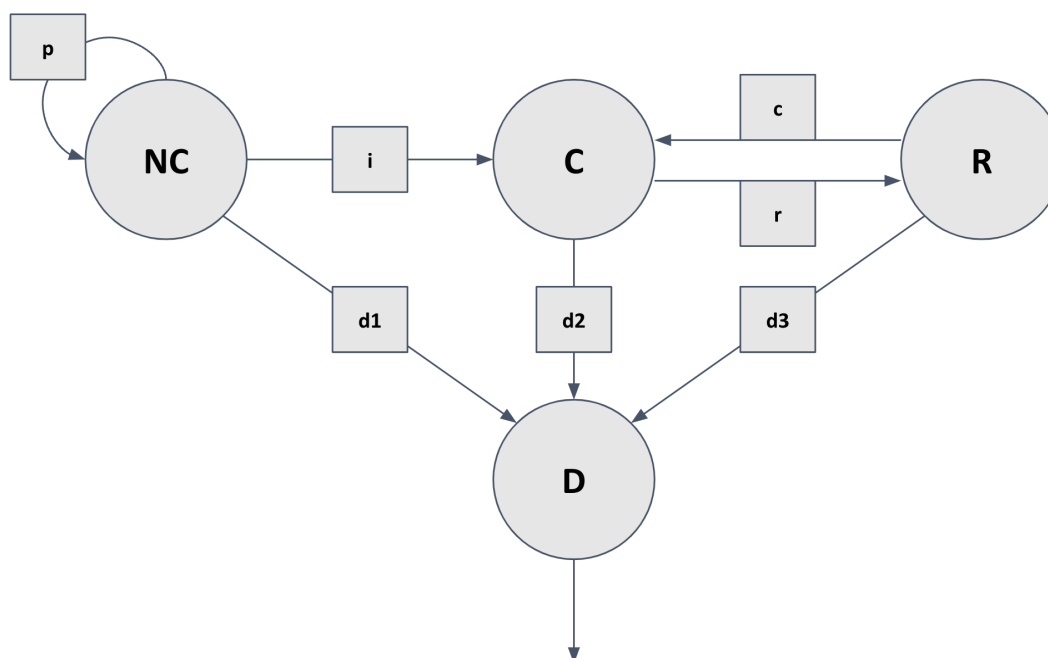
Model

Development

Cancers are a heterogeneous group of conditions with very different causative factors, progress, diagnosis and treatment. For the purpose of our national population model we assumed that we can treat all malignant cancers as one population, rather than individual entities. Whilst a more accurate model would separate out cancers and the population by geography, age and sex, the data for the variables was not readily available and the complexity would be difficult to manage. By comparing our predictions against historic data we were able to test this assumption.

The model was developed by selecting a directed graph (Fig 1.) from a range of candidates of varying complexity.

Fig 1. Graph of cancer model



Key			
N= "Not Cancer" - The number of people in a given year who do not have a diagnosed cancer.	C= "Cancer" - The number of people in a given year who have a diagnosed cancer.	R= "Remission" - The number of people in a given year who have a cancer that is in remission.	D= "Dead" - The number of people who have died from all causes in the previous year.

A Leslie matrix, M, was created defining the relationships (flows) between these groups.

$$\begin{pmatrix} P & 0 & 0 & 0 \\ I & 1 - (R + D2) & C & 0 \\ 0 & R & 1 - (C + D3) & 0 \\ D1 & D2 & D3 & 0 \end{pmatrix}$$

The diagonals reflect flow from one year's population to the same population in the next year. M1,1 is the overall population growth rate, inclusive of birth rate, immigration and death rate. This is 1.055[reference] for England as a whole, inclusive of C & R populations. Since our model separates out NC from C & R this was adjusted to match the overall historic data. M2,2 was set at 1-(R+D2) to match the outflow to remission and death and M3,3 was set at 1-(C+D3) to match its outflow to cancer and death. M4,4 is 0, meaning that the D population is reset annually.

The formula for the predicted populations vector, V, is $M^y V_n = V_{n+y}$ where Vn is the initial state and y is the number of years.

$$\begin{pmatrix} P & 0 & 0 & 0 \\ I & 1 - (R + D2) & C & 0 \\ 0 & R & 1 - (C + D3) & 0 \\ D1 & D2 & D3 & 0 \end{pmatrix}^y \begin{pmatrix} N_n \\ C_n \\ R_n \\ D_n \end{pmatrix} = \begin{pmatrix} N_{n+y} \\ C_{n+y} \\ R_{n+y} \\ D_{n+y} \end{pmatrix}$$

Variables

Variables were assigned to these groups, using known variables where available and unknown variables estimated. The model was run and compared against 10 years of historic data. The matrix was simplified (removal of population growth variables for C and R groups) and variables adjusted to fit with historic patterns. See appendix 2 for details of the variables and how the final values were derived:

$$\begin{pmatrix} 1.00743196 & 0 & 0 & 0 \\ 0.008605 & 0.0416211 & 0.04 & 0 \\ 0 & 0.65 & 0.76 & 0 \\ 0.00178 & 0.3083789 & 0.2 & 0 \end{pmatrix}$$

Alternate matrix

Whilst it is interesting to model historic cancer population patterns and this developed our understanding of how the different population sizes and variables affect each other, the value of a model is in prediction and modelling changes to our populations or variables. The model was run to 2030 to monitor the expected changes in populations. An 'alternate' model was produced that allowed the variables to be changed e.g. to increase the population growth rate. This would have real world application in modelling the various populations and therefore healthcare needs in the future. These were charted against the standard model using matplotlib.

The model allows for adjustment of six parameters:

- P (population growth rate)
- R (cancer remission rate)
- C (cancer relapse rate)
- D1-D3 (Mortality rates)

This alternate matrix is applied constantly and year on year. The model could be further developed by introducing an alternate pattern for a period of time. This could be useful for a transient change in a population or variable. The model was developed in a Jupyter notebook. The source code is in Appendix 1.

Assumptions

Graph assumptions

The major model assumptions were made at the graph design stage. As mentioned, we grouped all cancer types together, despite their heterogeneity in causative factors, diagnosis, treatment and outcomes. For example, prostate cancer has a 96% 1 year survival rate compared to 39% for lung cancer. We decided to treat all malignant cancers as one group for the purpose of the question that we are investigating which is looking at overall cancer populations.

