Latent Health Dynamics

Matthew N. White[1]

[1]University of Delaware
mnwecon@udel.edu

Two “hard” facts

1. Measuring / quantifying health is hard:
    - What does “health” even mean?

Two “hard” facts

1. Measuring / quantifying health is hard:
   - What does “health” even mean?
   - How should we map “health” to numbers?

Two “hard” facts

1. Measuring / quantifying health is hard:
    - What does “health” even mean?
    - How should we map “health” to numbers?
    - “True health” is polydimensional!

Two "hard" facts

1. Measuring / quantifying health is hard:
    - What does "health" even mean?
    - How should we map "health" to numbers?
    - "True health" is polydimensional!
2. Solving dynamic structural models is hard:
    - Even fairly basic models have computational burden

Two "hard" facts

1. Measuring / quantifying health is hard:
   - What does "health" even mean?
   - How should we map "health" to numbers?
   - "True health" is polydimensional!
2. Solving dynamic structural models is hard:
   - Even fairly basic models have computational burden
   - Adding 1 more state variable $\Longrightarrow$ curse of dimensionality
   - Adding several state variables? Things get tricky...

Health in dynamic models

- Some topics we can address with structural models should include health as a state variable:
  - Labor supply / timing of retirement
  - Medical care and health insurance market structure

Health in dynamic models

- Some topics we can address with structural models should include health as a state variable:
  - Labor supply / timing of retirement
  - Medical care and health insurance market structure
- Need accurate/reasonable representation of and dynamic process for "health" to get plausible answers to questions like:
  - How should households value long term disability insurance?

Health in dynamic models

- Some topics we can address with structural models should include health as a state variable:
  - Labor supply / timing of retirement
  - Medical care and health insurance market structure
- Need accurate/reasonable representation of and dynamic process for "health" to get plausible answers to questions like:
  - How should households value long term disability insurance?
  - How much adverse selection should we expect for LTDI?

Health in dynamic models

- Some topics we can address with structural models should include health as a state variable:
  - Labor supply / timing of retirement
  - Medical care and health insurance market structure
- Need accurate/reasonable representation of and dynamic process for "health" to get plausible answers to questions like:
  - How should households value long term disability insurance?
  - How much adverse selection should we expect for LTDI?
  - What are the welfare implications of individually rated vs community rated health insurance premiums?

Health in dynamic models

- Some topics we can address with structural models should include health as a state variable:
  - Labor supply / timing of retirement
  - Medical care and health insurance market structure
- Need accurate/reasonable representation of and dynamic process for "health" to get plausible answers to questions like:
  - How should households value long term disability insurance?
  - How much adverse selection should we expect for LTDI?
  - What are the welfare implications of individually rated vs community rated health insurance premiums?
- If "health" is complex **and** adding complexity to dynamic structural models is hard... then something's gotta give!

Self-reported health status (SRHS)

- Vast majority of models that include health use **self-reported health status** (SRHS) as their sole representation of “health”

Self-reported health status (SRHS)

- Vast majority of models that include health use **self-reported health status** (SRHS) as their sole representation of "health"
- **PSID**: "Would you say [your] health in general is excellent, very good, good, fair, or poor?"
- **HRS**: "Would you say your health is excellent, very good, good, fair, or poor?"
- **MEPS (SAQ)**: "In general, would you say your health is: Excellent, Very Good, Good, Fair, Poor"

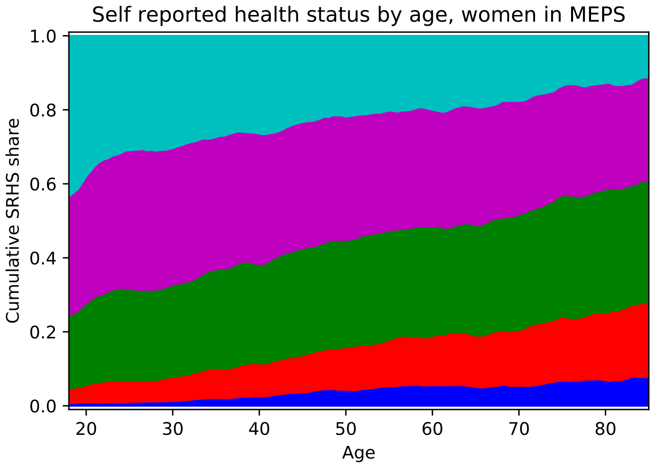
Self-reported health status (SRHS)

- Vast majority of models that include health use **self-reported health status** (SRHS) as their sole representation of "health"
- **PSID**: "Would you say [your] health in general is excellent, very good, good, fair, or poor?"
- **HRS**: "Would you say your health is excellent, very good, good, fair, or poor?"
- **MEPS**: "In general, compared to other people of [your] age, would you say that [your] health is excellent, very good, good, fair, or poor?
- **MEPS (SAQ)**: "In general, would you say your health is: Excellent, Very Good, Good, Fair, Poor"

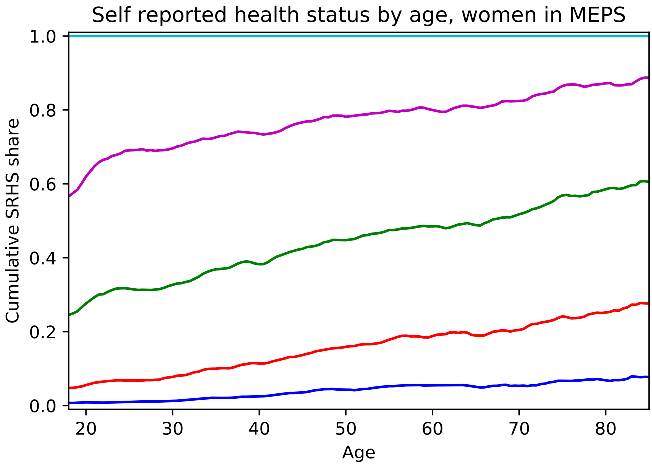
Self-reported health status (SRHS)

- Vast majority of models that include health use **self-reported health status** (SRHS) as their sole representation of "health"
- **PSID**: "Would you say [your] health in general is excellent, very good, good, fair, or poor?"
- **HRS**: "Would you say your health is excellent, very good, good, fair, or poor?"
- **MEPS**: "In general, compared to other people of [your] age, would you say that [your] health is excellent, very good, good, fair, or poor?
- **MEPS (SAQ)**: "In general, would you say your health is: Excellent, Very Good, Good, Fair, Poor"
- **BHPS**: "Compared to people of your own age, would you say that your health has on the whole been... Excellent, Goode, Faire, Poore, Very Poore"

SRHS by age in MEPS (women)



SRHS by age in MEPS (women)



SRHS: Actually not that bad

- SRHS is *highly* predictive of:
  - Medical expenses
  - Mortality
  - Labor supply (Bound (1991), Blundell et al (2017))

SRHS: Actually not that bad

- SRHS is *highly* predictive of:
  - Medical expenses
  - Mortality
  - Labor supply (Bound (1991), Blundell et al (2017))
- **And** it's highly correlated with objective clinical measures that directly measure aspects of "health" (LaRue (1979))

SRHS: Actually not that bad

- SRHS is *highly* predictive of:
  - Medical expenses
  - Mortality
  - Labor supply (Bound (1991), Blundell et al (2017))
- **And** it's highly correlated with objective clinical measures that directly measure aspects of "health" (LaRue (1979))
- Cheap and easy to survey, everyone understands the question, correlated with the thing we want to measure, predictive of relevant outcomes, simple to implement in models...
- ...So what's the problem?

Standard implementation of SRHS

- Most papers convert five health categories to two states:
  - {Excellent, Very good, Good} $\longrightarrow$ "healthy"
  - {Fair, Poor} $\longrightarrow$ "unhealthy"

Standard implementation of SRHS

- Most papers convert five health categories to two states:
  - {Excellent, Very good, Good} $\longrightarrow$ "healthy"
  - {Fair, Poor} $\longrightarrow$ "unhealthy"

- This is throwing out data! Losing relevant information

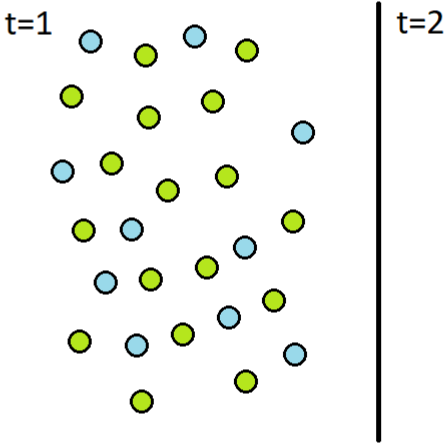
Standard implementation of SRHS

- Most papers convert five health categories to two states:
    - {Excellent, Very good, Good} $\longrightarrow$ "healthy"
    - {Fair, Poor} $\longrightarrow$ "unhealthy"
- And papers take transitions between the two states "straight from the data":
    - Markov(1) process with transition probabilities calculated as a simple fraction of obs
    - Or maybe a reduced form model using one period transitions (smoothing probs)
- This is throwing out data! Losing relevant information

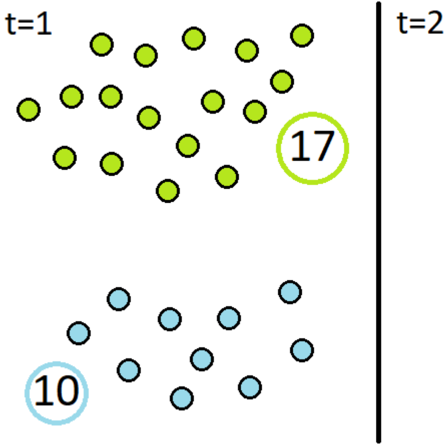
Standard implementation of SRHS

- Most papers convert five health categories to two states:
  - {Excellent, Very good, Good} $\longrightarrow$ "healthy"
  - {Fair, Poor} $\longrightarrow$ "unhealthy"
- And papers take transitions between the two states "straight from the data":
  - Markov(1) process with transition probabilities calculated as a simple fraction of obs
  - Or maybe a reduced form model using one period transitions (smoothing probs)
- This is throwing out data! Losing relevant information
- And taking SRHS **way** too literally

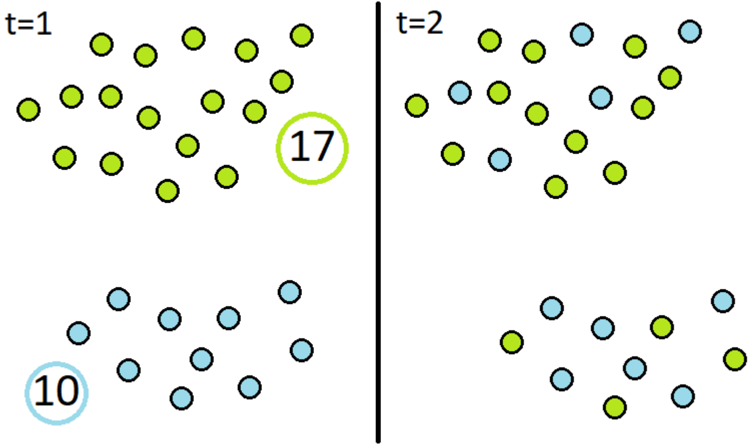
Taking SRHS transitions "straight from the data"



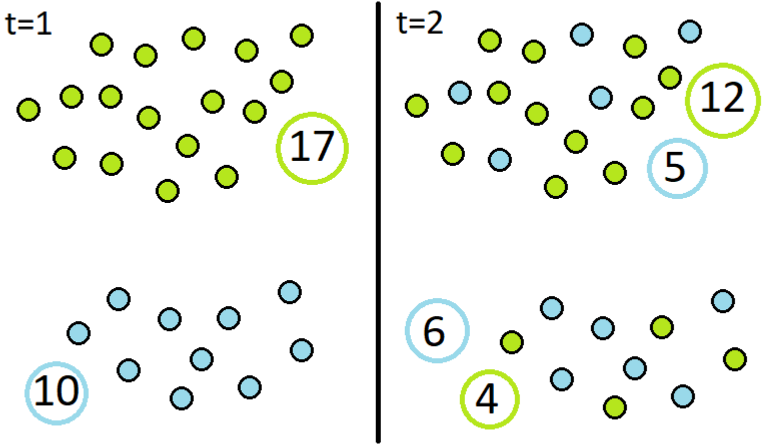
Taking SRHS transitions "straight from the data"



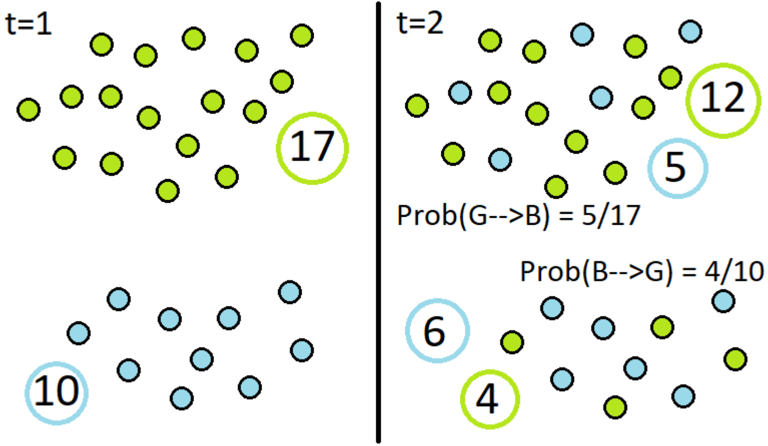
Taking SRHS transitions “straight from the data”



Taking SRHS transitions “straight from the data”



Taking SRHS transitions “straight from the data”



Taking SRHS transitions “straight from the data”

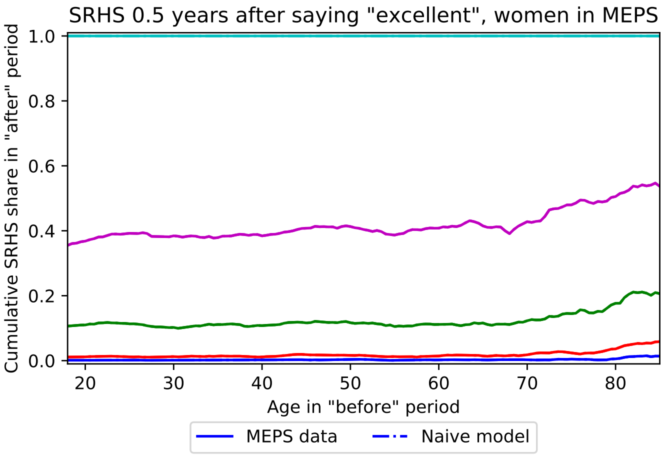
Probit or logit model:

$$y_t = \beta_0 + \beta_1 age_{t-1} + \beta_2 age_{t-1}^2 + \beta_3 sex + \beta_4 h_{t-1} + \epsilon_t,$$

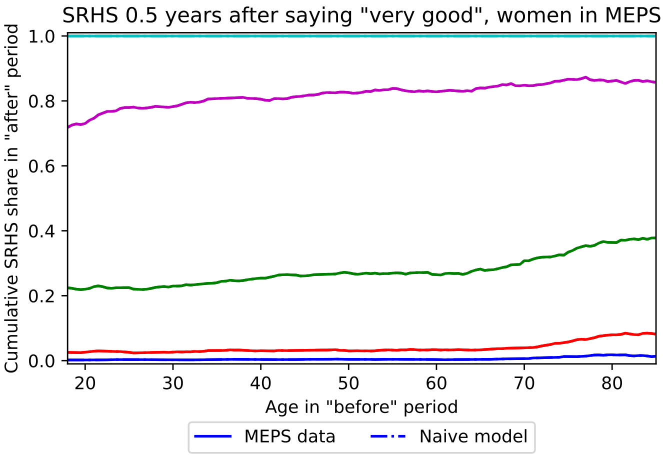
$$h_t = \begin{cases} 1 & \text{if } y_t < 0 \\ 2 & \text{if } y_t \geq 0 \end{cases}.$$

Models with more than two health states use multinomial logit (e.g. Khwaja (2010))...
but always on one period transitions

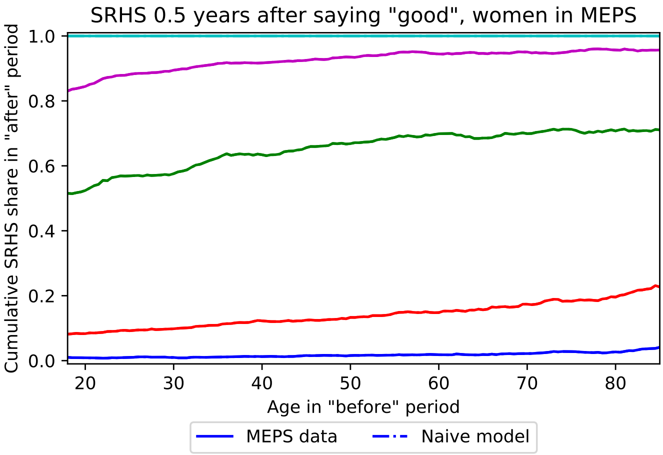
Naive SRHS transitions for MEPS women: 1 wave



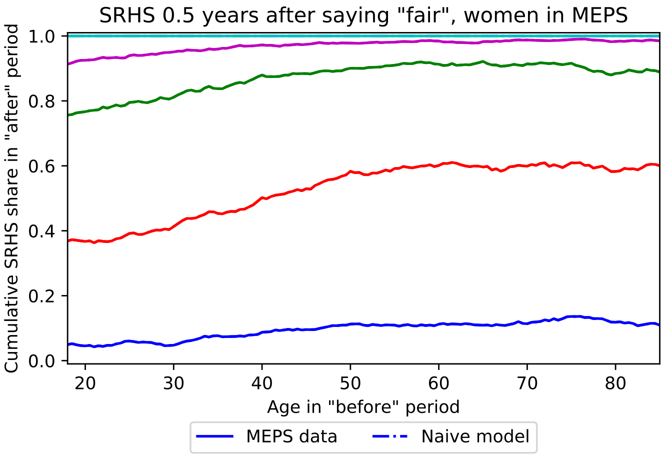
Naive SRHS transitions for MEPS women: 1 wave



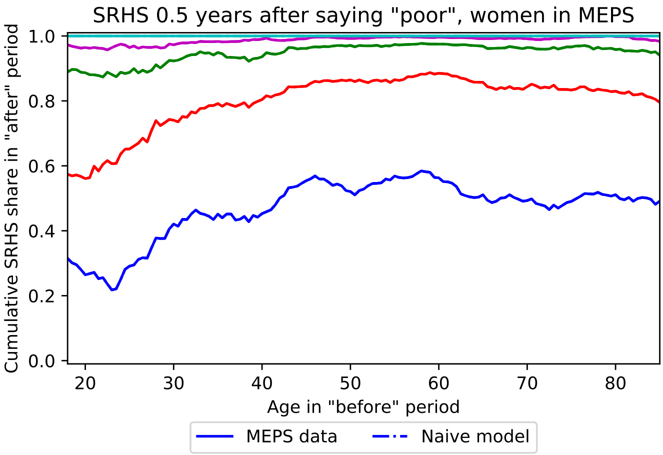
Naive SRHS transitions for MEPS women: 1 wave



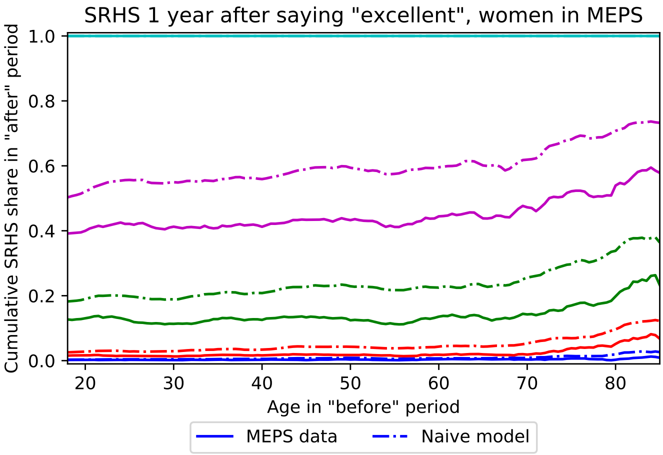
Naive SRHS transitions for MEPS women: 1 wave



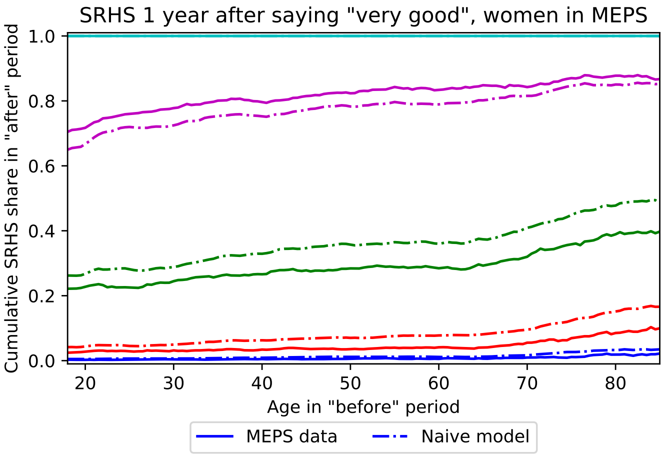
Naive SRHS transitions for MEPS women: 1 wave



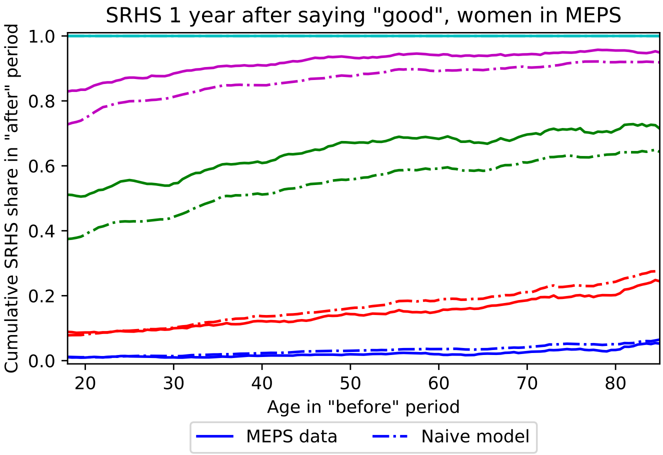
Naive SRHS transitions for MEPS women: 2 waves



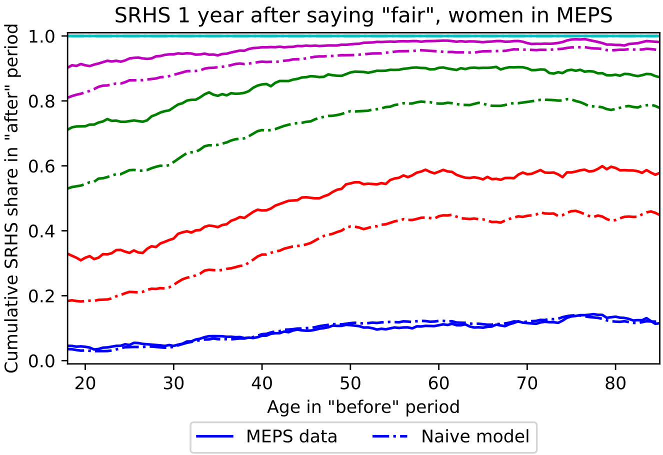
Naive SRHS transitions for MEPS women: 2 waves



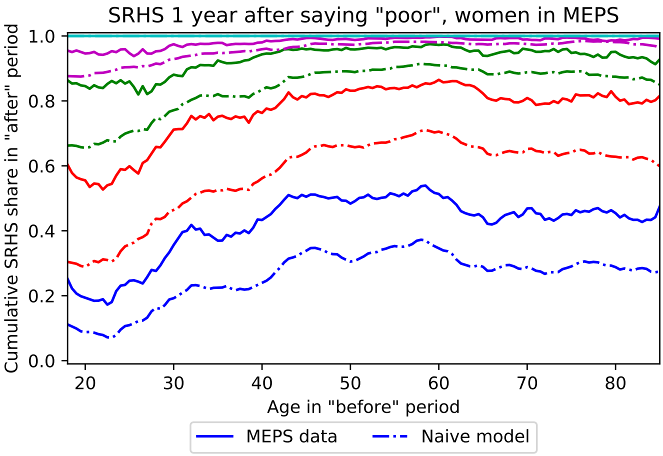
Naive SRHS transitions for MEPS women: 2 waves



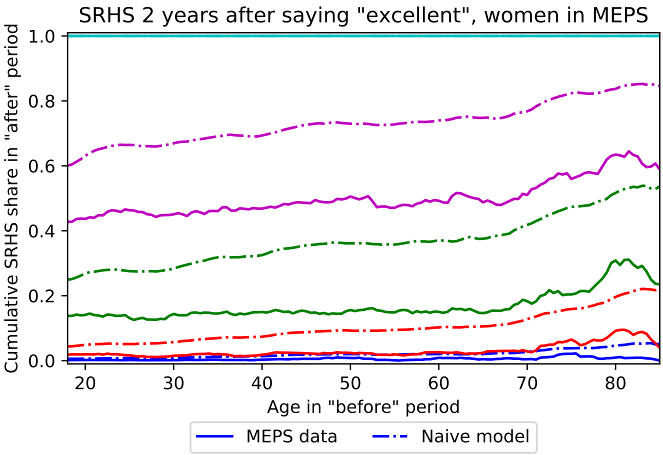
Naive SRHS transitions for MEPS women: 2 waves



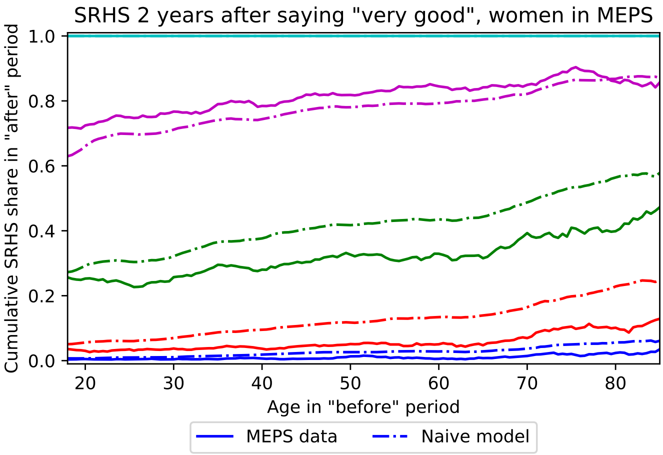
Naive SRHS transitions for MEPS women: 2 waves



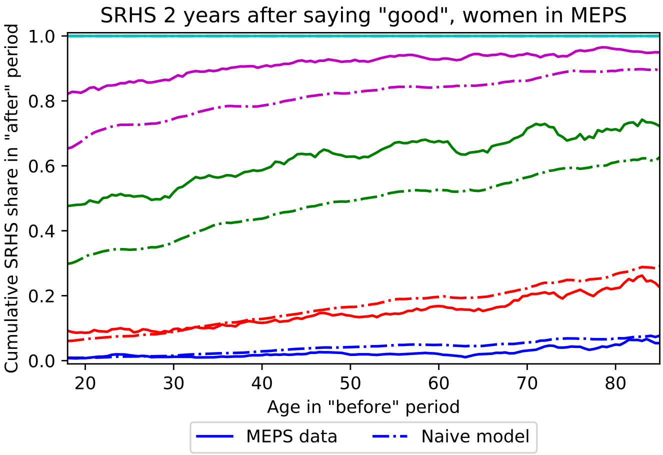
Naive SRHS transitions for MEPS women: 4 waves



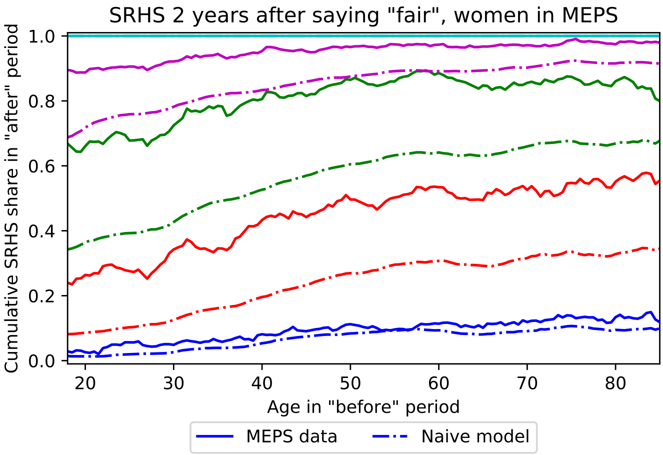
Naive SRHS transitions for MEPS women: 4 waves



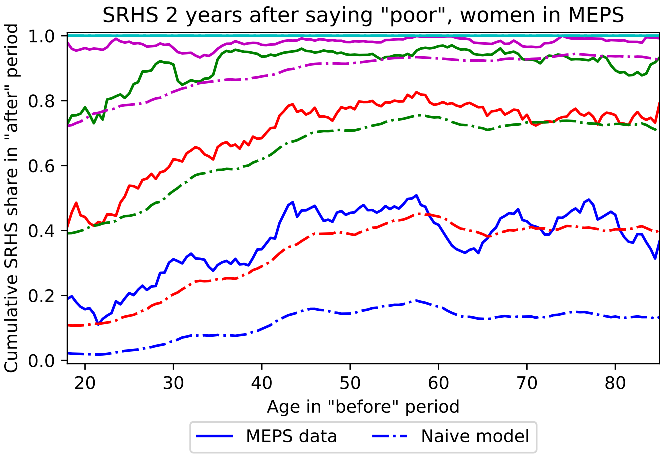
Naive SRHS transitions for MEPS women: 4 waves



Naive SRHS transitions for MEPS women: 4 waves



Naive SRHS transitions for MEPS women: 4 waves



Who uses naive SRHS dynamics?

