



Methodological challenges in life course studies

George B. Ploubidis & Bianca DeStavola

Website <http://pathways.lshtm.ac.uk>

Email pathways@lshtm.ac.uk

Twitter [@pathwaysNCRM](https://twitter.com/pathwaysNCRM)



Leading education
and social research
Institute of Education
University of London



Outline

- Methodological challenges in life course studies
- Two examples
- Some suggestions

“A life course approach is paradoxical as on the one hand it is intuitively obvious ... and yet on the other it is empirically complex”

Kuh and Ben-Shlomo, 2004

I would add

Methodologically complex too....

- Missing data
- Measurement error
- Unmeasured confounding

More challenges

- Life course mechanisms (accumulation, chains of risk, critical period, social drift) imply a mechanistic view of causality
- The process of reversibility implies both mediation and moderation
- All the hypothesised effects require the formal estimation and reliable quantification of direct, and indirect effects, as well as moderated mediation and/or interactions

- Repeated measures a strength of life course studies
- However, quantifying change and in particular parallel processes is not so straightforward
- If homogeneity over time is assumed then models that describe change with 1 – 2 parameters (single curve/trajectory) can be used (GEE, Multilevel models, Growth curve models)
- If the longitudinal patterns show evidence of heterogeneity some form of a mixture model can be used
- Either approach brings in additional assumptions and challenges (modelling time varying covariates and/or confounders for example)

ARTICLE IN PRESS

Social Science & Medicine xxx (2014) 1–8



Contents lists available at [ScienceDirect](#)

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed



Lifelong Socio Economic Position and biomarkers of later life health: Testing the contribution of competing hypotheses

George B. Ploubidis^{a,b,*}, Lenka Benova^a, Emily Grundy^c, Daniel Laydon^d,
Bianca DeStavola^b

^a Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, United Kingdom

^b Centre for Statistical Methodology, Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, United Kingdom

^c Department of Geography, University of Cambridge, United Kingdom

^d Faculty of Medicine, Imperial College, United Kingdom

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Health inequalities
Later life
Socio Economic Position
Biomarkers
Causal mediation
Structural equation modelling
Life course theory
Ageing

ABSTRACT

The relative contribution of early or later life Socio Economic Position (SEP) to later life health is not fully understood and there are alternative hypotheses about the pathways through which they may influence health. We used data from the English Longitudinal Study of Ageing with a formal approach for the identification of mediating factors in order to investigate alternative hypotheses about life course influences on biomarkers of later life health. We found that early life SEP predicts physical health at least 65 years later. However, a more complicated pattern of associations than that implied by previous findings was also observed. Age group specific effects emerged, with current SEP dominating the effect on later life physical health and fibrinogen levels in participants under 65, while early life SEP had a more prominent role in explaining inequalities in physical health for men and women over 75. We extend previous findings on mid adulthood and early old age, to old age and the beginnings of late old age. The complexity of our findings highlights the need for further research on the mechanisms that underlie the association between SEP and later life health.

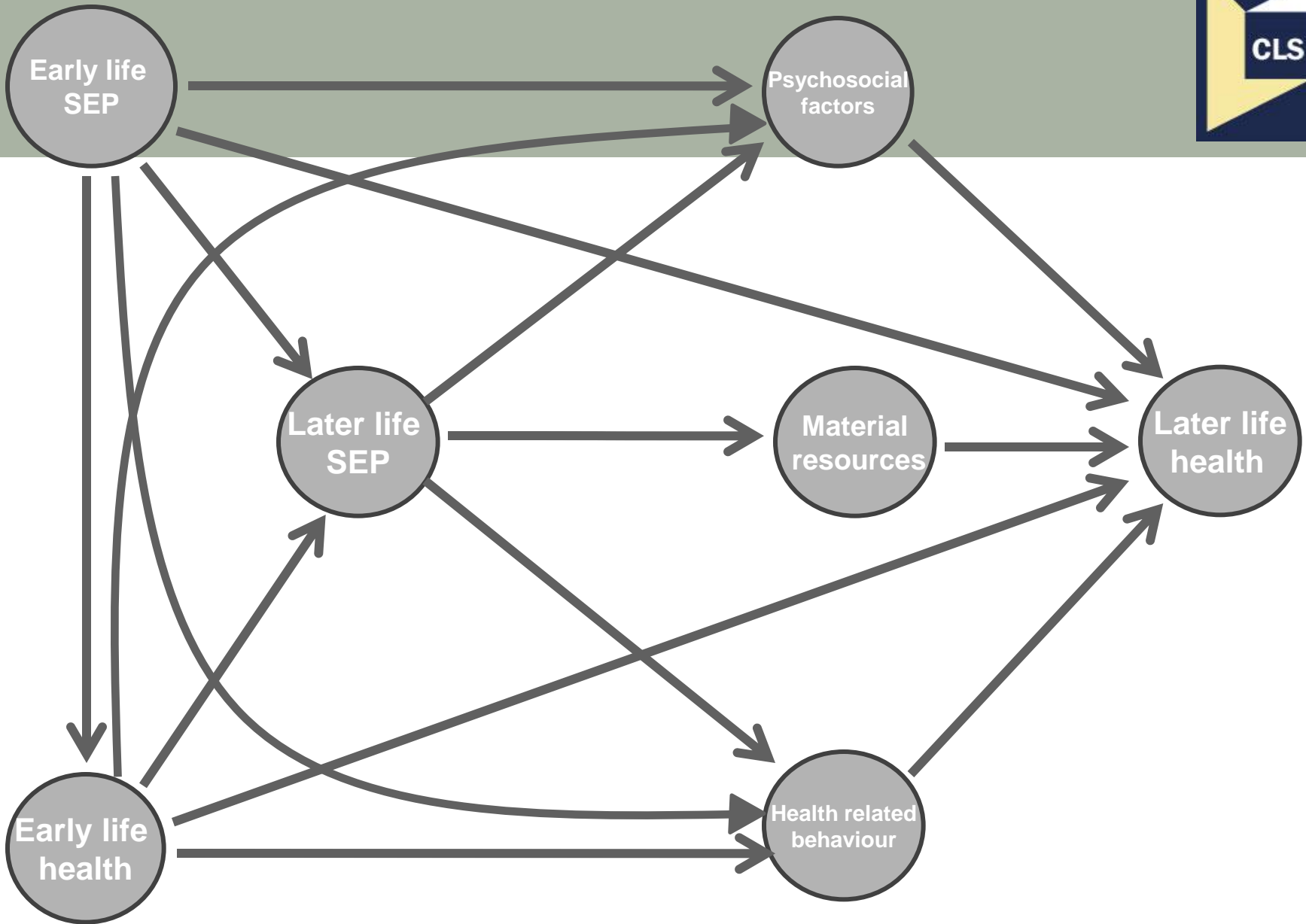
© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

Background

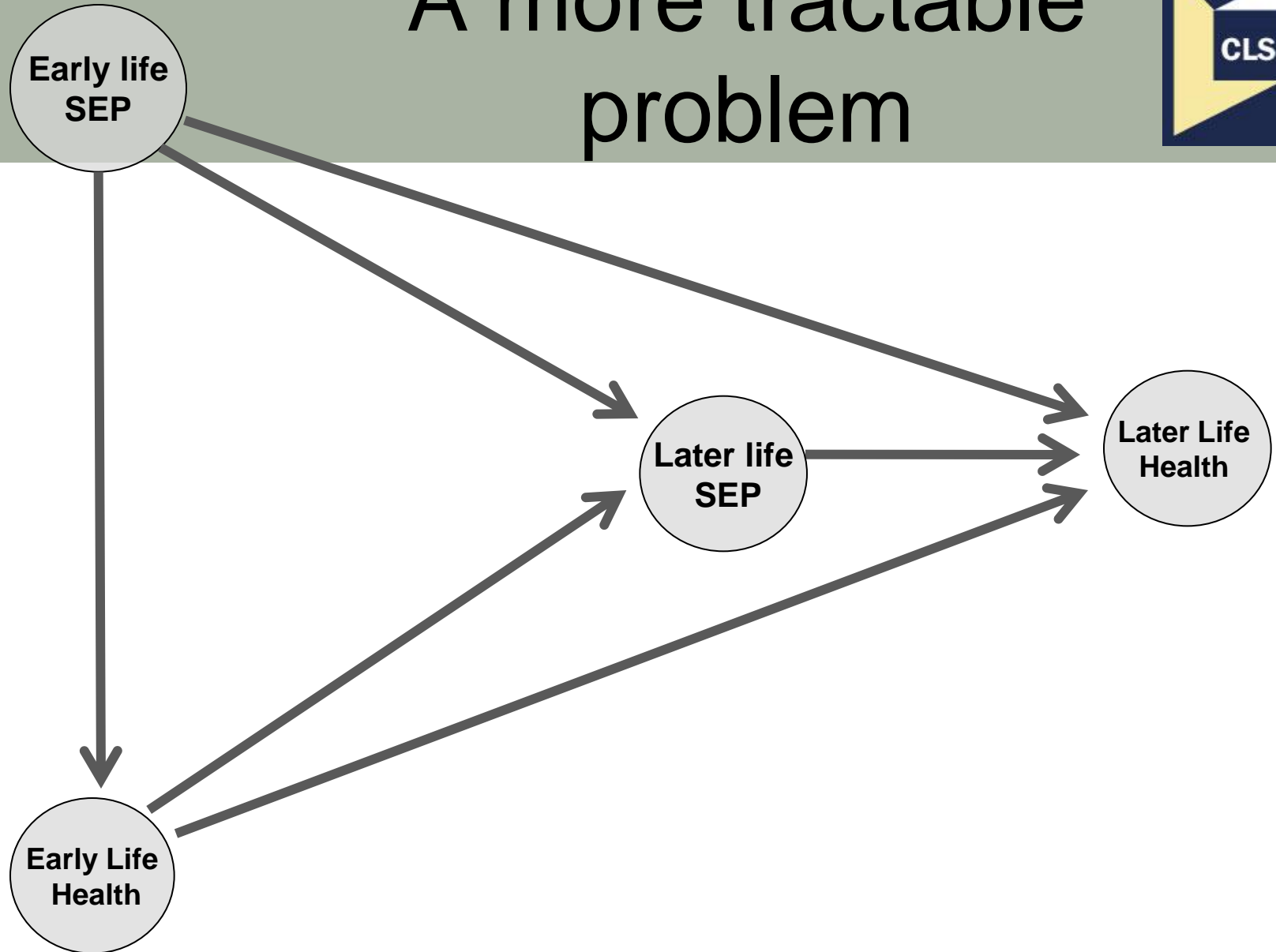
- In European countries with old age structures older people account for the majority of those in poor health
- Substantial inequalities in the health of different socio-economic groups persist in old age
- It has recently been shown that the economic costs of socioeconomic inequalities in health are in the order of €1000 billion, or **9.4% of European GDP**

- These observations suggest a particular need to investigate the influence of Socio Economic Position (SEP) on the well being of the older population
- There is a great potential for **shifting the overall distribution of risk** and improving average population health by eliminating or reducing the socioeconomic health gradient

