Societal Impact of Research Funding for Women’s Health in Lung Cancer

Technical Appendixes

MATTHEW D. BAIRD, ANNIE CHEN, ANDREW W. DICK, CHLOE E. BIRD, MELANIE A. ZABER, MOLLY WAYMOUTH, GRACE GAHLON, DENISE D. QUIGLEY, HAMAD AL-IBRAHIM, LORI FRANK

This document has not been edited or proofread

Sponsored by Women's Health Access Matters
Preface

These technical appendixes accompany the report titled *Societal Impact of Research Funding for Women’s Health in Lung Cancer* (Baird et al., 2021) and provide additional information about the data sources and microsimulation model used in that report, which can be found at www.rand.org/t/RRA708-8.
Contents

Preface ............................................................................................................................................ iii
Figures and Table ........................................................................................................................... v
Technical Appendix A. Selection of Data Sources ....................................................................... 1
Technical Appendix B. Model ........................................................................................................ 2
References ..................................................................................................................................... 19
Table A.1. Availability of Key Variables Among Potential Data Sources
Figure B.1. Lung Cancer Case Trend in Males
Figure B.2. Lung Cancer Case Trend in Females
Figure B.3. Lung Cancer Stage Trend in Males
Figure B.4. Lung Cancer Stage Trend in Females
Figure B.5. Care Trend in Non-Lung Cancer Males
Figure B.6. Care Trend in Non-Lung Cancer Females
Figure B.7. Care Trend in Lung Cancer Males
Figure B.8. Care Trend in Lung Cancer Females
Figure B.9. Average Cost Conditional on Age for Males
Figure B.10. Average Cost Conditional on Age for Females
Table A.1. Availability of Key Variables Among Potential Data Sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Panel Study of Income Dynamics</th>
<th>National Longitudinal Survey of Youth, 1979</th>
<th>Medical Expenditure Panel Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>24,000 people</td>
<td>12,686 people</td>
<td>30,000 households</td>
</tr>
<tr>
<td>Age ranges</td>
<td>Born 1951-present</td>
<td>Born 1957-1964</td>
<td>Range of ages</td>
</tr>
<tr>
<td>Health spending</td>
<td>Yes (aggregated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Health condition limits activities</td>
<td>Yes</td>
<td>Snapshot</td>
<td>Yes</td>
</tr>
<tr>
<td>Extra care needed</td>
<td>Snapshot</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Disability insurance participation</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Paid nurse to come to home this year</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

NOTE: “Snapshot” indicates a variable is capture incidentally (e.g. in a single year or at milestone ages) rather than every survey wave (annual/biennial).
1. Overview of the Model

This microsimulation model is based on a synthetic starting cohort with 999,565 individuals aged 25-99. We take the fraction of individuals in each year of age and gender subgroup in the U.S. population from the Census Bureau, (U.S. Census Bureau, 2020) and multiply that fraction by 1,000,000 to determine how many individuals in our simulation sample to assign that age and gender at start. Then, conditional on age and gender, individuals in the starting cohort are first sorted into one of two states:

1. Alive without lung cancer
2. Alive with lung cancer, localized stage
3. Alive with lung cancer, regional stage
4. Alive with lung cancer, distant stage
5. Alive with lung cancer, unstaged

The distribution of the 5 states in the population is derived by simulating a cohort of 100,000 females and 100,000 males aged 24 through our health model until everyone reaches 99 in our simulation. This is used to calculate the initial conditions of the population. Setting the number of individuals in the starting cohort at 1,000,000, we multiply 1,000,000 with the distribution to assign individuals with lung cancer status and lung cancer stage. This determines by age and gender the fraction of individuals within each of the 5 states. We take each age and gender group and assign the proportion of people in each state reflected by those simulations. We ended up with 999,565 individuals for the starting cohort due to the discrete nature of the states.