- Naive health transitions calculated straight from the data are **obviously** a bad idea

Who uses naive SRHS dynamics?

- Naive health transitions calculated straight from the data are **obviously** a bad idea
- Who would possibly use them in “serious” work?
- What journals would publish such papers?

Who uses naive SRHS dynamics?

"We estimate the probability of death and bad health as logistic functions of a cubic in age, sex, sex interacted with age, previous health status, health status interacted with age, a quadratic in permanent income rank, and permanent income rank interacted with age."

— DeNardi, French, & Jones (2010), J. of Political Economy

Who uses naive SRHS dynamics?

"In the first [estimation] step, we estimate or calibrate parameters that can be cleanly identified without explicitly using our model. For example, we estimate mortality rates and health transitions straight from demographic data."

— French & Jones (2011), Econometrica

Who uses naive SRHS dynamics?

"To construct the transition matrix we measure the fraction of people who move from one bin to another between two consecutive years"

— Pashchenko & Porapakkarm (2013), Review of Economic Dynamics

Who uses naive SRHS dynamics?

"[S]ome parameters are estimated outside the structure of the model. For some parameters, this is because no structure is needed: disability risk can be estimated directly from transitions between disability states because of the exogeneity assumption."

— Low & Pistaferri (2015), American Economic Review

Who uses naive SRHS dynamics?

"We estimate health transitions and mortality rates simultaneously by fitting the transitions observed in the HRS to a multinomial logit model [...] on age, sex, current health status, and PI."

— DeNardi, French, & Jones (2016), American Economic Review

Who uses naive SRHS dynamics?

"We estimated the transition probabilities using the logit method. We regressed next period's health status on a constant, age, age squared, age cubic, education level, current health status, and age times current health status."

— Ferreira & Gomes (2017), Review of Economic Dynamics

Who uses naive SRHS dynamics?

"The transition to next period health status is determined as a logistic function of [math notation]."

—Aizawa (forthcoming), Quantitative Economics

SRHS is not Markov(1) (1/3)

- Markov(1) assumption means that **all** relevant information about realization of SRHS h_t is contained in information at $t-1$ (Ω_{t-1})

SRHS is not Markov(1) (1/3)

- Markov(1) assumption means that **all** relevant information about realization of SRHS h_t is contained in information at $t-1$ (Ω_{t-1})
- Specifically: h_{t-n} for $n \geq 2$ is not predictive of h_t if h_{t-1} is known

SRHS is not Markov(1) (1/3)

- Markov(1) assumption means that **all** relevant information about realization of SRHS h_t is contained in information at $t-1$ (Ω_{t-1})
- Specifically: h_{t-n} for $n \geq 2$ is not predictive of h_t if h_{t-1} is known
- Problem: That's just not true

SRHS is not Markov(1) (2/3)

- MEPS: Respondents interviewed five times over two years (6 month waves); new Panel started each year from 1996 onward
- Example: Women between ages 40-45 in Panels 1-20

SRHS is not Markov(1) (2/3)

- MEPS: Respondents interviewed five times over two years (6 month waves); new Panel started each year from 1996 onward
- Example: Women between ages 40-45 in Panels 1-20
- Probability of reporting SRHS “unhealthy” at wave 5 conditional on reporting “unhealthy” at wave 4: 74.7%

SRHS is not Markov(1) (2/3)

- MEPS: Respondents interviewed five times over two years (6 month waves); new Panel started each year from 1996 onward
- Example: Women between ages 40-45 in Panels 1-20
- Probability of reporting SRHS "unhealthy" at wave 5 conditional on reporting "unhealthy" at wave 4: 74.7%
- ...**and** reporting "unhealthy" at wave 3: 80.1%

SRHS is not Markov(1) (2/3)

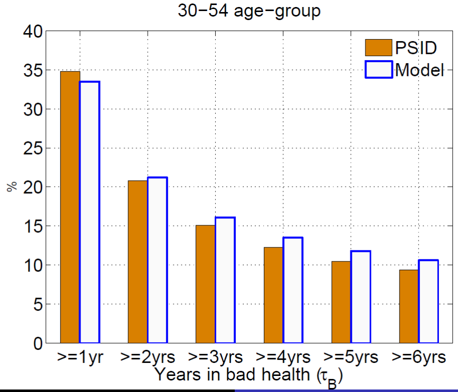
- MEPS: Respondents interviewed five times over two years (6 month waves); new Panel started each year from 1996 onward
- Example: Women between ages 40-45 in Panels 1-20
- Probability of reporting SRHS "unhealthy" at wave 5 conditional on reporting "unhealthy" at wave 4: 74.7%
- ...**and** reporting "unhealthy" at wave 3: 80.1%
- ...**and** reporting "unhealthy" at wave 2: 81.8%

SRHS is not Markov(1) (2/3)

- MEPS: Respondents interviewed five times over two years (6 month waves); new Panel started each year from 1996 onward
- Example: Women between ages 40-45 in Panels 1-20
- Probability of reporting SRHS "unhealthy" at wave 5 conditional on reporting "unhealthy" at wave 4: 74.7%
- ...**and** reporting "unhealthy" at wave 3: 80.1%
- ...**and** reporting "unhealthy" at wave 2: 81.8%
- ...**and** reporting "unhealthy" at wave 1: 84.6%

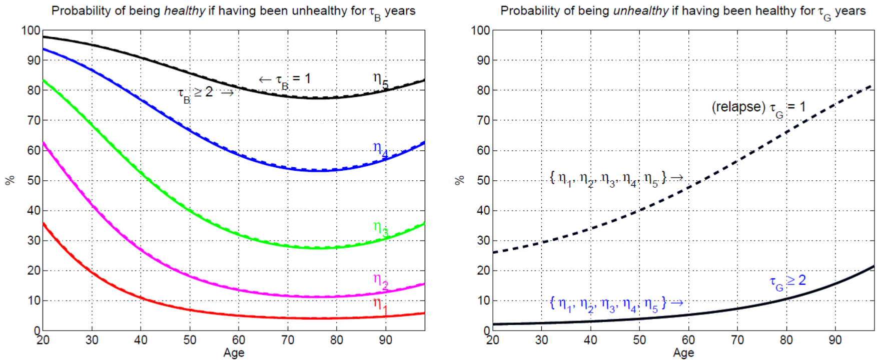
SRHS is not Markov(1) (3/3)

DeNardi, Pashchenko, and Porapakkarm (2018): This is evidence of *duration dependence* in health, so let's make transition probabilities dependent on duration (plus add heterogeneity)!



SRHS is not Markov(1) (3/3)

DeNardi, Pashchenko, and Porapakkarm (2018): This is evidence of *duration dependence* in health, so let's make transition probabilities dependent on duration (plus add heterogeneity)!



But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...

But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...
- ...But still treat "healthy" and "unhealthy" as the only states: mortality and medical expense distribution depend on h_t only!

But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...
- ...But still treat "healthy" and "unhealthy" as the only states: mortality and medical expense distribution depend on h_t only!
- This is easily falsifiable; quick example with early 40's women:
- Mean log medical costs for "unhealthy": 7.80
- Mean log medical costs for "healthy": 6.82

But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...
- ...But still treat "healthy" and "unhealthy" as the only states: mortality and medical expense distribution depend on h_t only!
- This is easily falsifiable; quick example with early 40's women:
- Mean log medical costs for "unhealthy": 7.80
  - ...and "unhealthy" one year prior: 7.95
  - ...and "healthy" one year prior: 7.57
- Mean log medical costs for "healthy": 6.82

But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...
- ...But still treat "healthy" and "unhealthy" as the only states: mortality and medical expense distribution depend on h_t only!
- This is easily falsifiable; quick example with early 40's women:
- Mean log medical costs for "unhealthy": 7.80
  - ...and "unhealthy" one year prior: 7.95
  - ...and "healthy" one year prior: 7.57
- Mean log medical costs for "healthy": 6.82
  - ...and "unhealthy" one year prior: 7.14
  - ...and "healthy" one year prior: 6.79

But it's not duration dependence (1/3)

- DeNardi et al model health transitions between two states as Markov(2): h_t probabilities depend on h_{t-1} and h_{t-2}...
- ...But still treat "healthy" and "unhealthy" as the only states: mortality and medical expense distribution depend on h_t only!
- This is easily falsifiable; quick example with early 40's women:
- Mean log medical costs for "unhealthy": 7.80
  - ...and "unhealthy" one year prior: 7.95
  - ...and "healthy" one year prior: 7.57
- Mean log medical costs for "healthy": 6.82
  - ...and "unhealthy" one year prior: 7.14
  - ...and "healthy" one year prior: 6.79
- **Lagged SRHS generates observably different medical expense distributions; "unhealthy" people are not all in the same health state!**

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%
- Probability of reporting SRHS "unhealthy" at wave 2 conditional on reporting "unhealthy" at wave 1: 57.9%

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%
- Probability of reporting SRHS "unhealthy" at wave 2 conditional on reporting "unhealthy" at wave 1: 57.9%
- Probability of reporting SRHS "unhealthy" at wave 3 conditional on reporting "unhealthy" at wave 1: 57.6%

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%
- Probability of reporting SRHS "unhealthy" at wave 2 conditional on reporting "unhealthy" at wave 1: 57.9%
- Probability of reporting SRHS "unhealthy" at wave 3 conditional on reporting "unhealthy" at wave 1: 57.6%
- Probability of reporting SRHS "unhealthy" at wave 4 conditional on reporting "unhealthy" at wave 1: 56.7%

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%
- Probability of reporting SRHS "unhealthy" at wave 2 conditional on reporting "unhealthy" at wave 1: 57.9%
- Probability of reporting SRHS "unhealthy" at wave 3 conditional on reporting "unhealthy" at wave 1: 57.6%
- Probability of reporting SRHS "unhealthy" at wave 4 conditional on reporting "unhealthy" at wave 1: 56.7%
- Probability of reporting SRHS "unhealthy" at wave 5 conditional on reporting "unhealthy" at wave 1: 55.8%

But it's not duration dependence (2/3)

- Let's take another look at those early 40's MEPS women
- Probability of reporting SRHS "unhealthy" at wave 1 conditional on reporting "unhealthy" at wave 1: 100.0%
- Probability of reporting SRHS "unhealthy" at wave 2 conditional on reporting "unhealthy" at wave 1: 57.9%
- Probability of reporting SRHS "unhealthy" at wave 3 conditional on reporting "unhealthy" at wave 1: 57.6%
- Probability of reporting SRHS "unhealthy" at wave 4 conditional on reporting "unhealthy" at wave 1: 56.7%
- Probability of reporting SRHS "unhealthy" at wave 5 conditional on reporting "unhealthy" at wave 1: 55.8%
- **Wave 1 SRHS is nearly as predictive of W5 SRHS as it is for W2 SRHS!**

But it's not duration dependence (3/3)

- Suppose you saw these examples, but the labels were "in bottom income quintile" and "above bottom income quintile"

But it's not duration dependence (3/3)

- Suppose you saw these examples, but the labels were "in bottom income quintile" and "above bottom income quintile"
- You would not hypothesize: "There must be duration dependence in being poor!"

But it's not duration dependence (3/3)

- Suppose you saw these examples, but the labels were "in bottom income quintile" and "above bottom income quintile"
- You would not hypothesize: "There must be duration dependence in being poor!"
- Income quintile is a discretization of continuous income
- SRHS is a discretization of continuous "health"

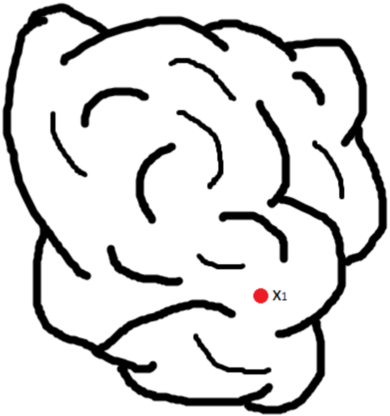
But it's not duration dependence (3/3)

- Suppose you saw these examples, but the labels were "in bottom income quintile" and "above bottom income quintile"
- You would not hypothesize: "There must be duration dependence in being poor!"
- Income quintile is a discretization of continuous income
- SRHS is a discretization of continuous "health"
- Income changes with both permanent and transitory shocks
- SRHS changes with both changes in "true health" and transitory **reporting** shocks

MODEL

Graphical model of latent health and SRHS (1/5)

"True health" lives in some weird polydimensional space



Graphical model of latent health and SRHS (2/5)

Dynamics of "true health" really **are** Markov(1)

Graphical model of latent health and SRHS (2/5)

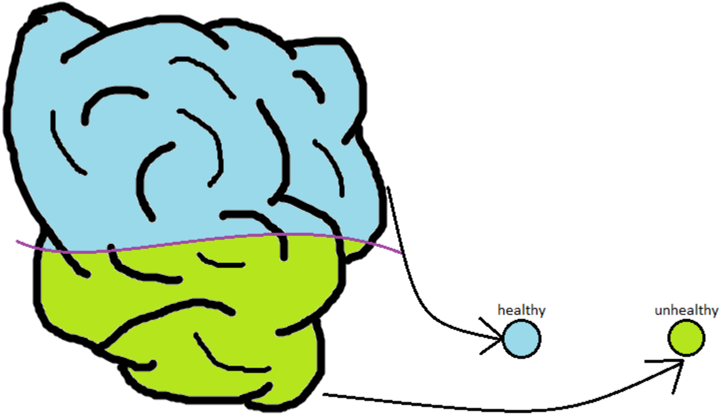
Dynamics of “true health” really **are** Markov(1)

Graphical model of latent health and SRHS (2/5)

Dynamics of “true health” really **are** Markov(1)

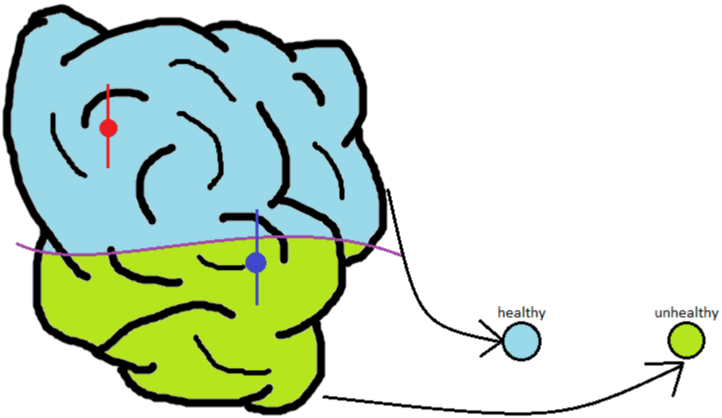
Graphical model of latent health and SRHS (3/5)

SRHS is a projection of that polydimensional space onto a discrete space...



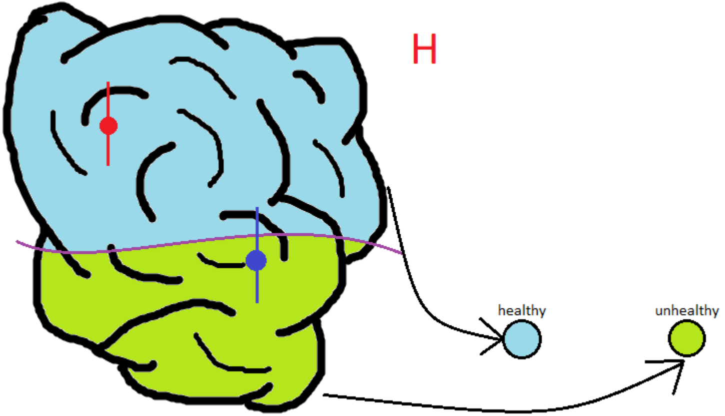
Graphical model of latent health and SRHS (3/5)

SRHS is a projection of that polydimensional space onto a discrete space... with reporting error



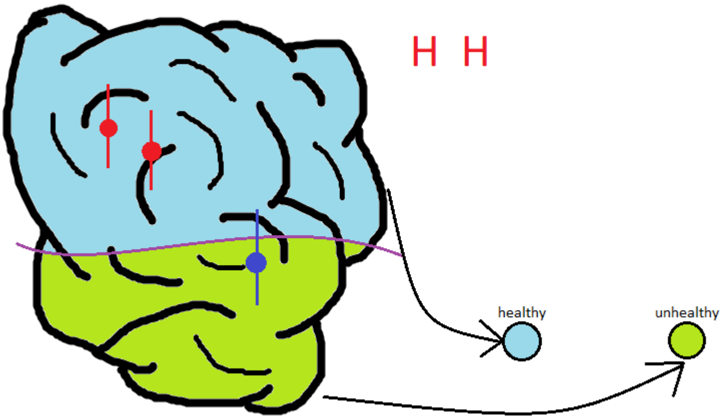
Graphical model of latent health and SRHS (4/5)

Individuals whose "true health" is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



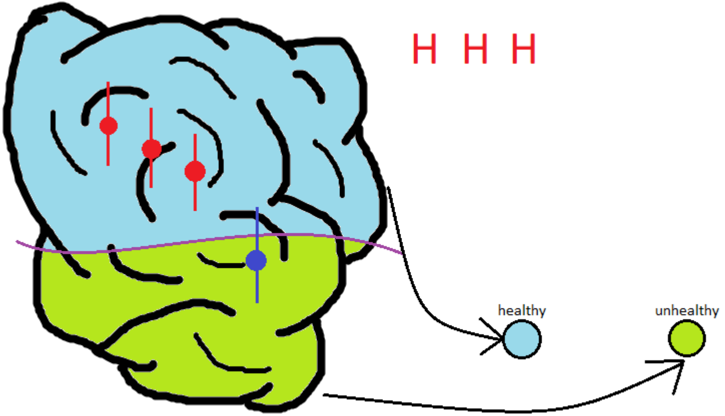
Graphical model of latent health and SRHS (4/5)

Individuals whose "true health" is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



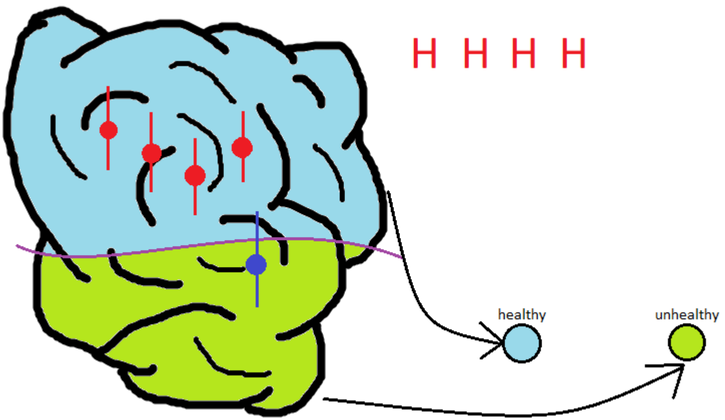
Graphical model of latent health and SRHS (4/5)

Individuals whose "true health" is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



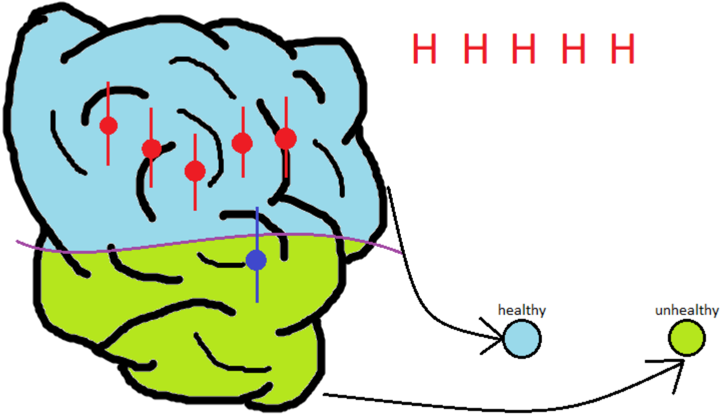
Graphical model of latent health and SRHS (4/5)

Individuals whose “true health” is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



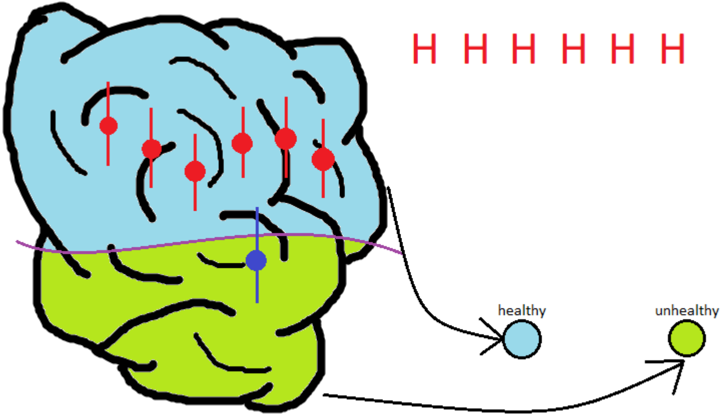
Graphical model of latent health and SRHS (4/5)

Individuals whose "true health" is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



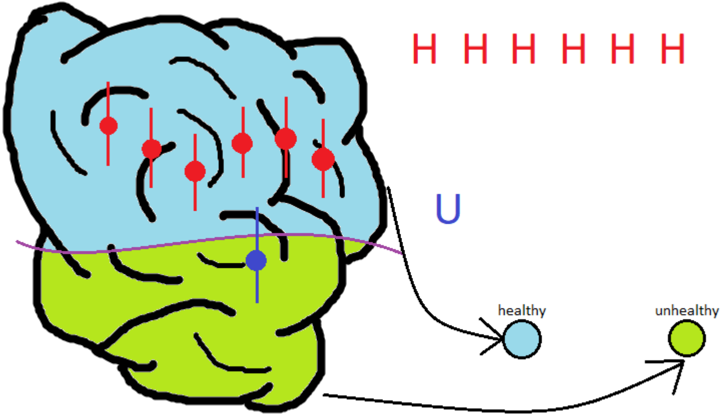
Graphical model of latent health and SRHS (4/5)

Individuals whose “true health” is far from the healthy/unhealthy threshold are likely to have the same SRHS period after period



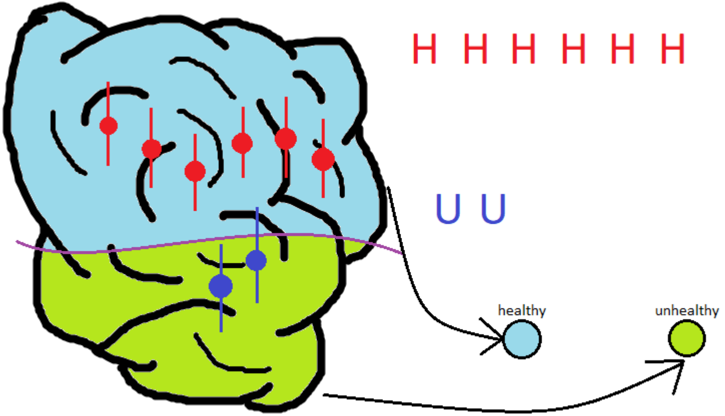
Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



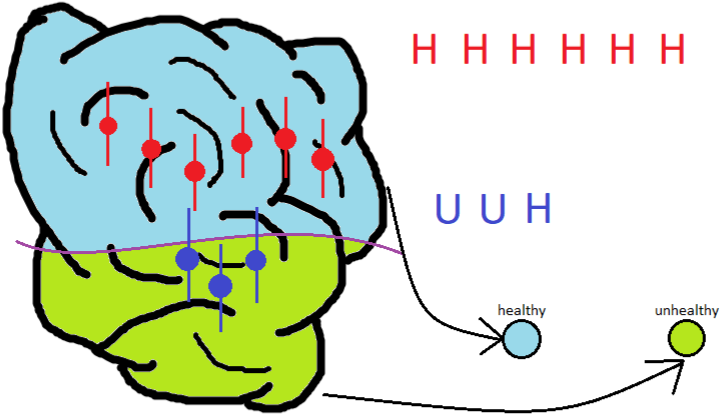
Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