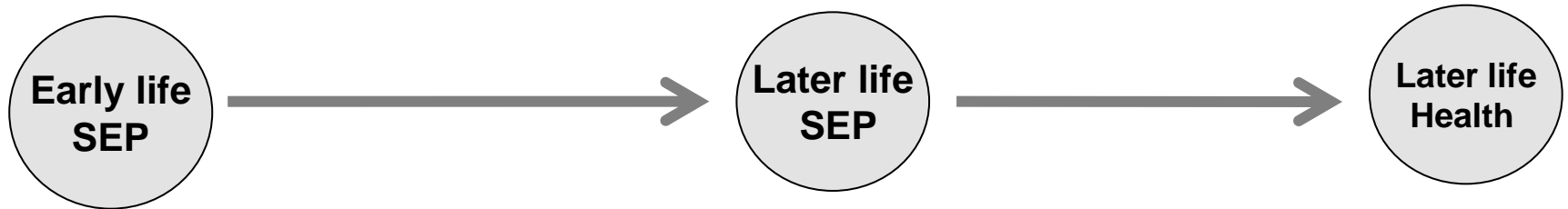
- We have **limited scientific evidence** on the individual and macro level mechanisms that underlie socio economic inequalities in health
- Although interventions on the exposure (SEP) are welcome, intervening on the mechanism that links SEP and health is a more realistic target
- In the current climate of financial austerity, efficient and cost effective policies are needed



A more tractable problem



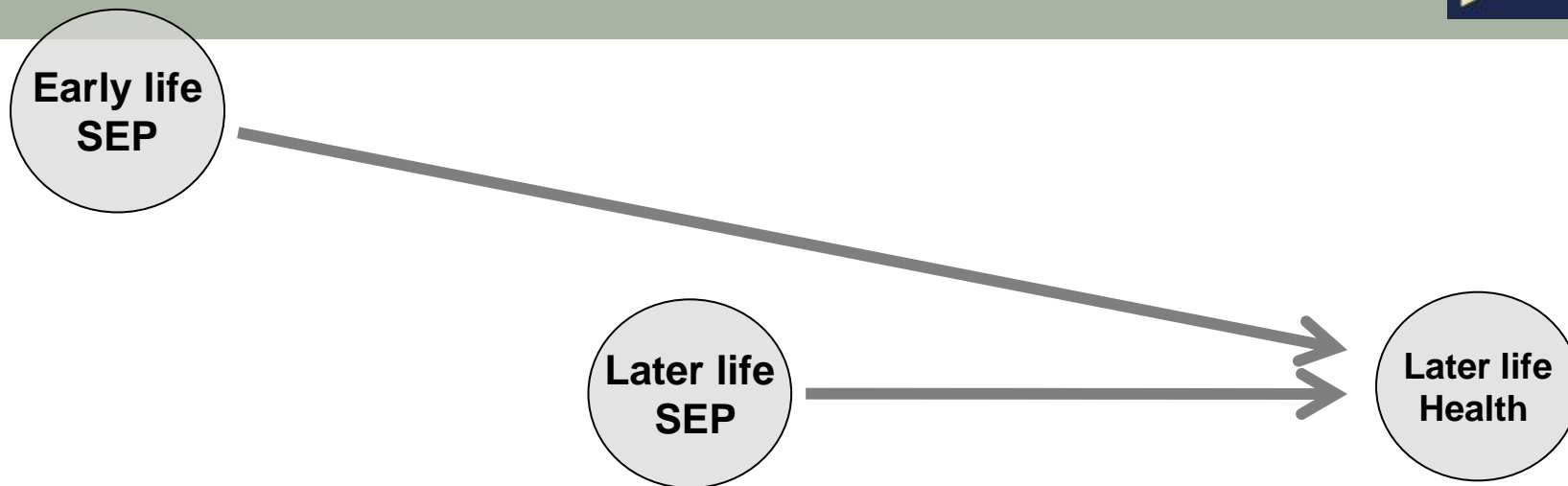
Chains of risk





Childhood/Early life

Accumulation



Social drift



Aims

- The major aim of the present study is to test **the relative contribution** of these hypotheses to later life health inequalities
- A first step in understanding the mechanism that underlies the association between SEP and health
- We need to assign reliable parameters to all these hypotheses – **it's all about mediation !!**

Causal Mediation

- Assigning reliable parameters and standard errors to direct and indirect effects is not so straightforward, especially if mediators and/or outcomes are binary/ordinal
- Natural vs controlled direct/indirect effects
- If interactions are present, they need to be taken into account
- Causal mediation provides meaningful quantities (parameters) that **potentially** have causal properties **under certain assumptions**
- Quantities derived from traditional methods do not have a causal interpretation under any assumption
- Within linear systems LSEMs can be used to formally quantify direct and indirect effects (actually all available methods give identical results)
- With binary outcomes and/or mediators other options such as G estimation, the G formula or Inverse Probability Weighting should be used

Not a free lunch

- Stronger assumptions about unmeasured confounding are needed for each additional mediator or moderator in the DAG.
- Within LSEM more than one mediators can be handled
- However, if the mediator and/or outcome are not continuous **only one mediator** can be handled (even in theory)
- A theoretical and practical issue
- Endogenous confounding, or the potential association of the exposure with a mediator – outcome measured confounder (the kite) can only be handled within LSEM

Sample

- We used data from the English Longitudinal Study of Ageing (ELSA), a nationally representative multi-purpose sample of the population aged 50 and over living in England
- We analysed a partially incomplete dataset (N = 7758), in which participants were included if they had at least one non missing observation in early life SEP indicators (ELSA Life history interview)
- Stratified by gender and age group (50-64, 65-74, 75+)

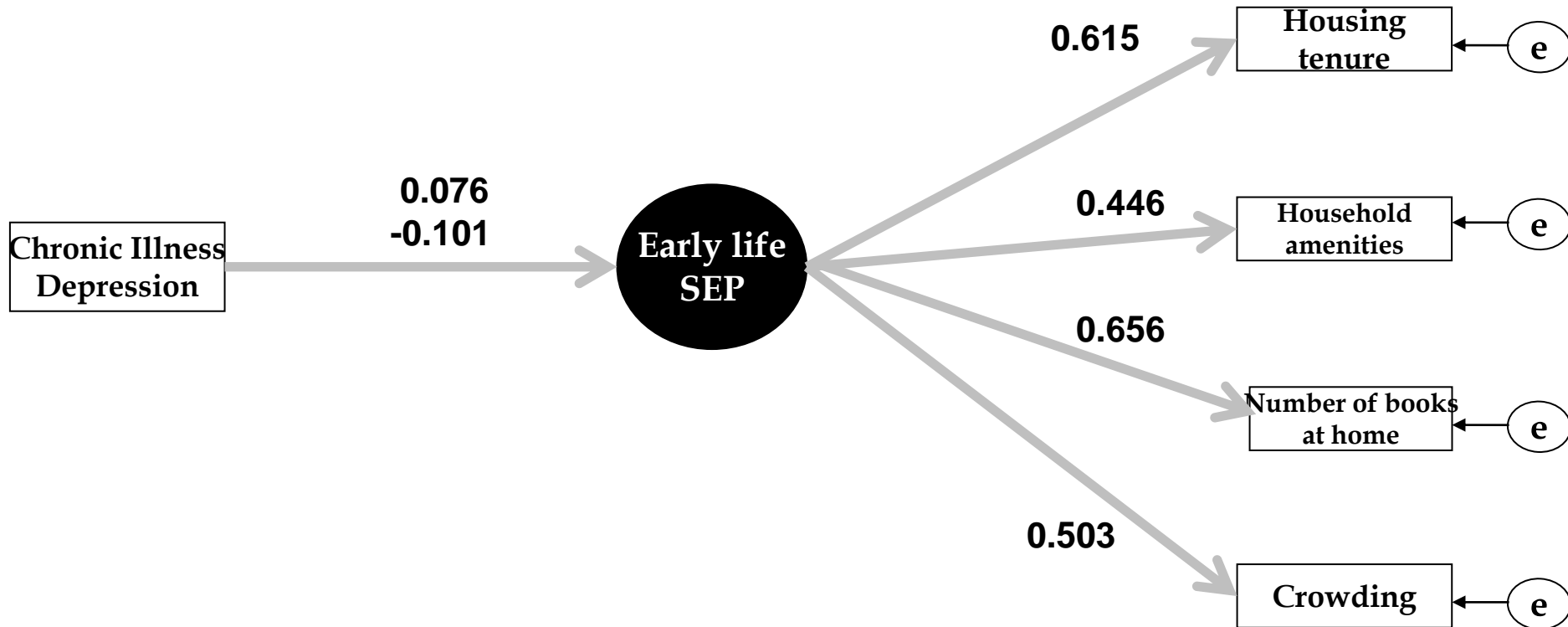
Measures

ELSA Life history interview – 2007

- Recollection of early life SEP (Age 10)
- Recollection of early life health (childhood - adolescence)

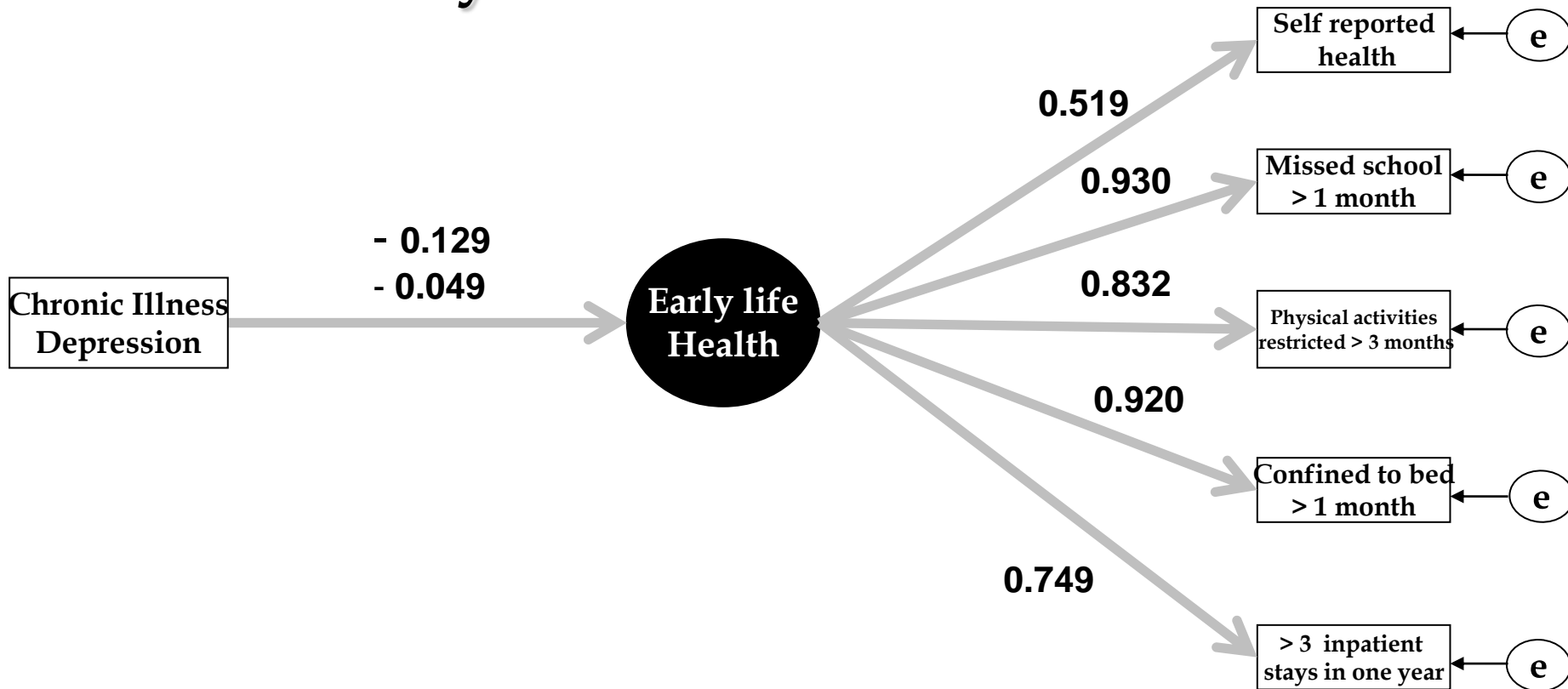
Recollection of early life SEP (age 10)