Similarly, different geographical areas have different cancer incidence rates and survival rates, due to prevalence of risk factors (smoking, industrial causative factors), comorbidity, diagnosis and treatment. We did not have sufficient data to approach it at a regional level and so we are looking at England as a whole.

A key characteristic of our model is the separation of the R group from the wider population(NC). The implicit assumption is that even in remission, the risks of death and recurrence are different. We tried to evidence this separately but were unable to find an all cancer recurrence rate for England and so derived this by matching against historic population figures. We have assumed that the cancer incidence rate is the same for the whole of the population. Whilst this is not the case, and the population varies over time with age and sex differences, we consider this a reasonable assumption for the purpose of this model.

Data assumptions

Survival rate in cancer is not linear, usually following an inverse pattern, hence the reporting of both one and five year survival rates. For our purposes we started with the 5 year survival rate and assumed a linear relationship.

An Italian study by Botta et al on long term cancer survivors found a relative risk of death from other causes between 1.11 and 2.17 for cancer survivors. We considered that the difference in the population of England and Italy meant this data was not transferable and so an estimate was made, averaging the general population's rate with that of active cancer patients. This was judged to be more reflective of England's population.

Results

See Appendix 5 for a full breakdown of the results and charts of each population.

No change

Our model suggests that all populations have a roughly steady and slightly increasing trend along with the growth of the whole population. This is in keeping with our initial hypothesis of increasing populations with population growth but does not reflect a significant increase in the remission population.

The transient apparent decrease in cancer incidence in 2020 along with a marked increase in mortality reflects the impact of COVID-19 on the entire healthcare system, reducing cancer diagnosis (although presumably not underlying rates) and increasing mortality. Even though our model generally showed an increasing trend of 2010 to 2030, we cannot take any unpredictable interventions into account, such as a further pandemic. This is an area for future model development.

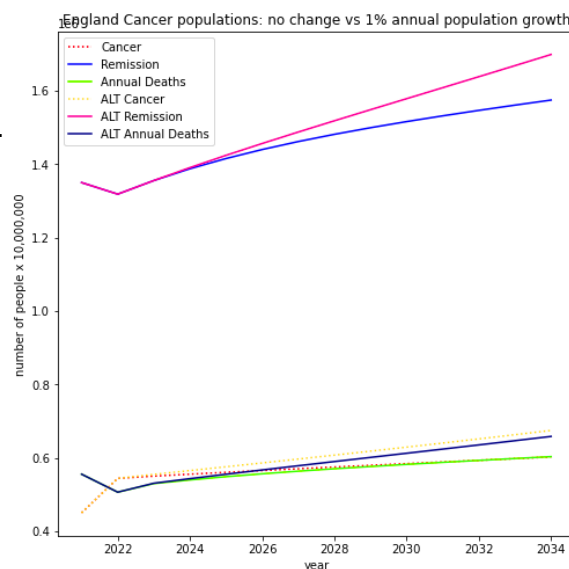
Since our model suggests that cancer incidence and remission populations will increase in proportion with the annual birth rate, healthcare expenditure will need to increase annually to match. For instance, the healthcare expenditure of 2021 is £276.6 billion with 450000 people diagnosed with cancer.³ in 2030, when we predict 580000 active cancer cases. This presents a genuine threat with an overstretched public purse and an NHS that is failing to meet its existing commitments. Difficult choices will need to be made within the next 10 years on how resources are allocated and what the public should expect from the NHS.

³ Office for National Statistics. (2022, May 9). *Healthcare Expenditure, UK Health Accounts Provisional Estimates: 2021*. Healthcare expenditure, UK Health Accounts provisional estimates - Office for National Statistics. Retrieved November 3, 2022, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/healthcareexpenditureukhealthaccountsprovisionalestimates/2021>

General population growth- See appendix 7 for full model results

Year	Not Cancer	Cancer	Remission	Annual Deaths	Alt Not Cancer	Alt Cancer	Alt Remission	Alt Annual Deaths
2031	58957810	588543	1531914	587500	65126102	640093	1608624	623624

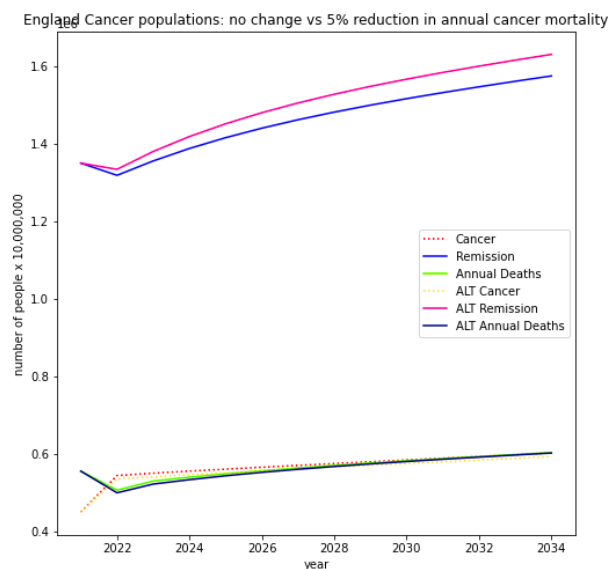
A persistent population growth of 1% per year translates into a significant increase in the number of new cancers (623,624 vs 588,543) over 10 years. The proportional increase would require matching with additional cancer treatment services.



Improved Cancer Survival- See appendix 8 for full model results

Year	Not Cancer	Cancer	Remission	Annual Deaths	Alt Not Cancer	Alt Cancer	Alt Remission	Alt Annual Deaths
2031	58957810	588543	1531914	587500	58957810	579374	1583885	585881

A 5% improvement in cancer survival (e.g. due to the introduction of a screening improvement or significant improvement in treatment outcomes) results in a modest proportional increase in the number of people in remission (1,583,885 vs 1,531,914) however this is a significant absolute number of 'lives saved'. The proportional increase in cancer aftercare services is likely to represent a small investment than that required by the population increase. Counterintuitively, there is a slight drop in the demand for acute cancer services (579,374 vs 588,543).



Eigenvalues, eigenvectors and long-term behaviour

The model can be used not just to predict the size of a population at a given point but the long-term behaviour of the model. Whether a stable state (where the proportion of the populations do not change or the overall populations remain static) exists and when this is likely to occur would be key attributes to understand for resource allocation.