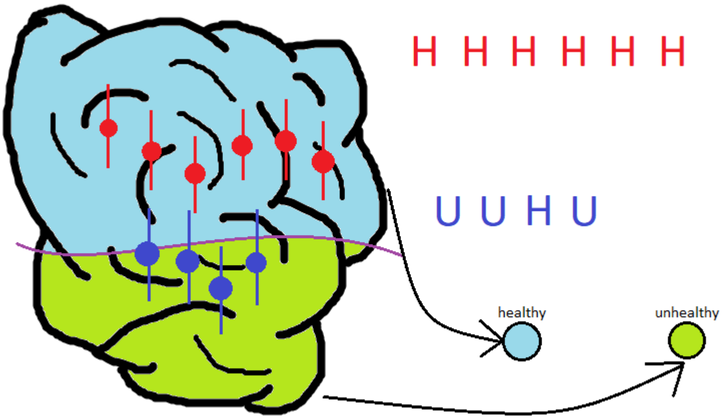
Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



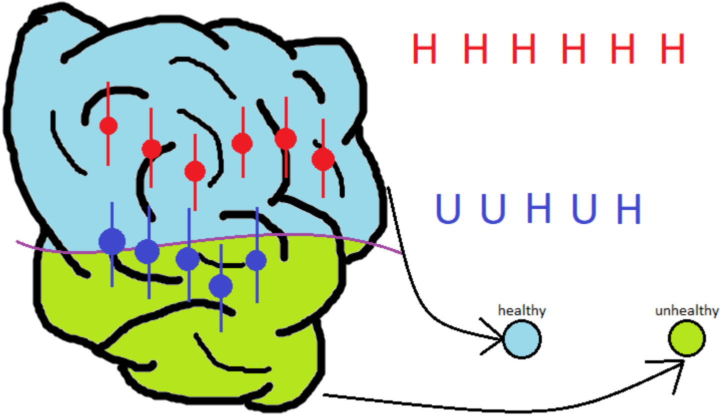
Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



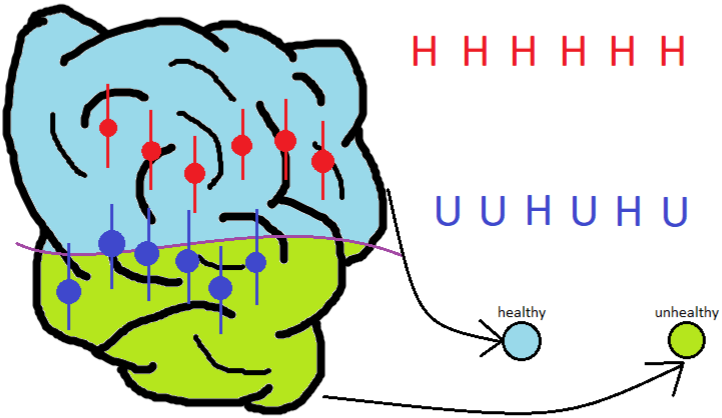
Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



Graphical model of latent health and SRHS (5/5)

Individuals whose "true health" is close to the healthy/unhealthy threshold are more likely to change their SRHS back and forth



Feasible modeling

- Polydimensional space would be hard to model
- No way to estimate that model with only SRHS anyway

Feasible modeling

- Polydimensional space would be hard to model
- No way to estimate that model with only SRHS anyway
- Feasible: one factor latent health model

Feasible modeling

- Polydimensional space would be hard to model
- No way to estimate that model with only SRHS anyway
- Feasible: one factor latent health model
- Will estimate by maximum likelihood: probability of observing **sequences** of SRHS in panel data
- Track (nonparametric) posterior distribution of latent health across waves for each respondent

Feasible modeling

- Polydimensional space would be hard to model
- No way to estimate that model with only SRHS anyway
- Feasible: one factor latent health model
- Will estimate by maximum likelihood: probability of observing **sequences** of SRHS in panel data
- Track (nonparametric) posterior distribution of latent health across waves for each respondent
- No (unobserved) heterogeneous types

Mathematical model of latent health and SRHS

One factor latent health model:

1. Distribution of latent health at model start x_0 is normal

Mathematical model of latent health and SRHS

One factor latent health model:

1. Distribution of latent health at model start x_0 is normal
2. Survival to next period is a probit on age j_t and x_t

Mathematical model of latent health and SRHS

One factor latent health model:

1. Distribution of latent health at model start x_0 is normal
2. Survival to next period is a probit on age j_t and x_t
3. Latent health dynamics are AR(1) with standard normal shock

Mathematical model of latent health and SRHS

One factor latent health model:

1. Distribution of latent health at model start x_0 is normal
2. Survival to next period is a probit on age j_t and x_t
3. Latent health dynamics are AR(1) with standard normal shock
4. SRHS is an ordered probit on a linear function of x_t with standard normal error

Mathematical model of latent health and SRHS

One factor latent health model:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

2. Survival to next period is a probit on age j_t and x_t
3. Latent health dynamics: AR(1) with standard normal shock
4. SRHS is an ordered probit on linear function of x_t with standard normal error

Mathematical model of latent health and SRHS

One factor latent health model:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

3. Latent health dynamics: AR(1) with standard normal shock
4. SRHS is an ordered probit on linear function of x_t with standard normal error

Mathematical model of latent health and SRHS

One factor latent health model:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \Phi. \tag{3}$$

4. SRHS is an ordered probit on linear function of x_t with standard normal error

Mathematical model of latent health and SRHS

One factor latent health model:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \Phi. \tag{3}$$

$$h_t^* = \alpha_0 + \alpha_1 x_t + \eta_t, \qquad \eta_t \sim \Phi, \tag{4}$$

$$h_t = 1 + \sum_{k=1}^{K-1} \mathbf{1}(h_t^* \geq \chi_k).$$

Why reporting error?

- Why model SRHS with reporting error? Or: Why is it not transitory shocks to true health?

Why reporting error?

- Why model SRHS with reporting error? Or: Why is it not transitory shocks to true health?
- Crossley & Kennedy (2002): 28% of respondents **change their SRHS** if asked again later in the same interview

Why reporting error?

- Why model SRHS with reporting error? Or: Why is it not transitory shocks to true health?
- Crossley & Kennedy (2002): 28% of respondents **change their SRHS** if asked again later in the same interview
- MEPS mails short self-administered questionnaire (SAQ) to all respondents shortly after waves 2 and 4; about 75% response rate
- Includes **nearly** identical SRHS question as main MEPS

Why reporting error?

- Why model SRHS with reporting error? Or: Why is it not transitory shocks to true health?
- Crossley & Kennedy (2002): 28% of respondents **change their SRHS** if asked again later in the same interview
- MEPS mails short self-administered questionnaire (SAQ) to all respondents shortly after waves 2 and 4; about 75% response rate
- Includes **nearly** identical SRHS question as main MEPS
- Early 40's women: 37-38% change SRHS from their answer a couple weeks prior if reported poor to very good health; 44% change answer if reported excellent health

Identifying assumptions and functional forms

- Mortality probit $f(\cdot)$ is polynomial of age and latent health plus linear interaction term (parameterized by θ).

Identifying assumptions and functional forms

- Mortality probit $f(\cdot)$ is polynomial of age and latent health plus linear interaction term (parameterized by θ).
- Latent health correlation coefficient ρ_j is polynomial of age (parameterized by γ).

Identifying assumptions and functional forms

- Mortality probit $f(\cdot)$ is polynomial of age and latent health plus linear interaction term (parameterized by θ).
- Latent health correlation coefficient ρ_j is polynomial of age (parameterized by γ).
- Latent health regression term $g(j_t)$ is polynomial of age (parameterized by β).

Identifying assumptions and functional forms

- Mortality probit $f(\cdot)$ is polynomial of age and latent health plus linear interaction term (parameterized by θ).
- Latent health correlation coefficient ρ_j is polynomial of age (parameterized by γ).
- Latent health regression term $g(j_t)$ is polynomial of age (parameterized by β).
- Scale of x_t already pinned down because $\epsilon_{t+1} \sim \Phi$.
- Fix location of x_t by setting $\alpha_0 = 0$ and $\chi_1 = 0$.

Identifying assumptions and functional forms

- Mortality probit $f(\cdot)$ is polynomial of age and latent health plus linear interaction term (parameterized by θ).
- Latent health correlation coefficient ρ_j is polynomial of age (parameterized by γ).
- Latent health regression term $g(j_t)$ is polynomial of age (parameterized by β).
- Scale of x_t already pinned down because $\epsilon_{t+1} \sim \Phi$.
- Fix location of x_t by setting $\alpha_0 = 0$ and $\chi_1 = 0$.
- All parameters can differ between men and women.
- No reporting error for mortality.
- No selection in attrition or missing SRHS.

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood
- Slice space of latent health x_t into many bins (150)
- Choose some very old age that no one in data reaches

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood
- Slice space of latent health x_t into many bins (150)
- Choose some very old age that no one in data reaches
- Given a vector of parameter guess $(\theta, \gamma, \beta, \alpha, \chi, \mu_0, \sigma_0)$, can precompute arrays:
  1. Initial pmf of x_t at model start (age "0")

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood
- Slice space of latent health x_t into many bins (150)
- Choose some very old age that no one in data reaches
- Given a vector of parameter guess $(\theta, \gamma, \beta, \alpha, \chi, \mu_0, \sigma_0)$, can precompute arrays:
  1. Initial pmf of x_t at model start (age "0")
  2. Survival probability for every x_t and age j_t
  3. Transition probability from every x_t to every x_{t+1} at every age

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood
- Slice space of latent health x_t into many bins (150)
- Choose some very old age that no one in data reaches
- Given a vector of parameter guess $(\theta, \gamma, \beta, \alpha, \chi, \mu_0, \sigma_0)$, can precompute arrays:
  1. Initial pmf of x_t at model start (age "0")
  2. Survival probability for every x_t and age j_t
  3. Transition probability from every x_t to every x_{t+1} at every age
  4. Probability of reporting each SRHS h_t conditional on x_t

Log likelihood function: Precomputation

- Can estimate latent health model using maximum likelihood
- Slice space of latent health x_t into many bins (150)
- Choose some very old age that no one in data reaches
- Given a vector of parameter guess $(\theta, \gamma, \beta, \alpha, \chi, \mu_0, \sigma_0)$, can precompute arrays:
  1. Initial pmf of x_t at model start (age "0")
  2. Survival probability for every x_t and age j_t
  3. Transition probability from every x_t to every x_{t+1} at every age
  4. Probability of reporting each SRHS h_t conditional on x_t
- Using (1-3), can precompute **unconditional** pmf of x_t at every age (i.e. simulation by multiplication)

Log likelihood function: Observation likelihood

For each individual in the data:

1. Initialize "survival probability since last observation" to 1.0
2. Get precomputed pmf of x_t at first observed age

Log likelihood function: Observation likelihood

For each individual in the data:

1. Initialize "survival probability since last observation" to 1.0
2. Get precomputed pmf of x_t at first observed age

4. Get probability of observing reported h_t at every x_t
5. Update pmf of x_t using those probs (and include in likelihood)

Log likelihood function: Observation likelihood

For each individual in the data:

1. Initialize "survival probability since last observation" to 1.0
2. Get precomputed pmf of x_t at first observed age

4. Get probability of observing reported h_t at every x_t
5. Update pmf of x_t using those probs (and include in likelihood)
6. Get survival probability at every x_t at this age
7. Update pmf of x_t using those probs, assuming survival; accumulate survival prob

Log likelihood function: Observation likelihood

For each individual in the data:

1. Initialize "survival probability since last observation" to 1.0
2. Get precomputed pmf of x_t at first observed age

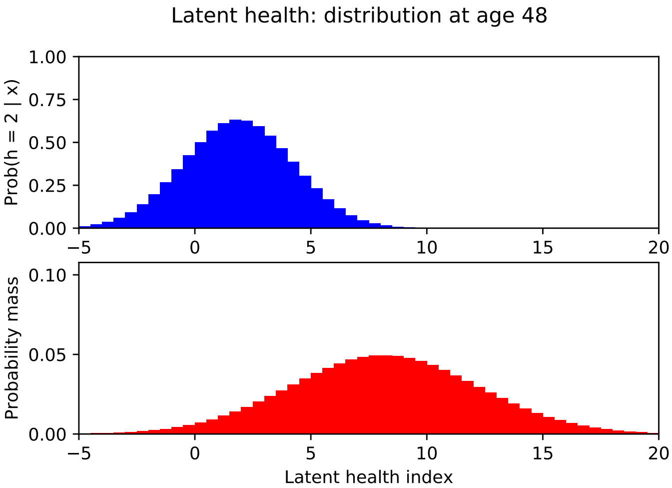
4. Get probability of observing reported h_t at every x_t
5. Update pmf of x_t using those probs (and include in likelihood)
6. Get survival probability at every x_t at this age
7. Update pmf of x_t using those probs, assuming survival; accumulate survival prob
8. Update pmf of x_t using transition probs from this age to next
9. If not last observed wave, go to (3); else END

Log likelihood function: Observation likelihood

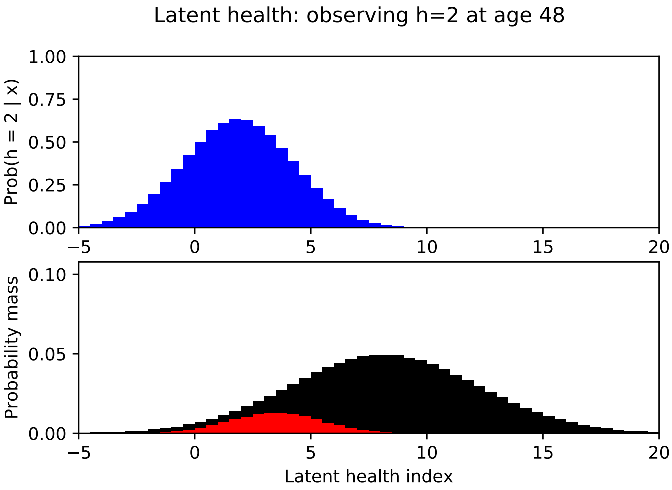
For each individual in the data:

1. Initialize “survival probability since last observation” to 1.0
2. Get precomputed pmf of x_t at first observed age
3. If alive, include survival probability in likelihood, reset to 1.0; if dead, include complementary probability and END
4. Get probability of observing reported h_t at every x_t
5. Update pmf of x_t using those probs (and include in likelihood)
6. Get survival probability at every x_t at this age
7. Update pmf of x_t using those probs, assuming survival; accumulate survival prob
8. Update pmf of x_t using transition probs from this age to next
9. If not last observed wave, go to (3); else END

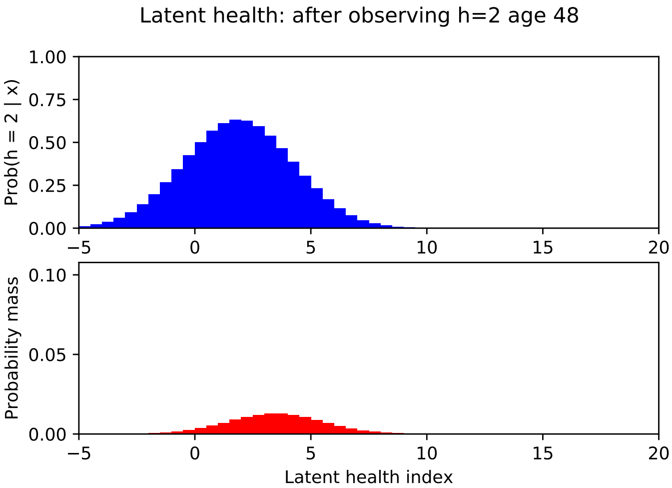
Log likelihood function: self-reported health status



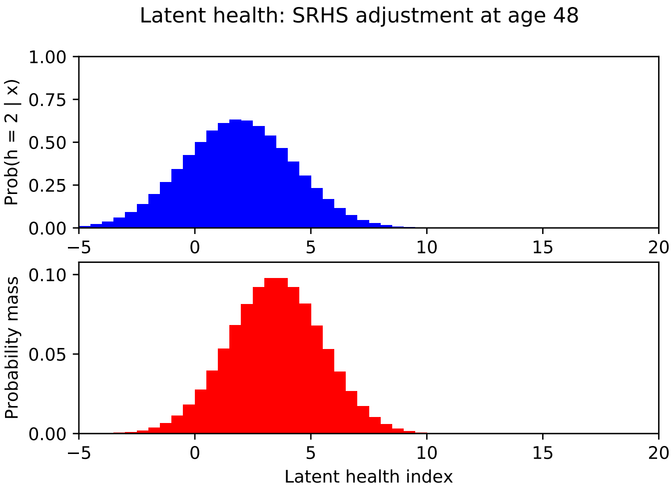
Log likelihood function: self-reported health status



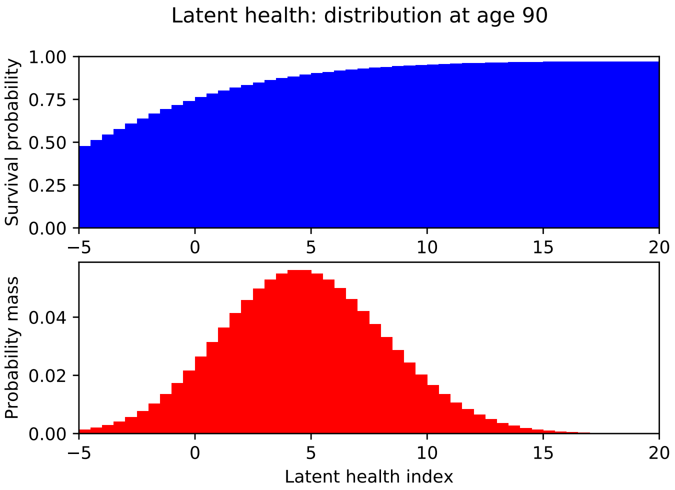
Log likelihood function: self-reported health status



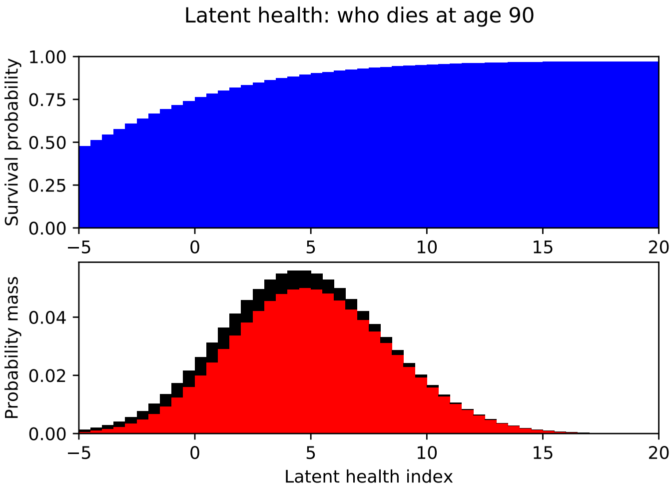
Log likelihood function: self-reported health status



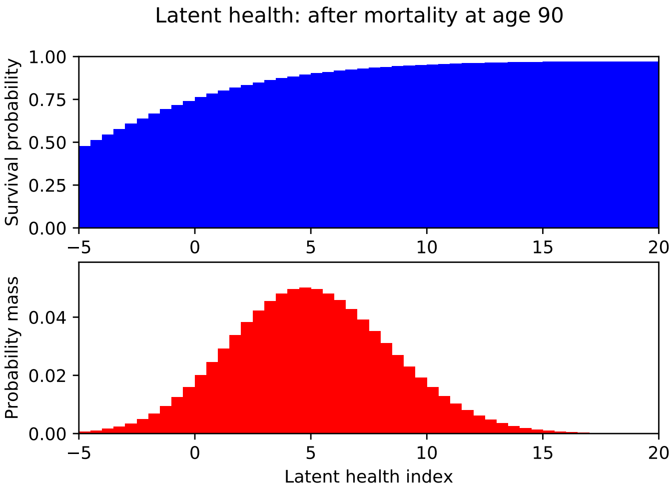
Log likelihood function: mortality



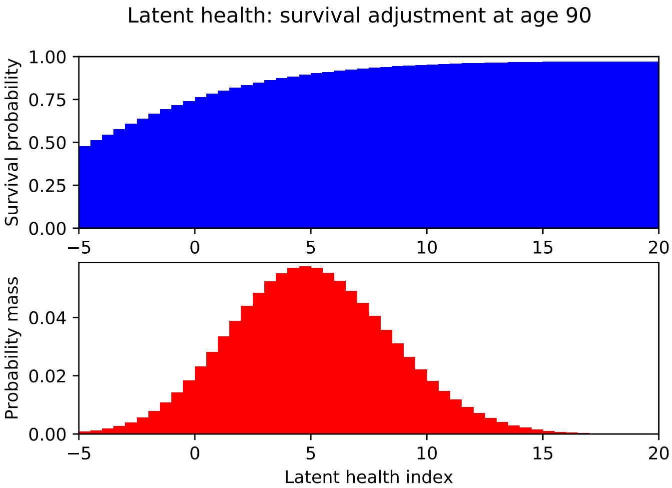
Log likelihood function: mortality



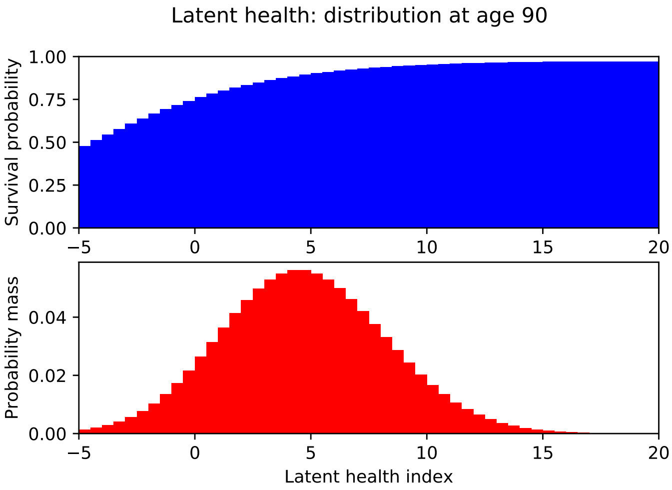
Log likelihood function: mortality



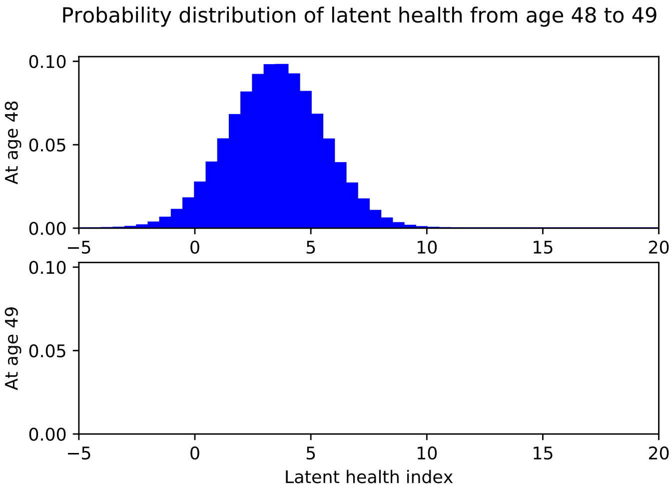
Log likelihood function: mortality



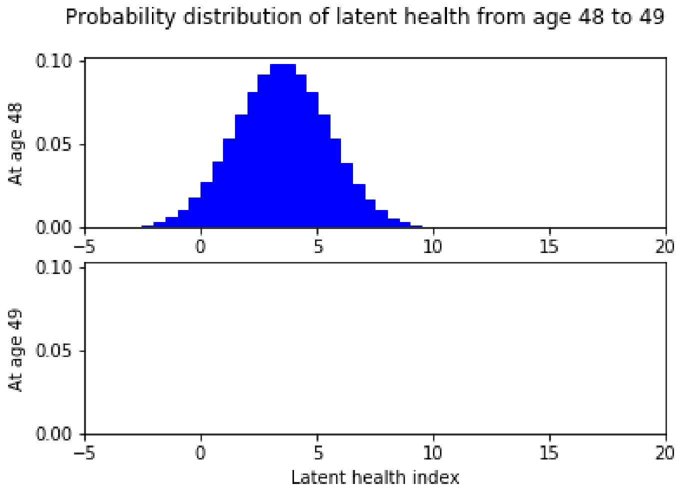
Log likelihood function: mortality



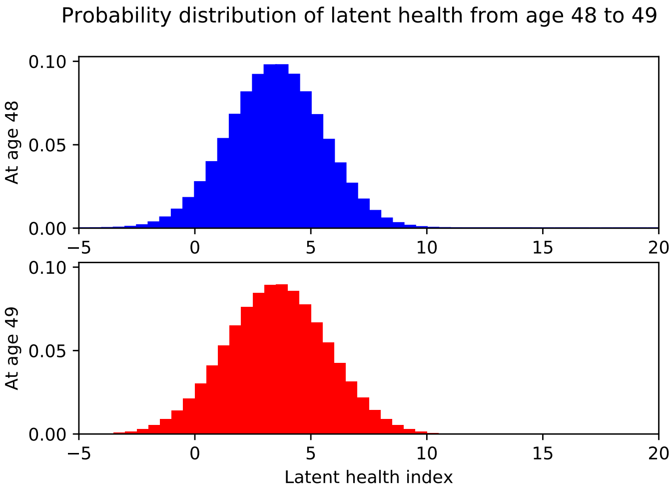
Log likelihood function: latent health transitions



Log likelihood function: latent health transitions



Log likelihood function: latent health transitions



Identification challenge

- α_1 near zero means SRHS is mostly reporting error
- High α_1 means SRHS is almost perfect report of health

Identification challenge

- α_1 near zero means SRHS is mostly reporting error
- High α_1 means SRHS is almost perfect report of health
- Holding ρ_j fixed, lower α_1 means SRHS changes more between t and $t+1$

Identification challenge

- α_1 near zero means SRHS is mostly reporting error
- High α_1 means SRHS is almost perfect report of health
- Holding ρ_j fixed, lower α_1 means SRHS changes more between t and $t+1$
- ρ_j near zero means latent health is (almost) iid
- ρ_j near one means it's (almost) a random walk

Identification challenge

- α_1 near zero means SRHS is mostly reporting error
- High α_1 means SRHS is almost perfect report of health
- Holding ρ_j fixed, lower α_1 means SRHS changes more between t and $t+1$
- ρ_j near zero means latent health is (almost) iid
- ρ_j near one means it's (almost) a random walk
- Holding α_1 fixed, lower ρ_j means SRHS changes more between t and $t+1$

Identification challenge

- α_1 near zero means SRHS is mostly reporting error
- High α_1 means SRHS is almost perfect report of health
- Holding ρ_j fixed, lower α_1 means SRHS changes more between t and $t+1$
- ρ_j near zero means latent health is (almost) iid
- ρ_j near one means it's (almost) a random walk
- Holding α_1 fixed, lower ρ_j means SRHS changes more between t and $t+1$
- How to tell apart low correlation in latent health from large reporting shocks?