ELSA Life History Interview - Wave 3



Recollection of early life health

ELSA Life History Interview - Wave 3



ELSA Wave 4 - 2009

Mediator

- Later life SEP

Health outcomes

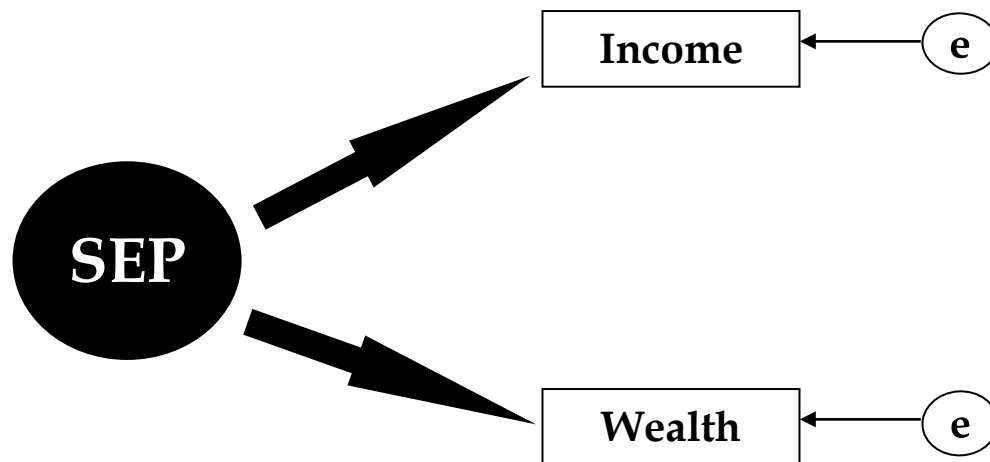
- Later life physical health
- Fibrinogen

Fibrinogen is the major coagulation protein in blood by mass; it is the precursor of fibrin and an important determinant of blood viscosity and platelet aggregation. Fibrinogen level is associated with an approximate doubling in risk of major cardiovascular disease outcomes (such as coronary heart disease and stroke) and of aggregate nonvascular mortality (mainly comprising cancer deaths)

Confounders

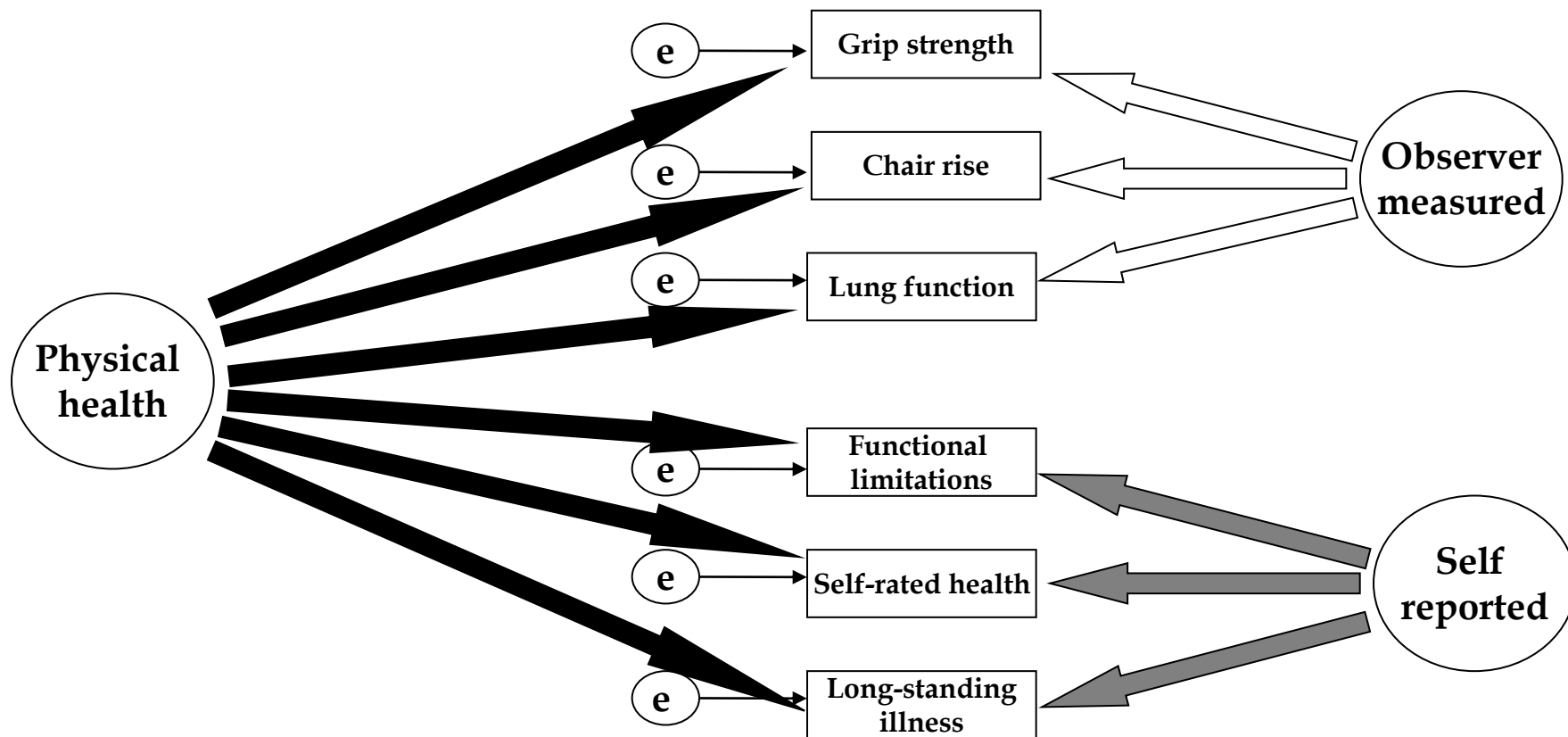
- Age, retirement status, marital status, number of children and cognitive ability were included in the structural model

Later life SEP, Wave 4 - 2009



Ploubidis, G., DeStavola, B., & Grundy, E. (2011). Health differentials in the older population of England: An empirical comparison of the materialist, lifestyle and psychosocial hypotheses. *BMC Public Health*, 11(1), 390.

Later life Physical Health, Wave 4 -2009

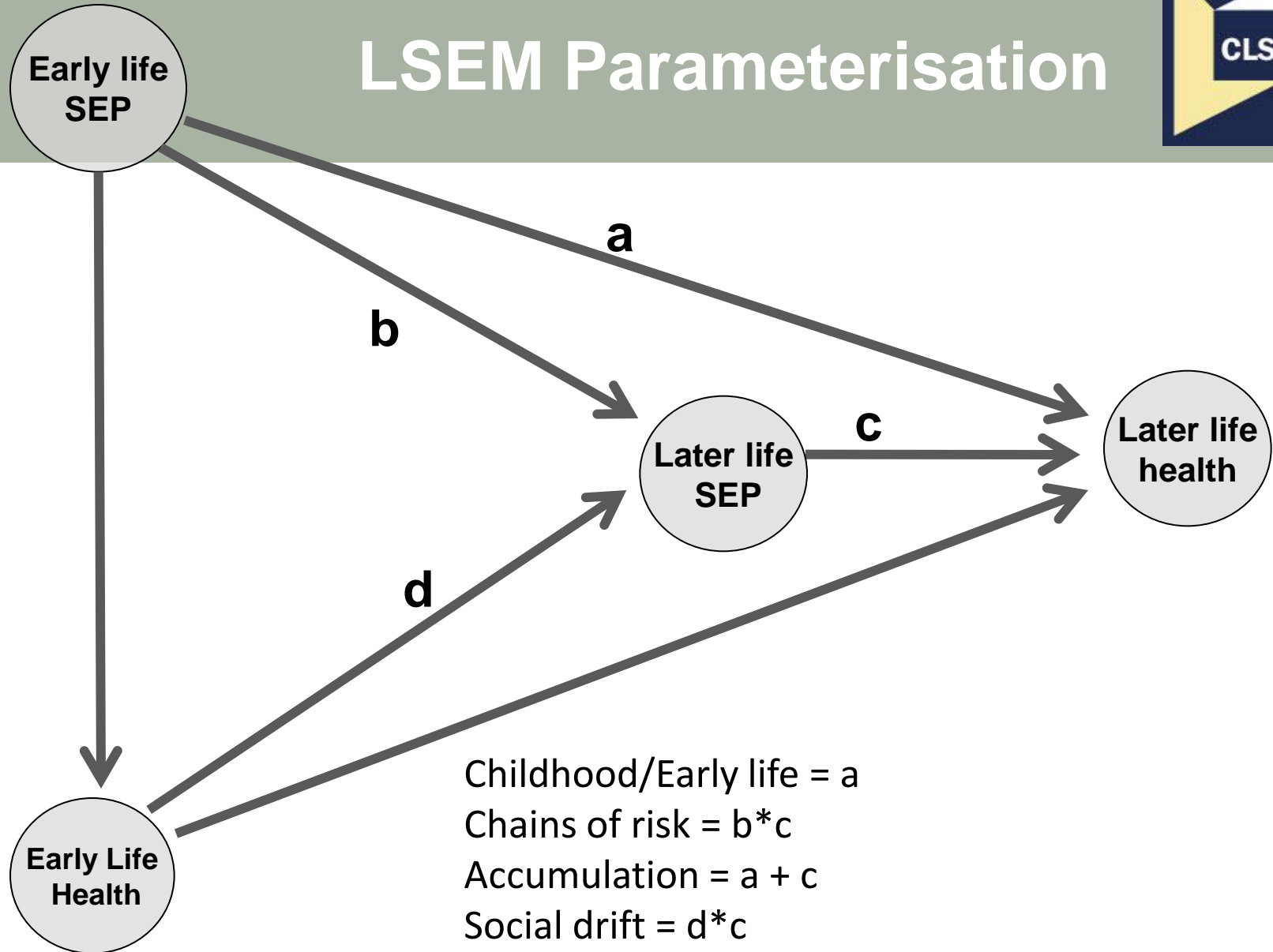


Ploubidis, GB., & Grundy, E. (2011). Health Measurement in Population Surveys: Combining Information from Self-reported and Observer Measured Health Indicators. *Demography*, 48(2), 699-724.