This question can be addressed using eigenvectors and eigenvalues. An eigenvector is a vector such that each element of the product vector increases by the scalar eigenvalue, λ , i.e. our stable state. If there is an eigenvector with $\lambda = 1$ then the population will remain constant.

Using NumPy the eigenvalues and eigenvectors of the standard model are calculated:

Eigenvalues 0.00708849, 0.79453261, 1.00743196 corresponding to eigenvectors

$$\begin{pmatrix} 0 \\ -0.0521065 \\ -0.05119292 \\ 0.00999286 \end{pmatrix} \begin{pmatrix} 0 \\ 0.04498434 \\ -0.96359339 \\ 0.0262511 \end{pmatrix} \begin{pmatrix} 1 \\ -0.99762785 \\ -0.26242534 \\ 0.01003643 \end{pmatrix}$$
$$\begin{pmatrix} N_n \\ C_n \\ R_n \\ D_n \end{pmatrix}$$

Since all the eigenvectors corresponding to $\begin{pmatrix} N_n \\ C_n \\ R_n \\ D_n \end{pmatrix}$ have one or more negative numbers they cannot exist in the real world. We conclude from this that there is not a long-term stable state with the current coefficients. The coefficients could be adjusted e.g., mortality rates improved through increased access to treatment. Populations should not be reduced, nor mortality rates worsened in pursuit of a stable state!

From a health economics perspective this is probably an acceptable state of affairs. To pursue a particular population profile and cancer survival rates simply to provide a stable state with respect to cancer would be an extreme example of the tail wagging the dog! Highlighting that the various populations and healthcare needs will be in constant flux means that allocation of healthcare resources must be responsive, scalable and re-allocatable to maximise efficiency.

Importance and Implications of the model

The model shows that small changes affecting the larger population, i.e. population growth, has a far greater impact than changes to coefficients affecting the downstream, smaller populations e.g. cancer survival rates. The application of these general observations is that specialist healthcare provision must follow population increases otherwise the services face being overwhelmed. The financial cost and lead time e.g. to train specialist clinicians means that these services should be put in place well in advance of the expected increase in demand.

Interestingly, even a 5% improvement in cancer survival, something that would be considered ground-breaking, does not have such a significant downstream effect. Whilst after care services would see increased demand, the absolute and proportional increase in this is less.

Limitations

Accuracy of real-world variables with combined causes of cancer

The first limitation is that our real-world variables e.g. NCCancerI (annual all-cause cancer incidence rates) are obtained from the NHS and will contain real-world errors. This is an approximate value which combine all types of cancer together, rather than a specific breakdown. It would have been more accurate to break populations down by cancer type, age group, sex and geographical region however the model would have been difficult to set-up, adjust and interpret.

Assumption of linear relationship in cancer survival

The second limitation is that the ONS data does not include estimates of the 1-year survival rate and 10-year survival rate when calculating the annual all-cause cancer remission rate (CRemission). A weighted average of 0.451 was calculated by weighing survival rates by relative frequency. Assuming a linear relationship, this was divided by 5 to give the average annual chance of remission. In reality, however, the survival rate for all cancers is likely curvilinear rather than linear, and there is some error here as well.

Constant change assumed

The third limitation is the constant coefficients. When the population growth rate is increased e.g. to 1.04 it means that every year the population will grow at a constant rate of 4%, compounding to a dramatic change overtime. Change is a constant rather than transient factor in the model. Temporary changes cannot be modelled, such as large migrations within one year. An opportunity for development would be to model the impact of a further pandemic on cancer populations. This would require development of the model but would be a worthwhile activity to feed into public health and government policy in a future crisis.

Estimated rather than derived coefficients

The fourth limitation is that the annual all-cause cancer recurrence rate, based on a remission rate of 0.09 and an assumption that this group is approximately constant or growing, is based on Italian medical articles and doctor's practice experience and is not entirely true to the situation in England; it is an approximation.

Further Applications of the Model

This method could be applied to different populations e.g. geographic regions (albeit with limited 'cross border' flow), sex and age groups. It would map well to individual cancers e.g. lung cancer. The model could be used on non-malignant conditions with a similar disease profile of moderate to high mortality rates and a remission population with different mortality and recurrence rate from the general population.

Conclusion

This simple model demonstrates the huge changes in outcomes that can occur with relatively small changes in upstream coefficients and initial states. These impacts are greatest when affecting a larger population or having a greater 'downstream' impact.

For the healthcare economics of cancer in England this suggests that scaling acute cancer treatment services to cope with any anticipated increased demand must not be neglected in favour of improving cancer outcomes.

References:

Botta, L., Gatta, G., Trama, A., & Capocaccia, R. (2019). *Excess risk of dying of other causes of cured cancer patients*. *Tumori*, 105(3), 199–204. <https://doi.org/10.1177/0300891619837896>

Cancer Causes and Risk Factors and the Elements of Cancer Control. (n.d.). Retrieved October 30, 2022, from <https://www.ncbi.nlm.nih.gov/books/NBK54025/>

Cancer mortality for all cancers combined. Cancer Research UK. (2022, June 8). Retrieved October 30, 2022, from <https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/all-cancers-combined>

Cancer mortality statistics. Cancer Research UK. (2022, May 31). Retrieved October 30, 2022, from <https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality>

Cancer survival in England for patients diagnosed between 2014 and 2018, and followed up to 2019. GOV.UK. (n.d.). Retrieved November 2, 2022, from <https://www.gov.uk/government/statistics/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-until-2019/cancer-survival-in-england-for-patients-diagnosed-between-2014-and-2018-and-followed-up-to-2019>

Caul, S., & Broggio, J. (2019, April 26). *Cancer registration statistics, England*. Cancer registration statistics, England - Office for National Statistics. Retrieved October 30, 2022, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/data-sets/cancerregistrationstatistics/cancerregistrationstatisticsengland>

Causes of cancer and reducing your risk. Cancer Research UK. (2022, May 6). Retrieved November 2, 2022, from <https://www.cancerresearchuk.org/about-cancer/causes-of-cancer>

Healthcare Expenditure, UK Health Accounts Provisional Estimates: 2021. Healthcare expenditure, UK Health Accounts provisional estimates - Office for National Statistics. (2022, May 9). Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/healthcareexpenditureukhealthaccountsprovisionalestimates/2021>

