Health status over the life-cycle

Raun van Ooijen [a,c] Rob Alessie [a,c] Marike Knoef [b,c]

[a] *University of Groningen, the Netherlands*
[b] *Leiden University, the Netherlands*
[c] *Netspar, the Netherlands*

Leiden, 27 January 2016

Netspar International pension workshop

university of groningen

Motivation

- What is health and how should we measure it?
- Self-Reported Health (SRH) has proven to be a very useful measure of health BUT
  - Time inconsistency in reporting (Benitez-Silva and Ni, 2008)
  - Health perception (e.g. Lindeboom and Van Doorslaer, 2004)
  - Justification bias (Bound et al., 2001)

Contributions

1. Health measurement model. This 'advanced' regression model...
   - relates self-reported health to objective medical conditions
     - Self-reported health is obtained from a panel survey
     - This survey is merged with administrative data on medical conditions
     - Admin. data is much less prone to measurement error (Baker et al., 2004).
   - allows for differences in health perception
   - accounts for unobserved heterogeneity and the persistence in **unobserved health shocks**
   - accounts for inconsistent reporting over time in self-reported health

Contributions (II)

2. We use the model to impute health status in a large admin. data set solely based on medical conditions

3. This large data set allows for a detailed investigation into the relationship between socio-economic status (SES) and health

Data: two samples

- Estimation sample (2007-2010)
  - 'Enriched' LISS panel for self-assessed health with admin. health data
  - Admin. health data
    - Prescription drugs (no inpatient medications)
    - Hospital discharge register (88% of all inpatient stays)
    - Home care
- Prediction sample (2006-2010)
  - Administrative data for the whole population
  - Admin. health data, education, income, wealth

LISS panel for self-assessed health (2007-2010)

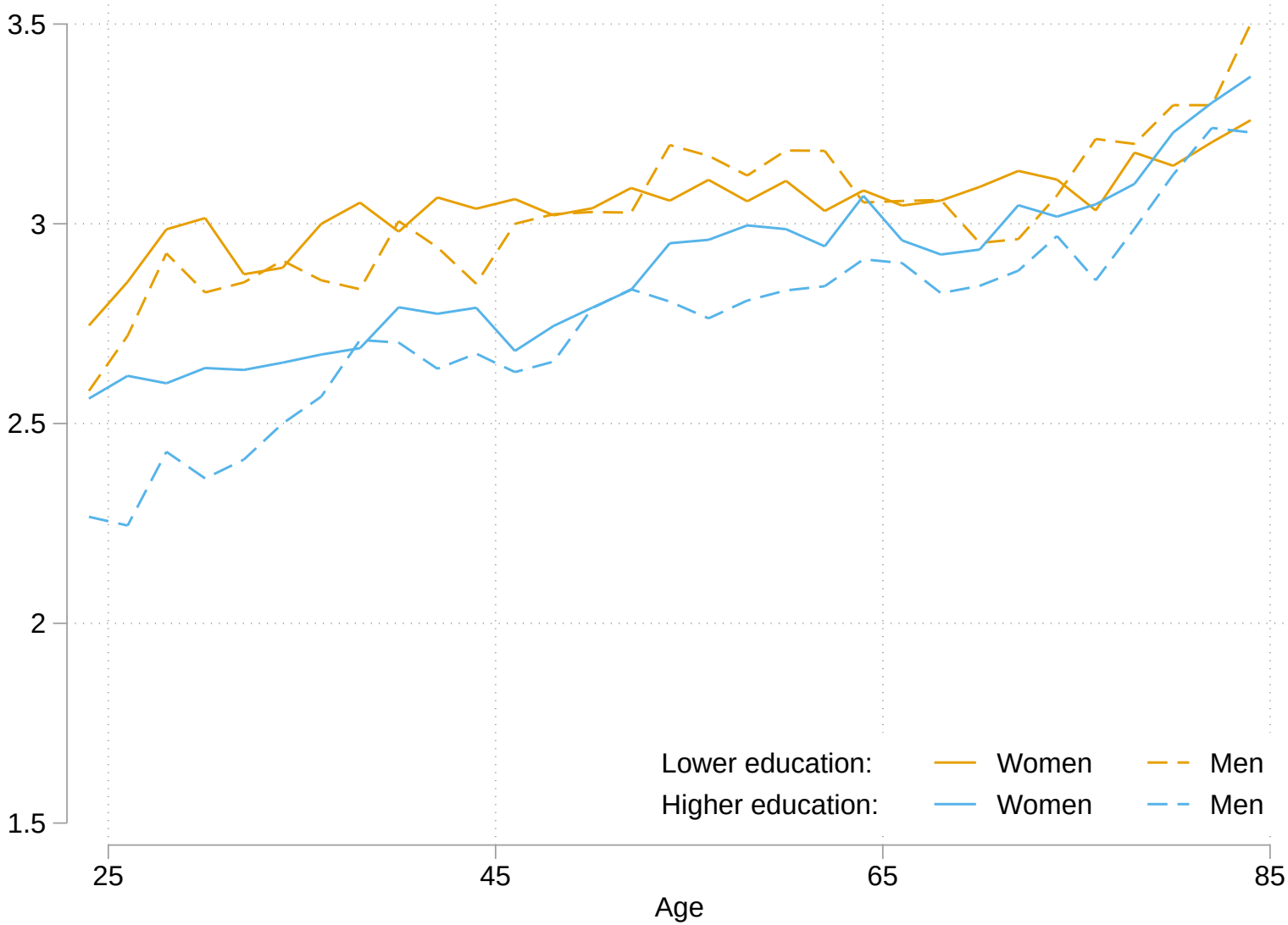
- Longitudinal Internet Study in the Social Sciences, gathered by CentERdata, University of Tilburg
- Internet-based panel
- Representative for the non-institutionalized population
- 16,720 individual-year observations (age 15+)
- Consent to link to administrative data in 2011
  - On average 70 percent can be linked