Identification strategy (1/2)

- How to tell apart low correlation in latent health from large reporting shocks?

Identification strategy (1/2)

- How to tell apart low correlation in latent health from large reporting shocks?
- Fit **sequences** of SRHS, not just transitions
- Separate correlation of latent health from reporting error by t to $t+n$ transitions

Identification strategy (1/2)

- How to tell apart low correlation in latent health from large reporting shocks?
- Fit **sequences** of SRHS, not just transitions
- Separate correlation of latent health from reporting error by t to $t+n$ transitions
- Compare SRHS of fair-to-poor R's vs fair-to-good R's in $t+2$

Identification strategy (1/2)

- How to tell apart low correlation in latent health from large reporting shocks?
- Fit **sequences** of SRHS, not just transitions
- Separate correlation of latent health from reporting error by t to $t+n$ transitions
- Compare SRHS of fair-to-poor R's vs fair-to-good R's in $t+2$
- More similar distributions of SRHS: low α_1, high ρ_j
- Less similar distributions of SRHS: high α_1, low ρ_j

Identification strategy (1/2)

- How to tell apart low correlation in latent health from large reporting shocks?
- Fit **sequences** of SRHS, not just transitions
- Separate correlation of latent health from reporting error by t to $t+n$ transitions
- Compare SRHS of fair-to-poor R's vs fair-to-good R's in $t+2$
- More similar distributions of SRHS: low α_1, high ρ_j
- Less similar distributions of SRHS: high α_1, low ρ_j
- Need more than 2 SRHS categories to do this!

Identification strategy (2/2)

- Rest of the parameters are straightforward to identify
- Main difficulty: Changing α_1 and γ parameters changes scale of x_t... requiring many other parameters to adjust!

Identification strategy (2/2)

- Rest of the parameters are straightforward to identify
- Main difficulty: Changing α_1 and γ parameters changes scale of x_t... requiring many other parameters to adjust!
- Identify mortality params θ off of deaths by age and SRHS

Identification strategy (2/2)

- Rest of the parameters are straightforward to identify
- Main difficulty: Changing α_1 and γ parameters changes scale of x_t... requiring many other parameters to adjust!
- Identify mortality params θ off of deaths by age and SRHS
- Identify expected health parameters β, initial health distribution (μ_0, σ_0^2), and probit cut points χ from distribution of SRHS by age

DATA

Medical Expenditure Panel Survey

- Began in 1996, focused on medical events & expenses
- Nationally representative sample (weights provided)
- Short panel: respondents interviewed five times over 2 years
- New panel started every year (overlapping panels)

Medical Expenditure Panel Survey

- Began in 1996, focused on medical events & expenses
- Nationally representative sample (weights provided)
- Short panel: respondents interviewed five times over 2 years
- New panel started every year (overlapping panels)
- Use Panels 1-20, spanning 1996-2017
- Drop under 18's and over 84's (age is topcoded at 85)
- Relabel age to have half-years based on age in W1 and W2

Medical Expenditure Panel Survey

- Began in 1996, focused on medical events & expenses
- Nationally representative sample (weights provided)
- Short panel: respondents interviewed five times over 2 years
- New panel started every year (overlapping panels)
- Use Panels 1-20, spanning 1996-2017
- Drop under 18's and over 84's (age is topcoded at 85)
- Relabel age to have half-years based on age in W1 and W2
- 126,486 women observed for 601,713 waves (1,168 deaths)
- 110,044 men observed for 518,577 waves (1,314 deaths)
- 89% of respondents observed all 5 waves; most others drop after 3 (maybe planned drops?)

Health and Retirement Survey

- Begins "in earnest" in 1996– oldest cohorts started in 1993
- Respondents enter around age 55; nationally representative
- Long panel: R's interviewed every two years (until death)

Health and Retirement Survey

- Begins "in earnest" in 1996– oldest cohorts started in 1993
- Respondents enter around age 55; nationally representative
- Long panel: R's interviewed every two years (until death)
- Use 1996 to 2016 data waves (11 waves)
- Drop R's under 50's: not enough spouses this young
- Drop R's first observed older than age 95 (not enough)
- Observe SRHS every 2 years; recoded to 1 year periods

Health and Retirement Survey

- Begins "in earnest" in 1996– oldest cohorts started in 1993
- Respondents enter around age 55; nationally representative
- Long panel: R's interviewed every two years (until death)
- Use 1996 to 2016 data waves (11 waves)
- Drop R's under 50's: not enough spouses this young
- Drop R's first observed older than age 95 (not enough)
- Observe SRHS every 2 years; recoded to 1 year periods
- 20,108 women observed for 107,542 waves (5,794 deaths)
- 16,094 men observed for 81,091 waves (5,286 deaths)
- 50% observed alive at least 5 waves, 11% obs all 11 waves

Panel Study of Income Dynamics

- Began in 1968 with nationally representative sample of HH's
- Snowballing sample (with recent immigrants added)
- Long panel: annual interviews until 1997, biannual since

Panel Study of Income Dynamics

- Began in 1968 with nationally representative sample of HH's
- Snowballing sample (with recent immigrants added)
- Long panel: annual interviews until 1997, biannual since
- Use data from 1997 to 2017 (11 waves, recode to 21 years)
- SRHS **only** asked of HH heads and spouses
- Selection bias: college students aren't heads yet!
- Enter my data at age 23; also drop R's who start 90+

Panel Study of Income Dynamics

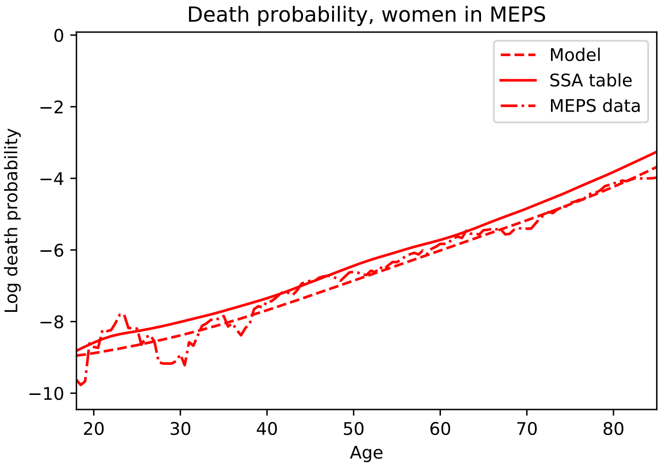
- Began in 1968 with nationally representative sample of HH's
- Snowballing sample (with recent immigrants added)
- Long panel: annual interviews until 1997, biannual since
- Use data from 1997 to 2017 (11 waves, recode to 21 years)
- SRHS **only** asked of HH heads and spouses
- Selection bias: college students aren't heads yet!
- Enter my data at age 23; also drop R's who start 90+
- 11,787 women observed for 69,329 waves (1,012 deaths)
- 10,791 men observed for 58,316 waves (995 deaths)
- 32% observed 2 waves or fewer, but 22% obs all 11 waves

RESULTS

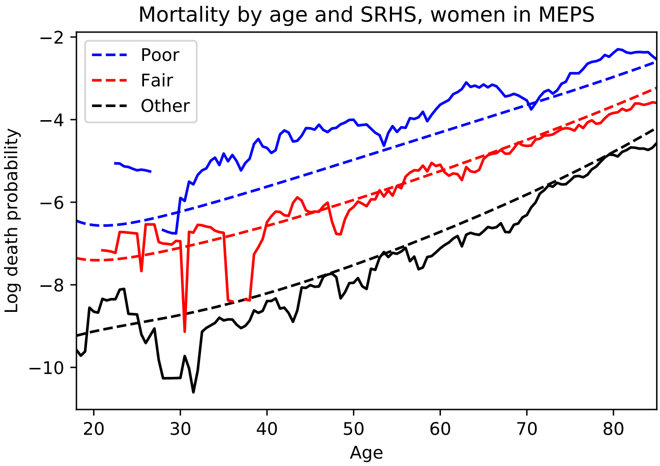
Estimated parameters: MEPS women (mortality)

Param	Description	MEPS
θ_0	Mortality probit: constant term	2.667
		(9.96e-2)
θ_{x1}	Mortality probit: linear coefficient on health	0.116
		(4.65e-3)
θ_{x2}	Mortality probit: quadratic coefficient on health	-5.51e-4
		(2.51e-4)
θ_{j1}	Mortality probit: linear coefficient on age	1.29e-2
		(3.38e-3)
θ_{j2}	Mortality probit: quadratic coefficient on age	-3.43e-4
		(3.19e-5)
θ_{j3}	Mortality probit: cubic coefficient on age	2.80e-7
		(1.10e-7)
θ_{xj}	Mortality probit: coefficient on age$\times$health	1.24e-4
		(3.77e-5)

Model fit: Mortality by age for MEPS women



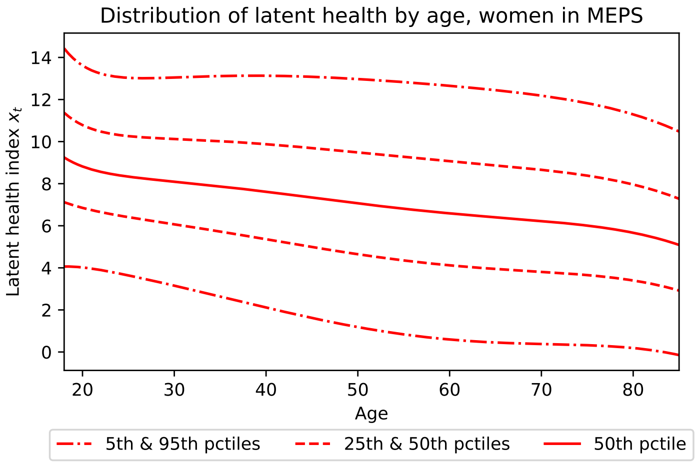
Model fit: Mortality by age and SRHS for MEPS women



Estimated parameters: MEPS women (latent health)

Param	Description	MEPS
β_0	Expected health: constant term	-1.067
		(0.295)
β_{j1}	Expected health: linear coefficient on age	0.974
		(3.19e-2)
β_{j2}	Expected health: quadratic coefficient on age	-3.57e-2
		(1.27e-3)
β_{j3}	Expected health: cubic coefficient on age	5.18e-4
		(2.13e-5)
β_{j4}	Expected health: quartic coefficient on age	-2.68e-6
		(1.25e-7)
μ_0	Mean of latent health at model start	9.383
		(8.80e-2)
σ_0	Standard deviation of latent health at model start	3.160
		(4.47e-2)

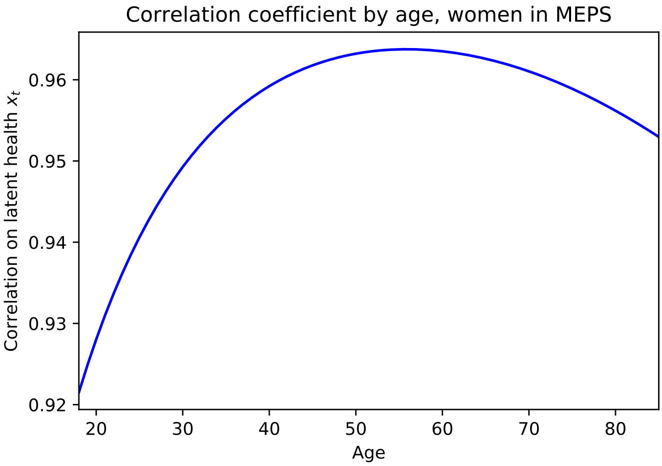
Estimated model: Distribution of latent health (MEPS women)



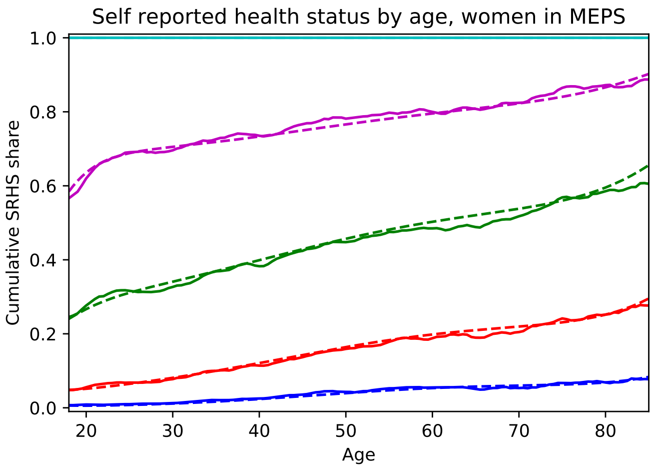
Estimated parameters: MEPS women (correlation and SRHS)

Param	Description	MEPS
γ_0	Correlation factor: constant term	1.262 (9.87e-2)
γ_{j1}	Correlation factor: linear coefficient on age	8.54e-2 (6.71e-3)
γ_{j2}	Correlation factor: quadratic coefficient on age	-1.11e-3 (1.43e-4)
γ_{j3}	Correlation factor: cubic coefficient on age	4.15e-6 (9.65e-7)
α_1	SRHS: Linear coefficient on latent health	0.503 (5.33e-3)
χ_2	SRHS: Cut b/w reporting "fair" and "good"	1.596 (6.90e-3)
χ_3	SRHS: Cut b/w reporting "good" and "very good"	3.389 (9.82e-3)
χ_4	SRHS: Cut b/w reporting "very good" and "excellent"	5.112 (1.28e-2)

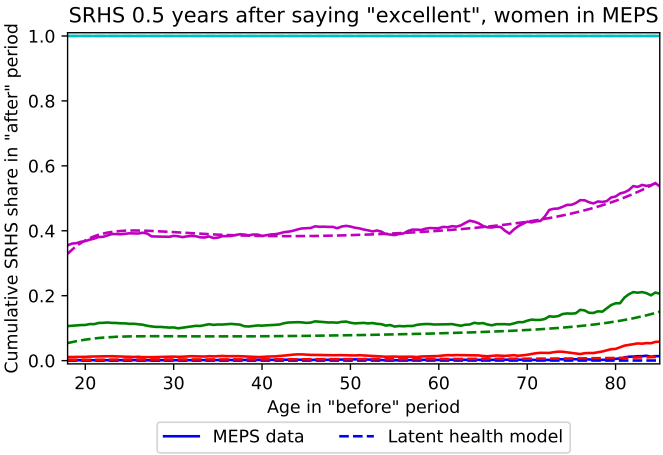
Estimated model: Latent health serial correlation (MEPS women)



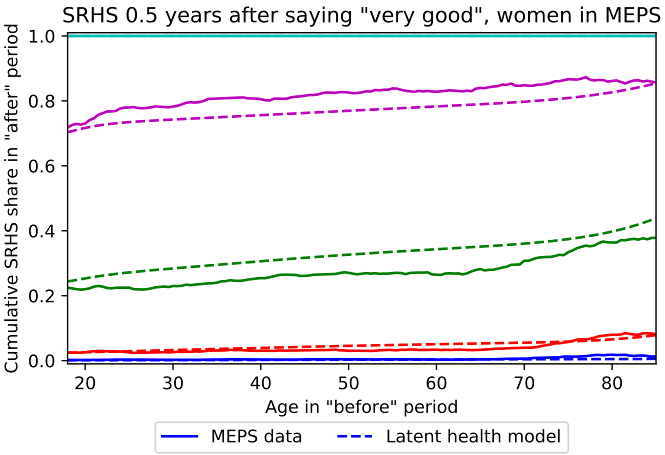
Model fit: SRHS by age for MEPS women



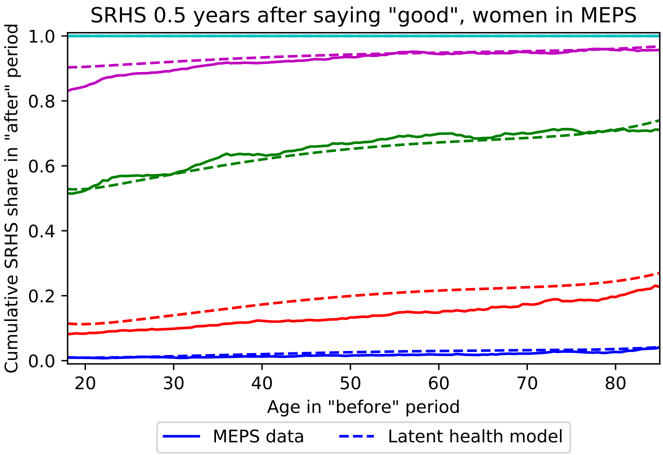
Model fit: SRHS transitions for MEPS women: 1 wave



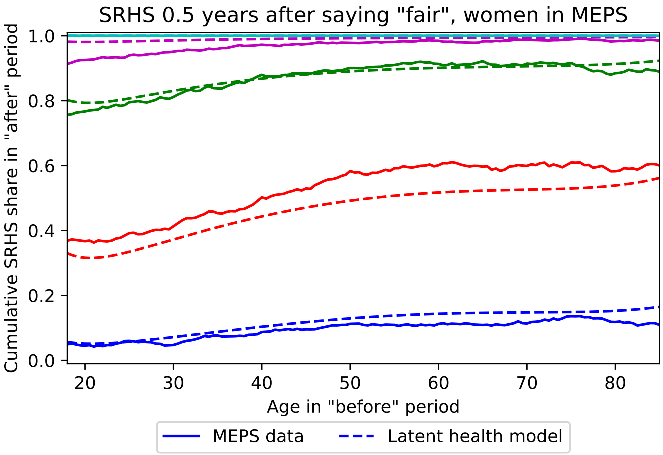
Model fit: SRHS transitions for MEPS women: 1 wave



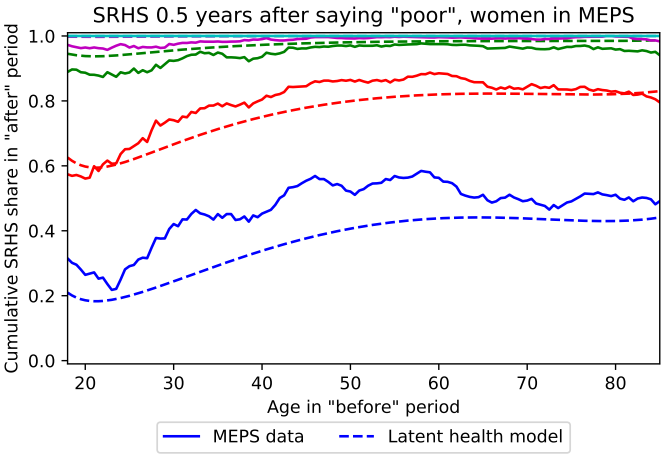
Model fit: SRHS transitions for MEPS women: 1 wave



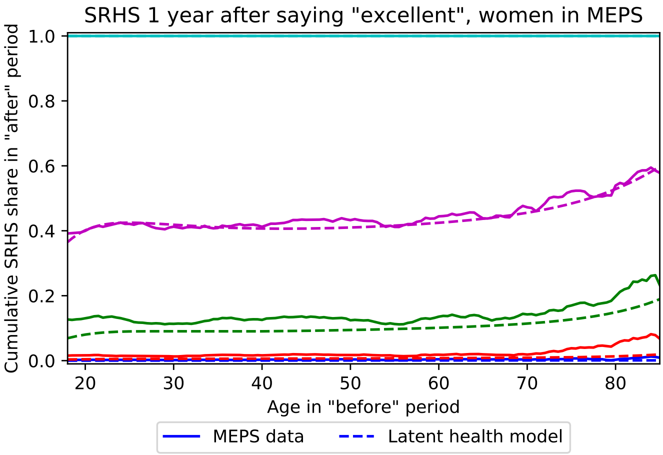
Model fit: SRHS transitions for MEPS women: 1 wave



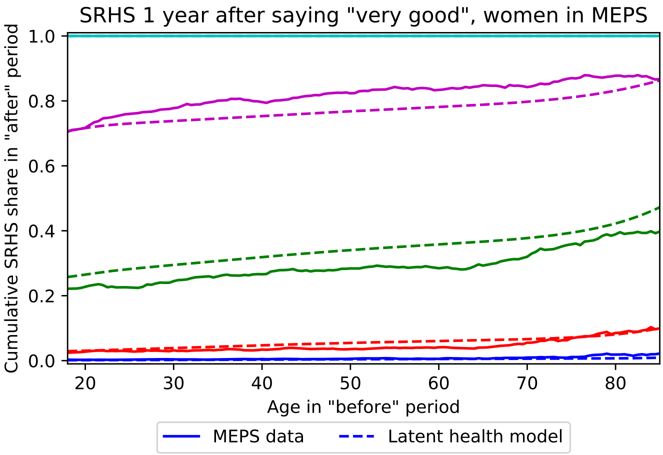
Model fit: SRHS transitions for MEPS women: 1 wave



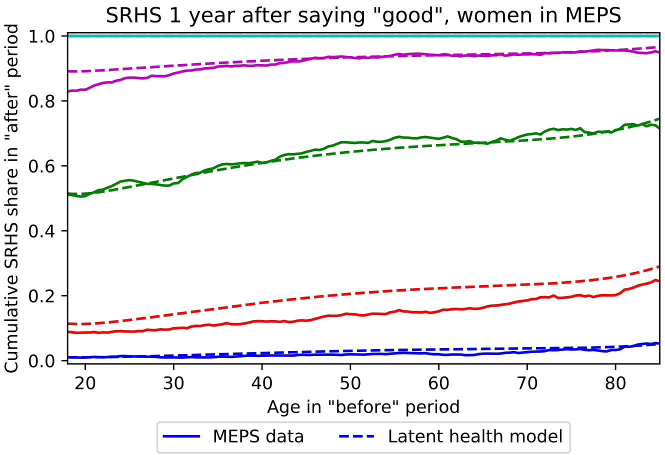
Model fit: SRHS transitions for MEPS women: 2 waves



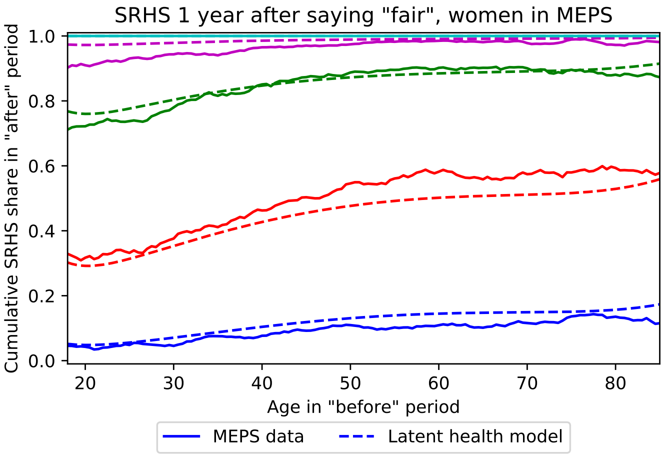
Model fit: SRHS transitions for MEPS women: 2 waves



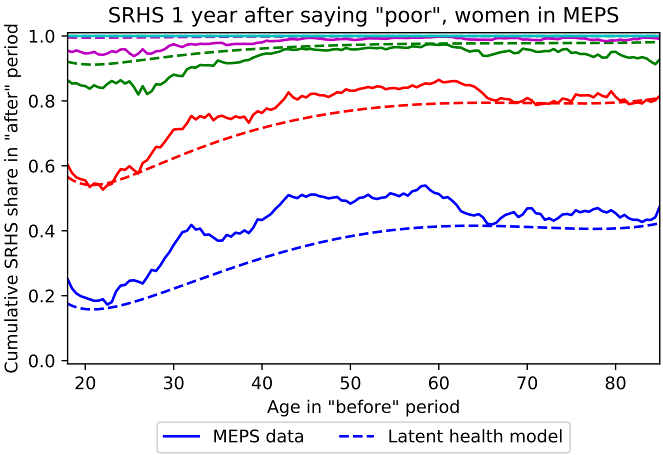
Model fit: SRHS transitions for MEPS women: 2 waves



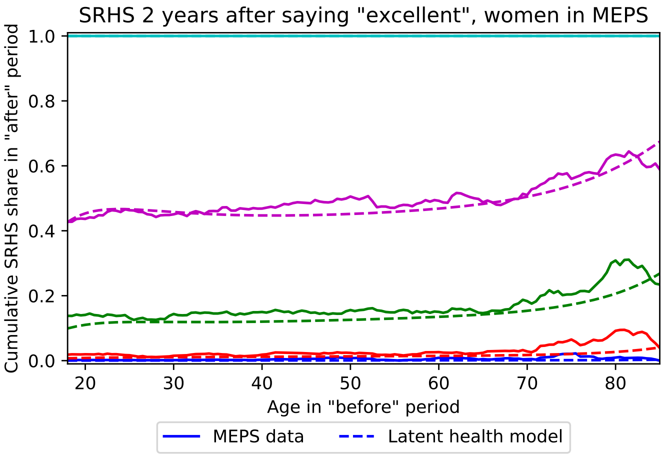
Model fit: SRHS transitions for MEPS women: 2 waves



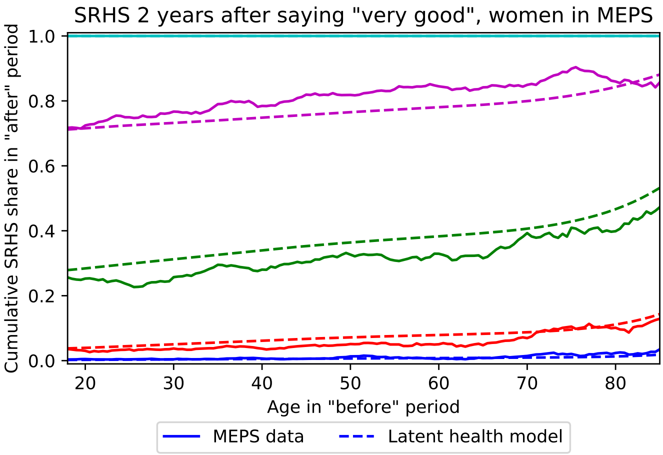
Model fit: SRHS transitions for MEPS women: 2 waves



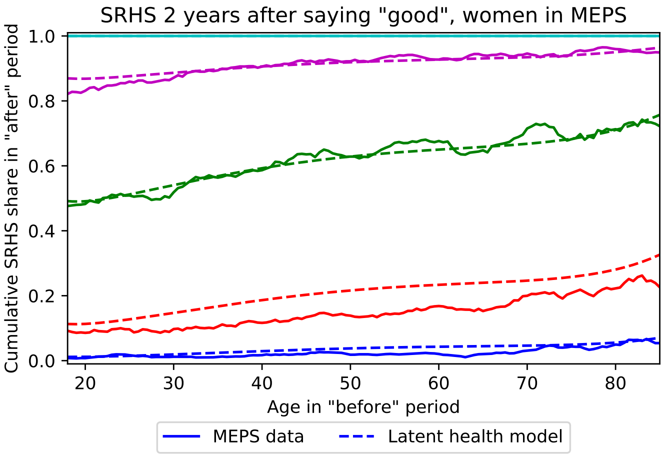
Model fit: SRHS transitions for MEPS women: 4 waves