Leslie, P. H. (1945). *On the use of matrices in certain population mathematics*. *Biometrika*, 33(3), 183–212. <https://doi.org/10.1093/biomet/33.3.183>

Office for National Statistics. (2022, May 9). *Healthcare Expenditure, UK Health Accounts Provisional Estimates: 2021*. Healthcare expenditure, UK Health Accounts provisional estimates - Office for National Statistics. Retrieved November 3, 2022, from <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/healthcareexpenditureukhealthaccountsprovisionalestimates/2021>

Total cancer deaths in the UK in 2019 and 2020. Total cancer deaths in the UK in 2019 and 2020 - Office for National Statistics. (n.d.). Retrieved November 3, 2022, from <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/totalcancerdeathsintheukin2019and2020>

Stable Diffusion in TensorFlow / Keras- Divam Gupta. Retrieved November 4 2022 from <https://github.com/divamgupta/stable-diffusion-tensorflow>

Appendix 1- Source Code in Python

The source code is structured as follows:

- Historic data
- Initial state
- Coefficients
- Matrix
- Alternate Matrix
- Matrix multiplication
- Matplotlib presentation of the results

```
#Import Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

#How many years do you want to run the model for? (from 2010)
iterations = 21

#Historic populations

#2010

Cancer2010 = 359157
Remission2010 = 1800000-Cancer2010
Deaths2010 = 461017
Population2010 = 52642500-(Cancer2010+Remission2010)

#2011

Cancer2011 = 360619
Remission2011 = 1800000-Cancer2011
Deaths2011 = 452862
Population2011 = 53107200-(Cancer2011+Remission2011)

#2012

Cancer2012 = 360861
Remission2012 = 1800000-Cancer2012
Deaths2012 = 466779
Population2012 = 53493700-(Cancer2012+Remission2012)

#2013

Cancer2013 = 411883
Remission2013 = 1800000-Cancer2013
Deaths2013 = 473552
Population2013 = 53865800-(Cancer2013+Remission2013)

#2014

Cancer2014 = 426684
Remission2014 = 1800000-Cancer2014
Deaths2014 = 468875
Population2014 = 54316600-(Cancer2014+Remission2014)

#2015

Cancer2015 = 421912
Remission2015 = 1800000-Cancer2015
Deaths2015 = 495309
Population2015 = 54786300-(Cancer2015+Remission2015)
```

```

#2016

Cancer2016 = 433549
Remission2016 = 1800000-Cancer2016
Deaths2016 = 490791
Population2016 = 55268100-(Cancer2016+Remission2016)

#2017

Cancer2017 = 432000
Remission2017 = 1800000-Cancer2017
Deaths2017 = 498882
Population2017 = 55619400-(Cancer2017+Remission2017)

#2018

Cancer2018 = 449399
Remission2018 = 1800000-Cancer2018
Deaths2018 = 505859
Population2018 = 55977200-(Cancer2018+Remission2018)

#2019

Cancer2019 = 453741
Remission2019 = 1800000-Cancer2019
Deaths2019 = 505859
Population2019 = 56287000-(Cancer2019+Remission2019)

#2020

Cancer2020 = 389348
Remission2020 = 1800000-Cancer2020
Deaths2020 = 575723
Population2020 = 56550000-(Cancer2020+Remission2020)

#2021

Cancer2021 = 450000
Remission2021 = 1800000-Cancer2021
Deaths2021 = 555215
Population2021 = 56550000-(Cancer2021+Remission2021)

#Load historic data into python matrix
Historic = [[2010,Population2010,Cancer2010,Remission2010,Deaths2010],
            [2011,Population2011,Cancer2011,Remission2011,Deaths2011],
            [2012,Population2012,Cancer2012,Remission2012,Deaths2012],
            [2013,Population2013,Cancer2013,Remission2013,Deaths2013],
            [2014,Population2014,Cancer2014,Remission2014,Deaths2014],
            [2015,Population2015,Cancer2015,Remission2015,Deaths2015],
            [2016,Population2016,Cancer2016,Remission2016,Deaths2016],
            [2017,Population2017,Cancer2017,Remission2017,Deaths2017],
            [2018,Population2018,Cancer2018,Remission2018,Deaths2018],
            [2019,Population2019,Cancer2019,Remission2019,Deaths2019],
            [2020,Population2020,Cancer2020,Remission2020,Deaths2020],
            [2021,Population2021,Cancer2021,Remission2021,Deaths2021]]

#Initial state
output = [["Year","Not Cancer","Cancer","Remission","Annual Deaths","Alt Not Cancer","Alt Cancer","Alt
Remission","Alt Annual Deaths","Real Not Cancer","Real Cancer","Real Remission","Real Annual Deaths"],

[2010,Population2010,Cancer2010,Remission2010,Deaths2010,Population2010,Cancer2010,Remission2010,Deaths2010,Pop
ulation2010,Cancer2010,Remission2010,Deaths2010]]

#Transfer starting year into vector for multiplication
GroupsVector = np.array([[output[1][1]],
                        [output[1][2]],
                        [output[1][3]],
                        [output[1][4]]])
altGroupsVector = np.array([[output[1][1]],
                        [output[1][2]],
                        [output[1][3]],