Statistical modelling

- The specification of each of the latent dimensions was carried out with models appropriate for combinations of binary, ordinal and continuous indicators
- A LSEM was then estimated in order to jointly model the predictors, mediators and health outcomes (adjusted for confounders)
- Preliminary results showed no evidence of interactions
- Missing data on Wave 4 mediators and health outcomes handled with FIML assuming a MAR mechanism
- Estimation with MLR in Mplus 6.12

LSEM Parameterisation



Childhood/Early life = a

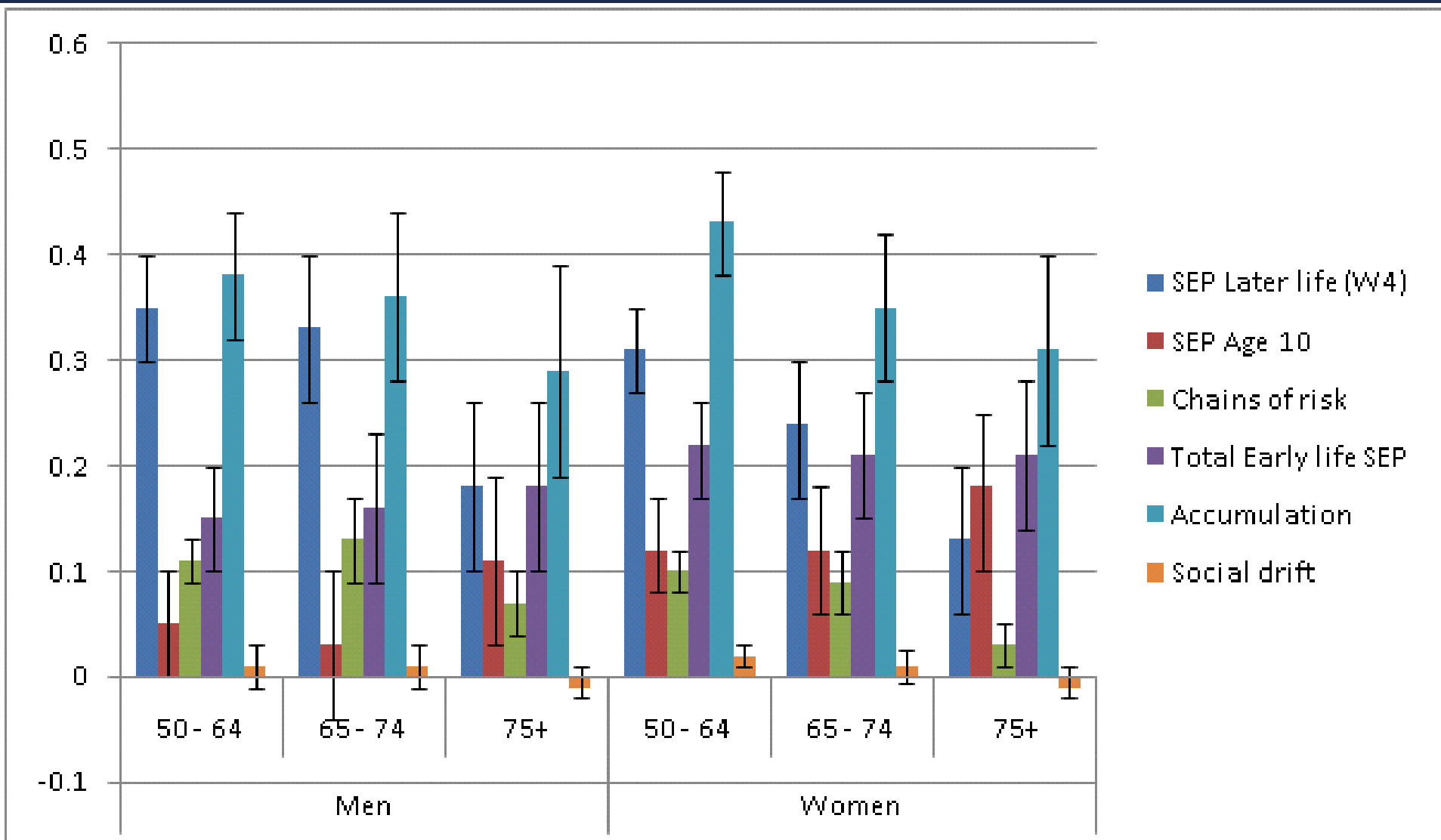
Chains of risk = $b * c$

Accumulation = $a + c$

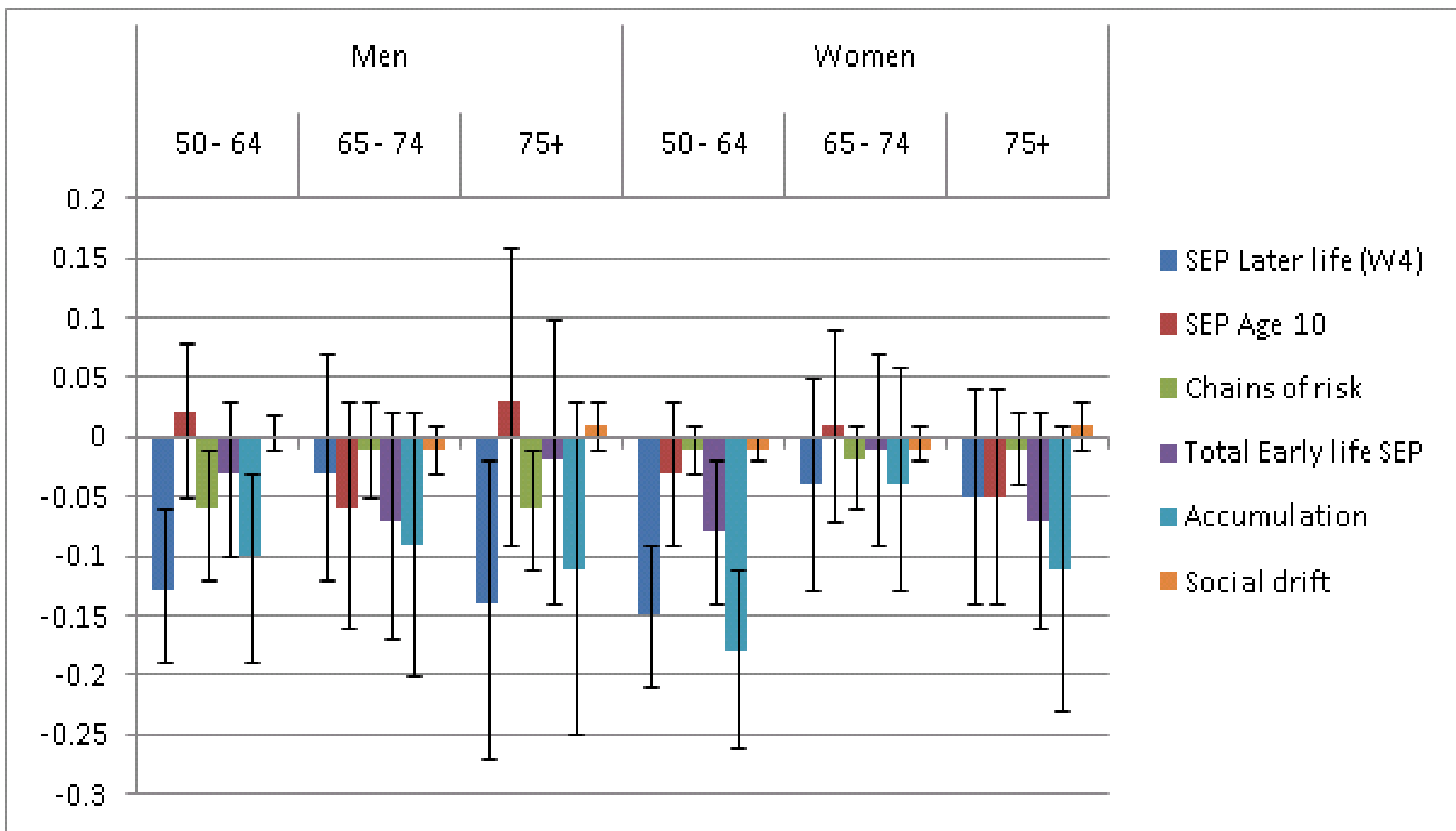
Social drift = $d * c$

Total effect of **Early life SEP** = $a + (b * c)$

Results - Physical Health

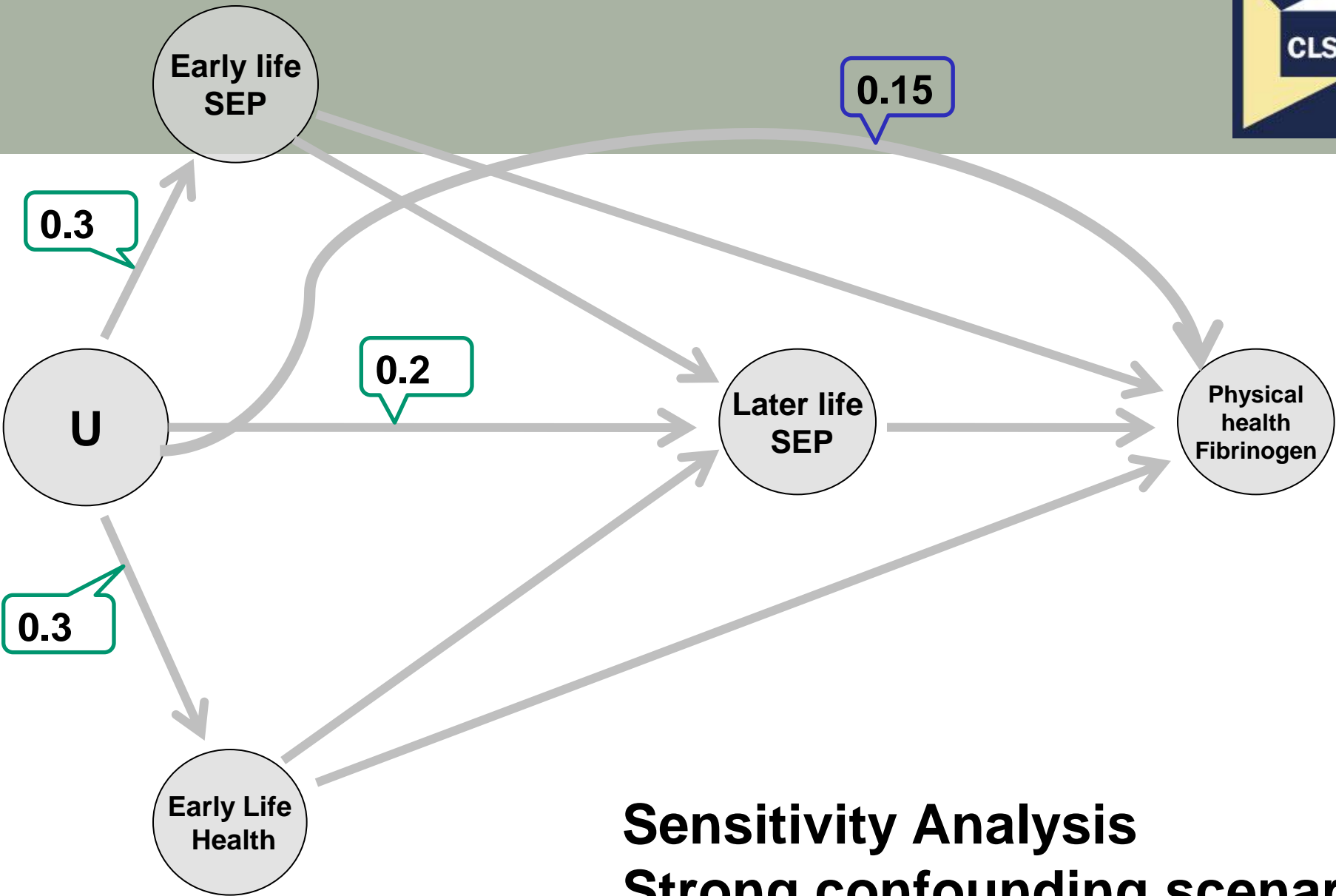


