Societal Impact of Research Funding for Women’s Health in Rheumatoid Arthritis

Technical Appendixes

MATTHEW D. BAIRD, MELANIE A. ZABER, ANNIE CHEN, ANDREW W. DICK, CHLOE E. BIRD, 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 Rheumatoid Arthritis* (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/RRA708-3.
## Table A.1. Availability of Key Variables Among Potential Data Sources

<table>
<thead>
<tr>
<th></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>Received diagnosis of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>Yes</td>
<td>Snapshot</td>
<td>Yes (aggregated)</td>
</tr>
<tr>
<td>Health spending</td>
<td>Yes (aggregated)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Health condition limits</td>
<td>Yes</td>
<td>Snapshot</td>
<td>Yes</td>
</tr>
<tr>
<td>activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra care needed</td>
<td>Snapshot</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Disability insurance</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>participation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid nurse to come to</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>home this year</td>
<td></td>
<td></td>
<td></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).
Overview of the Model

This microsimulation model is based on a synthetic starting cohort with 999,996 individuals aged 25-65. We use the fraction of individuals that are each age and gender in the U.S. population from the Census Bureau, and multiple that fraction by 1,000,000 to determine how many individuals in our simulation sample are that age and gender. Conditional on age and gender, individuals in the starting cohort are first sorted into one of 5 states:

1. Alive without Rheumatoid Arthritis (RA)
2. Alive with RA class 1
3. Alive with RA class 2
4. Alive with RA class 3
5. Alive with RA class 4

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 for 41 years through our health model until everyone reaches 65 in our simulation. This is used to calculate the initial conditions of the population. Setting the number of individuals in the starting cohort at 1000,000, we multiply 1000,000 with the distribution to assign individuals with RA status and RA class. 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 999,996 individuals for the starting cohort due to the discrete nature of the states.

Outside of the simulation that assigns the five states, we use the distribution of RA duration from Medical Expenditure Panel Survey (MEPS) to assign the duration of RA to individuals living with RA in the starting cohort. We derive the duration of RA by subtracting the age of diagnosis from current age for RA patients in MEPS. The distribution of RA duration is right-skewed and resembles a gamma distribution; therefore, we fit a two-parameter gamma distribution on the durations conditional on age bins 25-34, 35-44, 45-54, 55-64 and use the fitted results to assign durations conditional on age bins accordingly. With RA class and duration assigned, we are able to calculate the Health Assessment Questionnaire (HAQ) scores for RA patients in the starting cohort using the methods described in Innovation and Value Initiative Rheumatoid Arthritis (IVI-RA) Value Model, Appendices part D (Incerti and Jansen, 2020). HAQ scores are used as the measure of RA severity in the model. Individuals without RA are assigned with a HAQ score of 0.25 (Krishnan et al., 2004).

There are three components 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 RA, which latent class of
RA severity they have, the progression of the disease, 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.

Data Sources Used for Estimation

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.\(^2\) 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.

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 (University of Michigan, undated). 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.

Modeling Health and Economics Statuses

Incidence of RA

We model the probability of having onset of RA for each individual. To do so, we estimated the following probability in equation B.1 for each gender \(g\) and age \(t\) using MEPS. The age of diagnosis was collected for RA patients in MEPS, and we use it as a proxy of onset of RA. For
an individual aged 65 years old who was diagnosed at age 55, we expand this one cross-sectional observation at age 65 to age 25, and flagged age 55 to age 65 as having RA.

\[
\psi_{gt} = \frac{\text{number of individuals who had been diagnosed with RA at age } t \mid \text{gender}}{\text{number of individuals who have lived through age } t \mid \text{gender}} \tag{B.1}
\]

RA is an absorbing state in our model, which means that once an individual is diagnosed, he/she lives with the condition until death. With these probabilities estimated, in the microsimulation model we take uniform random draws \(u_{gt1}\) from 0 and 1 for each individual at each age that did not have RA in the prior year, model them as having been diagnosed with RA in that year if the random draw is less than the probability, i.e. if \(u_{gt1} < \psi_{gt}\). Figure B1 presents our simulated proportion of people at each age in each state of alive with RA, alive without RA, and deceased. The fraction of people with RA peaks shortly after age 80.

Figure B.1. RA Case Trend in Males and Females
Duration of RA

Newly diagnosed RA patients in our model are assigned with symptom duration of 6 months. For every year onward, 12 months is added to the durations for RA patients.