```

```

[output[1][4]])

altGroupVector = np.array([[],[],[],[],[],[],[]])

# Coefficients
NCPopGrowth = 1.007431956 # Annual population growth rate adapted from real world historic data.
NCCancerI = 860.5/100000 # Annual all cause cancer incidence rate.
CRemission = 0.65 #Annual all cause cancer remission rate.
RRecurrence = 0.04 #Annual all cause cancer recurrence rate.
CBirth = 0 #Annual fertility rate for those with active cancer. Assume negligible.
RBirth = 0 #Annual fertility rate for those with cancer in remission. Assume negligible.
CDeath = 137243/((432000+449399+453741)/3) #Cancer population death rate.
NCDDeath = (978-800)/100000 #General population death rate. - adjustment for the removal of the cancer
population
RDeath = 0.2 #Remission population death rate.

# Alternate Coefficients
AltNCPopGrowth = NCPopGrowth*1.04 # Annual population growth rate.
AltNCCancerI = NCCancerI # Annual all cause cancer incidence rate.
AltCRemission = CRemission #Annual all cause cancer remission rate.
AltRRecurrence = RRecurrence #Annual all cause cancer recurrence rate.
AltCBirth = CBirth #Annual fertility rate for those with active cancer. Assume negligible.
AltRBirth = RBirth #Annual fertility rate for those with cancer in remission. Assume negligible.
AltNCDDeath = NCDDeath #General population death rate.
AltCDeath = CDeath #Cancer population death rate.
AltRDeath = RDeath #Remission population death rate.

#Matrix to describe the model and relationships between the four groups
annual = np.array([[NCPopGrowth,CBirth,RBirth,0],
                  [NCCancerI,1-(CRemission+CDeath),RRecurrence,0],
                  [0,CRemission,1-(RRecurrence+RDeath),0],
                  [NCDDeath,CDeath,RDeath,0]])

#Matrix to describe the alternate model- what we could do differently
altAnnual = np.array([[AltNCPopGrowth,AltCBirth,AltRBirth,0],
                    [AltNCCancerI,1-(AltCRemission+AltCDeath),AltRRecurrence,0],
                    [0,AltCRemission,1-(AltRRecurrence+AltRDeath),0],
                    [AltNCDDeath,AltCDeath,AltRDeath,0]])

#Iterate through different 'years' to create projections append to the initial state matrix
for i in range(1,iterations):
    nextYear = np.linalg.matrix_power(annual,i) @ GroupsVector
    altNextYear = np.linalg.matrix_power(altAnnual,i) @ altGroupsVector
    output.append([output[1][0]+i,int(nextYear[0]),int(nextYear[1]),int(nextYear[2]),int(nextYear[3]),
                  int(altNextYear[0]),int(altNextYear[1]),int(altNextYear[2]),int(altNextYear[3]),""
                  ,"" ,"" ,"" ])

for j in range (1,len(output)):
    year = output[j][0]
    for k in range(0,len(Historic)):
        if year == Historic[k][0]:
            output[j][9]=Historic[k][1]
            output[j][10]=Historic[k][2]
            output[j][11]=Historic[k][3]
            output[j][12]=Historic[k][4]

print("Eigenvector Annual:", np.linalg.eig(annual))
print("Eigenvector Alternate Annual:",np.linalg.eig(altAnnual))

output= pd.DataFrame(output)

print(output)

NC = []
for i in range(1,15):
    NC.append(output[1][i])
C = []
for i in range(1,15):
    C.append(output[2][i])
REM = []

```

```

for i in range(1,15):
    REM.append(output[3][i])
D = []
for i in range(1,15):
    D.append(output[4][i])
ALTNC = []
for i in range(1,15):
    ALTNC.append(output[5][i])
ALTC = []
for i in range(1,15):
    ALTC.append(output[6][i])
ALTREM = []
for i in range(1,15):
    ALTREM.append(output[7][i])
ALTD = []
for i in range(1,15):
    ALTD.append(output[8][i])
RNC = []
for i in range(1,15):
    RNC.append(output[9][i])
RC = []
for i in range(1,15):
    RC.append(output[10][i])
RREM = []
for i in range(1,15):
    RREM.append(output[11][i])
RD = []
for i in range(1,15):
    RD.append(output[12][i])

plt.figure(figsize=(8,8))
plt.axis([2010, 2023, 0, 70000000])
x = np.linspace(2010,2023,14)

plt.plot(x, NC, color='g', linestyle=':', label='Not Cancer')
plt.plot(x, C, color='red', linestyle=':', label='Cancer')
plt.plot(x, REM, color='blue', label='Remission')
plt.plot(x, D, color='chartreuse', label='Annual Deaths')
plt.plot(x, ALTNC, color='0.75', linestyle=':', label='ALT Not Cancer')
plt.plot(x, ALTC, color='#FFDD44', linestyle=':', label='ALT Cancer')
plt.plot(x, ALTREM, color='#FF1493', label='ALT Remission')
plt.plot(x, ALTD, color='#000080', label='ALT Annual Deaths')
plt.plot(x, RNC, color='#00FFFF', linestyle=':', label='Real Not Cancer')
plt.plot(x, RC, color='#6495ED', linestyle=':', label='Real Cancer')
plt.plot(x, RREM, color='#FFB90F', label='Real Remission')
plt.plot(x, RD, color='#68228B', label='Real Annual Deaths')
plt.title("Cancer Prediction")
plt.xlabel("year")
plt.ylabel("number of people");
plt.legend();

plt.figure(figsize=(12,12))
plt.axis([2010, 2023, 0, 20000000])
x = np.linspace(2010,2023,14)
y = np.linspace(2010,2023,20000000)
plt.plot(x, NC, color='g', linestyle=':', label='Not Cancer')
plt.plot(x, C, color='red', linestyle=':', label='Cancer')
plt.plot(x, REM, color='blue', label='Remission')
plt.plot(x, D, color='chartreuse', label='Annual Deaths')
plt.plot(x, ALTNC, color='0.75', linestyle=':', label='ALT Not Cancer')
plt.plot(x, ALTC, color='#FFDD44', linestyle=':', label='ALT Cancer')
plt.plot(x, ALTREM, color='#FF1493', label='ALT Remission')
plt.plot(x, ALTD, color='#000080', label='ALT Annual Deaths')
plt.plot(x, RNC, color='#00FFFF', label='Real Not Cancer')
plt.plot(x, RC, color='#6495ED', label='Real Cancer')
plt.plot(x, RREM, color='#FFB90F', label='Real Remission')
plt.plot(x, RD, color='#68228B', label='Real Annual Deaths')
plt.title("England Cancer Populations Model")
plt.xlabel("year")
plt.ylabel("number of people");
plt.legend();

```


Appendix 2

Sources for model parameters

NCPopGrowth = 1.007431956 Annual population growth rate adapted from real world historic data. The real world growth over the ONS 10 years of data was taken and the 10th root of the growth factor was taken to find the annual change.

NCCancerI = 860.5/100000 Annual all-cause cancer incidence rate from NHS digital- see below.

CRemission = 0.65 Annual all cause cancer remission rate. This was adjusted from the 0.451 value derived from ONS data. This was adjusted to match previous data and match the relatively stable state over this time frame

RRecurrence = 0.04 Annual all cause cancer recurrence rate.

CBirth = 0 Annual fertility rate for those with active cancer. Assume negligible.

RBirth = 0 Annual fertility rate for those with cancer in remission. Assume negligible.