Results - Fibrinogen



Sensitivity analysis

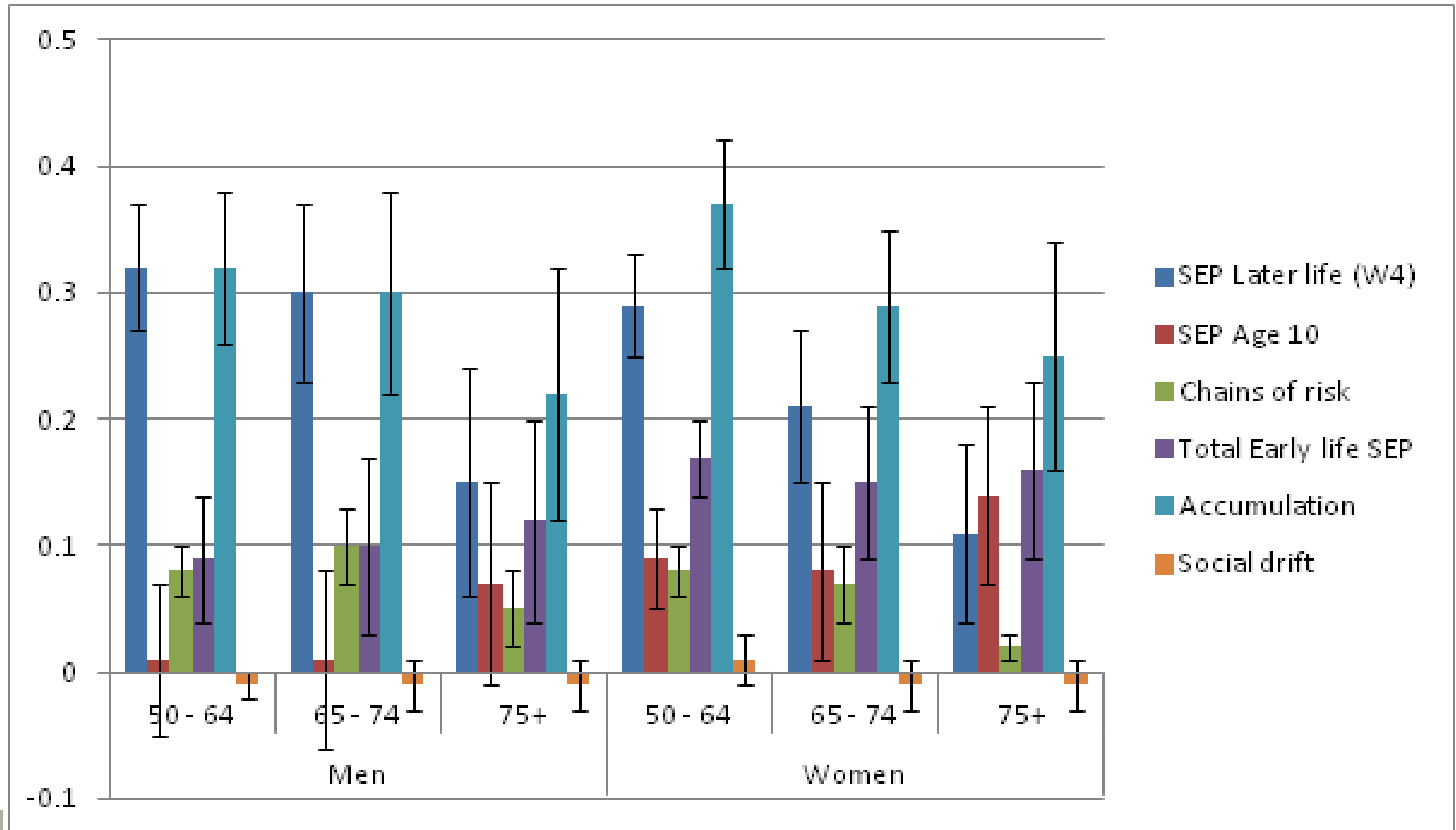
- Nice parameter estimates, but sequential ignorability implies no unmeasured confounders
- Sufficiently approximated* for the later life SEP – health association, but not for other parts of the DAG
- No data on parental characteristics such as cognitive ability and health status
- Results could reflect the effect of unmeasured parental characteristics



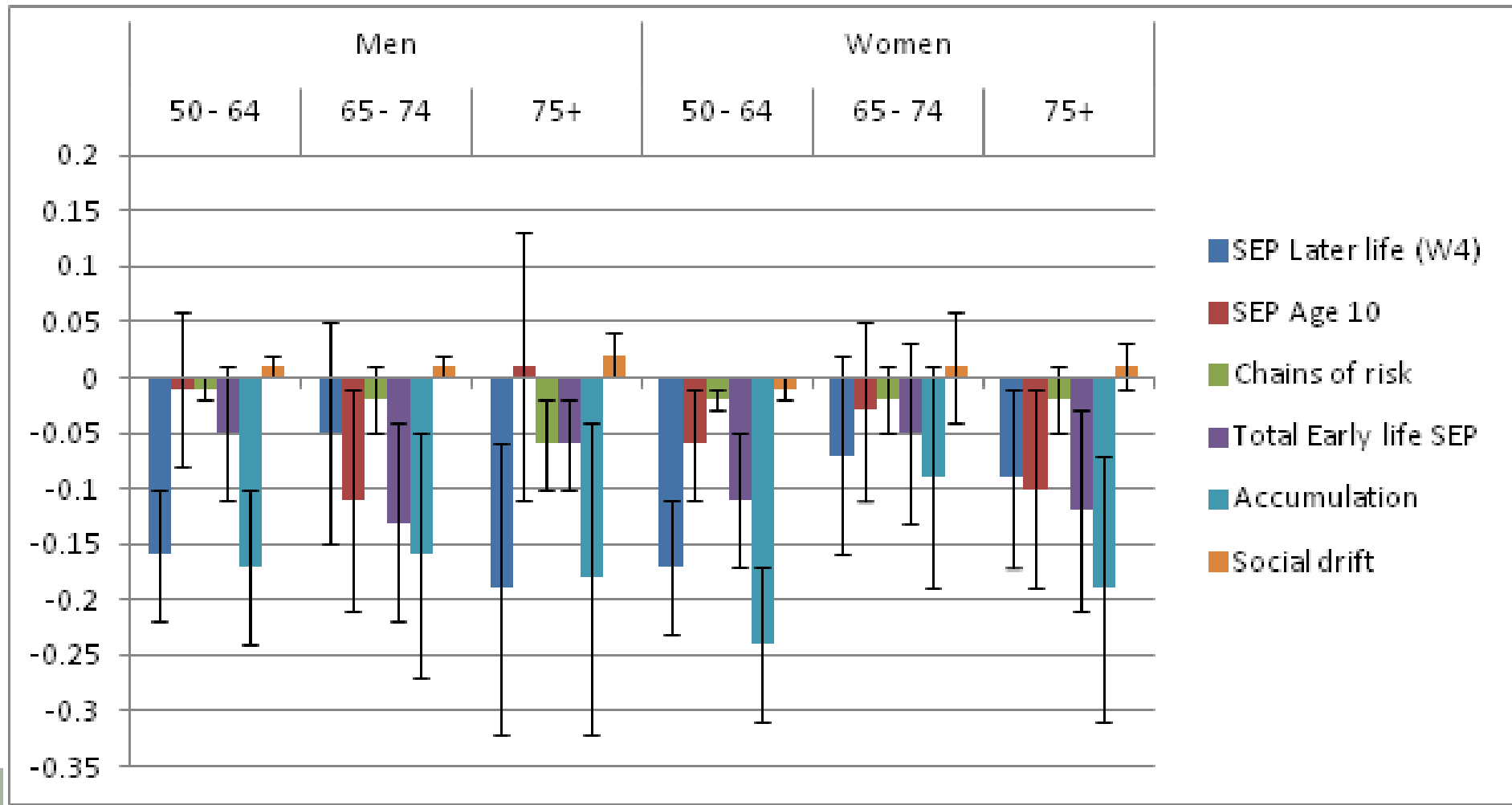
Sensitivity Analysis

Strong confounding scenario

Physical health sensitivity analysis



Fibrinogen sensitivity analysis

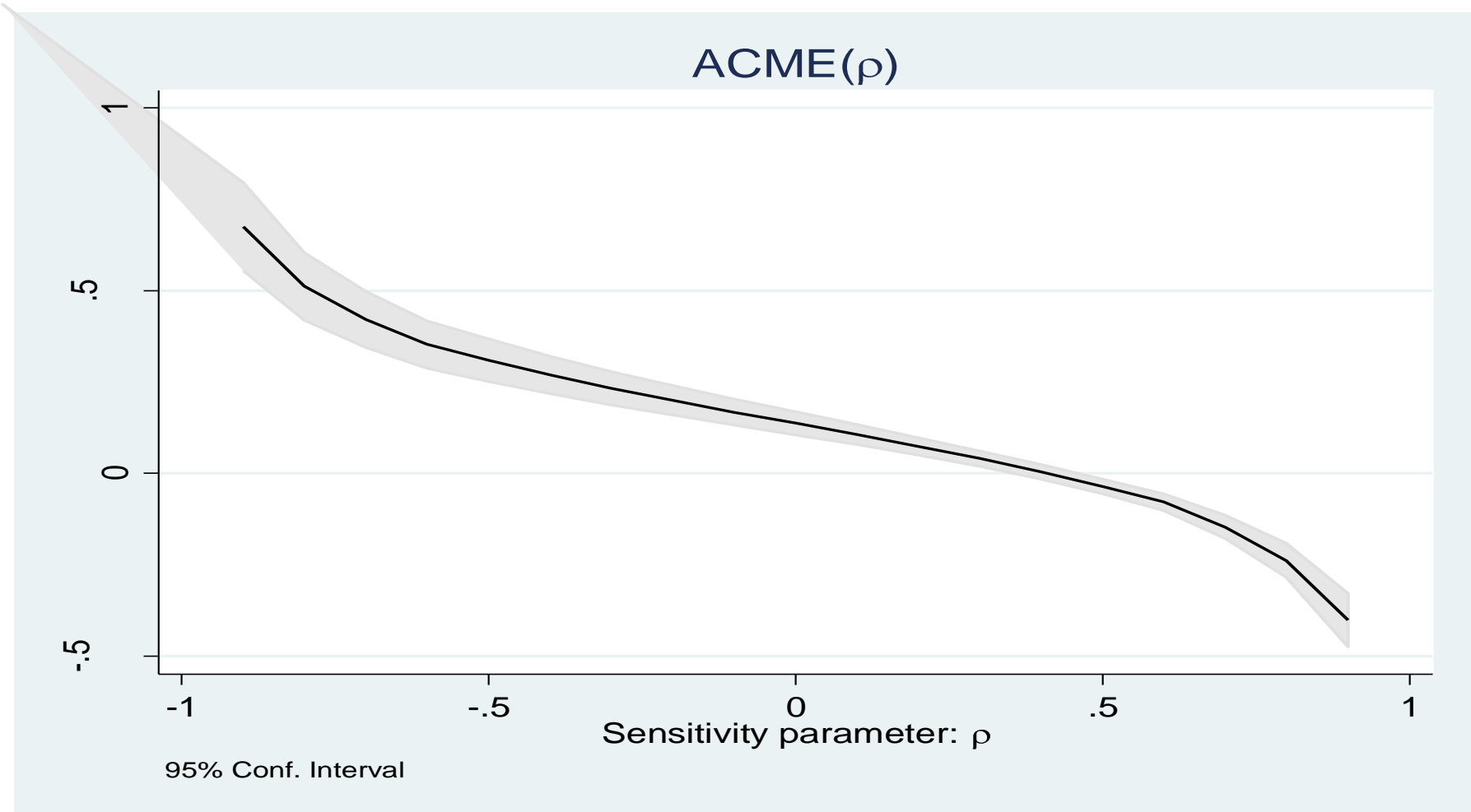


- We have included possible confounders of the later life SEP physical health/fibrinogen associations, but how about running some more sensitivity analyses?