SRH questions in the LISS

- *How would you describe your health in general?* With response options: Poor, Moderate, Good, Very Good and Excellent.

- *Can you indicate whether your health is poorer or better, compared to last year?* With response options: considerably poorer, somewhat poorer, the same, somewhat better and considerably better.

Self reported health by education 1 = Excellent — 5 =Poor



Source: LISS 2007-2012

Administrative records 2006-2010 (Statistics Netherlands)

- Administrative data on SES
  - Educational attainment (register data & labor force survey)
  - Net Worth, Income, Employment status (tax records)
  - Demographic variables (gender, age, marital status, household size)
  - Admin. health data
    - Prescription drugs (no inpatient medications)
    - Hospital discharge register (88% of all inpatient stays)
    - Home care
- Representative sample of 163.695 Dutch residents alive in 2006.
- Age 15 and older who are not institutionalized.

Mapping prescription drugs → chronic conditions

- Objective measures in administrative data

Disease	ATC-code in medical register
Coronary disease	B01A, C04A
Cardiac disease	C01, C03C
Hypertension	C02, C03A, C07, C08, C09A,B
Rheumatic conditions	H02, M01, M02
High blood cholesterol	C10A
Diabetes	A10A, A10B
Glaucoma and cataract	S01E
Peptic Ulcers	A02A, A02B
Chronic bronchitis, asthma	R03
Anxiety and depression	N05B, N06A
Osteoporosis	M05

- Matching Self-reported condition in LISS.