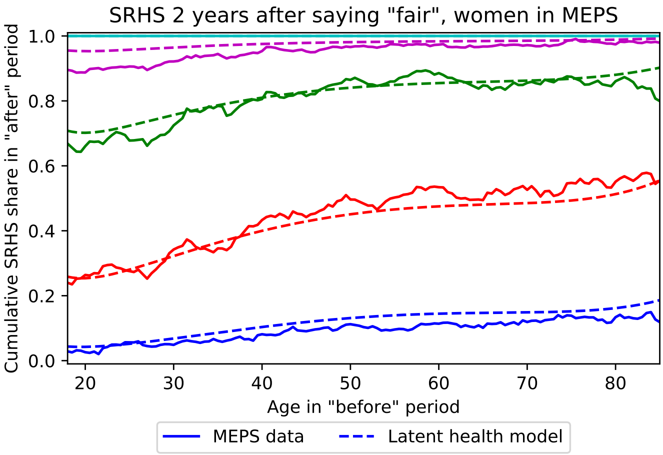
Model fit: SRHS transitions for MEPS women: 4 waves



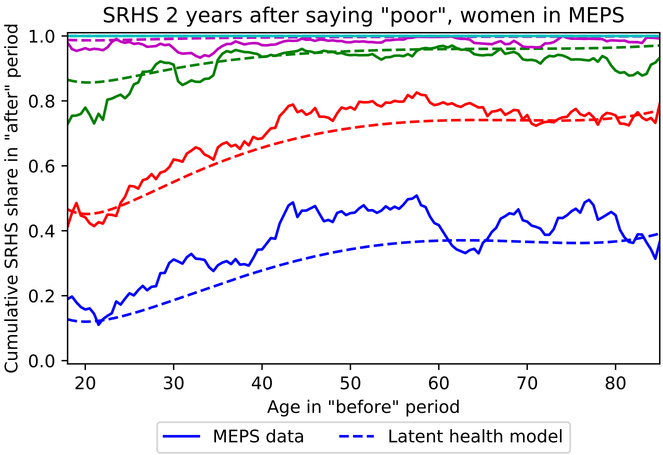
Model fit: SRHS transitions for MEPS women: 4 waves



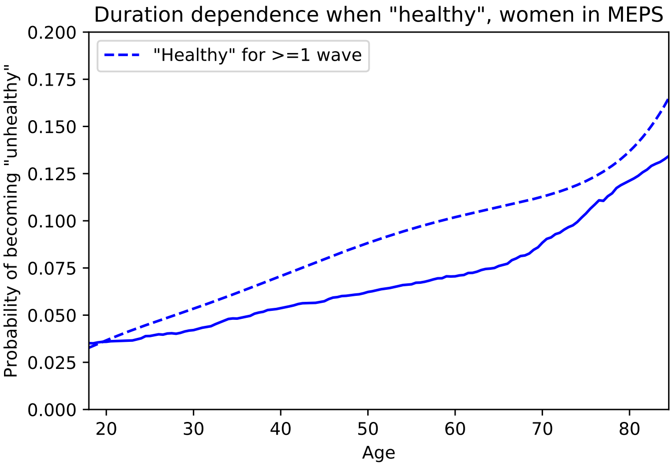
Model fit: SRHS transitions for MEPS women: 4 waves



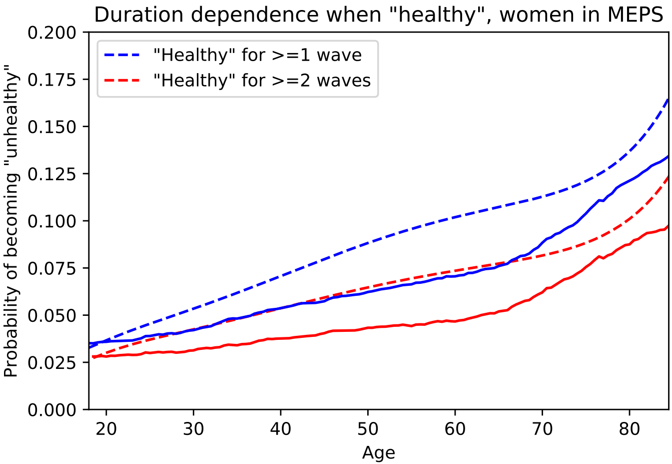
Model fit: SRHS transitions for MEPS women: 4 waves



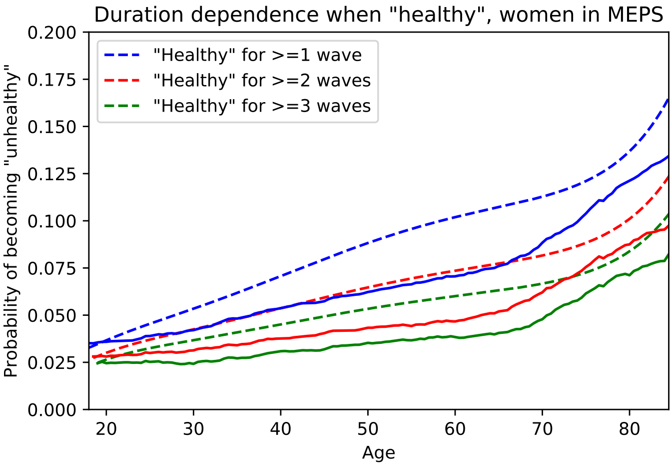
Model fit: Duration dependence, “healthy” to “unhealthy”



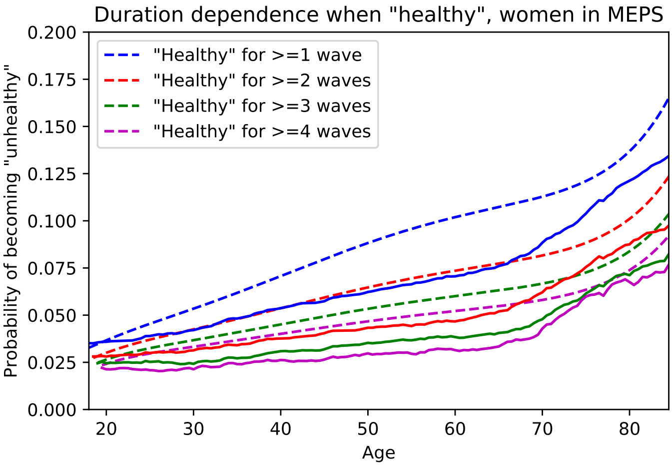
Model fit: Duration dependence, “healthy” to “unhealthy”



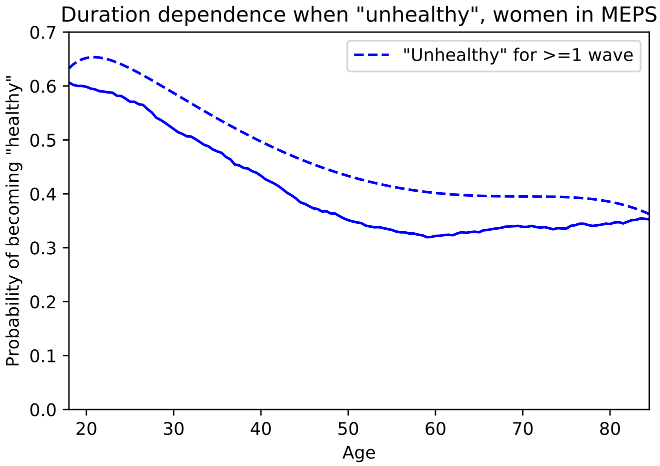
Model fit: Duration dependence, “healthy” to “unhealthy”



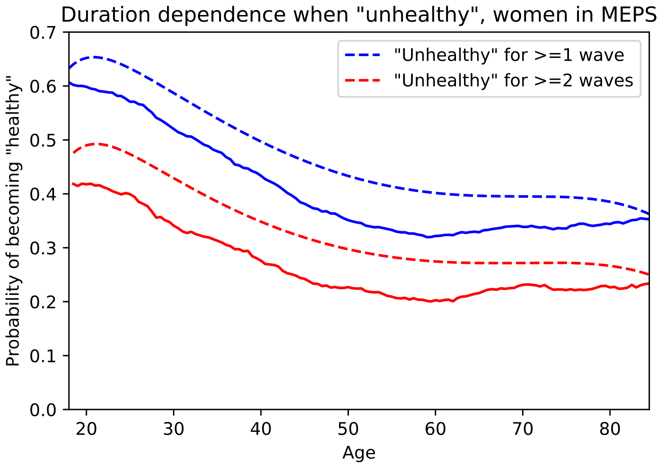
Model fit: Duration dependence, “healthy” to “unhealthy”



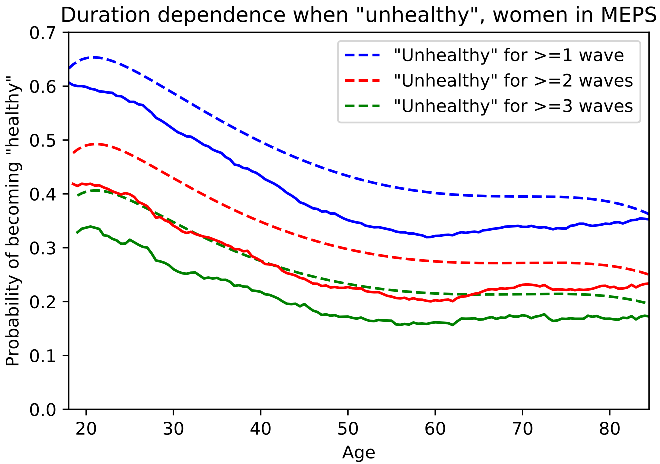
Model fit: Duration dependence, “unhealthy” to “healthy”



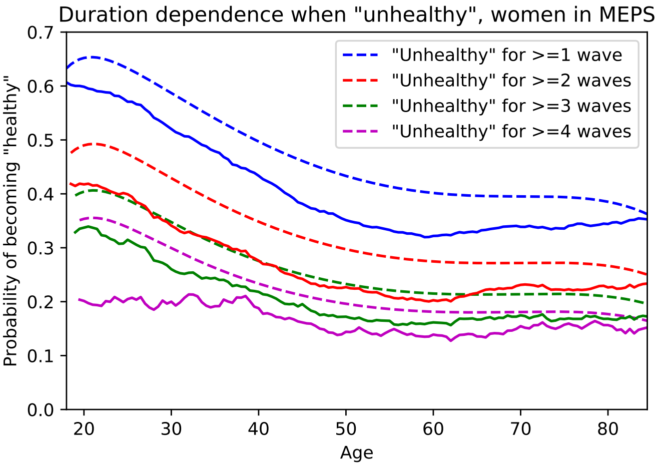
Model fit: Duration dependence, “unhealthy” to “healthy”



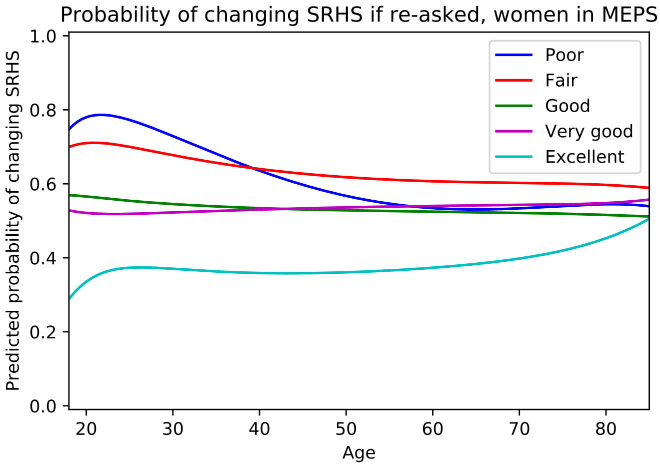
Model fit: Duration dependence, “unhealthy” to “healthy”



Model fit: Duration dependence, “unhealthy” to “healthy”



Estimated model: Changing SRHS if asked twice (MEPS women)



Estimated model: Predicting (log) medical expenses

- SRHS is known to be predictive of (log) medical costs
- How does model-predicted latent health x_t do?

Estimated model: Predicting (log) medical expenses

- SRHS is known to be predictive of (log) medical costs
- How does model-predicted latent health x_t do?
- For all 125 possible sequences of SRHS in waves 1-3 at each age, use estimated model to calculate $\hat{x}_t = \mathrm{E}[x_t]$ and $\widehat{x^2}_t = \mathrm{E}[x_t^2]$
- Get total medical expenses between waves 3 and 5 (1 year)

Estimated model: Predicting (log) medical expenses

- SRHS is known to be predictive of (log) medical costs
- How does model-predicted latent health x_t do?
- For all 125 possible sequences of SRHS in waves 1-3 at each age, use estimated model to calculate $\hat{x}_t = \mathrm{E}[x_t]$ and $\widehat{x^2}_t = \mathrm{E}[x_t^2]$
- Get total medical expenses between waves 3 and 5 (1 year)
- Regress $\log(\mathit{TotalMed})$ on cubic in age, $\hat{x}_t$, and $\widehat{x^2}_t$: $R^2 = 0.1600$
- Regress $\log(\mathit{TotalMed})$ on **full interaction** of age and SRHS: $R^2 = 0.1602$

Estimated model: Predicting (log) medical expenses

- SRHS is known to be predictive of (log) medical costs
- How does model-predicted latent health x_t do?
- For all 125 possible sequences of SRHS in waves 1-3 at each age, use estimated model to calculate $\hat{x}_t = \mathrm{E}[x_t]$ and $\widehat{x^2}_t = \mathrm{E}[x_t^2]$
- Get total medical expenses between waves 3 and 5 (1 year)
- Regress $\log(TotalMed)$ on cubic in age, $\hat{x}_t$, and $\widehat{x^2}_t$: $R^2 = 0.1600$
- Regress $\log(TotalMed)$ on **full interaction** of age and SRHS: $R^2 = 0.1602$
- Same exercise for men: $R^2 = 0.2326$ (latent health) vs $R^2 = 0.2326$ (full interact)

Estimated model: Out-of-sample fit

- Also estimated model on PSID and HRS, men and women; all fit well
- Model was estimated on **short panel** MEPS data and fit it well...
- ...But does its long term predictions fit **long panel** datasets?

Estimated model: Out-of-sample fit

- Also estimated model on PSID and HRS, men and women; all fit well
- Model was estimated on **short panel** MEPS data and fit it well...
- ...But does its long term predictions fit **long panel** datasets?
- PSID and HRS ask for SRHS every 2 years over same timeframe
- Can project estimated model out to see fit of 4-20 year ahead transitions

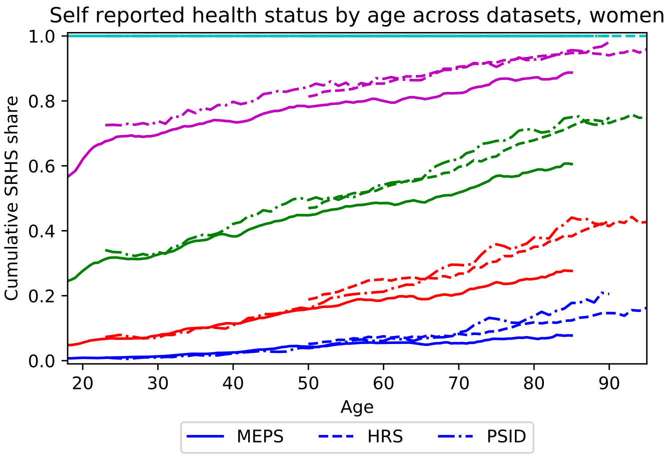
Estimated model: Out-of-sample fit

- Also estimated model on PSID and HRS, men and women; all fit well
- Model was estimated on **short panel** MEPS data and fit it well...
- ...But does its long term predictions fit **long panel** datasets?
- PSID and HRS ask for SRHS every 2 years over same timeframe
- Can project estimated model out to see fit of 4-20 year ahead transitions
- Result: **No, it absolutely does not fit**

Estimated model: Out-of-sample fit

- Also estimated model on PSID and HRS, men and women; all fit well
- Model was estimated on **short panel** MEPS data and fit it well...
- ...But does its long term predictions fit **long panel** datasets?
- PSID and HRS ask for SRHS every 2 years over same timeframe
- Can project estimated model out to see fit of 4-20 year ahead transitions
- Result: **No, it absolutely does not fit**
- Probably too much of a stretch to project a model 10x beyond data
- MEPS has a selection bias problem: too few respondents in poor health
- And asks respondents to compare to people of their own age

SRHS across datasets: MEPS, HRS, and PSID (women)



Cross-dataset sample construction

- Can fix two of those problems (but selection bias remains)

Cross-dataset sample construction

- Can fix two of those problems (but selection bias remains)
- Can create merged dataset across MEPS-SAQ, HRS, and PSID
- MEPS-SAQ: Mail-in supplemental survey that includes SRHS (waves 2 & 4)
- Use 1 year periods for model: 1 period transitions only seen in MEPS!

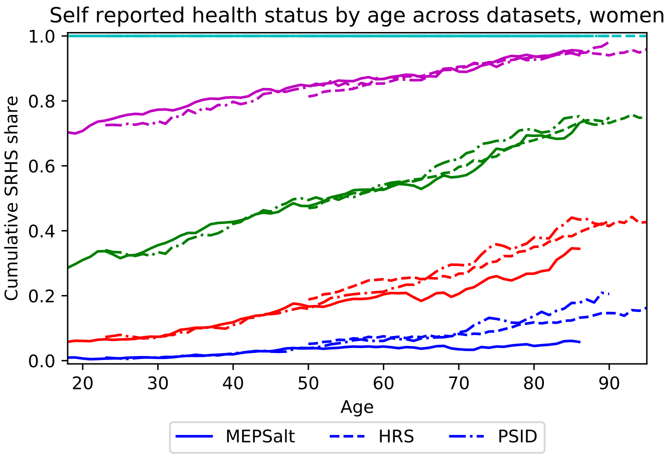
Cross-dataset sample construction

- Can fix two of those problems (but selection bias remains)
- Can create merged dataset across MEPS-SAQ, HRS, and PSID
- MEPS-SAQ: Mail-in supplemental survey that includes SRHS (waves 2 & 4)
- Use 1 year periods for model: 1 period transitions only seen in MEPS!
- Short run transitions from MEPS, long run paths from HRS & PSID
- Model starts at age 18; drop R's who start over age 90 (84 for MEPS)
- Only admit PSID respondents when they turn 23

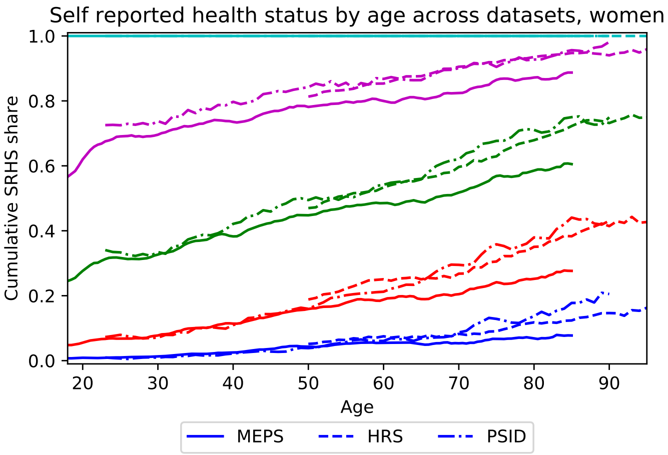
Cross-dataset sample construction

- Can fix two of those problems (but selection bias remains)
- Can create merged dataset across MEPS-SAQ, HRS, and PSID
- MEPS-SAQ: Mail-in supplemental survey that includes SRHS (waves 2 & 4)
- Use 1 year periods for model: 1 period transitions only seen in MEPS!
- Short run transitions from MEPS, long run paths from HRS & PSID
- Model starts at age 18; drop R's who start over age 90 (84 for MEPS)
- Only admit PSID respondents when they turn 23
- 120,055 women observed for 342,796 waves (6,918 deaths)
- 102,965 men observed for 281,681 waves (6,568 deaths)
- 82% of respondents observed 2 or fewer waves (mostly MEPS)

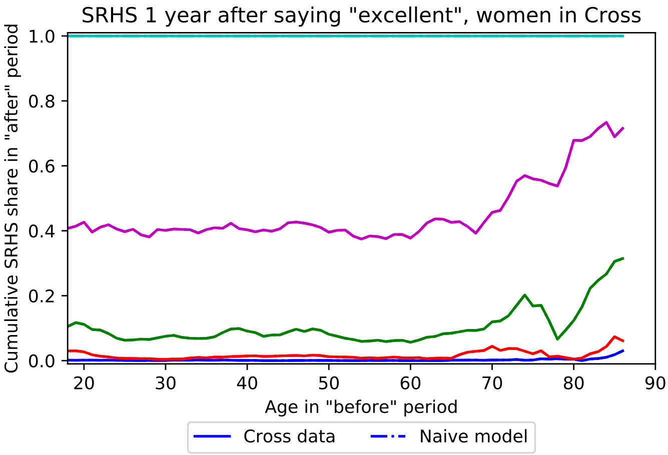
SRHS across datasets: MEPS-SAQ, HRS, and PSID (women)



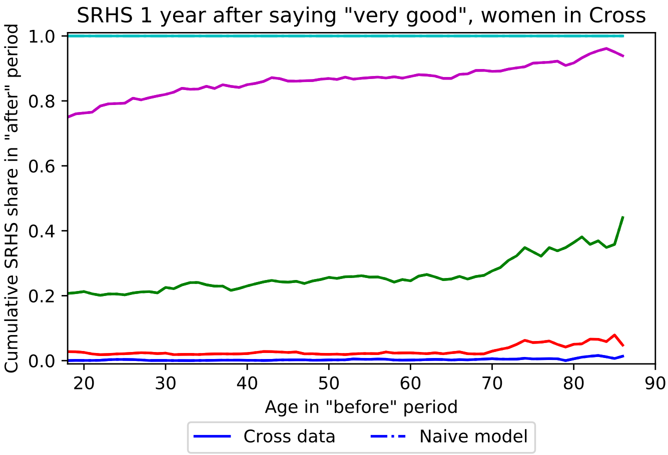
SRHS across datasets: MEPS, HRS, and PSID (women)



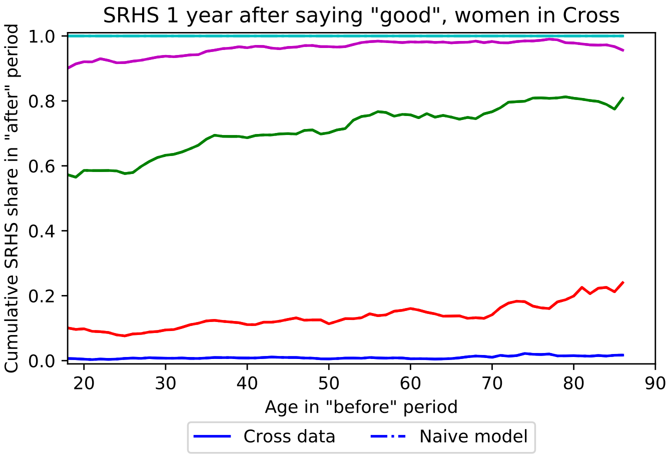
Naive SRHS transitions for cross-data women: 1 year



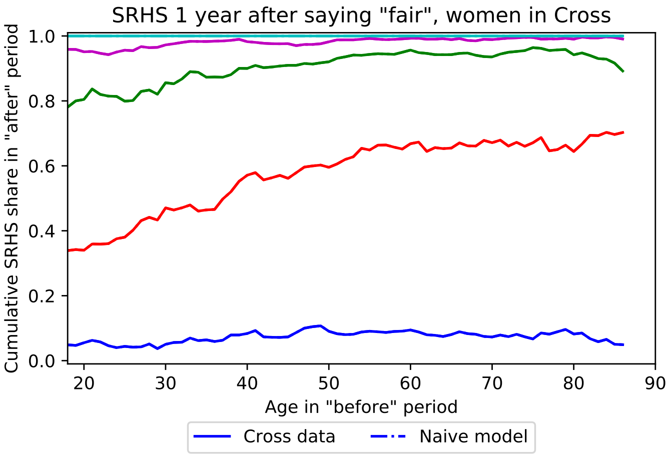
Naive SRHS transitions for cross-data women: 1 year



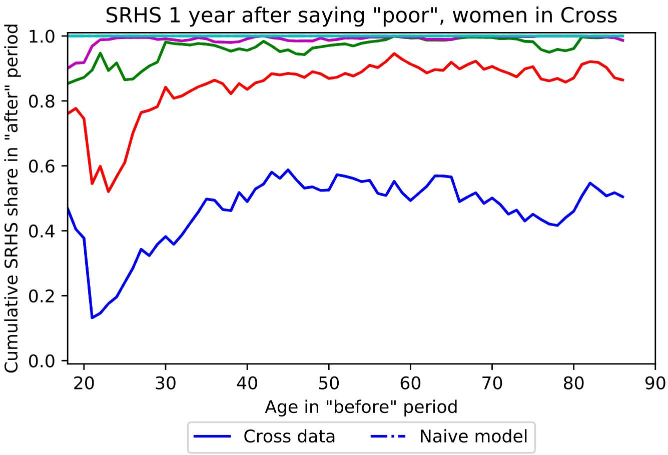
Naive SRHS transitions for cross-data women: 1 year



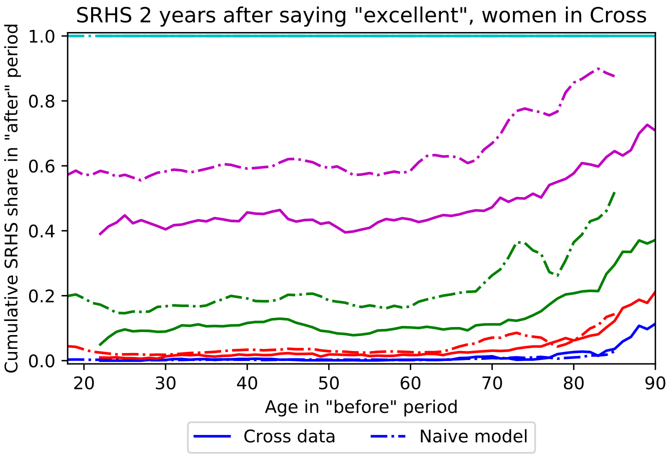
Naive SRHS transitions for cross-data women: 1 year