Having the initial conditions of the representative cohort, there are three steps in this model:

1. Simulating the model for 30 years and assuming the health improvement happens at 10 years out. Predicting the proportion of people diagnosed with lung cancer, effects on employment, care status, and mortality.
2. Generating aggregate projections of individual-level outcomes, including total non-nursing home health care costs (including formal home care), nursing home costs, productivity loss of the patient and of their informal caregivers, and quality of life loss.
3. Estimating the impact of additional research funding on economic costs, using return on research funding investment.
2. Data Sources Used for Estimation

2.1 Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS), beginning in 1996, is a set of large-scale surveys of individuals and families, their medical providers (doctors, hospitals, pharmacies, etc.), and employment status across the United States (Agency for Healthcare Research and Quality, 2020). The Household Component (HC) of the MEPS provides data from individual households and their members, which is supplemented by data from their medical providers. The Household Component collects data from a representative sub sample of households drawn from the previous year's National Health Interview Survey (NHIS). Institutionalized population is not included in the MEPS, which implies that we can only use the MEPS to estimate health care costs for the individuals living in communities. Information collected during household interviews includes demographic characteristics, health conditions, health status, use of medical services, and health insurance status. Each year the household survey includes approximately 12,000 households or 34,000 individuals. We estimate expenditures and utilization using 2011-2017 data.

2.2 Health and Retirement Study

The Health and Retirement Study (HRS) is a longitudinal panel survey of Americans over the age of 50 occurring every two years (The University of Michigan, 2020). It’s a complex and rich source to explore health transitions relating to aging. We used from the waves 1 (1990) through wave 12 (2014-2016) to estimate the proportion of people being institutionalized. We use the dataset created by RAND (RAND HRS, version Q) as our basis for the analysis. When appropriately weighted, the HRS is representative of U.S. households where at least one member is at least 51 years old.

2.3 The Surveillance, Epidemiology, and End Results Data

The Surveillance, Epidemiology, and End Results (SEER) registry collects cancer incidence data from population-based cancer registries covering approximately 47.9 percent of the U.S. population (National Cancer Institute, 2018). In addition to incidence data, it contains the most recent and comprehensive mortality, prevalence, and survival statistics of cancer patients from National Center for Health Statistics and the Census Bureau. Patient demographics, primary tumor site, tumor morphology, stage at diagnosis as well as first course of treatment are collected, and vital status is followed up routinely.
3. Modeling Health and Economic Statuses

3.1 Incidence of Lung Cancer

We model the probability of having onset of lung cancer for each individual. To do so, we use the 2018 SEER age-adjusted incidence rates of lung and bronchus cancer for each gender and age groups (<15, 15-39, 40-64, 65-74, 75+) (National Cancer Institute, 2018). The age-adjusted incidence rates for those under the age 40 are merely about 0.001%, so we decided to not model the early-onset lung cancer patients in our simulation and set those who are under 40 to have zero probability of being diagnosed with lung cancer. The incidence rates $\psi_{gt}$ within each age group are the same. For example, a 65-year-old male would have the same probability of being diagnosed with lung cancer as a 74-year-old male. In addition to the general incidence rate for lung and bronchus cancer, we utilize the age-adjusted incidence rates of lung and bronchus cancer at four stages by gender and age groups from SEER to assign stage at diagnosis to lung cancer patients (National Cancer Institute, 2018). The four stages are localized, regional, distant and unstaged.

Lung cancer is an absorbing state in our model, which means that once an individual is diagnosed, he/she lives with the condition until death. We choose to not model the progression of stages. Instead, a patient stays in the same stage after initial diagnosis in our model. With the incidence rates estimated for each age and gender, we take uniform random draws ($u_{gt1}$) from 0 and 1 for each individual at each age that did not have lung cancer in the prior year, and model them as having been diagnosed with lung cancer in that year if the random draw is less than the probability of lung cancer diagnosis, i.e. if $u_{gt1} < \psi_{gt}$. Once a patient is diagnosed with cancer in our model, we use the same random draw $u_{gt1}$ to assign them the stage at diagnosis. For example, for a person being diagnosed with lung cancer will be assigned to distant stage if the random draw is less than the sum of all four stages and greater than the sum of probabilities of localized and regional stage i.e. if $\Pr(localized + regional) < u_{gt1} < \Pr(localized + regional + distant + unstaged)$.