Prevalence of chronic conditions and health care utilization, 2007-2010

	Men				Women			
	Born $\geq$ 1945		Born $<$ 1945		Born $\geq$ 1945		Born $<$ 1945	
	Liss	Admin	Liss	Admin	Liss	Admin	Liss	Admin
High blood cholesterol	6.87	8.40	32.16	34.15	3.75	4.28	22.72	25.26
Cardiovascular disease	10.42	13.15	46.88	51.51	10.15	13.49	43.30	48.39
Rheumatism / osteoporosis	3.63	19.49	7.74	21.88	8.14	25.84	22.01	37.58
Cataract / glaucoma	0.73	7.26	6.78	14.92	0.56	8.85	10.29	23.32
Diabetes	3.03	3.27	9.81	10.52	1.61	2.05	8.56	8.32
Peptic Ulcers	4.72	10.38	12.12	26.13	4.85	12.97	14.27	31.88
Chronic bronchitis / asthma	4.39	15.64	5.82	22.68	5.16	21.37	6.09	23.32
Depression and anxiety	3.34	6.82	2.36	7.43	4.69	10.45	4.39	14.39
Cancer	0.48	0.31	4.28	2.23	0.91	0.57	2.19	1.82
Medication use	32.98	59.67	73.41	86.51	41.65	82.54	74.59	89.92
Home care	1.35	0.53	3.22	1.38	2.20	0.25	8.20	3.04
Hospital stay	7.88	6.95	13.59	17.90	11.26	10.29	12.64	17.61
Length of stay (days)	0.48	0.26	0.74	0.95	0.48	0.35	0.68	0.89
N	5824		2083		7293		1828	

Health measurement model - Level (Jürges, 2007)

$$y_{it}^{*} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}, \;\; t = 1, \ldots, T$$

$\boldsymbol{x}$, is a vector of chronical health conditions,
$\varepsilon_{it}|\boldsymbol{x}_{it} \sim N(0, 1)$

$$\varepsilon_{it} = c_i + u_{it}$$
$$c_i \sim NID(0, \sigma_c^2)$$

$$u_{it} = \gamma u_{it-1} + \zeta_{it}$$
$$\zeta_{it} \sim NID(0, \sigma_\zeta^2)$$
$$\sigma_\zeta^2 = (1 - \sigma_c^2) \cdot (1 - \gamma^2)$$

$$\text{cov}(c_i, u_{it}) = 0, \;\; t = 1, \ldots, T$$

- *SRH* reflects latent health status.

$$SRH_{it} = L \text{ if } \lambda^g_{L-1} < y^*_{it} \leqslant \lambda^g_L, \; L = 1, \cdots, 5; \; g = 1, \ldots, G$$

- $\boldsymbol{\lambda}^g = (\lambda^g_1, \lambda^g_2, \lambda^g_3, \lambda^g_4)'$ are the threshold parameters for demographic group g ($\lambda^g_0 = -\infty$ and $\lambda^g_5 = \infty$).
  - Males born before 1945
  - Males born after 1945
  - Females born before 1945
  - Females born after 1945

- Account for 'cut-point' shifting
- No 'index shifting' to have a unique health index.

- Estimate the 'structural' parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_c^2, \gamma, \boldsymbol{\lambda}^g)'$.

$$
\begin{aligned}
y_{i1}^* &= \boldsymbol{x}_{i1}'\boldsymbol{\beta}_1^g + \varepsilon_{i1} \\
y_{i2}^* &= \boldsymbol{x}_{i2}'\boldsymbol{\beta}_2^g + \varepsilon_{i2} \\
y_{i3}^* &= \boldsymbol{x}_{i3}'\boldsymbol{\beta}_3^g + \varepsilon_{i3} \\
y_{i4}^* &= \boldsymbol{x}_{i4}'\boldsymbol{\beta}_4^g + \varepsilon_{i4}
\end{aligned}
$$

$$
\boldsymbol{\varepsilon}_i | \boldsymbol{x}_i \sim NID\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{21}^g & \rho_{31}^g & \rho_{41}^g \\ \rho_{21}^g & 1 & \rho_{32}^g & \rho_{42}^g \\ \rho_{31}^g & \rho_{32}^g & 1 & \rho_{43}^g \\ \rho_{31}^g & \rho_{32}^g & \rho_{43}^g & 1 \end{pmatrix} \right)
$$

$$
\begin{aligned}
\rho_{21} = \rho_{32} = \rho_{43} &= (1-\gamma)\sigma_c^2 + \gamma \\
\rho_{31} = \rho_{42} &= (1-\gamma^2)\sigma_c^2 + \gamma^2 \\
\rho_{41} &= (1-\gamma^3)\sigma_c^2 + \gamma^3
\end{aligned}
$$