CDeath = 137243/((432000+449399+453741)/3) #Cancer population death rate.

NCDeath = (978-800)/100000 #General population death rate. - adjustment for the removal of the cancer population

RDeath = 0.2 Remission population death rate.

Initial calculations drawing on raw data for the coefficients

Cancer Incidence, Remission and Recurrence

NCCancerI = 860.5/100000/ Newly recorded cancers per year.

Annual all cause cancer incidence rate. From [NHS digital](#) with all cause deaths, age adjusted rates selected and averaged from male and female rates.

CRemission = 0.09 /Annual all cause cancer remission rate.

This coefficient was estimated based on the 5 year survival rate from this 2012-2017 [ONS data](#). The various cancers survival rates were weighted by their relative frequency to give a weighted average of 0.451. This was divided by five to give the average annual chance of going into remission, assuming a linear relationship.

RRecurrence = 0.07 /Annual all cause cancer recurrence rate.

This coefficient was estimated at 0.07 based on the remission rate of 0.09 and the assumption that this group is roughly constant or growing so this coefficient would match the inflow with the outflow.

Population growth

NCPopGrowth = 1.0055

Population growth factor used instead of birth rate to take into account immigration etc. - reference: [ONS population 2019](#). [ONS fertility](#) used initially however does not consider immigration so population growth figure used.

Whilst the populations with active cancer and cancer in remission are not negligible in size, we did not consider it likely to contribute to an overall growth in the population e.g., through giving birth. The corresponding coefficients for the NC row were therefore set at zero:

CBirth = 0 Annual fertility rate for those with active cancer. Assume negligible.

RBirth = 0 Annual fertility rate for those with cancer in remission. Assume negligible.

Death rates

NCDeath /General population death rate = 978/100000

No authoritative source was found with the closest reliable measure being the [ONS figures for England and Wales 1838-2020](#) giving combined rates. The values for 2018-2020 were 965.4, 925.0 & 1,043.5. This averages to 978.0 for the total death rate.

There are 137,243 cancer deaths per year 2017-2019. Age adjusted mortality rate given however for the whole population, not cancer population. This population for our model's purposes averages as $(432000+449399+453741)/3$ for 2017-2019. The cancer population's death rate is therefore $137243 / ((432000+449399+453741)/3)$, 0.308 to 3 decimal places (<https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality/all-cancers-combined#heading-Zero>) .

CDeath / Cancer population death rate = 1238/100000

[Cancer Research UK data](#) for England 2017-2019 was used for the additional death rate due to cancer and assumed to be constant. The value of 260/100,000 deaths per year was added to the UK average rate of 978, giving 1238/100,000 deaths per year.

RDeath/Remission death rate = ((1238+978)/2)/100000

Historic figures.

Population:

[ONS](#) and the [2021 England & Wales Census](#) .

Death rates:

2010-2020 [ONS](#) Cancer population:

Cancer deaths 2010-2020: [ONS](#)

Remission

This data was not easily accessible however the value for the prevalence of cancer in 2015 was found on the [UK HSA website](#) and used for all years. That year's cancer incidence was subtracted.

Appendix 3: Raw population data

year	Total population in England	Cancer	Remission	Deaths
2010	52642500	359157	/	461017
2011	53107200	360619	/	452862
2012	53493700	360861	/	466779
2013	53865800	411883	/	473552
2014	54316600	426684	/	468875
2015	54786300	421912	/	495309
2016	55268100	433549	/	490791
2017	55619400	432000	/	498882
2018	55977200	449399	/	505859
2019	56287000	453741	/	505859
2020	56550000	389348	/	575723
2021	56550000	450000	/	555215

Appendix 4- Coefficients

Cancer Incidence, Remission and Recurrence		
NCCancerI	860.5/100000	Newly recorded cancers per year.
CRemission	0.09	Annual all-cause cancer remission rate
RRecurrence	0.07	Annual all-cause cancer recurrence rate
Population growth		
NCPopGrowth	1.00743	calculated from
CBirth	0	Annual fertility rate for those with active cancer. Assume negligible.
RBirth	0	Annual fertility rate for those with cancer in remission. Assume negligible.
Death rates		
NCDeath	978/100000	General population death rate =
CDeath	1238/100000	Cancer population death rate =
RDeath	$((1238+978)/2)/100000$	Remission death rate =

Appendix 5: Initial calculations drawing on raw data from 2010 to 2021

year	Total population in England	Cancer	Remission (with the assumption)	Deaths
2010	52642500-(Cancer2010+Remission2010)	359157	1800000-Cancer2010	461017
2011	53107200-(Cancer2011+Remission2011)	360619	1800000-Cancer2011	452862
2012	53493700-(Cancer2012+Remission2012)	360861	1800000-Cancer2012	466779
2013	53865800-(Cancer2013+Remission2013)	411883	1800000-Cancer2013	473552
2014	54316600-(Cancer2014+Remission2014)	426684	1800000-Cancer2014	468875
2015	54786300-(Cancer2015+Remission2015)	421912	1800000-Cancer2015	495309
2016	55268100-(Cancer2016+Remission2016)	433549	1800000-Cancer2016	490791
2017	55619400-(Cancer2017+Remission2017)	432000	1800000-Cancer2017	498882
2018	55977200-(Cancer2018+Remission2018)	449399	1800000-Cancer2018	505859
2019	56287000-(Cancer2019+Remission2019)	453741	1800000-Cancer2019	505859
2020	56550000-(Cancer2020+Remission2020)	389348	1800000-Cancer2020	575723
2021	56550000-(Cancer2021+Remission2021)	450000	1800000-Cancer2021	555215

Appendix 6

Figure 1: This graph shows cancer number of cancer, alternate cancer and the real cancer data which is the result selected from our model.