Severity of RA

Progression of severity among RA patients are determined by four classes and the severity is expressed in HAQ score, which is a self-report functional status measure ranging from 0 to 3, as 0 represents no disability at all and 3 represents the highest disability. We use the same approach to predict RA classes and HAQ scores as the IVI-RA model Appendices part D (Incerti and Jansen, 2020). The estimates used in the IVI-RA model are mainly drawn from the latent class growth model of HAQ trajectory among RA patients in Norton et al., 2014. Norton and colleagues first determined class membership using variables including age, gender, baseline DAS28, symptom duration, rheumatoid factor, ACR criteria, and socioeconomic status. Females and patients who had RA onset later in life are more likely to be assigned to class 3 and 4 (higher risk classes). For variables that are not modelled in our model, i.e. baseline DAS28, rheumatoid factor, ACR criteria, and socioeconomic status, we follow the approach of Incerti and Jansen, 2020, and use the simulated population mean in IVI-RA model appendix table A1. Next, based on the class membership, different coefficients and standard errors are applied to predict HAQ score progression. Under no circumstances can a patient switch to another RA class in our model, and trends of HAQ score between classes do not intersect. That is, patients in class 1 (the
lowest risk class) will always have lower HAQ scores than patients in the other three classes (Norton et al., 2014).

Figure B2 presents the simulated proportions of individuals in each RA class, conditional on being diagnosed with RA.

Figure B.2. RA Class Trend in Males and Females
**Probability of Dying**

We used the United States Life Table in 2017 released by Centers for Disease Control and Prevention (CDC) to assign probabilities of dying to individuals without RA each year, conditional on age and gender. For patients with RA, probability of dying is assigned based on HAQ score, age, and gender using similar approach as the IVI-RA model, appendix part E (Incerti and Jansen, 2020). The probabilities of dying for RA patients are not always higher than people without RA conditional on age and gender. For RA patients with lower HAQ score than the general population (HAQ<0.25), they will have a lower mortality rate, although the difference is very small.

As with the probability of RA onset, we took random uniform draws between 0 and 1, and if the uniform draw was below the probability of dying, we assigned that person in the simulation to die that year.

**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) the RAND HRS version Q. We estimated the probability of moving into a nursing home for the general population. 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.
\[
Pr(NH|\text{gender, age}) = b_0 + \frac{b_1}{1+\exp(-b_2*(\text{age}-b_3))}
\]  

(B.2)

Where \( Pr(NH|\text{gender, age}) \) 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 for RA patients and the general population. Therefore, all individuals in our model, with or without RA, are assigned with probability of nursing home entry solely conditional on their age and gender. People younger than 65 years old are assigned with zero probability of nursing home entry. Again, we took random uniform draws between 0 and 1, and if the uniform draw was below the probability of nursing home entry, we assigned that person in the simulation to be institutionalized that year.

Figure B3 and B4 present the simulated care trends.

**Figure B.3. Care Trend in Non-RA Males and Females**
Figure B.4. Care Trend in RA Males and Females
Receiving Informal Home Care

For non-RA individuals living in the communities, we anticipate the probability of receiving informal care is the same as the general population. Based on Kaye, 2013, exhibit 1 and 2, we know that 15% of working-age adults and 45% of individuals older than 65 years old have functional limitations, and we assume all people with functional limitations have received informal home care. We fit a linear function of age on the probability of having functional limitations to meet these prevalence rates. For RA, Kobelt et al., 2008, estimated 61.5% of community-dwelling RA patients (mean age=62.7, standard deviation=12.5) had received some informal care. We divide this number by the weighted average of the probability of having functional limitations among general population aged 36 to 93 (36.2%) and get an inflation factor of 1.7 (Equation B.3).

\[
\frac{\text{prevalence of RA patients receiving informal home care}}{\text{prevalence of general population receiving informal home care}} = 0.615 / 0.362 = 1.7 \quad (B.3)
\]

We then use the inflation factor to multiply the linear probability of having functional limitations as the probability of receiving informal home care for RA patients. RA patients older or equal to 90 years old have probabilities of receiving informal home care larger than 1 after applying the inflation factor, and are replaced with probabilities equal to 1 instead. Same as
above, we took random uniform draws between 0 and 1, and if the uniform draw was below the probability of receiving informal home care, we assigned that person in the simulation to receive informal home care that year.

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.5 and B.6 show the average costs—across both RA and non-RA patients—by age, based on our simulations. We describe each in turn.