- No 'index shifting': $\boldsymbol{\beta}^g = \boldsymbol{\beta}$ and $\boldsymbol{\rho}^g = \boldsymbol{\rho}$

Health measurement model - Level & Change

Latent level model: How would you describe your health in general?

$$y_{it}^* = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

$$\varepsilon_{it} = c_i + \zeta_{it} + \xi_{it},$$

where c_i is a random effect, ζ_{it} is an idiosyncratic health shock, and ξ_{it} captures measurement error.

1. $\text{var}(\varepsilon_{it}) = \sigma_c^2 + \sigma_\zeta^2 + \sigma_\xi^2 = 1$
2. $c_i|\boldsymbol{x}_i \sim NID(0, \sigma_c^2)$
3. $\xi_{it},\ t = 1, \ldots, T$
4. $\xi_{it}|\boldsymbol{x}_i \sim NID(0, \sigma_\xi^2)$.

$$\zeta_{it} = \gamma\zeta_{it-1} + \upsilon_{it},\ 0 \leq |\gamma| < 1$$

$$\upsilon_{it}|\boldsymbol{x}_i \sim NID(0, \sigma_\upsilon^2)$$

$$\zeta_{it}|\boldsymbol{x}_i \sim NID\left(0, \frac{\sigma_\upsilon^2}{1-\gamma^2}\right)$$

Latent difference model: How is your health compared to last year?

$$dy_{it}^{*} = \Delta \boldsymbol{x}_{it}' \boldsymbol{\beta} + u_{it}$$

$$u_{it} = \Delta \zeta_{it} + \theta_{it} = (\gamma - 1)\zeta_{it-1} + \upsilon_{it} + \theta_{it},$$

where θ_{it} captures measurement error.

- $\theta_{it} = d_i + \eta_{it}$
- We allow for a random effect d_i in the measurement error because the reduced form evidence suggests $\text{cov}(u_{it}, u_{i\tau}) > 0, t \neq \tau$.

1. $\text{var}(u_{it}) = 1$
2. $\text{cov}(\eta_{it}, \xi_{i\tau}) = \text{cov}(c_i, d_i) = \text{cov}(\eta_{it}, d_i) = \text{cov}(\eta_{it}, c_i) = \text{cov}(\xi_{it}, d_i) = \text{cov}(\theta_{it}, \xi_{i\tau}) = 0$ conditional upon $\boldsymbol{x}_i$.
3. $d_i | \boldsymbol{x}_i \sim NID(0, \sigma_d^2)$
4. $\eta_{it} \sim NID(0, \sigma_\eta^2)$.

- Eight variate ordered probit model
- These assumptions imply that $(t > \tau)$

$$\begin{aligned}
\text{var}(u_{it}) &= \sigma_d^2 + 2(1-\gamma)\sigma_\zeta^2 + \sigma_\eta^2 = 1 \\
\text{cov}(u_{it}, u_{i\tau}) &= \sigma_d^2 - \gamma^{t-1-\tau}(\gamma-1)^2\sigma_\zeta^2 \\
\text{cov}(u_{it}, \varepsilon_{it}) &= (1-\gamma)\sigma_\zeta^2 \\
\text{cov}(u_{i\tau}, \varepsilon_{it}) &= \gamma^{t-\tau}(1-\gamma)\sigma_\zeta^2 \\
\text{cov}(u_{it}, \varepsilon_{i\tau}) &= \gamma^{t-1-\tau}(\gamma-1)\sigma_\zeta^2
\end{aligned}$$

- Estimate the 'structural' parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_c^2, \sigma_\zeta^2, \sigma_d^2, \gamma, \boldsymbol{\lambda}_{change}^g \boldsymbol{\lambda}_{level}^g)'$.