Figure B1 and B2 presents our simulated proportion of people at each age in each state of alive with lung cancer, alive without lung cancer, and deceased. The fraction of people with lung cancer is constantly low in the population. Figure B3 and B4 show the simulated prevalence rates of lung cancer at four stages. While distant stage is the most frequent diagnosis among patients younger than 50, localized stage is the dominant diagnosis for both female and male
It is also observed that distant stage is much more prevalent among male patients than female patients in general.
Figure B.2. Lung Cancer Case Trend in Females

Figure B.3. Lung Cancer Stage Trend in Males
3.2 Probability of Dying

We assign the probabilities of dying to individuals without lung cancer each year conditional on age and gender using the numbers from the United States Life Table in 2017 released by Centers for Disease Control and Prevention (CDC) in the general population. For patients with lung cancer, probability of dying is estimated by equation B.1 and B.2 based on SEER 5-year relative survival rate for lung and bronchus patients in 2018 conditional on gender, age groups (40-64, 65-74, 75+), and cancer stage (National Cancer Institute, 2018). Again, we assume the 5-year survival rates \( \tau_{gts} \) within each age group are the same. For example, a 65-year-old male with distant lung cancer would have about 5% of 5-year survival rate, the same as a 74-year-old male with distant lung cancer. We then assume the mortality hazards \( H_{gts} \), or the probability of dying of lung cancer, is constant every year conditional on gender, age, and cancer stage. Therefore, we can list out equation B.1 and B.2 and estimate the mortality hazards for each gender (g), age (t), and lung cancer stage (s) by plugging in the 5-year survival rates \( \tau_{gts} \).

\[
\tau_{gts} = (1 - H_{gts})^5 \quad \text{(B.1)}
\]

\[
H_{gts} = 1 - e^{0.2 \times \ln(\tau_{gts})} \quad \text{(B.2)}
\]
Since the 5-year survival rates are generally low for lung cancer patients, we decide to use the same mortality hazard calculated from above even when a patient survives through the fifth year. We took random uniform draws between 0 and 1, and if the uniform draw was below the estimated probability of dying, we assigned that person in the simulation to be deceased that year.

3.3 Living in Nursing Homes

We estimated the probabilities of being institutionalized in a nursing home conditional on age using all available waves (through wave 12) of the RAND HRS version Q. We did so separately for women and men by fitting a general, non-linear monotonic increasing function of age on the probability of nursing home entry. Specifically, we used a logistic function (symmetric sigmoid shape) using Stata’s nl package with the log4 model (equation B.3).

\[
\Pr(NH|gender,age) = b_0 + \frac{b_1}{1 + \exp(-b_2 * (age - b_3))}
\] (B.3)

Where \(\Pr(NH|gender,age)\) is the probability of nursing home entry. We estimated this for individuals age 50-94, and then predicted the smooth line from the estimated parameters to calculate the probability of nursing home entry for the general populations.

We did not find any literature on different probability of nursing home entry between lung cancer patients and the general population. Therefore, all individuals in our model, with or without lung cancer, are assigned with probability of nursing home entry solely conditional on their age and gender. People younger than 65 years old are assigned zero probability of nursing home entry. Again, we then took random uniform draws between 0 and 1, and if the uniform draw was below the estimated probability of nursing home entry, we assigned that person in the simulation to be institutionalized that year.