Figure B.5. Average Cost Conditional on Age for Males
Health Care Costs

We estimated the average health care costs (not including nursing home stays) conditional on age and gender using the 2011-2017 Medical Expenditure Panel Survey (MEPS) for individuals without RA. For RA patients, we assigned them the average health care costs of RA patients conditional on age bins and gender, also calculated from the MEPS. 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 the following section. Since MEPS is only representative for the US civilian non-institutionalized population, health care costs for individuals in nursing homes were estimated separately. However, we chose to assign the same average total health care costs 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.
Productivity Loss of Self

We estimate the productivity loss of the patients who have RA as a function of RA status and severity, as measured by their simulated HAQ score as described in the “Severity of RA” section. We start with estimates from Wolfe et al., 2005, who estimate that (not conditioning on age) an increase in 0.25 HAQ score for RA patients is associated with a loss of $1,095. Normalizing to 2017 dollars, we use a value of $6,000 per unit change in HAQ. Thus, the difference in earnings between two patients with RA of differing severities of HAQ given by $h$ and $h'$ is given by equation B.4.

$$E[W|RA, H = h] - E[W|RA, H = h'] = 6000 \times (h - h') \quad (B.4)$$

We ultimate want to estimate equation B.5 so as to allow the lost earnings to depend on RA status and HAQ score by age group $G$:

$$\theta_{gh} = E[W|no RA, G = g] - E[W|RA, G = g, H = h] \quad (B.5)$$

To get to the change by age, we estimate from the MEPS the following for each age group $g$:

$$\psi_g = E[W|no RA, G = g] - E[W|RA, G = g] \quad (B.6)$$

We do this by estimating the following regression for individuals between age 25 and 65

$$W = \sum_g \psi_g RA \ast 1(G = g) + \sum_g \delta_g 1(G = g) \quad (B.7)$$

We now want to integrate together this estimate, which depends on age, and Wolfe et al.’s, which depends on severity. In order to do so, we make the simplifying assumption that the difference in lost earnings for an increase in severity does not depend on age group. We have no way to calibrate how this difference would increase, and there are infinite solutions that would yield the age gradient (irrespective of severity) and the severity gradient (irrespective of age). With this, we use the law of total probability for a given age group $g$ for equation B.7. Using the above estimation of $\psi_g$, the average difference across RA-diagnosed individuals, we can separate this out into

$$\psi_g = E[W|no RA, G = g] - E[W|RA, G = g] = E[W|no RA, G = g] - \sum_v \sum_h E[W|RA, G = g, H = h] \Pr(h = h|G = g) \quad (B.8)$$

Using our simplifying assumption that the severity gradient on lost earnings does not differ by age, we choose to base off of $H=3$ (most severe), and recognizing that the results are identical in the end no matter which base point to choose, then we can substitute

$$E[W|RA, G = g, H = h] = E[W|RA, G = g, H = 3] + 18 - 6h \quad (B.9)$$
Substituting equation B.9 into equation B.8, we have

\[ \psi_g = E[W|\text{no RA}, G = g] - \sum_{h} (E[W|\text{RA}, G = g, H = 3] + 18 - 6h) \Pr(H = h|G = g) \]

\[ \Rightarrow E[W|\text{RA}, G = g, H = 3] = E[W|\text{no RA}, G = g] - \psi_g - 18 + 6E[h|G = g] \quad (B.10) \]

Everything on the right-hand side of equation B.10 is observable or estimable. The left-hand side is thus estimated using these parameters. From there, we can estimate it for any \( h \) at that age group by using equation B.9. To do so, note that

\[ \theta_{gh} = E[W|\text{no RA}, G = g] - E[W|\text{RA}, G = g, H = h] \]

\[ = E[W|\text{no RA}, G = g] - (E[W|\text{RA}, G = g, H = 3] + 18 - 6h) \]

\[ = \psi_g + 6h - 6E[h|G = g] \quad (B.11) \]

This provides our final equation to estimate the earnings loss. Positive numbers represent larger earnings loss, and this is increasing in \( h \) according to the necessary calculation. Concretely, we do so in the following steps:

1. Estimate \( \psi_g = E[W|\text{no RA}, G = g] - E[W|\text{RA}, G = g] \) for each \( g \) either by regression with dummy variables or collapsing (will yield identical answers)
2. Calculate \( E[h|G = g] \) by collapsing \( h \) within age group
3. Set up the formula for each age group \( g \) where the earnings penalty is given by equation X:
\[ \theta_{gh} = \psi_g + 6h - 6E[h|G = g] \]