Results (1)

	Level		Level Change	
	Coeff	SE	Coeff	SE
Diabetes	-.567	.046	-.488	.040
Depression and anxiety	-.324	.023	-.281	.017
Cardiovascular disease	-.290	.022	-.256	.017
Malignancies	-.270	.066	-.383	.052
Peptic Ulcers	-.216	.018	-.198	.014
High blood cholesterol	-.203	.029	-.204	.025
Home care	-.150	.066	-.325	.048
Medication use	-.114	.015	-.070	.012
Rheumatism / osteoporosis	-.094	.014	-.080	.010
Cataract / glaucoma	-.079	.019	-.040	.014
Chronic bronchitis / Asthma	-.071	.016	-.054	.012
Hospital stay	-.063	.018	-.069	.014
Length of stay (days)	-.012	.002	-.011	.001
γ	.269	.022	.923	.021
σ^2_ζ	.624	.010	.004	.190
σ^2_ζ			.789	.189
σ^2_d			.211	.008

For the threshold parameters we distinguish on basis of the variables 'gender' and 'year-of-birth' four different demographic groups ($G = 4$): (1) males born before 1945, (2) males born after 1945 (3) females born before 1945, (4) females born after 1945

Results (2)

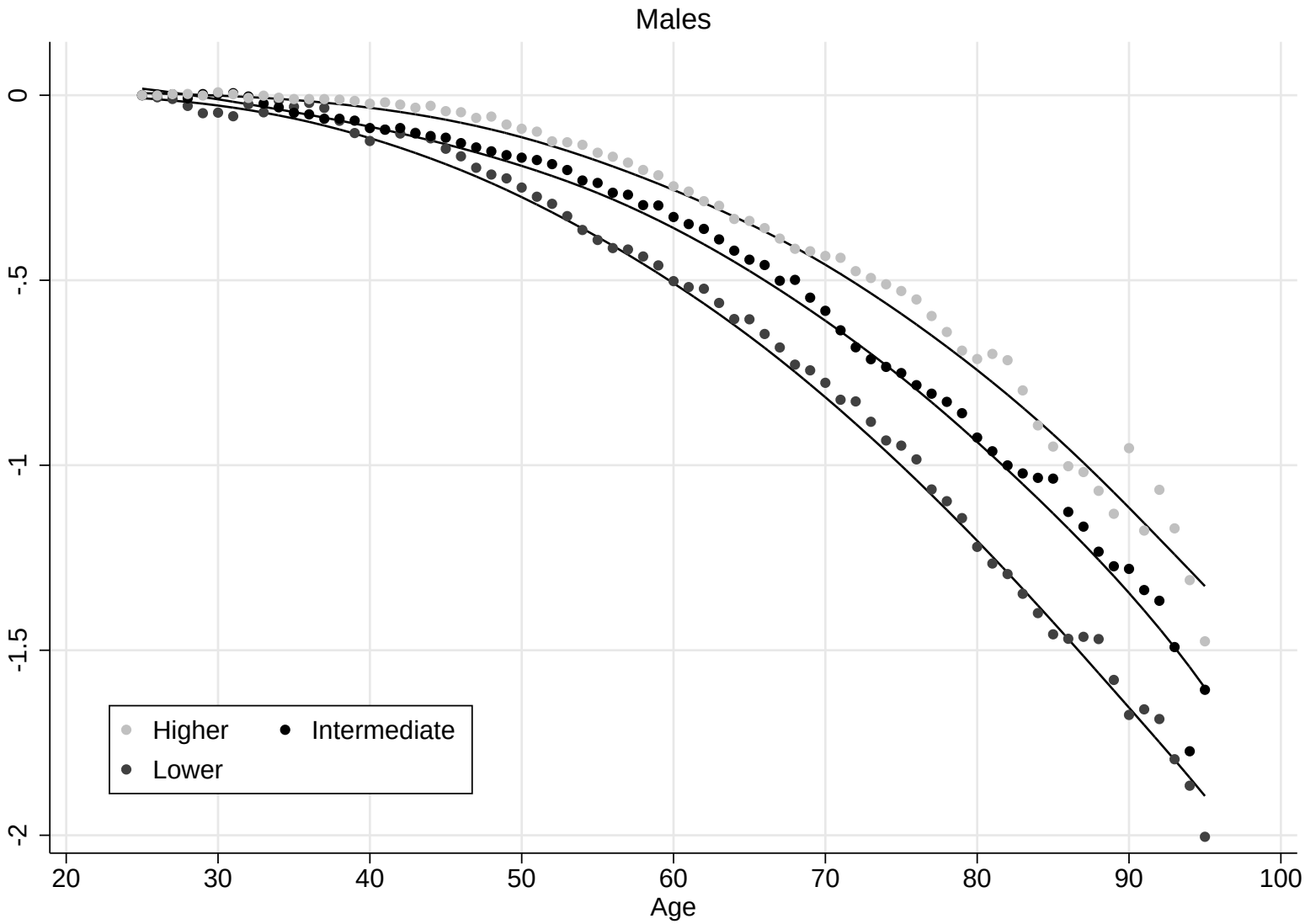
1. Using both measures of SRH substantially increases the estimated persistence in health status.
2. Differences in health perception -
   - Elderly ($<$ 1945) males born are more positive about their health status than elderly females (keeping health index constant).
   - The younger individuals (born after 1944) has a higher tendency to report that their health is 'poor' than the older generation keeping health index constant.