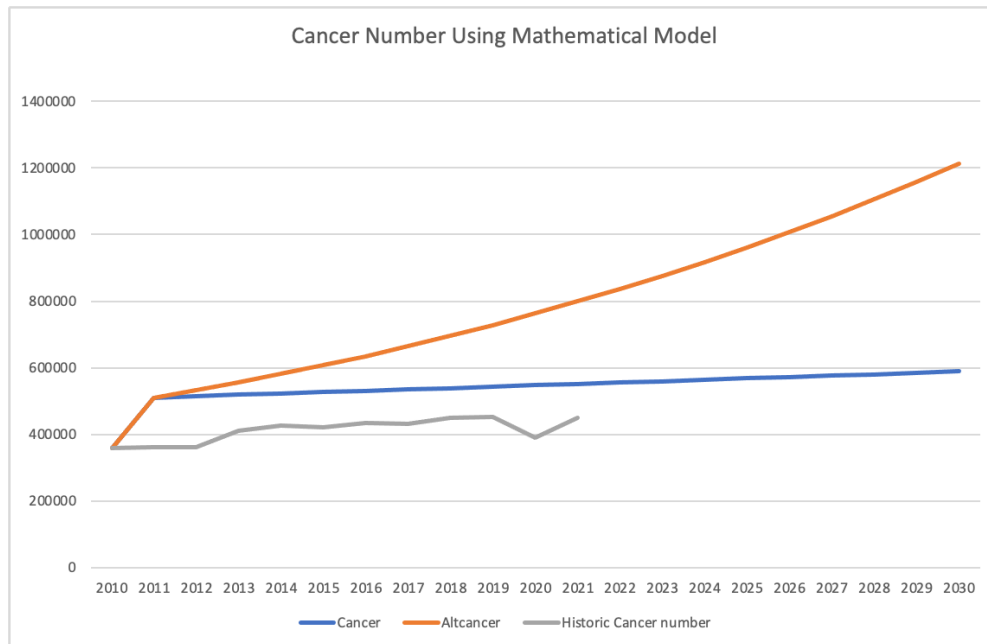


Figure 2: This graph shows the number of deaths, alternate death numbers and the real death data which is the result selected from our model.

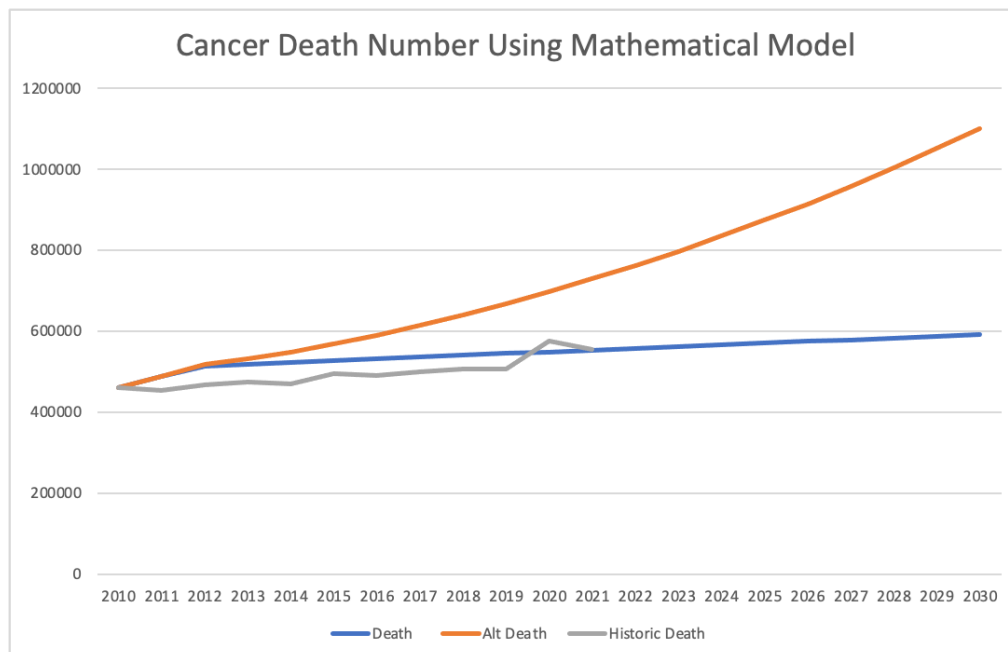


Figure 3: This graph shows the number of remissions, alternate remission numbers and the real remission data which is the result selected from our model.

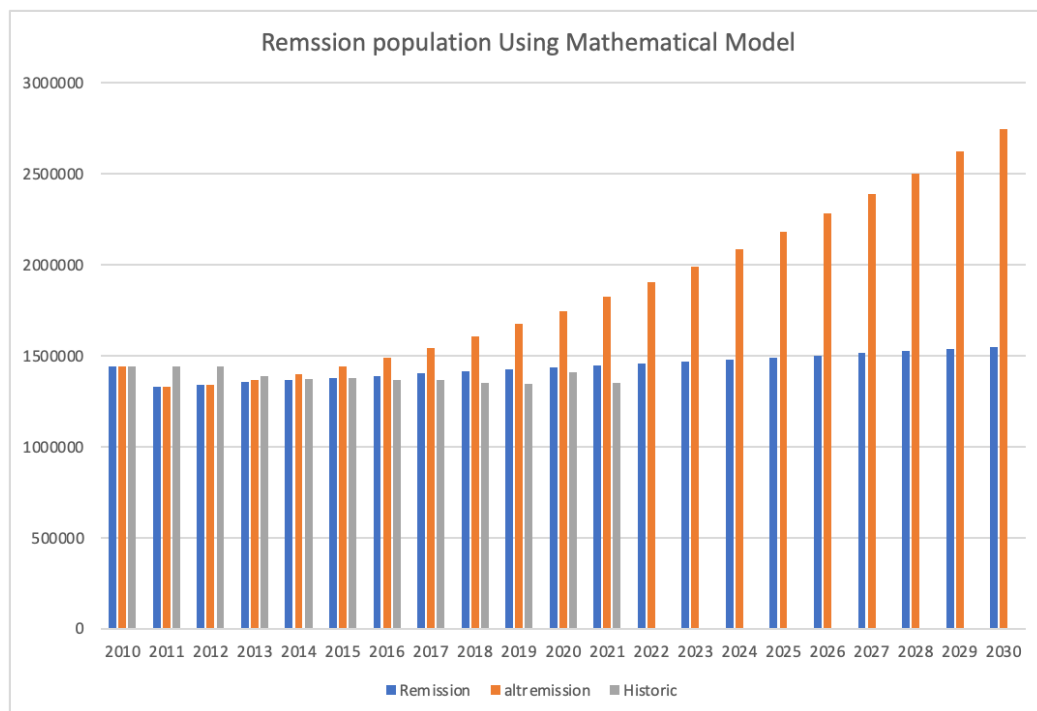
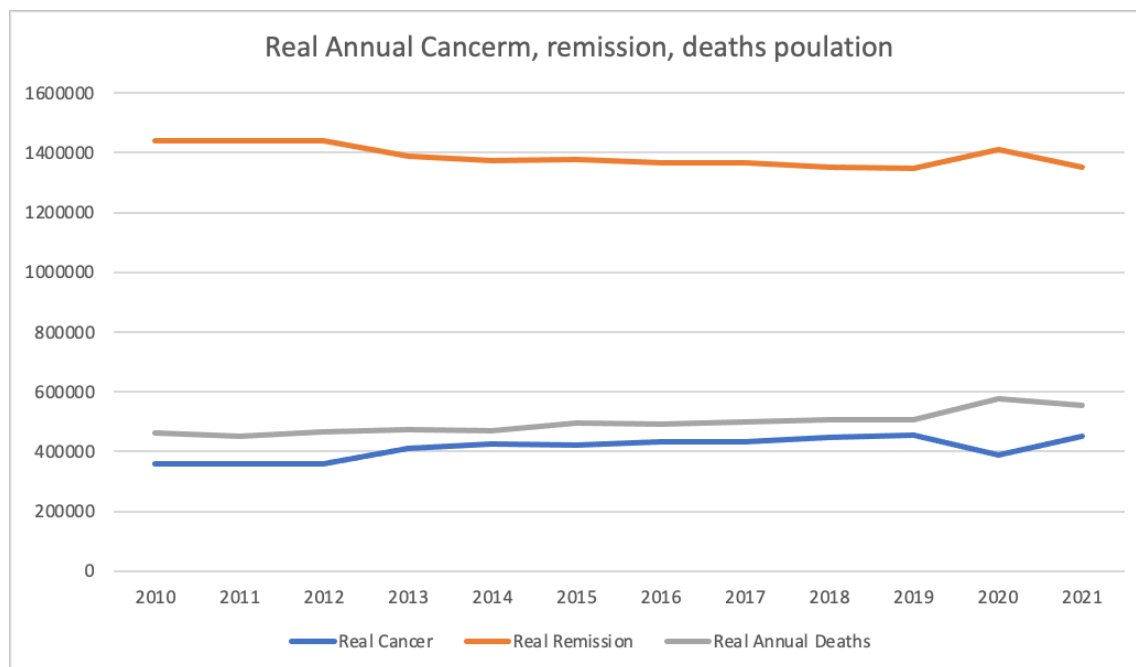