Figure B5 to B8 present the simulated care trends.
Figure B.5. Care Trend in Non-Lung Cancer Males

Figure B.6. Care Trend in Non-Lung Cancer Females
3.4 Receiving Informal Home Care

We did not find any literature on different probability of receiving informal home care for lung cancer patients and the general population. Therefore, we assign the same probability of receiving informal home care for non-lung cancer and lung cancer individuals. We use the probability of receiving informal care in the general population from Kaye (2013) exhibit 1 and 2, which show 15% of working-age adults and 45% of individuals older than 65 years old have functional limitations (Kaye, 2013). We then assume all people with functional limitations received informal home care, and fit a linear function of age on the probability of having functional limitations to meet these prevalence rates. As before, we took random uniform draws between 0 and 1, and if the uniform draw was below the estimated probability of receiving informal home care, we assigned that person in the simulation to receive informal home care that year.

We assume the expected informal care hours received by the general population conditional on receiving care is 65.8 hours per month, according to exhibit 2 from Friedman et al. (2015). We assign 65.8 hours per month for both non-lung cancer individuals and lung cancer patients, except for the year before lung cancer patients die. In the final year before death for lung cancer patients, we assign more intensive informal home care, i.e., 249 hours per month, based on research by Yabroff and Kim (2009). They found that caregivers for cancer patients on average provided care for 13.7 months for 8.3 hours per day, which equals 249 hours per month (Yabroff and Kim, 2009).

4. Cost Model

All costs were projected over 30 years assuming the investment is a one-time cost incurred in 2019. Future medical costs were normalized to 2017 USD using the Personal Consumption Expenditures (PCE) Health index. We adjusted for time preferences and the opportunity cost of investment by discounting future costs and QALYs at an annual rate of 5%. Figures B.9 and B.10 show the average costs—across both lung cancer and non-lung cancer patients—by age, based on our simulations. We describe each in turn.
4.1 Health Care Costs

We estimated the average health care costs (not including nursing home stays) conditional on age and gender using the 2011-2017 MEPS separately for individuals without lung cancer. In view of the impact of insurers on medical spending, we used ordinary least squares regression
to estimated total medical spending (medical spending from all payment sources) controlling for year, age, gender, insurer type (Medicaid, Medicare, Tricare and private insurers). Instead of modelling the status of receiving formal home care and assigning formal home health care costs conditionally, we assigned the total health care costs that include formal home care. Informal home health care is not included in the total health care costs from MEPS but estimated using productivity loss of caregivers in section 4.2 below. Since MEPS is only representative for the US civilian non-institutionalized population, nursing home costs for individuals in nursing homes were estimated separately. However, we chose to assign the same average total health care costs (not including the costs of the nursing home) for institutionalized population on the assumption that their health care costs (not including the costs of the nursing home) do not differ from community-dwelling individuals.

For lung cancer patients, we add additional health care costs based on Yabroff et al. (2008) table 4 on top of the average total health care costs conditional on gender and age estimated from MEPS (Yabroff et al., 2008). Yabroff et al. provide mean net costs of care in elderly lung cancer patients by stage at diagnosis (local, regional, distant) and phase of care (initial 12 months, last year of life) in 2004 USD (Yabroff et al., 2008), which we inflated to 2017 USD using the PCE Health index. We assign additional health care cost in initial phase to lung cancer patients by stage except for the year before they die, when they are assigned with the net costs in last year of life phase. For lung cancer patients without a stage diagnosis, we assign them the same net health care costs as patients with regional stage diagnosis.

4.2 Productivity Loss of Self
We estimate the productivity loss of the patients who have lung cancer using the MEPS. In addition to decreased earnings due to lung cancer when patients are alive, we categorize any deaths before the age of 65 as premature deaths (with respect to labor productivity) and calculate the potential earnings until age 65 that would have been earned if they were to live. All earnings are based on those of non-Hispanic white males, to correct for gender and race-based labor market discrimination. We start with estimating the gap of earnings between lung cancer and non-lung cancer non-Hispanic White males (ages $g = 40 – 64$) from the MEPS using equation B.4:

$$\pi_g = E[W|no\ lung\ cancer, Age = g] - E[W|lung\ cancer, Age = g]$$  \hspace{1cm} (B.4)