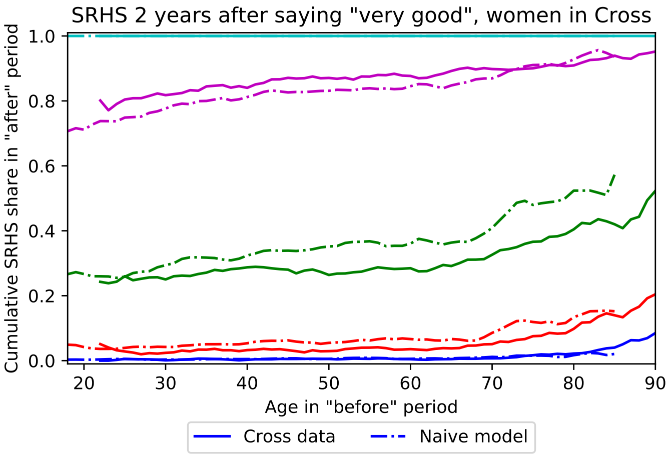
Naive SRHS transitions for cross-data women: 1 year



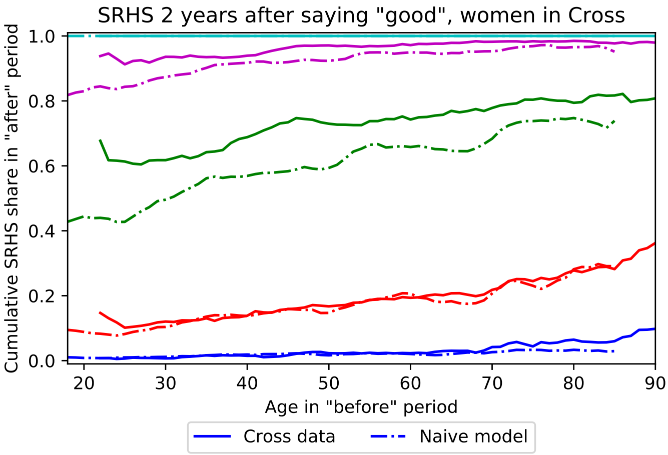
Naive SRHS transitions for cross-data women: 2 years



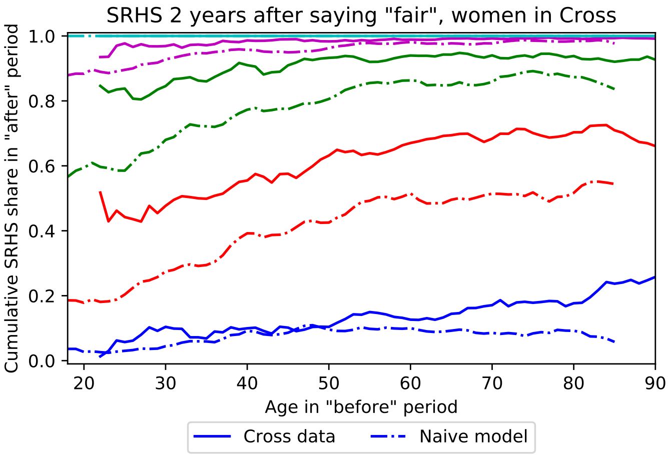
Naive SRHS transitions for cross-data women: 2 years



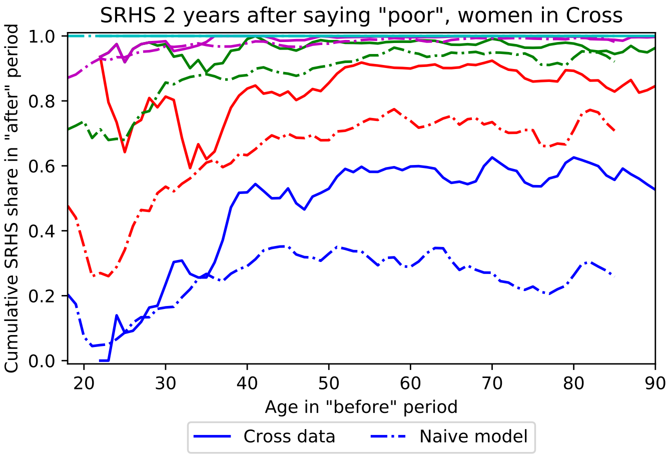
Naive SRHS transitions for cross-data women: 2 years



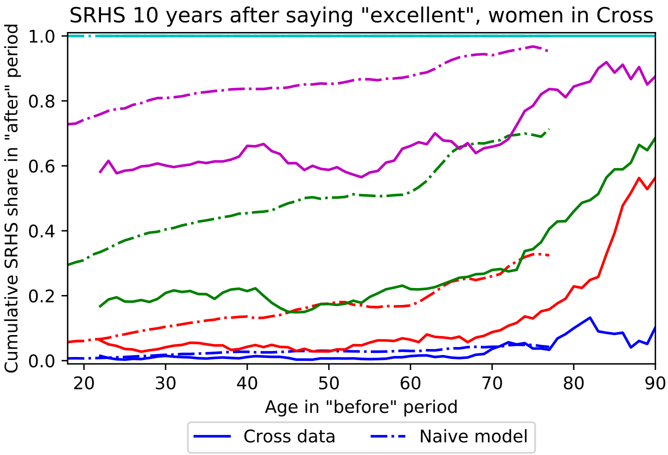
Naive SRHS transitions for cross-data women: 2 years



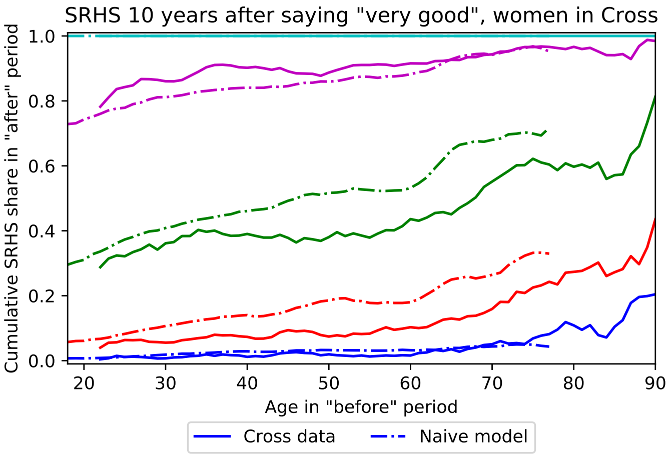
Naive SRHS transitions for cross-data women: 2 years



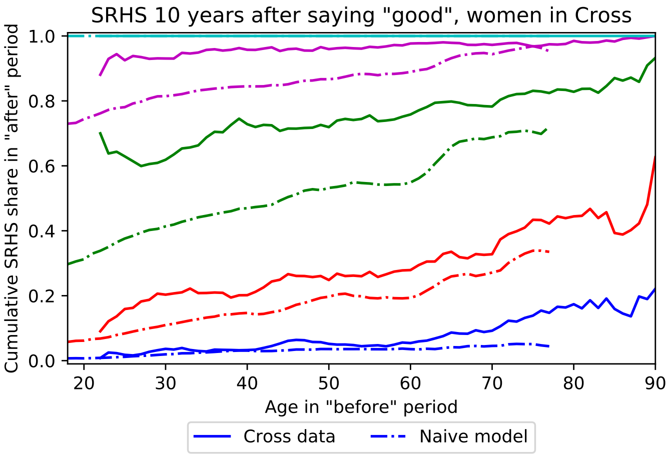
Naive SRHS transitions for cross-data women: 10 years



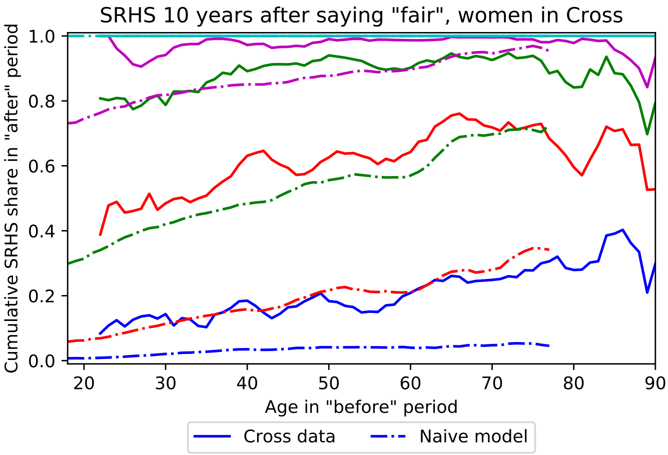
Naive SRHS transitions for cross-data women: 10 years



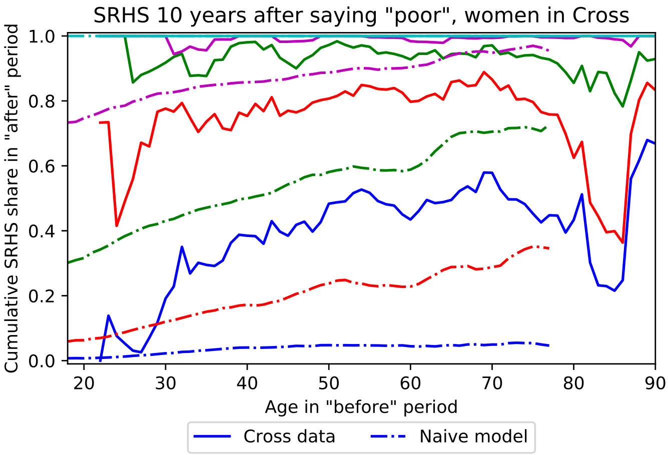
Naive SRHS transitions for cross-data women: 10 years



Naive SRHS transitions for cross-data women: 10 years



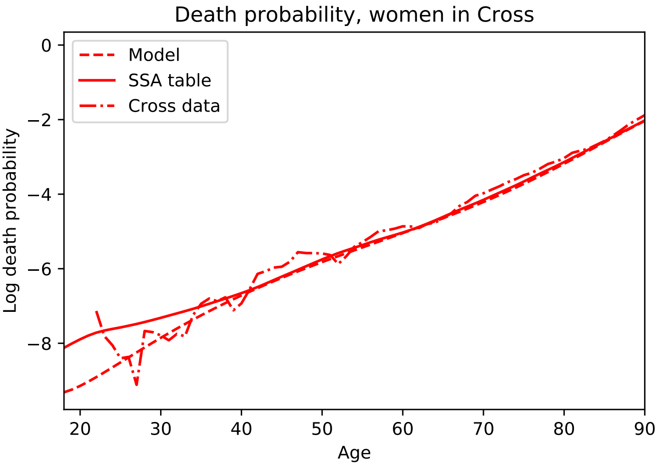
Naive SRHS transitions for cross-data women: 10 years



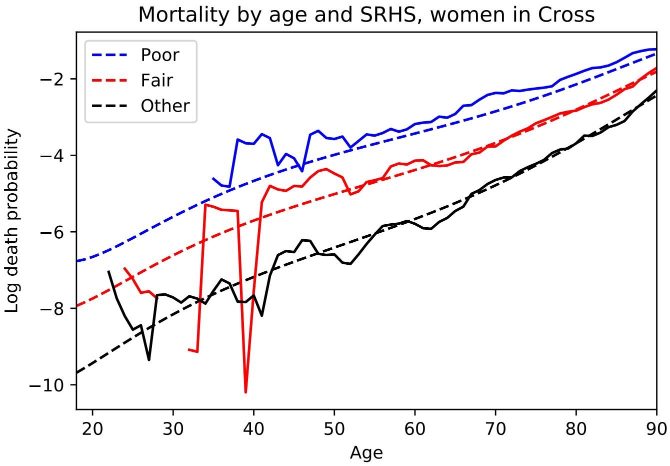
Estimated parameters: Cross-data women (mortality)

Param	Description	Cross
θ_0	Mortality probit: constant term	3.622
		(0.353)
θ_{x1}	Mortality probit: linear coefficient on health	0.141
		(4.96e-3)
θ_{x2}	Mortality probit: quadratic coefficient on health	-3.88e-3
		(4.66e-4)
θ_{j1}	Mortality probit: linear coefficient on age	-5.70e-2
		(1.74e-2)
θ_{j2}	Mortality probit: quadratic coefficient on age	8.86e-4
		(2.75e-4)
θ_{j3}	Mortality probit: cubic coefficient on age	-6.85e-6
		(1.41e-6)
θ_{xj}	Mortality probit: coefficient on age$\times$health	-2.90e-5
		(2.69e-5)

Model fit: Mortality by age for cross-data women



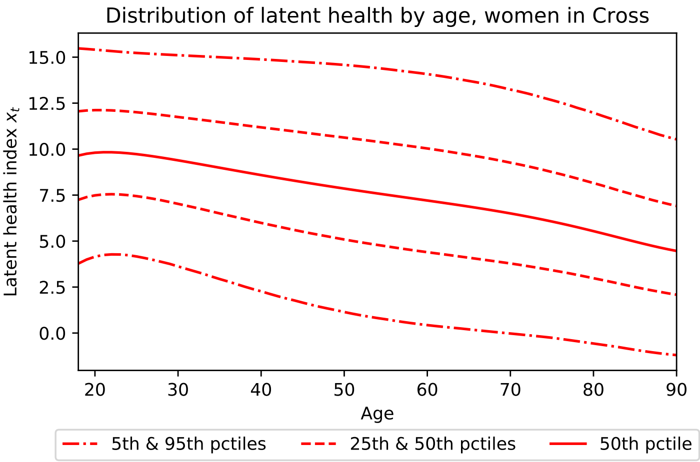
Model fit: Mortality by age-SRHS for cross-data women



Estimated parameters: Cross-data women (health)

Param	Description	Cross
β_0	Expected health: constant term	21.369
		(1.317)
β_{j1}	Expected health: linear coefficient on age	-0.752
		(9.78e-2)
β_{j2}	Expected health: quadratic coefficient on age	9.25e-3
		(2.62e-3)
β_{j3}	Expected health: cubic coefficient on age	2.98e-5
		(3.37e-5)
β_{j4}	Expected health: quartic coefficient on age	-9.16e-7
		(1.86e-7)
μ_0	Mean of latent health at model start	9.875
		(0.110)
σ_0	Standard deviation of latent health at model start	3.563
		(6.36e-2)

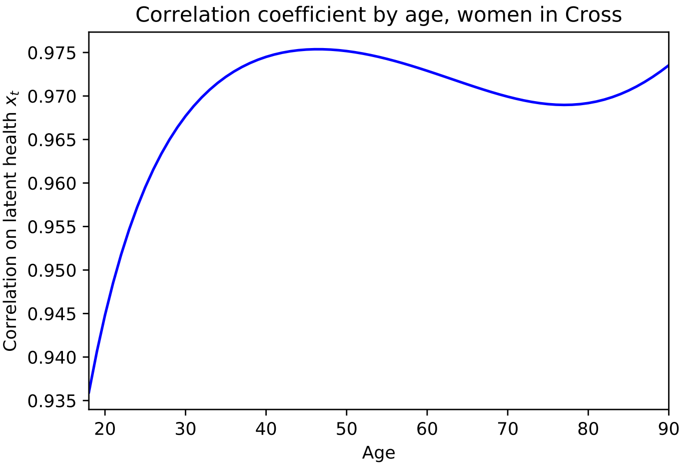
Estimated model: Distribution of latent health (cross-data women)



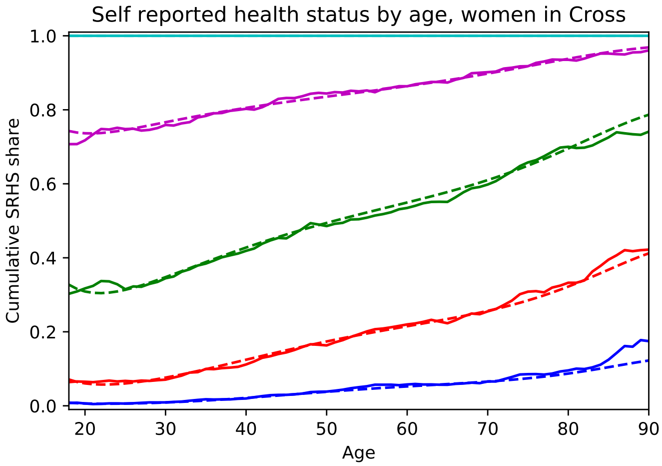
Estimated parameters: Cross-data women (correlation and SRHS)

Param	Description	Cross
γ_0	Correlation factor: constant term	0.337 (0.186)
γ_{j1}	Correlation factor: linear coefficient on age	0.178 (1.21e-2)
γ_{j2}	Correlation factor: quadratic coefficient on age	-3.08e-3 (2.41e-4)
γ_{j3}	Correlation factor: cubic coefficient on age	1.67e-5 (1.52e-6)
α_1	SRHS: Linear coefficient on latent health	0.480 (5.23e-3)
χ_2	SRHS: Cut b/w reporting "fair" and "good"	1.805 (9.61e-3)
χ_3	SRHS: Cut b/w reporting "good" and "very good"	3.814 (1.41e-2)
χ_4	SRHS: Cut b/w reporting "very good" and "excellent"	5.974 (1.95e-2)

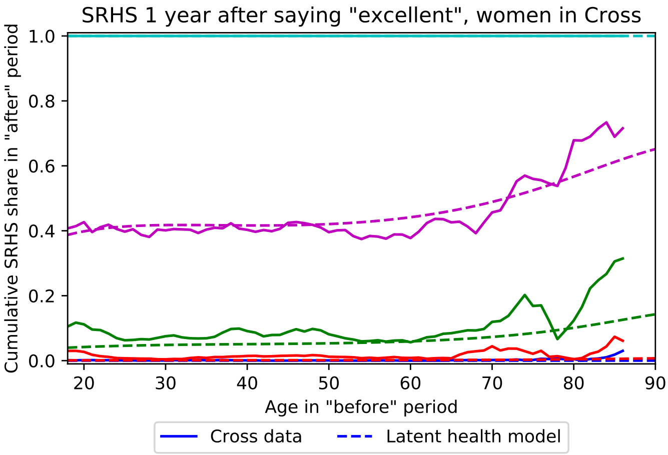
Estimated model: Latent health serial correlation (cross-dataset women)



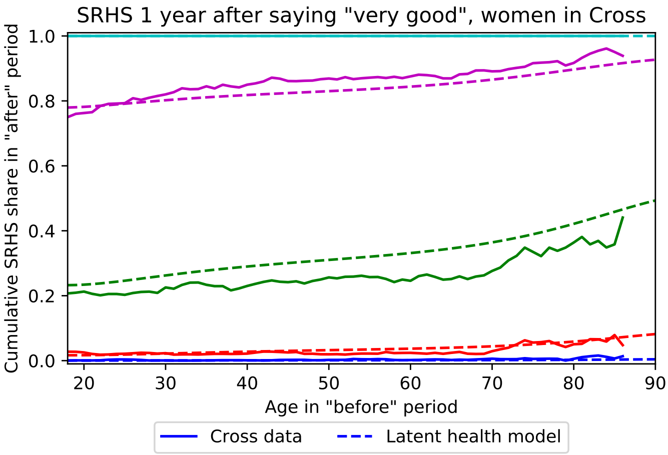
Model fit: SRHS by age for cross-data women



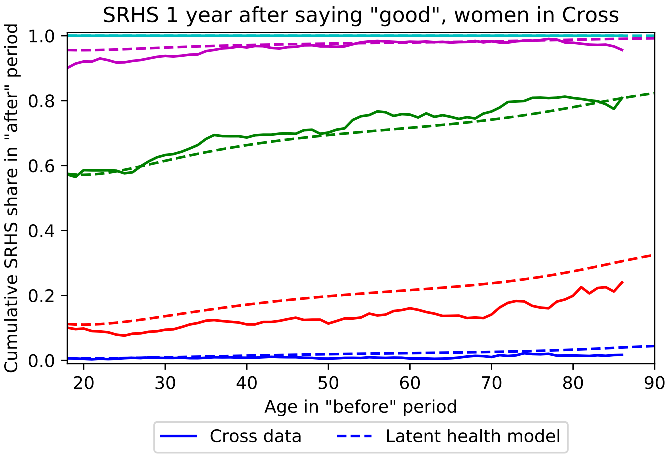
Model fit: SRHS transitions for cross-data women: 1 year



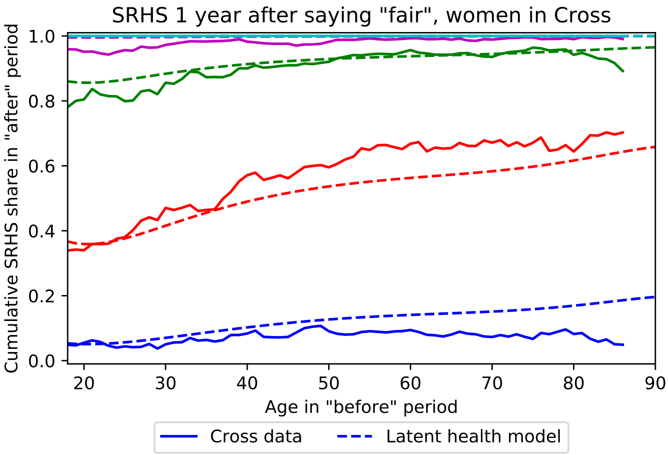
Model fit: SRHS transitions for cross-data women: 1 year



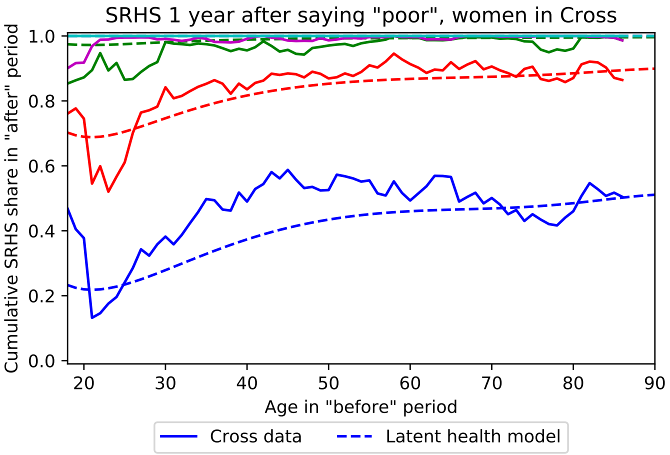
Model fit:e SRHS transitions for cross-data women: 1 year



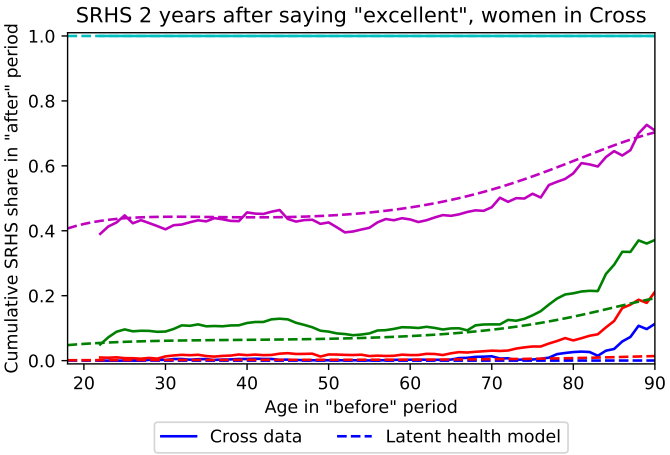
Model fit: SRHS transitions for cross-data women: 1 year



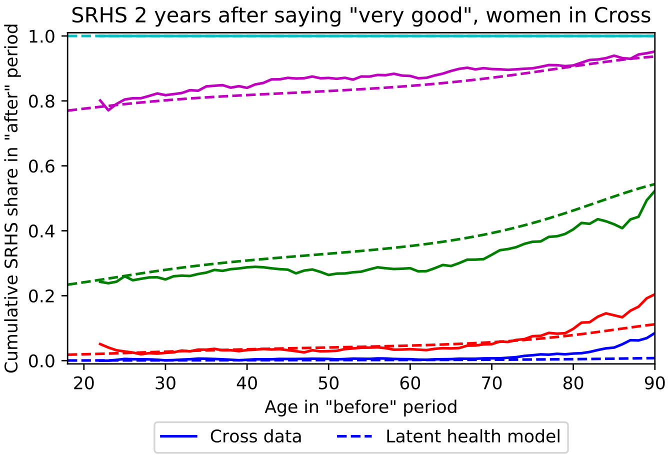
Model fit: SRHS transitions for cross-data women: 1 year



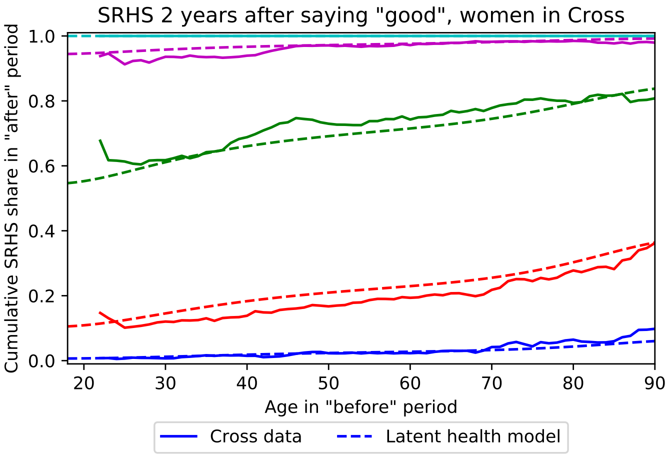
Model fit: SRHS transitions for cross-data women: 2 years



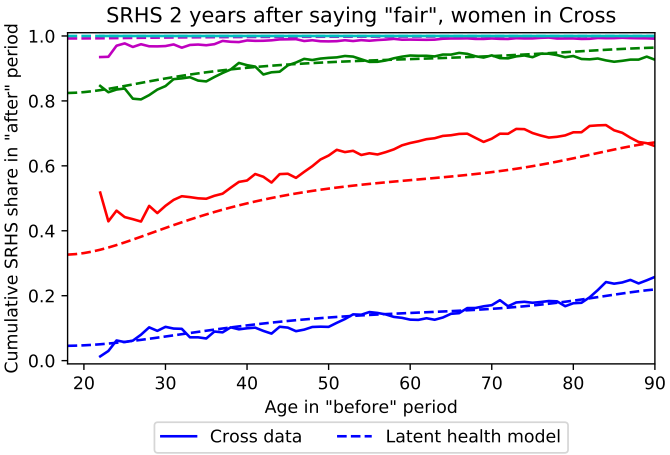
Model fit: SRHS transitions for cross-data women: 2 years