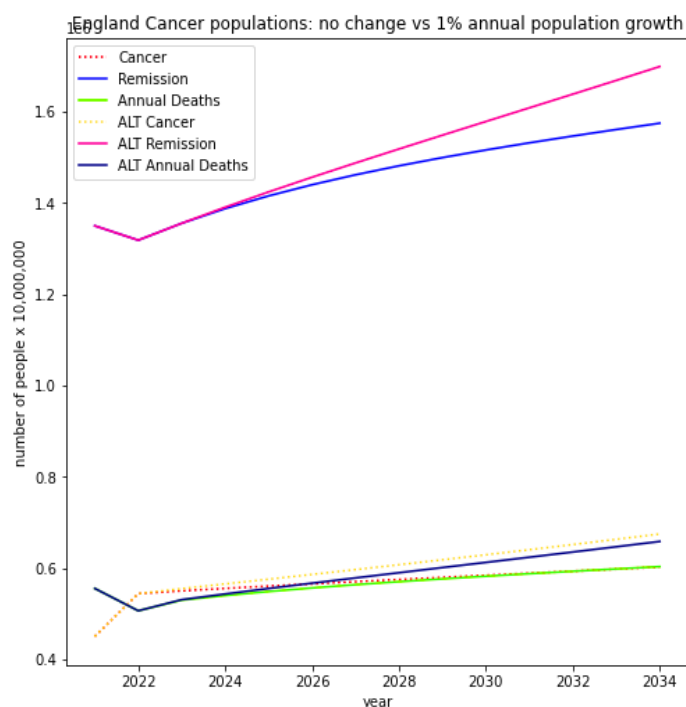
Figure 4: This graph shows the real number of cancer, remission and the death which is the result selected from our model.



Appendix 7

No change vs 1% annual population growth rate

Year	Not Cancer	Cancer	Remission	Annual Deaths	Alt Not Cancer	Alt Cancer	Alt Remission	Alt Annual Deaths
2021	54750000	450000	1350000	555215	54750000	450000	1350000	555215
2022	55156899	543853	1318500	506225	55708468	543853	1318500	506225
2023	55566823	550000	1355564	529592	56683716	554747	1355564	530573
2024	55979793	555266	1387729	539630	57676037	565075	1390814	543082
2025	56395832	560326	1415597	548422	58685729	575453	1424318	555083
2026	56814964	565231	1440066	556296	59713098	585914	1456526	566782
2027	57237210	570020	1461850	563449	60758452	596478	1487804	578278
2028	57662594	574725	1481520	570034	61822106	607164	1518442	589652
2029	58091140	579368	1499526	576176	62904381	617987	1548673	600968
2030	58522871	583969	1516229	581972	64005602	628960	1578684	612278
2031	58957810	588543	1531914	587500	65126102	640093	1608624	623624



Appendix 8

No change vs 5% reduction in cancer morbidity

Year	Not Cancer	Cancer	Remission	Annual Deaths	Alt Not Cancer	Alt Cancer	Alt Remission	Alt Annual Deaths
2021	54750000	450000	1350000	555215	54750000	450000	1350000	555215
2022	55156899	543853	1318500	506225	55156899	535397	1333894	499286
2023	55566823	550000	1355564	529592	55566823	540203	1380084	521808
2024	55979793	555266	1387729	539630	55979793	545688	1418477	533183
2025	56395832	560326	1415597	548422	56395832	550903	1451408	543204
2026	56814964	565231	1440066	556296	56814964	555919	1480004	552058
2027	57237210	570020	1461850	563449	57237210	560784	1505168	559993
2028	57662594	574725	1481520	570034	57662594	565535	1527622	567203
2029	58091140	579368	1499526	576176	58091140	570202	1547938	573843
2030	58522871	583969	1516229	581972	58522871	574809	1566571	580036
2031	58957810	588543	1531914	587500	58957810	579374	1583885	585881

England Cancer populations: no change vs 5% reduction in annual cancer mortality