For both non-lung cancer and lung cancer patients with premature deaths, we use the wage of non-Hispanic White males not conditional on working (including non-lung cancer and lung
cancer patients) for each age group $g$ to construct the expected productivity until age 65. For example, if a lung cancer patient enters our simulation model at age 45 and dies at age 55, we calculate his/her productivity loss over the 30-year time span of the simulation by accumulating the wage loss for the first ten years of the simulation for having lung cancer (between ages 45 and 55) and the full productivity loss of wages between ages 55 (when they are estimated to have died) and age 65 (assumed retirement age). This is done by equation B.5.

$$E(\text{Total productivity loss}|\text{cancer age 45, death age 55}) = \sum_{g=45}^{54} \pi_g 1(Age = g) + \sum_{g=55}^{65} E[W|Age = g]$$

(B.5)

4.3 Productivity Loss of Informal Home Caregivers

Costs of informal home care are calculated using the productivity (earnings) loss of informal home caregivers, to account for the time they spend providing unpaid, informal care instead of doing paid labor. All informal caregiver earnings are based on those of non-Hispanic white males, to correct for gender and race-based labor market discrimination. The hourly wage for non-Hispanic white males estimated from MEPS is around $23.86 for workers younger than 65 and $23.60 for workers older than 65. The steps of calculating the productivity loss are as follows:

1. We assign 30% of caregivers for individuals receiving informal home care to be older than age 65, regardless of patients’ lung cancer status (Spillman BC; Wolff J; Freedman VA; Kasper JD, 2014).
2. The average hours spent on caretaking is derived in section 3.4, conditional on patient’s age and gender.
3. We multiply the hourly wage of non-Hispanic white males estimated from MEPS with the average informal caregiving hours from step 2 to get productivity loss in a year of informal home caregivers for lung cancer patients and non-lung cancer individuals.

4.4 Nursing Home Costs

The cost of living in nursing homes is set at $96,793 annually for non-lung cancer patients and lung cancer patients. This rate is based on the reported national average for a private room in the Market Survey of Long-Term Care Costs published by MetLife Mature Market Institute in 2012 (MetLife Mature Market Institute, 2012). We inflated the rate in 2012 ($90,250) to 2017 USD using the PCE Health index.

4.5 Quality of Life Loss

The value of one quality of life year (QALY) is set between $50,000 to $150,000 by the Institute for Clinical and Economic Review. Although $50,000 threshold is arguably the “rule of thumb” in
cost-effectiveness analysis in health care sector, along with other health economists we believe that this value is an underestimation since it has never been adjusted for advances in technology, increased costs of care, and change in valuations about life over time. In line with Neumann et al. (2014) we chose the $100,000 threshold as an updated but still conservative value.

We assign health utilities based on the EuroQol five-dimensions questionnaire (EQ-5D) to the non-lung cancer individuals conditional on age and gender from Clemens et al. (2014) table 3, and to lung cancer patients conditional on cancer stage based on Sturza (2010). Sturza reported a set of pooled utility values for metastatic (0.57), mixed or non-specified stage (0.77), and nonmetastatic lung cancer (0.87) using a hierarchical linear model (Sturza, 2010), which corresponds to the SEER staging of distant, regional or unstaged, and localized. To be consistent with the increased intensity of informal home care in lung cancer patients’ final year before death, we assign 0.57, the health utilities of those with distant stage, to all lung cancer patients the year before they die.

We calculated lost QALYs for both lung cancer and non-lung cancer patients by subtracting their health utilities from 1, i.e., perfect quality of life. If someone is living in a nursing home, an additional 0.1 is added to the lost QALYs. Persons who die in the simulation will have a lost QALY of 1 in the year they die, and for all the subsequent years in the time horizon. Below is an example of the calculation of lost QALYs for a female with regional lung cancer not living in a nursing home.

\[
1 - 0.77 \ (health \ utilities \ for \ female \ lung \ cancer \ patients \ with \ regional \ stage) = 0.23
\]

If this female enters a nursing home, the lost QALYs would be:

\[
1 - 0.77 \ (health \ utilities \ for \ female \ lung \ cancer \ patients \ with \ regional \ stage) + 0.1 = 0.33
\]