Mediator – Outcome Confounding

- Medsens (Stata, R, Mplus)
- Employs the correlation between the residual variances (errors) of the models for the mediator and outcome
- Effects are computed given different fixed values of the residual covariance.
- The proposed sensitivity analysis asks the question of how large does ρ have to be for the mediation effect (Average Causal Mediation Effect – ACME) to disappear

Medsens Results



Summary

- Physical health more socially patterned than fibrinogen
- Accumulation, chains of risk and early life/childhood hypotheses confirmed
- No support for the social drift hypothesis
- Cohort and gender differences emerged with respect to the relative contribution of the confirmed hypotheses

Cohort & gender differences

- Accumulation of risk was dominated by the effect of later life SEP in those under 75
- Early life SEP (Critical period & Chains of risk) had the most prominent effect in those over 75
- Later life SEP the sole contributor to later life inequalities in fibrinogen levels
- Early life SEP more important for women

- Ageing process: younger cohorts expected to exhibit similar patterns of associations as they grow older
- Cohort specific effects due to the observed differences in early life SEP. Lower Early life SEP of older cohorts supports this idea. Those over 75 were born during the great depression of the 1930's
- Selection: Over 75's a selected sample of higher SEP within cohort survivors. Less SEP variance in this group supports this explanation

- Observational data – Causal inference a nearly alchemic task
- However, sensitivity analyses where confounders were simulated supported our results – but bias due to unknown unmeasured confounders cannot be ruled out
- Sensitivity parameters dependent on distribution of exogenous unmeasured confounder(s) – not identified non parametrically
- Retrospective data, only two timepoints

Another example

Life course partnership status and
biomarkers in mid-life: Evidence from
the 1958 British birth cohort

George B. Ploubidis, Richard J. Silverwood, Bianca
DeStavola & Emily Grundy

- Previous studies have shown that marital status is associated with health outcomes and mortality.
- With a few exceptions studies of marital status and health have considered only current marital status or transitions over relatively short periods
- The accumulated benefits and risks of marital status trajectories over the lifecourse have not been studied

- Furthermore, only a few studies have considered the association between non-marital cohabitation and health, a topic of increasing importance given that non –marital cohabitation is becoming more common
- Of those studies which have used measures of health, most have employed self-reported measures
- In the few studies where objective health indicators were used, sample sizes were relatively small
- In this study we are employing a population based birth cohort and a modelling approach that allows us to capture stability as well as change in partnership status over the lifecourse

Objectives

- Investigate the cumulative effect that different trajectories of partnership status over the life-course have on biomarkers in mid-life
- To what extent smoking accounts for the association between life course partnership status and biomarkers in mid-life?

- The British 1958 birth cohort includes all persons born in England, Scotland and Wales during one week in March 1958
- Cohort members have been followed-up periodically from birth into adulthood. Our outcomes are derived from the clinical examination in their home undertaken in 2002 – 2004
- Marital status and cohabitation have been recorded from sweep 4 (1981) when participants were 23 years old
- We are using data from sweep4 (1981, age 23), sweep5 (1991, age 33), sweep6 (2000 age 42) and the biomedical survey (2002-2004 age 44 – 46) to derive the partnership status trajectories
- Early life SEP and health are derived from sweeps 0 – 3 (ages 1 – 16)

Outcomes:

- Inflammatory and haemostatic biomarkers: Fibrinogen, C – Reactive Protein (CRP), Von Willebrand Factor (VWF), Tissue plasminogen activator antigen (TPA) and Fibrin D- dimer (Ddimer).
- Metabolic syndrome: MS was characterized using the International Diabetes Federation definition
- Respiratory function: Scores on Force Vital Capacity - the maximum amount of air a person can expel from the lungs after a maximum inhalation.

Measured confounders:

- All models adjusted for early life SEP, cognitive ability @ 10, early life health status, education @ 23, self reported health status @ 23, BMI @ 23 and various lab processing related variables.

Measures II – Partnership status indicators



		Men		Women				Men		Women	
		f	%	f	%			f	%	f	%
Married at 23	No	4083	65.2	2861	45.6	Cohabiting at 23	No *	5923	94.6	5813	92.8
	Yes	2179	34.8	3409	54.4		Yes	336	5.4	454	7.2
Married at 33	No	1660	31.0	1569	27.9	Cohabiting at 33	No	4780	89.2	5090	90.4
	Yes	3701	69.0	4063	72.1		Yes	581	10.8	542	9.6
Married at 40	No	1644	29.4	1667	28.9	Cohabiting at 40	No	5059	90.5	5251	91.1
	Yes	3948	70.6	4098	71.1		Yes	532	9.5	514	8.9
Married at 42	No	1220	27.0	1325	28.9	Cohabiting at 42	No	3921	87.8	3967	87.9
	Yes	3303	73.0	3262	71.1		Yes	543	12.2	547	12.1
Remarried by 42	No	8714	90.8	7945	88.7						
	Yes	881	9.2	1014	11.3						



Missing data - Rubin's framework

Missing Completely At Random (MCAR)

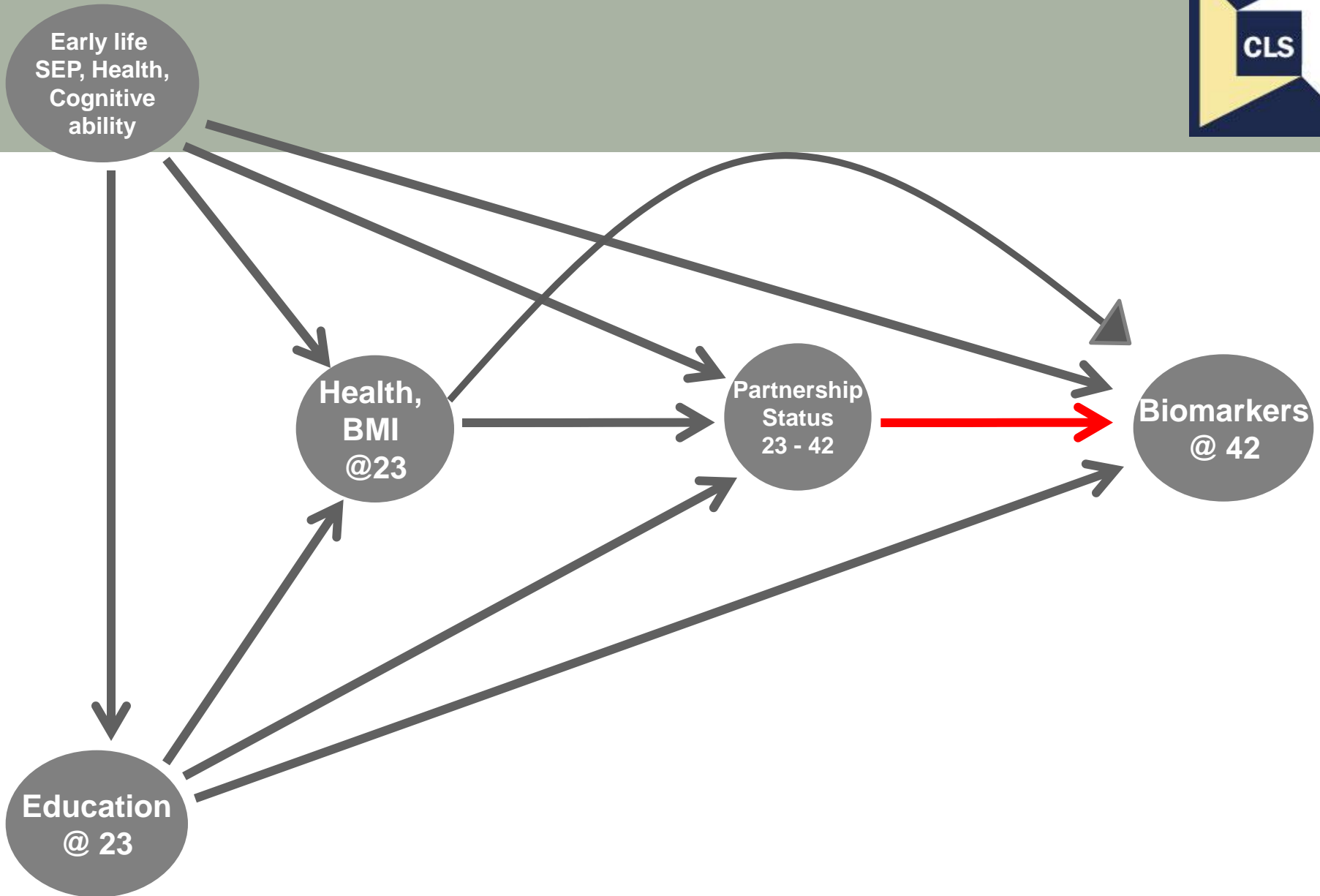
- Almost impossible in life course studies, some variables are known causes of attrition – partially testable
- In the 1958 cohort early life characteristics influence attrition
- Complete case analysis and non principled approaches (mean imputation, last observation forward) will likely be severely biased

Missing At Random (MAR)

- Implies that if all the variables that are responsible for the missing data generating mechanism are complete and included in the model, this "mechanism" can be ignored - Principled approaches (MI, FIML, IPW)
- Conditional ignorability is a reasonable assumption for life course studies

Not Missing At Random (NMAR)

- More complex scenario, solutions only for models with repeated measures. Pattern mixture model, Dingle – Kenward selection model.



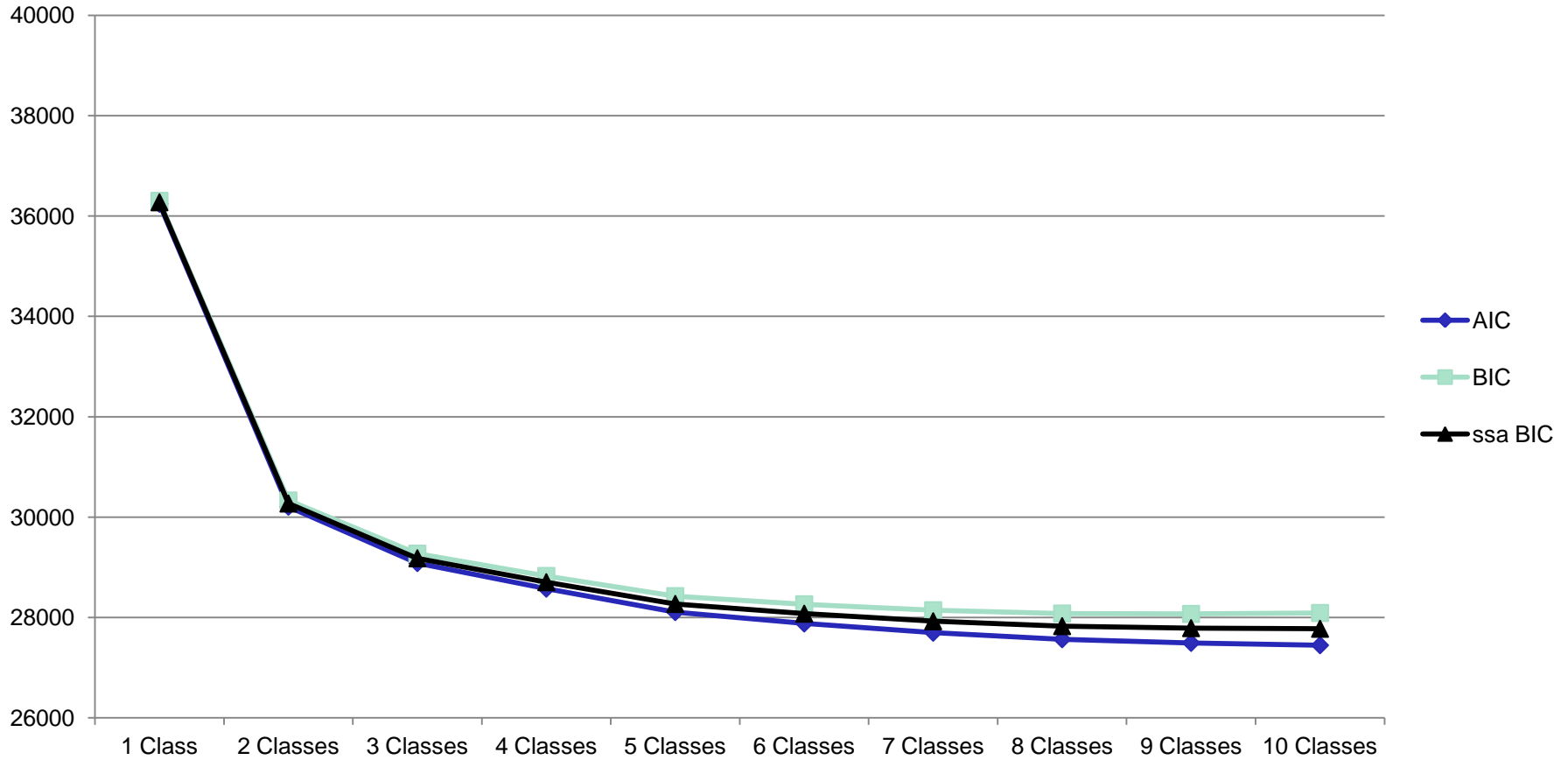
- Longitudinal Latent Class Analysis – Semi parametric model, introduces a discrete latent variable to capture common variation in the observed marital status and cohabitation indicators
- Forms latent classes - groups (trajectories) based on the pattern of responses to the observed indicators
- Captures heterogeneity – does not assume a single rate of change/single curve for all
- A data reduction method - in theory the number of possible response patterns in theory is $5^2 = 512$.
- However since participants who are married cannot simultaneously be non – married cohabiters, there are three responses available at each wave
- Distinct response patterns $2 \times (3^4) = 162$
- In this instance LCA is used to summarise these patterns creating longitudinal profiles in a parsimonious way that can be used in further analysis with appropriate link functions for the nature of the outcomes (linear and logit models)
- All models in Mplus 7.0, estimated with MLR, Monte Carlo integration.
- Missing data handled with FIML assuming MAR