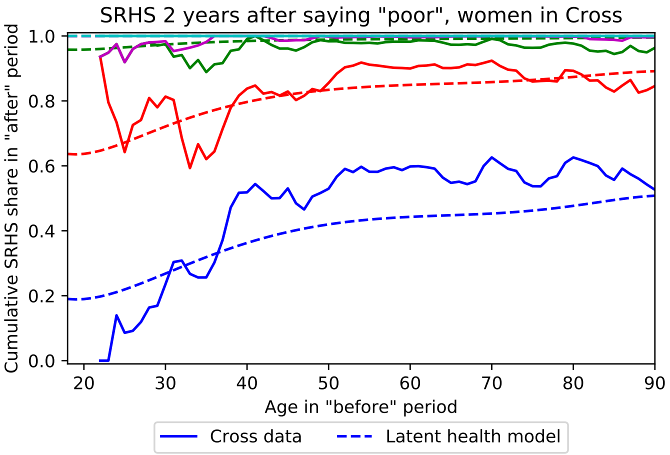
Model fit: SRHS transitions for cross-data women: 2 years



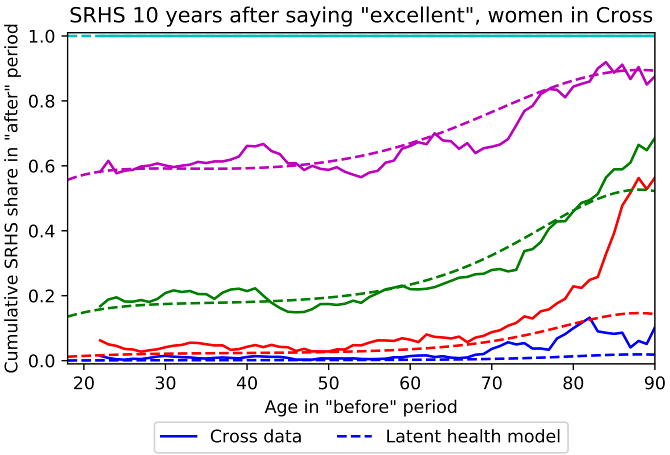
Model fit: SRHS transitions for cross-data women: 2 years



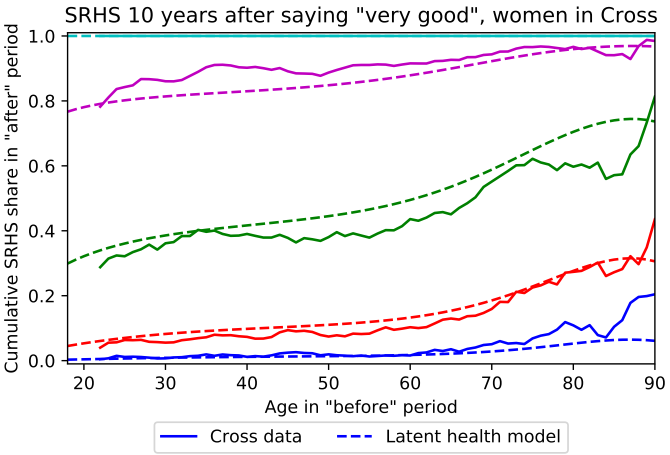
Model fit: SRHS transitions for cross-data women: 2 years



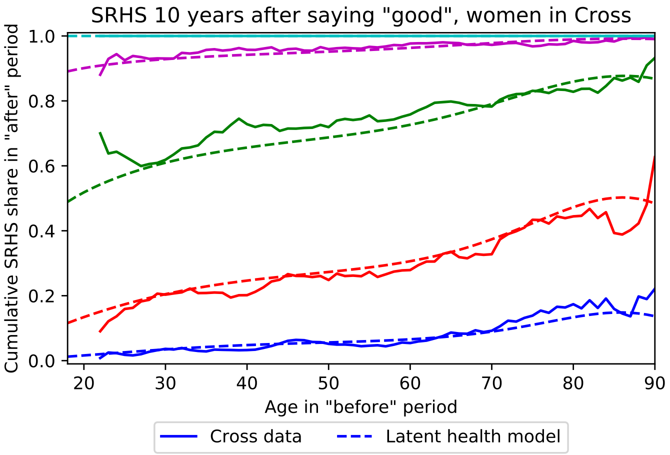
Model fit: SRHS transitions for cross-data women: 10 yrs



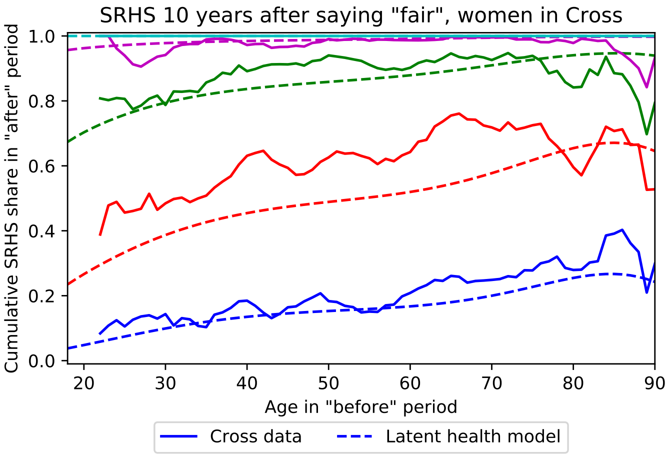
Model fit: SRHS transitions for cross-data women: 10 yrs



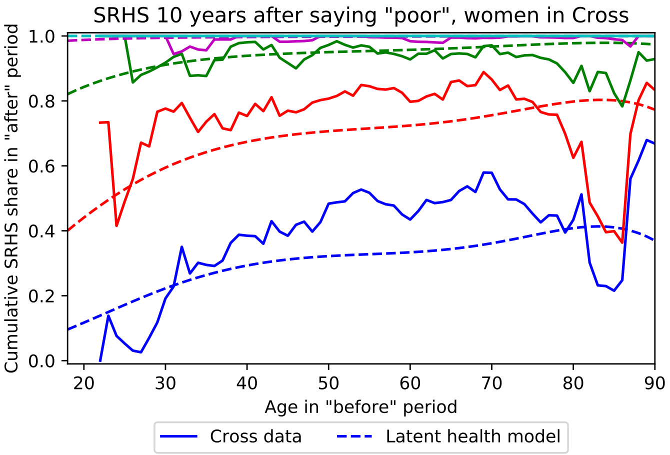
Model fit: SRHS transitions for cross-data women: 10 yrs



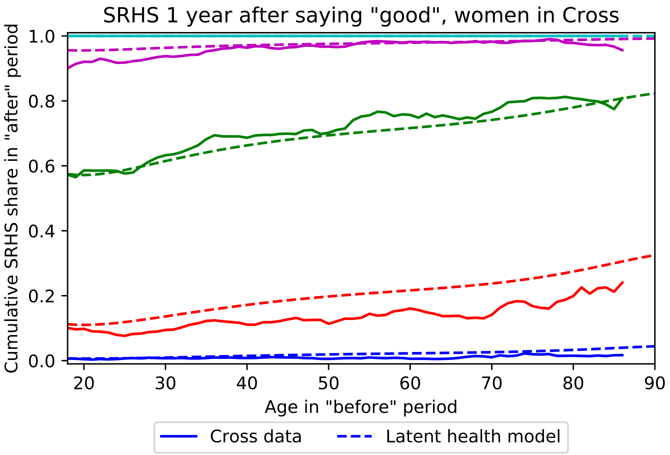
Model fit: SRHS transitions for cross-data women: 10 yrs



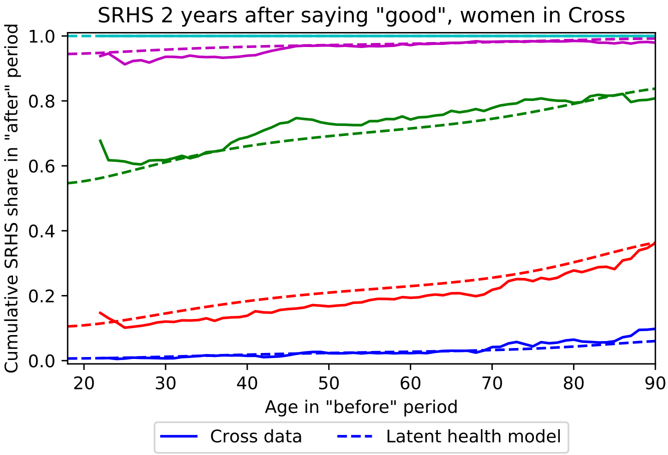
Model fit: SRHS transitions for cross-data women: 10 yrs



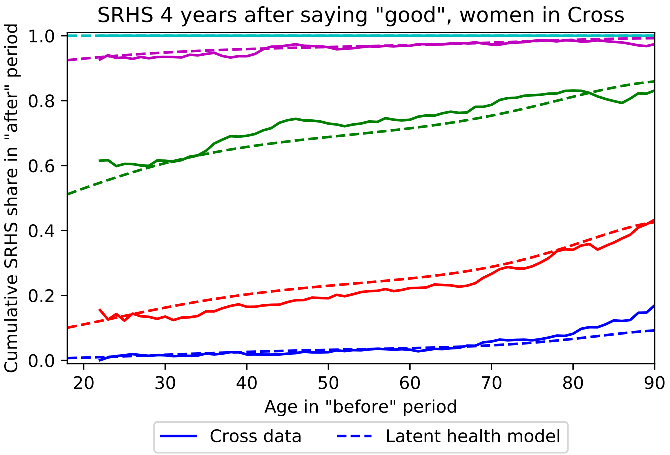
Model fit: SRHS transitions for cross-data women: 1 year



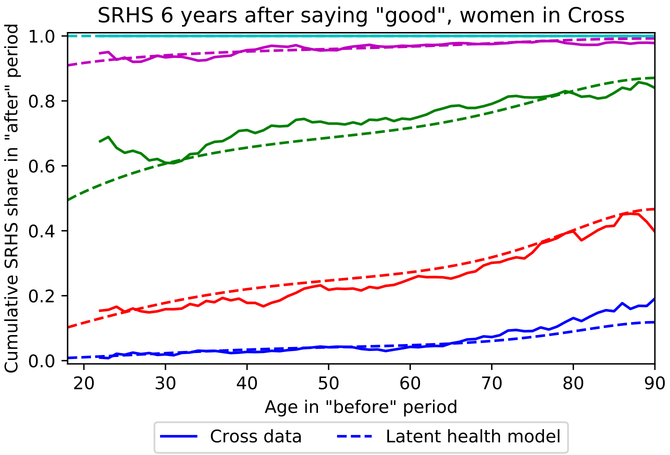
Model fit: SRHS transitions for cross-data women: 2 years



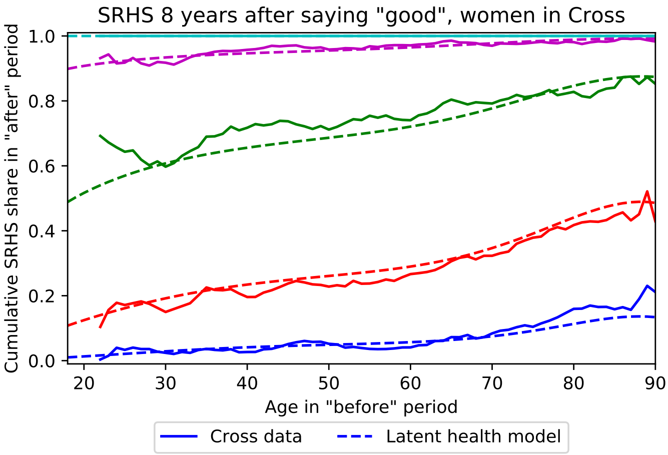
Model fit: SRHS transitions for cross-data women: 4 years



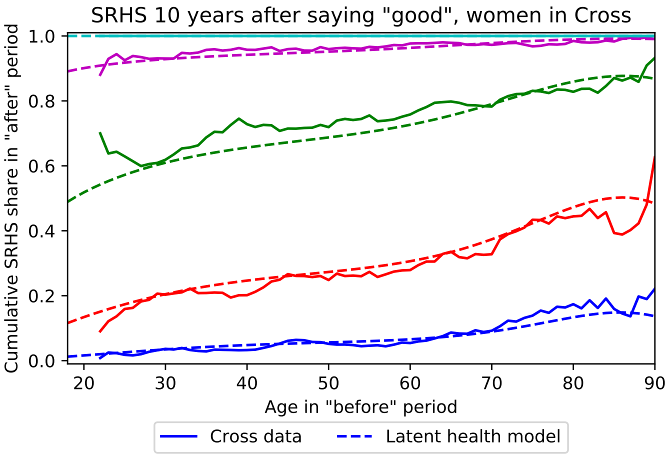
Model fit: SRHS transitions for cross-data women: 6 years



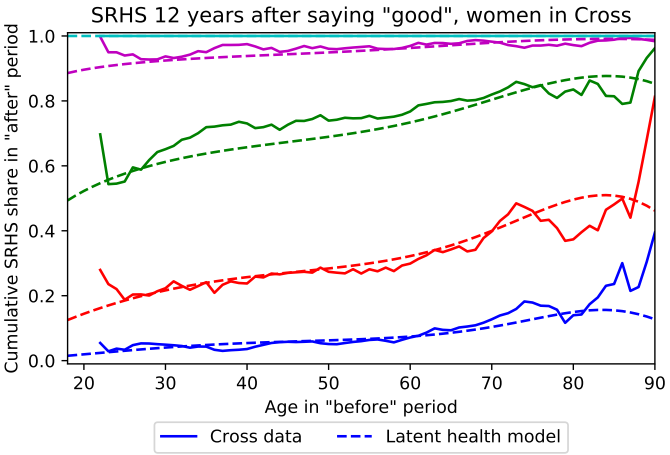
Model fit: SRHS transitions for cross-data women: 8 years



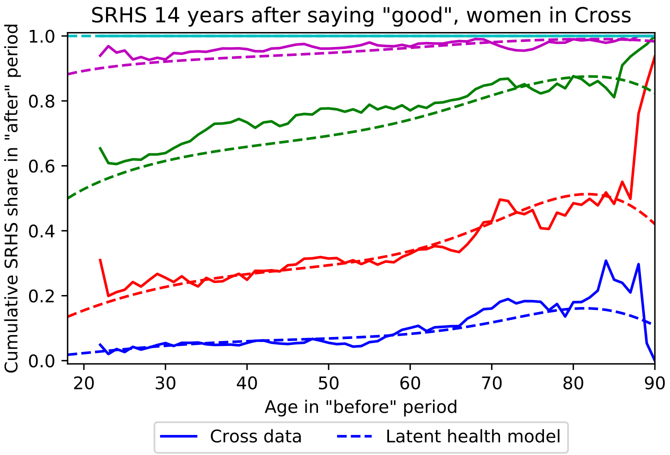
Model fit: SRHS transitions for cross-data women: 10 yrs



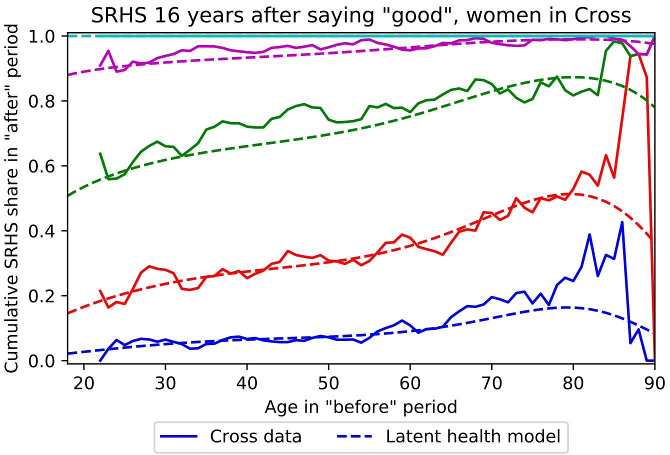
Model fit: SRHS transitions for cross-data women: 12 yrs



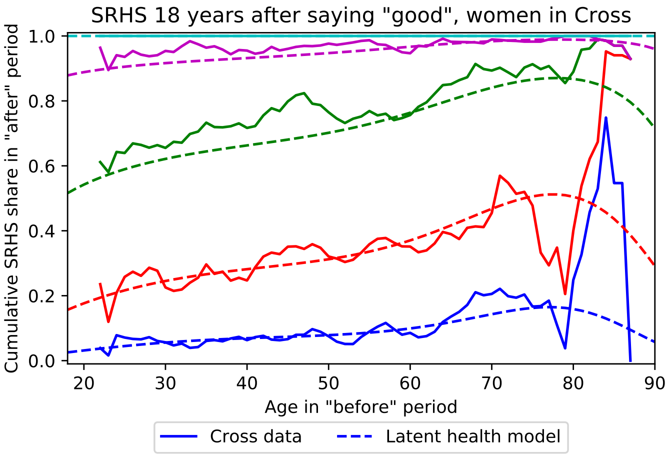
Model fit: SRHS transitions for cross-data women: 14 yrs



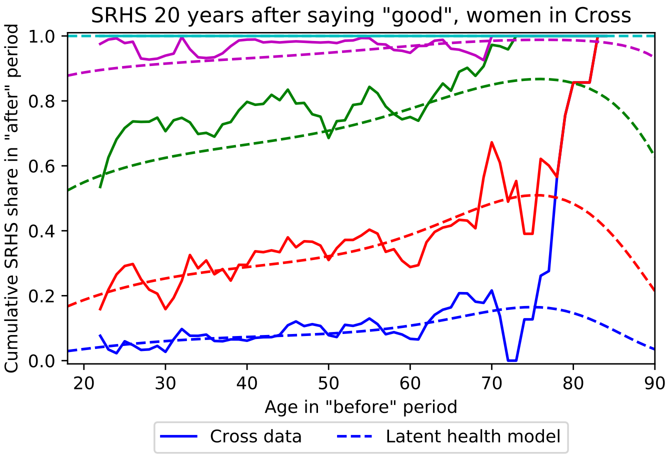
Model fit: SRHS transitions for cross-data women: 16 yrs



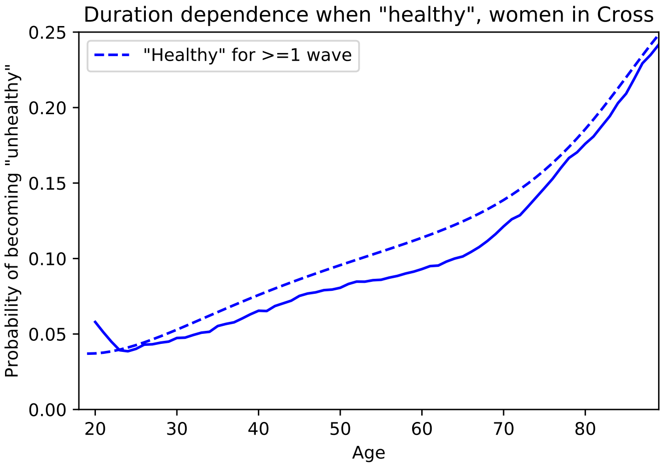
Model fit: SRHS transitions for cross-data women: 18 yrs



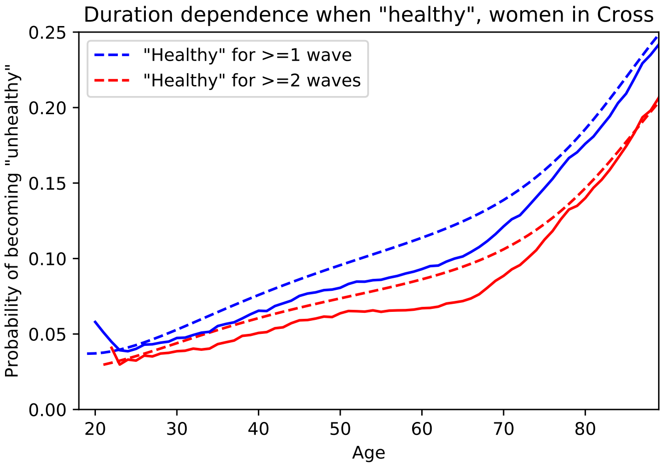
Model fit: SRHS transitions for cross-data women: 20 yrs



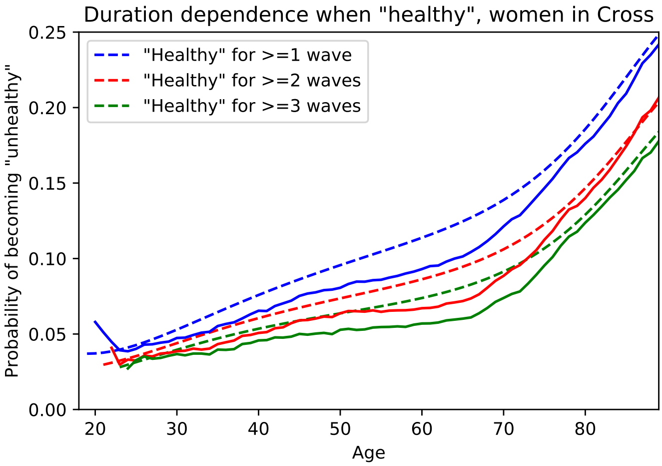
Model fit: Duration dependence, “healthy” to “unhealthy”



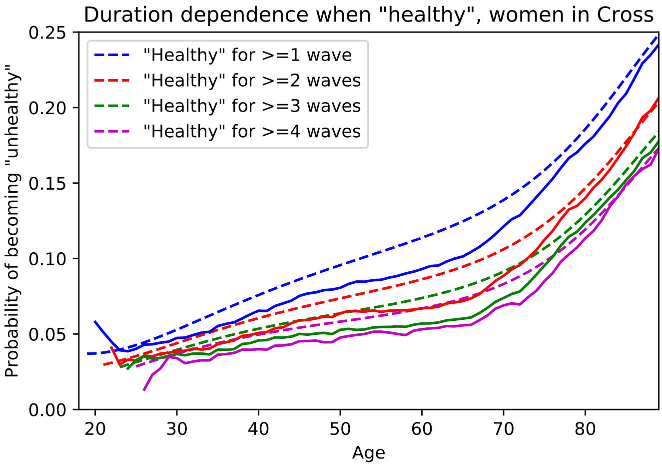
Model fit: Duration dependence, “healthy” to “unhealthy”



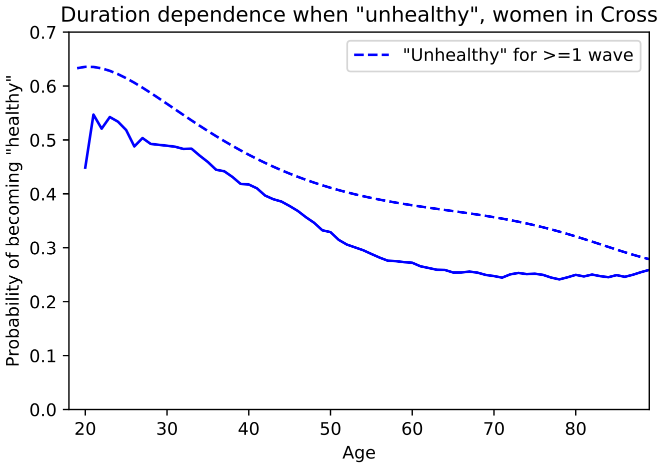
Model fit: Duration dependence, “healthy” to “unhealthy”



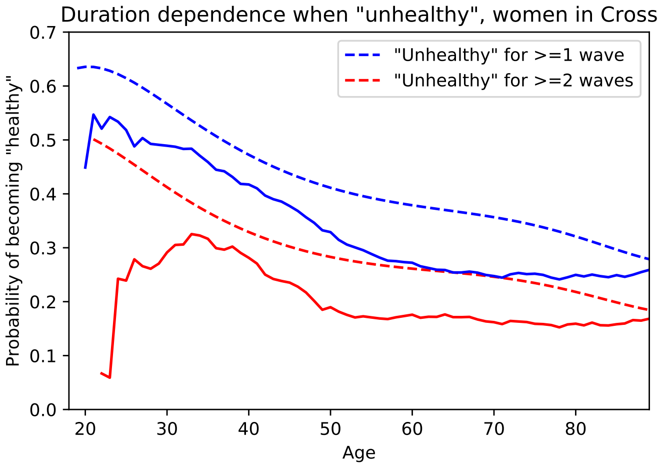
Model fit: Duration dependence, “healthy” to “unhealthy”



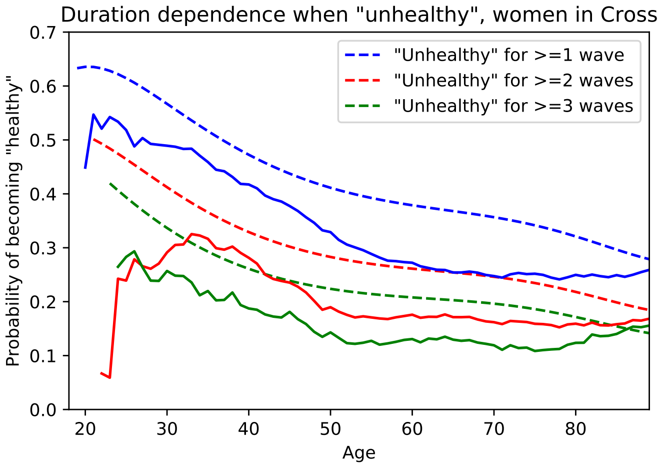
Model fit: Duration dependence, “unhealthy” to “healthy”



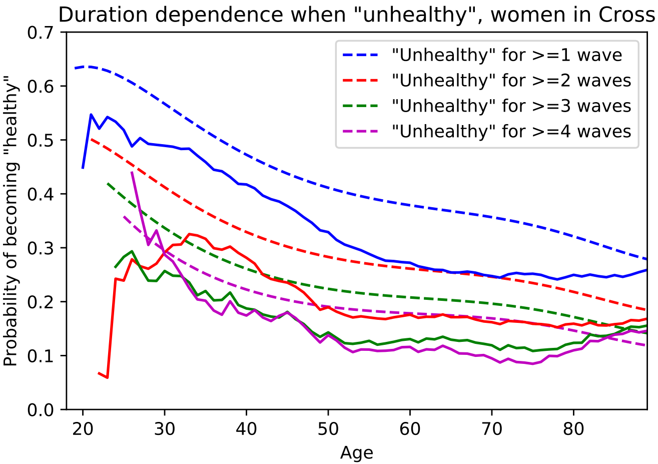
Model fit: Duration dependence, “unhealthy” to “healthy”



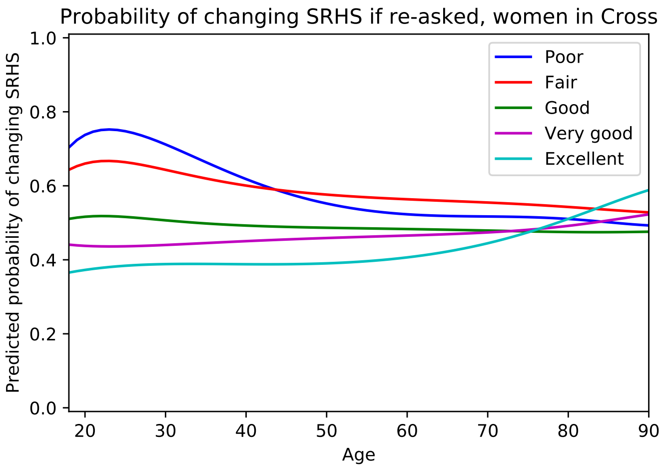
Model fit: Duration dependence, “unhealthy” to “healthy”



Model fit: Duration dependence, “unhealthy” to “healthy”

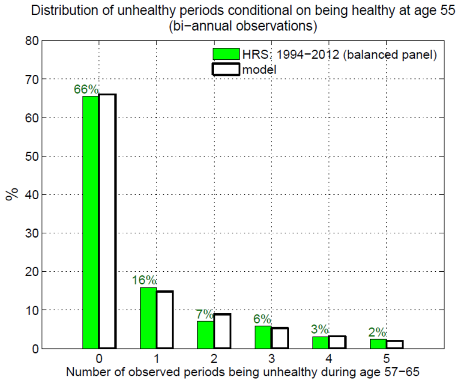


Estimated model: Changing SRHS if asked twice (cross-data women)

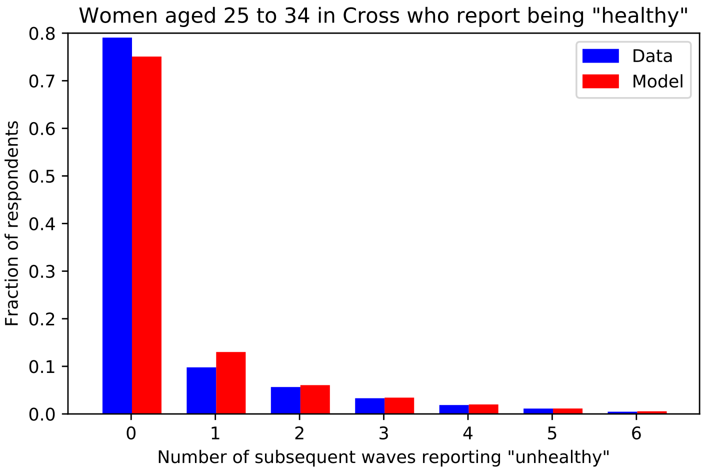


Model fit: Within-person distribution of “unhealthy” periods

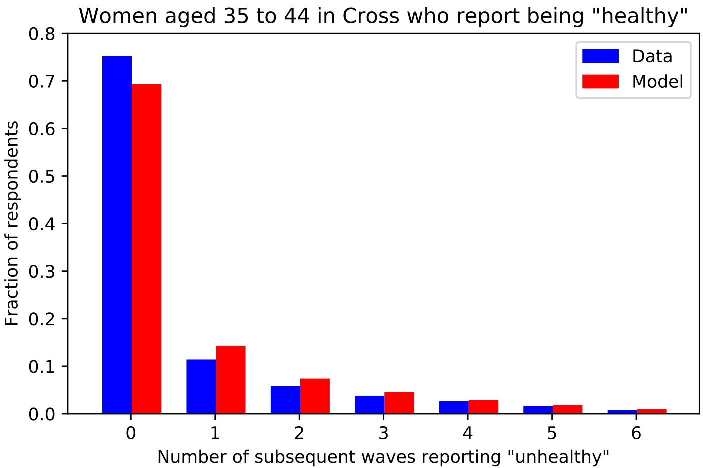
- DeNardi et al (2018) show that their duration dependence / unobserved heterogeneity model fits the distribution of “unhealthy” periods over next 5 waves
- Can the latent health model do that as well?