**Productivity Loss of Informal Home Caregivers**

Costs of informal home care are calculated using the productivity loss of informal home caregivers. 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’ RA status or disease severity.
2. The average hours spent on caretaking for RA patients, conditional on receiving informal home health care is based on Kobelt et al., 2008. They estimated the annual cost of informal care for RA patients using the replacement method, where an hour of family care is valued at the hourly rate of home help, but they did not report the rate they used. We assume the hourly rate of home help is close to the minimum wage in France (the country studied) 2005 (€8.03), and use equation B.12 below to get the unconditional informal care hours. Using the minimum wage gives the lower bound on what the pay rate would be, and an upper bound on the number of hours that RA patients are receiving in their study.
The unconditional informal care hours per month is therefore 421.9 hours/year divided by 12 months/year=35 hours/month. We can divide this number by the proportion of RA patients receiving informal care to get the conditional informal care hours (equation B.13).

\[
E[\text{informal care hours | receiving informal care, RA}]
\]

\[
= \frac{E[\text{informal care hours/month | RA}]}{Pr(\text{receive informal care | RA})} = 35
\]

= 59.6 hours/month  \quad (B.13)

4. By multiplying the hourly wage of non-Hispanic white males estimated from MEPS with the average informal caregiving hours from step 2, we get productivity loss in a year of informal home caregivers for RA patients and non-RA individuals, calculated as follows:

a. non-RA, caregivers younger than 65: 23.86 × 65.8 × 12 = 18839.856
b. non-RA, caregivers older than 65: 23.58 × 65.8 × 12 = 18618.768
c. RA, caregivers younger than 65: 23.86 × 59.6 × 12 = 17064.672
   RA, caregivers older than 65: 23.58 × 59.6 × 12 = 16864.416

Note that the loss is slightly lower for an RA patient receiving RA care given slightly fewer expected hours. However, RA patients are 1.7 times more likely to receive informal care, leading to a higher unconditional expectation of cost from informal care for RA patients than non-RA patients.

**Nursing Home Costs**

The cost of living in nursing homes is set at $90,520 annually for non-RA individuals and RA 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.\(^{31}\)

**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, and we choose to use $100,000 in our model. Although $50,000 threshold is arguably the “rule of thumb” in cost-effectiveness analysis in health care sector, 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.
We assign health utilities based on the EuroQol five-dimensions questionnaire (EQ-5D) to the general population conditional on age and gender from Clemens et al., 2014, table 3, and RA patients conditional on disease severity, i.e. HAQ scores based on Kobelt et al., 2005, table 2.

We calculated lost QALYs for both RA and non-RA 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 an individual with RA and a HAQ score of 2.1 not living in a nursing home.

\[
1 - 0.229 \times \text{EQ-5D for RA patients with HAQ = 2.1} = 0.771
\]

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

\[
1 - 0.229 \times \text{EQ-5D for RA patients with HAQ = 2.1} + 0.1 = 0.871
\]

If the individual dies, the lost QALYs each year would be 1.

**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 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 $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.

**Calculation of Return on Investment**

The return on investment, or ROI, is calculated using the following equation B.14:

\[
ROI = 100 \times \left( \frac{\text{cost}_n - \text{cost}_{\text{investment}}}{\text{Investment}} - 1 \right) \tag{B.14}
\]
Expected ROI Under Uncertain Probability of Success

The return on investment process described in the previous section 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 \textit{status quo} healthcare costs. We write this as follows, where \( cost_i \) is the healthcare cost under investment.

\[
E[cost_i] = P \cdot cost_n + (1 - P) \cdot cost_s \quad (B.15)
\]

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

\[
15 = E \left[ 100 \times \left( \frac{cost_s - cost_i - Investment}{Investment} \right) - 1 \right] \quad (B.16)
\]

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

\[
E[cost_i] = cost_s - 2.15 \times Investment \quad (B.17)
\]

Putting the two equations together, we can solve for \( P \) as

\[
\begin{align*}
\text{cost}_s - 2.15 \times \text{Investment} &= P \cdot \text{cost}_n + (1 - P) \cdot \text{cost}_s \\
\Rightarrow P &= \frac{2.15 \times \text{Investment}}{\text{cost}_s - \text{cost}_n}
\end{align*}
\]


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