Results I – Model selection



Men	Parameters	Log-Likelihood	AIC	BIC	ssa BIC	Entropy	BLRT	p
1 Class	9	-18113.063	36244.126	36302.557	36273.958	1.000		
2 Classes	19	-15085.063	30208.127	30331.480	30271.105	0.927	6056.001	0.001
3 Classes	29	-14513.203	29084.406	29272.682	29180.530	0.946	1143.721	0.001
4 Classes	39	-14248.346	28574.693	28827.892	28703.964	0.931	529.713	0.001
5 Classes	49	-14004.856	28107.713	28425.835	28270.130	0.909	486.981	0.001
6 Classes	59	-13881.343	27880.687	28263.731	28076.250	0.922	247.026	0.001
7 Classes	69	-13779.315	27696.629	28144.612	27925.339	0.925	204.058	0.001
8 Classes	79	-13704.204	27566.407	28079.298	27828.264	0.912	150.222	0.001
9 Classes	89	-13657.883	27493.767	28071.580	27788.770	0.921	92.641	0.001
10 Classes	99	-13624.711	27447.421	28090.156	27775.570	0.924	66.347	0.001

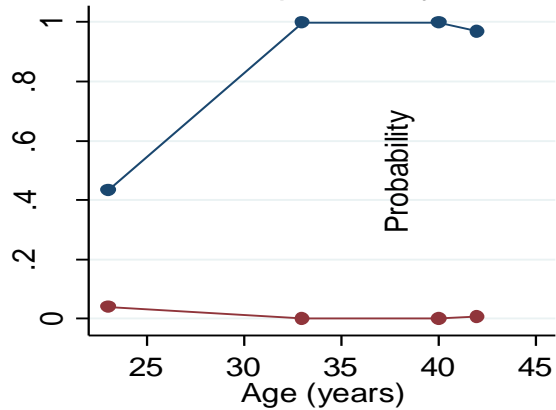
Men



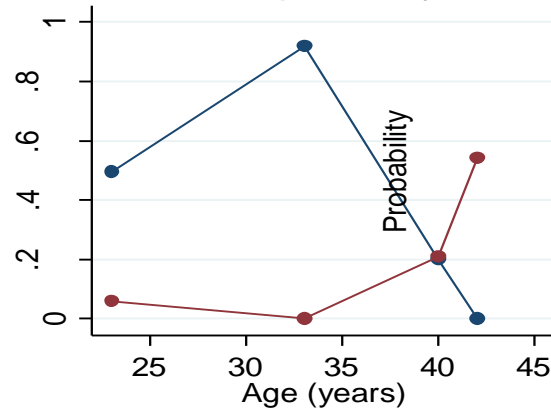
Men



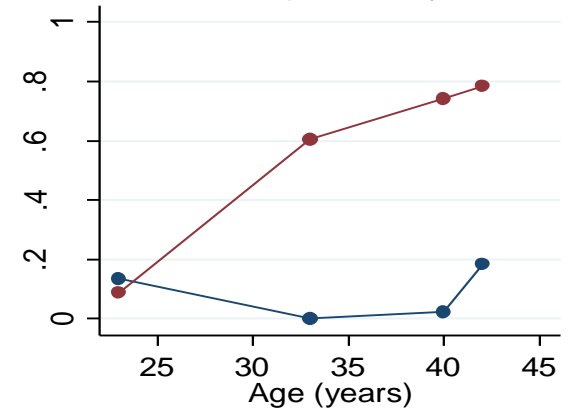
Class 1 (N = 3010, 61.7%)
Remarried probability: 0.125



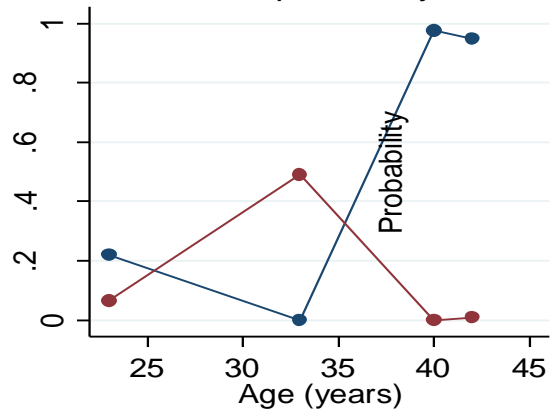
Class 2 (N = 401, 8.2%)
Remarried probability: 0.177



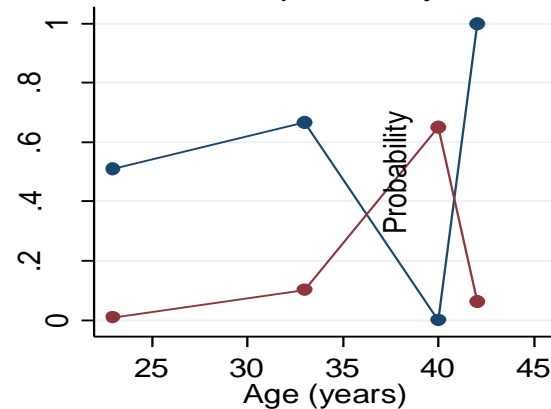
Class 3 (N = 362, 7.4%)
Remarried probability: 0.036



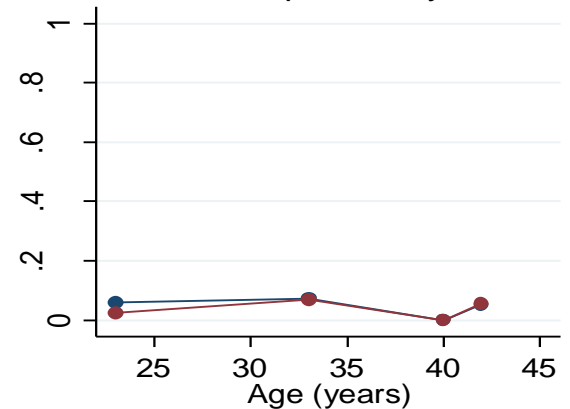
Class 4 (N = 462, 9.5%)
Remarried probability: 0.379



Class 5 (N = 100, 2.1%)
Remarried probability: 0.783



Class 6 (N = 542, 11.1%)
Remarried probability: 0.022

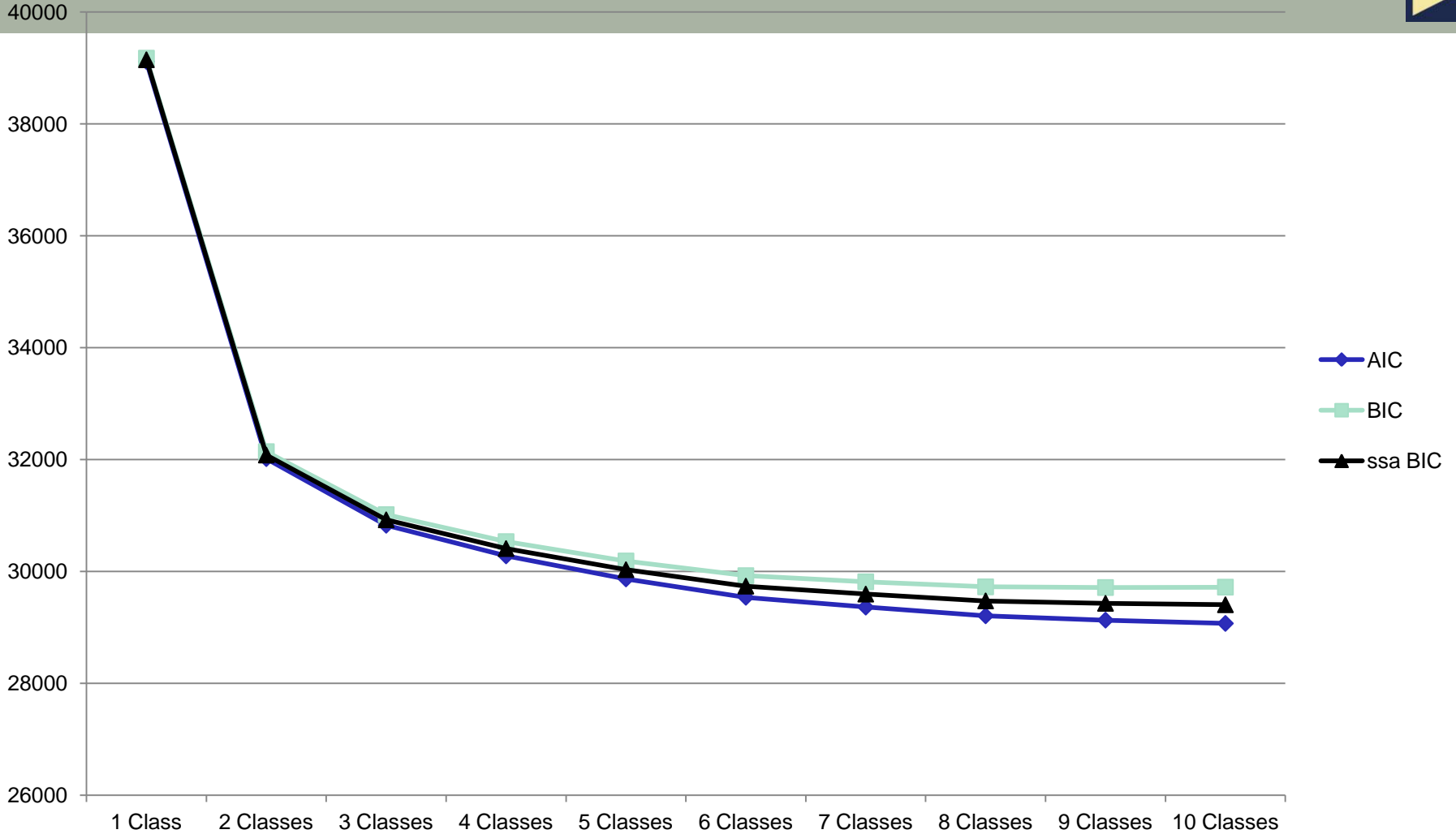


Results II – Model Selection



Women	Parameters	Log-Likelihood	AIC	BIC	ssa BIC	Entropy	BLRT	p
1 Class	9	-19548.128	39114.255	39173.193	39144.594	1.000		
2 Classes	19	-15989.385	32016.771	32141.196	32080.820	0.945	7117.485	0.001
3 Classes	29	-15383.938	30825.875	31015.787	30923.635	0.962	1210.895	0.001
4 Classes	39	-15100.217	30278.435	30533.834	30409.905	0.940	567.440	0.001
5 Classes	49	-14884.450	29866.899	30187.785	30032.079	0.918	431.536	0.001
6 Classes	59	-14710.590	29539.180	29925.553	29738.071	0.905	347.719	0.001
7 Classes	69	-14612.640	29363.279	29815.139	29595.880	0.916	195.901	0.001
8 Classes	79	-14524.971	29207.942	29725.289	29474.253	0.935	175.337	0.001
9 Classes	89	-14476.328	29130.656	29713.489	29430.677	0.933	97.286	0.001
10 Classes	99	-14436.959	29071.918	29720.239	29405.650	0.938	78.737	0.001