In-sample predictions self-reported: 'Level & change' , $N = 16,720$

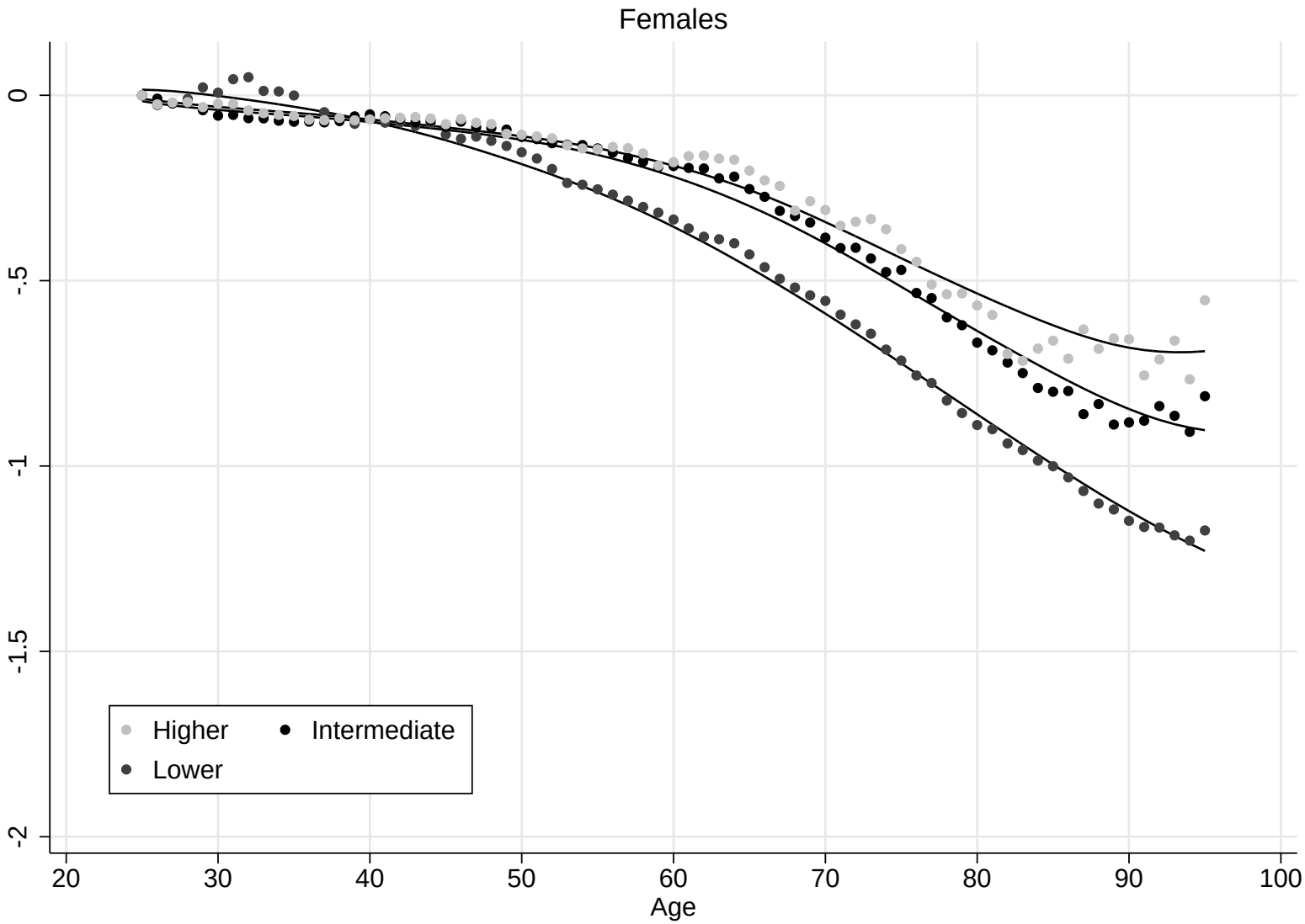
$t-1 \setminus t$	Poor	Moderate	Good	Very good	Excellent
Panel A: Empirical *SRH* level					
Poor	46.8	46.8	6.4	0.0	0.0
Moderate	3.3	59.5	36.3	0.9	0.1
Good	0.2	8.3	78.6	11.7	1.3
Very good	0.1	1.6	37.3	51.6	9.5
Excellent	0.2	0.7	17.6	37.1	44.5
Panel B: In-sample predictions *SRH* level					
Poor	34.2	60.2	5.6	0.0	0.0
Moderate	3.8	54.1	41.7	0.5	0.0
Good	0.1	10.8	74.4	13.5	1.2
Very good	0.0	0.4	42.8	43.7	13.1
Excellent	0.0	0.0	12.6	42.1	45.2
Panel C: In-sample predictions *SRH* 'Level & change'					
Poor	62.3	37.6	0.0	0.0	0.0
Moderate	2.9	73.5	23.6	0.0	0.0
Good	0.0	6.4	85.4	8.1	0.1
Very good	0.0	0.0	25.5	64.9	9.6
Excellent	0.0	0.0	0.6	29.9	69.6

We use the estimated parameters to construct a health index for a large random sample of the Dutch population