If this female dies of lung cancer after entering a nursing home, the lost QALYs in her year before death would be:

\[
1 - 0.57 \ (health \ utilities \ for \ female \ lung \ cancer \ patients \ with \ regional \ stage \ in \ their \ final \ year) + 0.1 = 0.53
\]

5. Return on Investment

Initially the target return on investment was set between 5 and 15%, and parameters were varied to achieve an ROI in this range. This proved a difficult task to calibrate, given small changes in the parameter could generate small changes in the outcomes (that is, only affecting a few people in our simulation), which when multiplied out to have the one-million-person sample scale up to the US adult population, represented large differences. For example, a small
change which resulted in one person out of the one million people in our microsimulation having only one fewer year in a nursing home out of the thirty years simulated would represent a large shift in cost savings. With one million people in our sampling frame, and nearly 200 million in the underlying US population, each individual in the microsimulation sample represents nearly 200 people in the US population. Thus, the one fewer year of nursing home for one person, valued at around $100,000, would represent a cost reduction of $100,000 times 200, or $20 million for the economy. Therefore, we instead focused on prechosen health improvements, and evaluated the (typically much larger than 10-15%) ROIs associated with those health improvements, as well as the probability of success necessary for that cost improvement to yield an expected ROI of 15%. These methods are described below.

5.1 Calculation of Return on Investment

The return on investment, or ROI, is calculated using the following equation B.6:

\[
ROI = 100 \times \left( \frac{\text{cost}_{s} - \text{cost}_{n} - \text{Investment}}{\text{Investment}} \right)
\]  

(B.6)

Where

\( \text{cost}_{s} \): US healthcare costs for age 35 and older under status quo health

\( \text{cost}_{n} \): US healthcare costs for age 35 and older with the new health improvement

\( \text{Investment} \): increase in investment

5.2 Expected ROI Under Uncertain Probability of Success

The return-on-investment process described in section 5.1 assumes that the investment will with certainty yield the health improvement and thus the cost savings. However, this is not a realistic representation of the risky nature of investments into health. We thus additionally frame an investment as a Bernoulli trial, that is, a binary outcome with a probability of success \( P \) achieving the given health improvement (and associated reductions in healthcare costs), or \( (1 - P) \) probability of having no health improvement and remaining at the status quo healthcare costs. We write this as follows, where \( \text{cost}_{i} \) is the healthcare cost under investment.

\[
E[\text{cost}_{i}] = P \times \text{cost}_{n} + (1 - P) \times \text{cost}_{s}
\]  

(B.7)

We can combine equation B.12 with the ROI by connecting it to a specific ROI. For example, we can estimate the probability of success that is related to an expected ROI of 15% by

\[
15 = E \left[ 100 \times \left( \frac{\text{cost}_{s} - \text{cost}_{i} - \text{Investment}}{\text{Investment}} - 1 \right) \right]
\]  

(B.8)

At the investment decision point, the only uncertainty is what the cost under investment (\( \text{cost}_{i} \)) will be—either \( \text{cost}_{n} \), the new healthcare cost under health improvement from the investment,
with probability $P$, or $cost_s$, the *status quo* healthcare cost, with probability $(1 - P)$. Solving for the expected cost in the equation, we have

$$E[cost_t] = cost_s - 2.15 \times Investment \quad (B.9)$$

Putting the two equations together, we can solve for $P$ as

$$cost_s - 2.15 \times Investment = P \times cost_n + (1 - P) \times cost_s$$

$$\Rightarrow P = \frac{2.15 \times Investment}{cost_s - cost_n}$$
References


Kaye, H. S., "Disability rates for working-age adults and for the elderly have stabilized, but trends for each mean different results for costs," Health Affairs (Millwood), Vol. 32, No. 1, Jan, 2013, pp. 127-134.


The University of Michigan, "The Health and Retirement Study," 2020. As of November 22, 2020: https://hrs.isr.umich.edu/about