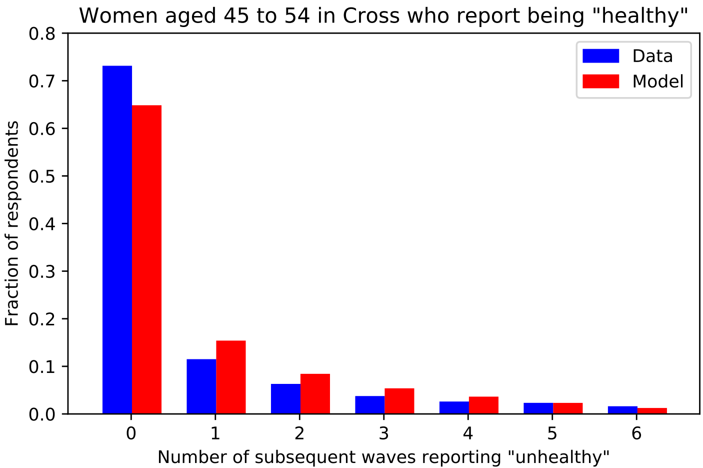
Model fit: Freqency of reporting “unhealthy” (cross-data women)



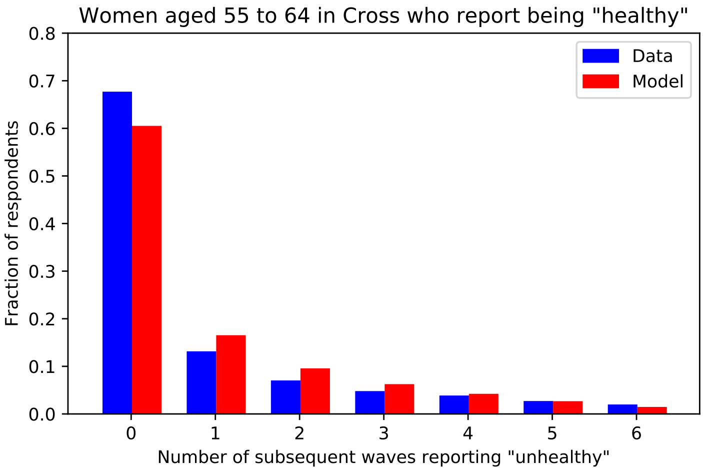
Model fit: Freqency of reporting "unhealthy" (cross-data women)



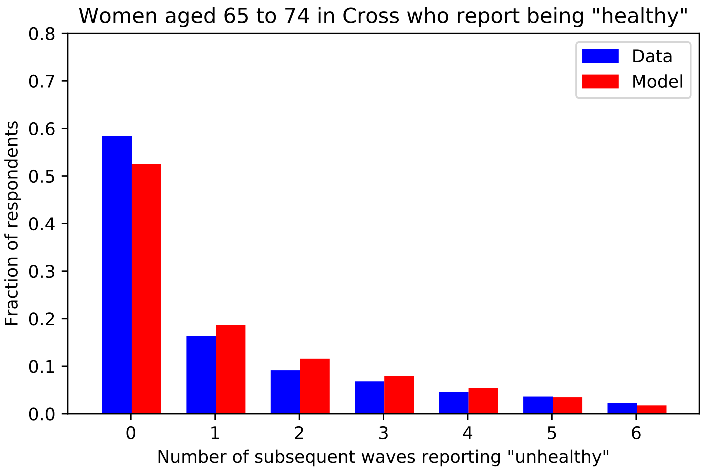
Model fit: Freqency of reporting “unhealthy” (cross-data women)



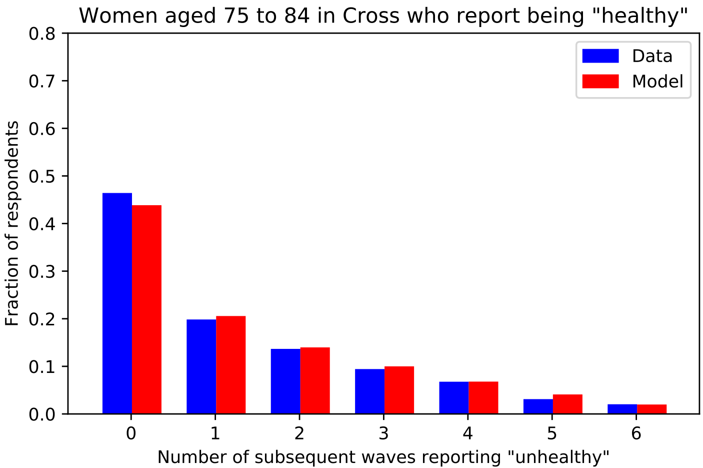
Model fit: Freqency of reporting “unhealthy” (cross-data women)



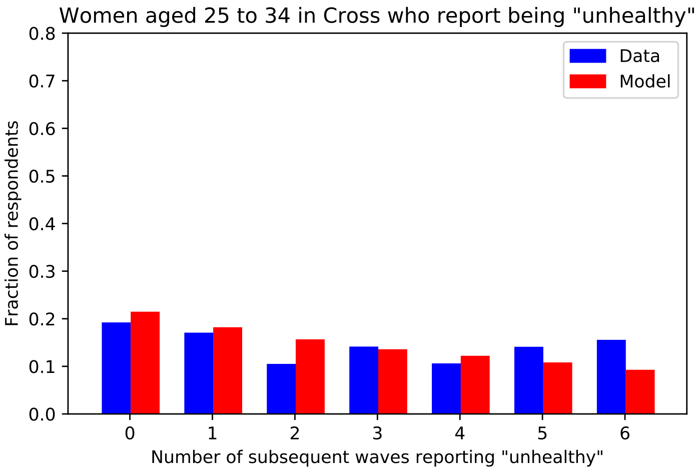
Model fit: Freqency of reporting "unhealthy" (cross-data women)



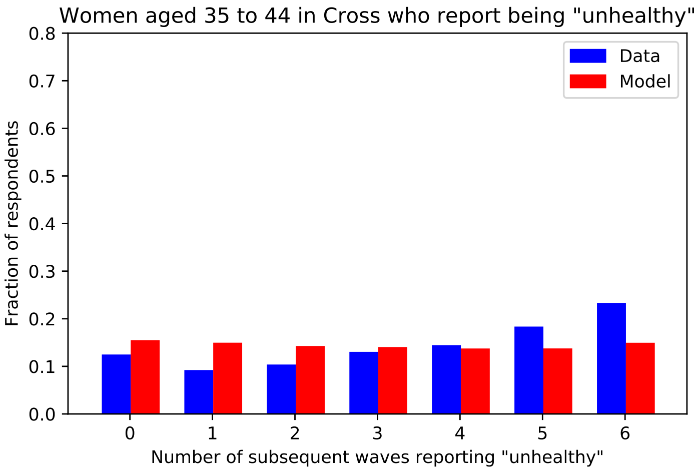
Model fit: Freqency of reporting “unhealthy” (cross-data women)



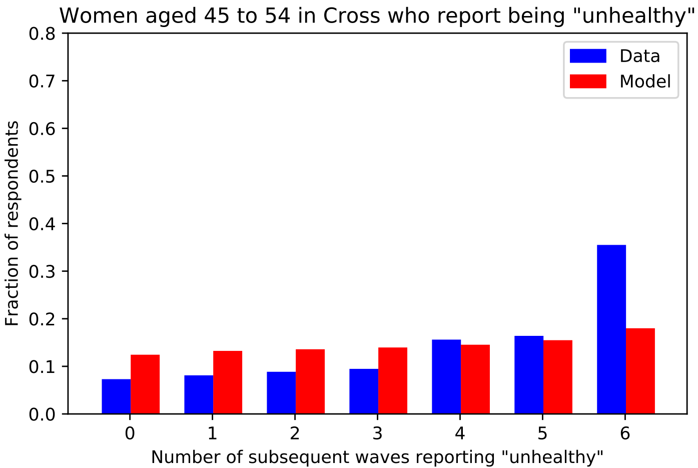
Model fit: Freqency of reporting “unhealthy” (cross-data women)



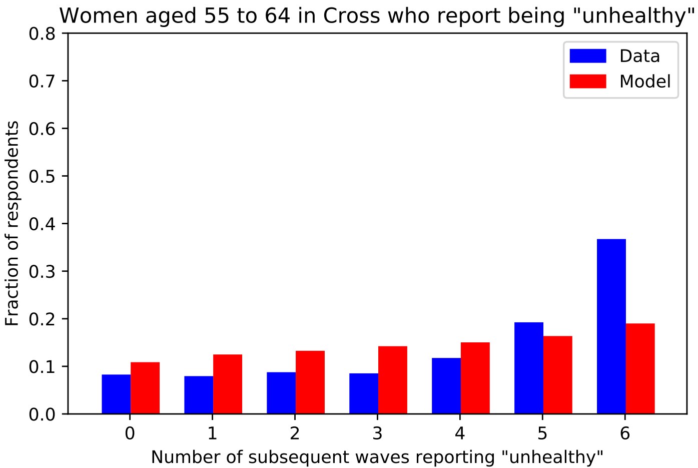
Model fit: Freqency of reporting “unhealthy” (cross-data women)



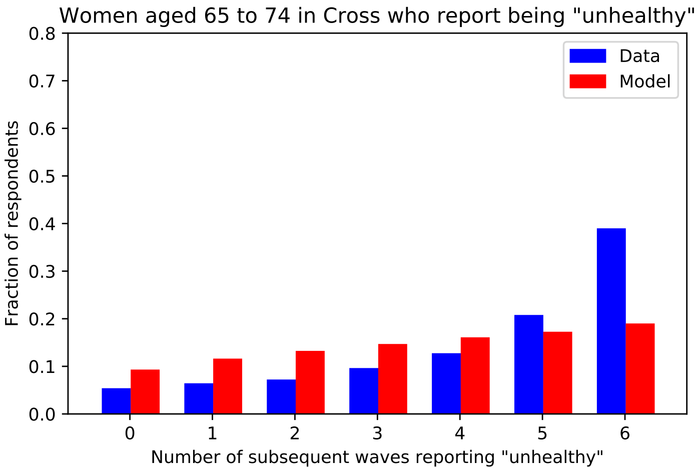
Model fit: Freqency of reporting “unhealthy” (cross-data women)



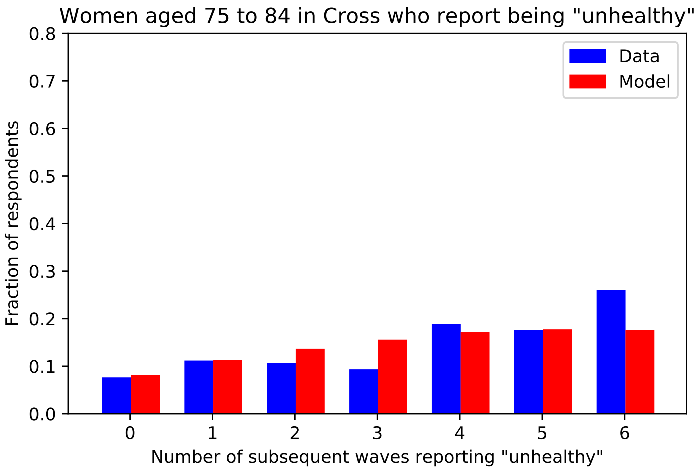
Model fit: Freqency of reporting “unhealthy” (cross-data women)



Model fit: Freqency of reporting “unhealthy” (cross-data women)



Model fit: Freqency of reporting “unhealthy” (cross-data women)



EXTENSIONS

Other work on latent health

- Modeling health as a latent variable is not novel; long history
- Actually new: estimating latent health dynamics on SRHS sequences

Other work on latent health

- Modeling health as a latent variable is not novel; long history
- Actually new: estimating latent health dynamics on SRHS sequences
- Lange & McKee (2012): Use **multiple measures** to construct latent health index, estimate dynamics in second step

Other work on latent health

- Modeling health as a latent variable is not novel; long history
- Actually new: estimating latent health dynamics on SRHS sequences
- Lange & McKee (2012): Use **multiple measures** to construct latent health index, estimate dynamics in second step
- That still interprets measurement error as true change in health!

Other work on latent health

- Modeling health as a latent variable is not novel; long history
- Actually new: estimating latent health dynamics on SRHS sequences
- Lange & McKee (2012): Use **multiple measures** to construct latent health index, estimate dynamics in second step
- That still interprets measurement error as true change in health!
- Blundell et al (2017): Investigate whether estimates of SRHS effect on retirement has self-justification bias ("I stopped working, so I must be unhealthy")

Other work on latent health

- Modeling health as a latent variable is not novel; long history
- Actually new: estimating latent health dynamics on SRHS sequences
- Lange & McKee (2012): Use **multiple measures** to construct latent health index, estimate dynamics in second step
- That still interprets measurement error as true change in health!
- Blundell et al (2017): Investigate whether estimates of SRHS effect on retirement has self-justification bias ("I stopped working, so I must be unhealthy")
- No: Similar results with SRHS or constructed index from multiple measures. Self-justification bias is offset by attenuation bias!

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?
- **Health types:** Some people don't recover from bad health because they have different transition probabilities

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?
- **Health types:** Some people don't recover from bad health because they have different transition probabilities
- **SRHS subjectivity:** Some people have lower standard for what constitutes "bad" health, so they report bad health a lot (shift in χ)

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?
- **Health types:** Some people don't recover from bad health because they have different transition probabilities
- **SRHS subjectivity:** Some people have lower standard for what constitutes "bad" health, so they report bad health a lot (shift in χ)
- **Skewed health shocks:** Model assumes normal shocks to x_t, but left skewed shock distribution would put some people "deeper" into bad health

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?
- **Health types:** Some people don't recover from bad health because they have different transition probabilities
- **SRHS subjectivity:** Some people have lower standard for what constitutes "bad" health, so they report bad health a lot (shift in χ)
- **Skewed health shocks:** Model assumes normal shocks to x_t, but left skewed shock distribution would put some people "deeper" into bad health
- **Second latent factor:** One factor model is insufficient; second dimension of latent health is required to accurately represent dynamics

Improving my latent health model (1/2)

- Why can't my model fit fraction of people who **consistently** report bad health?
- Why does model have more switching between good and fair health than the data?
- **Health types:** Some people don't recover from bad health because they have different transition probabilities
- **SRHS subjectivity:** Some people have lower standard for what constitutes "bad" health, so they report bad health a lot (shift in χ)
- **Skewed health shocks:** Model assumes normal shocks to x_t, but left skewed shock distribution would put some people "deeper" into bad health
- **Second latent factor:** One factor model is insufficient; second dimension of latent health is required to accurately represent dynamics
- Which of these hypotheses can my model address with extensions?

Improving my latent health model (2/2)

- Which of these hypotheses can model address with extensions? **ALL OF THEM**
- Which of them **should** we look at first (or last)?

Improving my latent health model (2/2)

- Which of these hypotheses can model address with extensions? **ALL OF THEM**
- Which of them **should** we look at first (or last)?
- Unobserved heterogeneity in health process should be the explanation of last resort

Improving my latent health model (2/2)

- Which of these hypotheses can model address with extensions? **ALL OF THEM**
- Which of them **should** we look at first (or last)?
- Unobserved heterogeneity in health process should be the explanation of last resort
- MLE can handle second latent factor, but would be computationally expensive
- And want representation of health that structural models can handle!

Improving my latent health model (2/2)

- Which of these hypotheses can model address with extensions? **ALL OF THEM**
- Which of them **should** we look at first (or last)?
- Unobserved heterogeneity in health process should be the explanation of last resort
- MLE can handle second latent factor, but would be computationally expensive
- And want representation of health that structural models can handle!
- Need multiple measures to be able to address subjective SRHS, skewed shocks, or second latent factor, so let's start there

Other measures of health

- Limitations on work, housework, or schoolwork due to health?

Other measures of health

- Limitations on work, housework, or schoolwork due to health?
- Activities of daily living: climbing stairs, carrying grocery bag

Other measures of health

- Limitations on work, housework, or schoolwork due to health?
- Activities of daily living: climbing stairs, carrying grocery bag
- Mental health: How often feel nervous, hopeless, worthless, etc?

Other measures of health

- Limitations on work, housework, or schoolwork due to health?
- Activities of daily living: climbing stairs, carrying grocery bag
- Mental health: How often feel nervous, hopeless, worthless, etc?
- Extent to which pain interferes with normal work activities

Other measures of health

- Limitations on work, housework, or schoolwork due to health?
- Activities of daily living: climbing stairs, carrying grocery bag
- Mental health: How often feel nervous, hopeless, worthless, etc?
- Extent to which pain interferes with normal work activities
- Clinical measures: hand grip strength, walking speed, lung capacity

Latent health model with single measure (SRHS)

Baseline latent health model:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \Phi. \tag{3}$$

$$h_t^* = \alpha_0 + \alpha_1 x_t + \eta_t, \qquad \eta_t \sim \Phi, \tag{4}$$

$$h_t = 1 + \sum_{k=1}^{K-1} \mathbf{1}(h_t^* \geq \chi_k).$$

Latent health model with multiple measures (1/2)

Latent health model with multiple independent measures:

$$x_0 \sim N(\mu_0, \sigma_0^2). \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \Phi. \tag{3}$$

$$m^*_{\ell t} = \alpha_{\ell 0} + \alpha_{\ell 1} x_t + \eta_{\ell t}, \qquad \eta_{\ell t} \sim \Phi, \tag{4}$$

$$m_{\ell t} = 1 + \sum_{k=1}^{K_\ell - 1} \mathbf{1}(m^*_t \geq \chi_{k\ell}), \qquad \ell \in \{0, \cdots, L\}, \qquad h_t \equiv m_{0t}.$$

Latent health model with multiple measures (2/2)

- Each wave, observe multiple categorical measures of latent health
- Measure 0 is set aside for SRHS (and mortality)
- Location assumption: $\chi_{1\ell} = 0 \ \forall \ell$, but only $\alpha_{00} = 0$

Latent health model with multiple measures (2/2)

- Each wave, observe multiple categorical measures of latent health
- Measure 0 is set aside for SRHS (and mortality)
- Location assumption: $\chi_{1\ell} = 0 \; \forall \ell$, but only $\alpha_{00} = 0$
- Identifying assumption: Independence of reporting errors $\eta_{\ell t}$
- Adding another measure **should not** affect estimates of other α coefficients

Latent health model with multiple measures (2/2)

- Each wave, observe multiple categorical measures of latent health
- Measure 0 is set aside for SRHS (and mortality)
- Location assumption: $\chi_{1\ell} = 0 \ \forall \ell$, but only $\alpha_{00} = 0$
- Identifying assumption: Independence of reporting errors $\eta_{\ell t}$
- Adding another measure **should not** affect estimates of other α coefficients
- Obvious violation: Second latent factor (omitted variable bias)

Latent health model with multiple measures (2/2)

- Each wave, observe multiple categorical measures of latent health
- Measure 0 is set aside for SRHS (and mortality)
- Location assumption: $\chi_{1\ell} = 0 \ \forall \ell$, but only $\alpha_{00} = 0$
- Identifying assumption: Independence of reporting errors $\eta_{\ell t}$
- Adding another measure **should not** affect estimates of other α coefficients
- Obvious violation: Second latent factor (omitted variable bias)
- My MLE code is already set up to estimate such a model, and appropriate datasets have been constructed...
- ...but I don't have a specification or results I'd like to share today

Adding other model extensions

- With multiple measures, have **more objective** observations of health
- Can identify whether respondents with consistently bad SRHS also report negatively on other measures of health...

Adding other model extensions

- With multiple measures, have **more objective** observations of health
- Can identify whether respondents with consistently bad SRHS also report negatively on other measures of health...
- ...or whether they're "pessimists" who have different labeling standard

Adding other model extensions

- With multiple measures, have **more objective** observations of health
- Can identify whether respondents with consistently bad SRHS also report negatively on other measures of health...
- ...or whether they're "pessimists" who have different labeling standard
- And can test whether normality of latent health shocks is well supported

Adding other model extensions

- With multiple measures, have **more objective** observations of health
- Can identify whether respondents with consistently bad SRHS also report negatively on other measures of health...
- ...or whether they're "pessimists" who have different labeling standard
- And can test whether normality of latent health shocks is well supported
- Nothing in core MLE code has to change to add these features!
- **Label heterogeneity:** Discretized x_t vector would include "type"; change precomputed report probability matrix, tile transition probabilities
- **Skewed health shocks:** Just change precomputed transition probability array for x_t, probably use mixed normal (impose mean 0, stdev 1)

Latent health model: Adding SRHS labeling heterogeneity

Latent health model with "labeling shifter" for SRHS:

$$x_0 \sim N(\mu_0, \sigma_0^2), \qquad \delta \sim \text{Discrete}. \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \Phi. \tag{3}$$

$$m^*_{\ell t} = \alpha_{\ell 0} + \alpha_{\ell 1} x_t + \delta \mathbf{1}(\ell = 0) + \eta_{\ell t}, \qquad \eta_{\ell t} \sim \Phi, \tag{4}$$

$$m_{\ell t} = 1 + \sum_{k=1}^{K_\ell - 1} \mathbf{1}(m^*_t \geq \chi_{k\ell}), \qquad \ell \in \{0, \cdots, L\}, \qquad h_t \equiv m_{0t}.$$

Latent health model: Adding skewed health shocks

Latent health model with "labeling shifter" for SRHS:

$$x_0 \sim N(\mu_0, \sigma_0^2), \qquad \delta \sim \text{Discrete}. \tag{1}$$

$$\text{Prob}(h_{t+1} = 0) = 1 - \Phi(f(j_t, x_t)). \tag{2}$$

$$x_{t+1} = \rho_j x_t + (1 - \rho_j) g(j_t) + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \text{Mixed Normal}. \tag{3}$$

$$m_{\ell t}^* = \alpha_{\ell 0} + \alpha_{\ell 1} x_t + \delta \mathbf{1}(\ell = 0) + \eta_{\ell t}, \qquad \eta_{\ell t} \sim \Phi, \tag{4}$$

$$m_{\ell t} = 1 + \sum_{k=1}^{K_\ell - 1} \mathbf{1}(m_t^* \geq \chi_{k\ell}), \qquad \ell \in \{0, \cdots, L\}, \qquad h_t \equiv m_{0t}.$$

CONCLUSION

Concluding remarks

- Structural models that use coarse categorical health get long run dynamics of health wrong; calculations of long term risk or welfare likely faulty

Concluding remarks

- Structural models that use coarse categorical health get long run dynamics of health wrong; calculations of long term risk or welfare likely faulty
- Bad news: we have to model health as a continuous(-ish) variable
- Can discretize health states (e.g. using Tauchen's method), but there have to be enough nodes that short run transitions don't "blur" long run transitions

Concluding remarks

- Structural models that use coarse categorical health get long run dynamics of health wrong; calculations of long term risk or welfare likely faulty
- Bad news: we have to model health as a continuous(-ish) variable
- Can discretize health states (e.g. using Tauchen's method), but there have to be enough nodes that short run transitions don't "blur" long run transitions
- Should we go back and re-do a bunch of the papers I cited?
- Computation **will be more intense**, but some papers are 10 years old

Concluding remarks

- Structural models that use coarse categorical health get long run dynamics of health wrong; calculations of long term risk or welfare likely faulty
- Bad news: we have to model health as a continuous(-ish) variable
- Can discretize health states (e.g. using Tauchen's method), but there have to be enough nodes that short run transitions don't "blur" long run transitions
- Should we go back and re-do a bunch of the papers I cited?
- Computation **will be more intense**, but some papers are 10 years old
- All hail Moore's law!