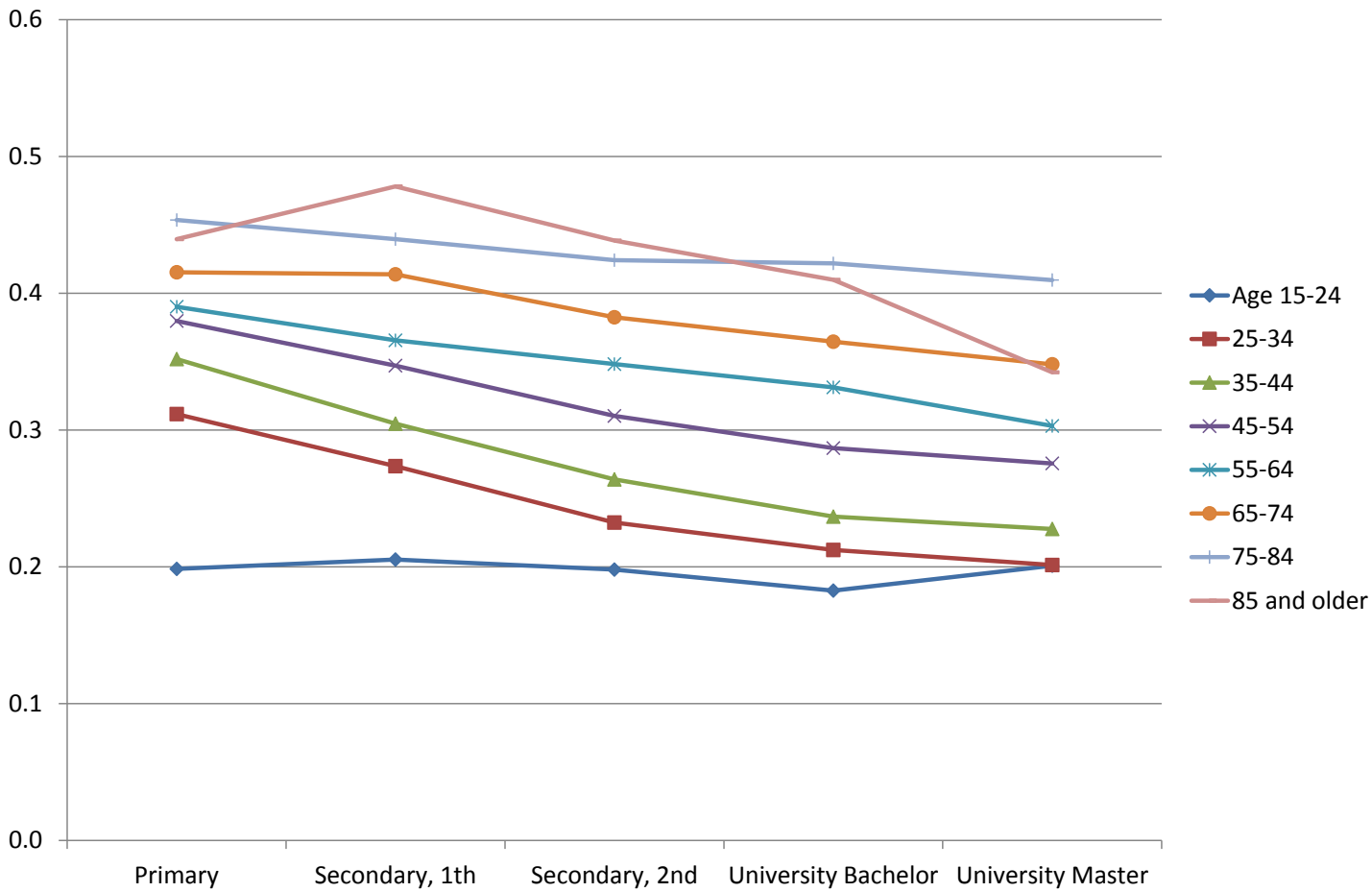
Predicted health by age and education - males



Predicted health by age and education - females



Standard deviation of predicted health by age and education



Main results - predicted health and SES

- Lower educated and income people are more likely to stay in poor health than higher educated people.
- The age at which health starts to decline at a greater rate arrives earlier for males and persons with a lower level of education.

Conclusion

- Evidence of reporting bias.
- Construct a health measurement model which takes this into account.
- Using both measures of SRH substantially increases the estimated persistence in health status.
- We use predicted health to study the evolution of 'true' health and the 'true' health SES relationship.
- We find plausible associations between socio-economics status and health.
- Compared with other studies we find that health decays at an faster rate, especially for males.
- This may be due to the fact that, compared to SRH, our objective health measure is less likely to suffer from reporting effects.