Women

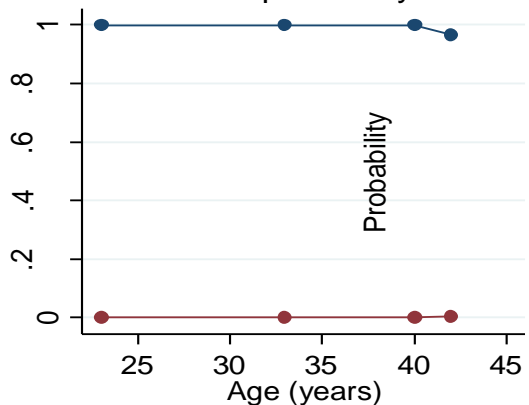


Women



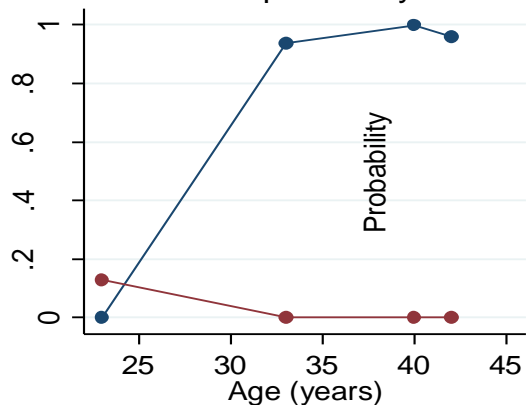
Class 1 (N = 2168, 42.0%)

Remarried probability: 0.139



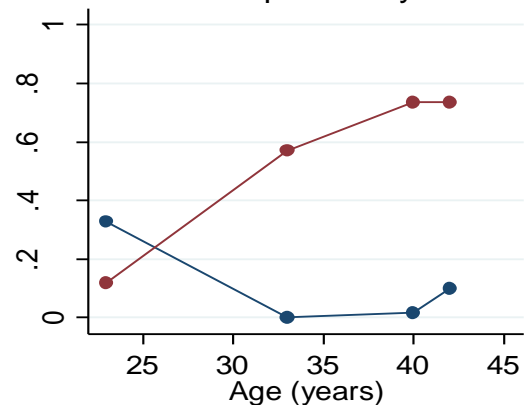
Class 2 (N = 1199, 23.2%)

Remarried probability: 0.121



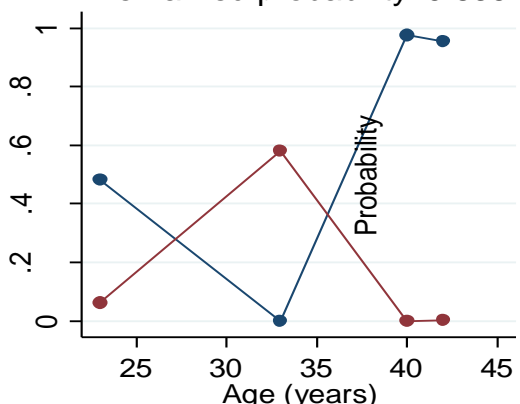
Class 3 (N = 415, 8.0%)

Remarried probability: 0.080



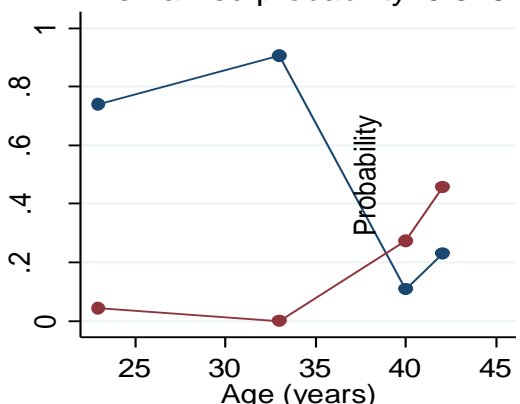
Class 4 (N = 291, 5.6%)

Remarried probability: 0.659



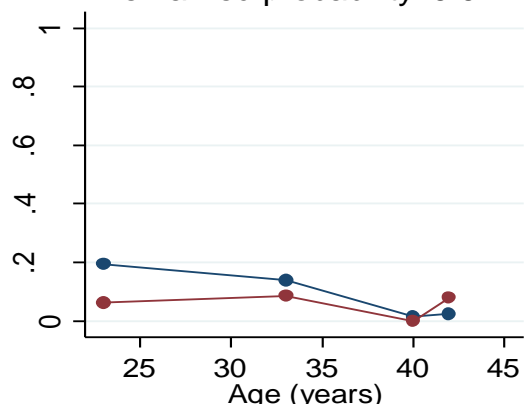
Class 5 (N = 446, 8.6%)

Remarried probability: 0.316



Class 6 (N = 641, 12.4%)

Remarried probability: 0.024



Results - Men



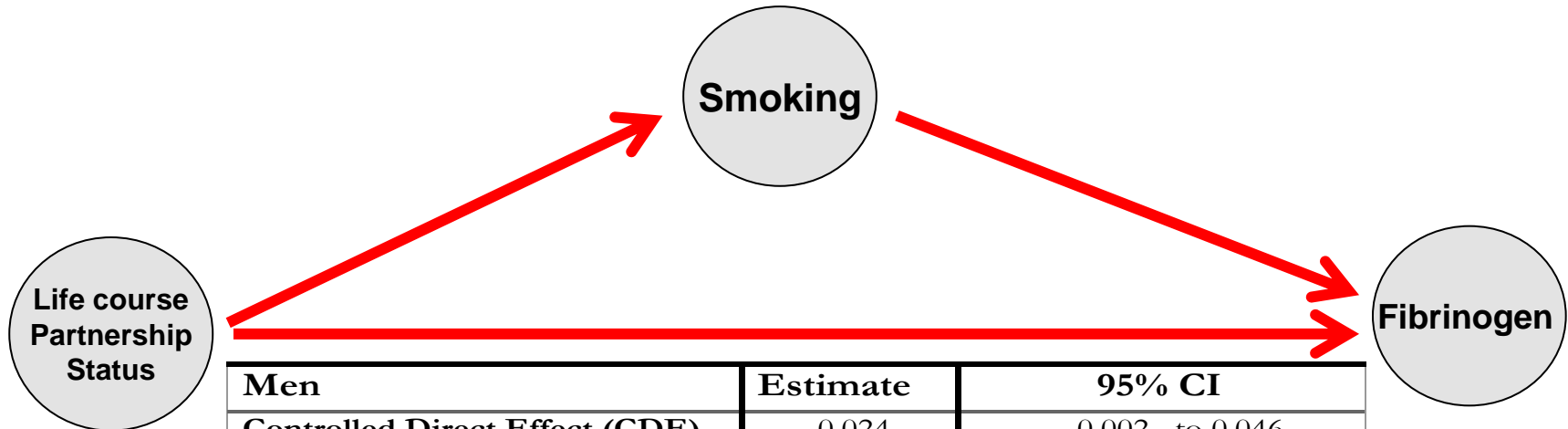
<i>Fibrinogen</i>		<i>CRP</i>		<i>VWF</i>		<i>TPA</i>	
Class1	0	0		0		0	
Class2	0.019 -0.006 to 0.045	0.131 -0.006 to 0.268		0.029 -0.010 to 0.069		0.036 -0.029 to 0.100	
Class3	0.010 -0.014 to 0.033	-0.013 -0.161 to 0.135		-0.001 -0.046 to 0.044		0.026 -0.041 to 0.093	
Class4	0.008 -0.015 to 0.031	0.008 -0.113 to 0.128		0.003 -0.036 to 0.041		0.003 -0.053 to 0.059	
Class5	0.028 -0.021 to 0.076	0.064 -0.215 to 0.342		-0.006 -0.081 to 0.070		0.045 -0.078 to 0.169	
Class6	0.034 0.012 to 0.056	0.148 0.025 to 0.270		0.020 -0.016 to 0.057		0.061 0.006 to 0.116	
<i>Ddimer</i>		<i>Metabolic Syndrome</i>		<i>FVC</i>			
Class1	0	1		0			
Class2	0.035 -0.036 to 0.105	0.756 0.575 to 0.993		0.071 -0.026 to 0.168			
Class3	0.043 -0.029 to 0.114	1.067 0.808 to 1.410		-0.112 -0.214 to -0.009			
Class4	0.054 -0.014 to 0.123	1.077 0.843 to 1.376		-0.076 -0.168 to 0.015			
Class5	0.016 -0.116 to 0.148	0.759 0.457 to 1.261		0.050 -0.129 to 0.229			
Class6	0.038 -0.029 to 0.105	0.867 0.677 to 1.111		-0.130 -0.225 to -0.035			

Results - Women



<i>Fibrinogen</i>			<i>CRP</i>		<i>VWF</i>		<i>TPA</i>	
Class1	0		0		0		0	
Class2	-0.018	-0.035 to -0.002	-0.087	-0.186 to 0.011	-0.011	-0.038 to 0.017	-0.036	-0.081 to 0.010
Class3	0.001	-0.023 to 0.023	-0.032	-0.173 to 0.110	-0.014	-0.055 to 0.027	-0.011	-0.075 to 0.053
Class4	0.010	-0.018 to 0.038	0.195	0.028 to 0.361	0.038	-0.008 to 0.085	-0.026	-0.111 to 0.058
Class5	-0.012	-0.037 to 0.014	-0.013	-0.161 to 0.134	0.003	-0.040 to 0.045	0.012	-0.058 to 0.082
Class6	0.028	0.006 to 0.050	0.029	-0.104 to 0.162	0.022	-0.015 to 0.058	-0.030	-0.088 to 0.028
<i>Ddimer</i>			<i>Metabolic Syndrome</i>		<i>FVC</i>			
Class1	0		1		0			
Class2	-0.002	-0.048 to 0.043	1.009	0.810 to 1.257	0.054	0.002 to 0.106		
Class3	0.016	-0.046 to 0.079	0.673	0.481 to 0.943	0.026	-0.046 to 0.098		
Class4	-0.037	-0.105 to 0.031	1.043	0.712 to 1.528	0.004	-0.094 to 0.101		
Class5	-0.064	-0.131 to 0.002	0.778	0.560 to 1.081	0.033	-0.043 to 0.109		
Class6	-0.012	-0.070 to 0.047	0.776	0.581 to 1.038	-0.051	-0.116 to 0.014		

Causal Mediation



Men	Estimate	95% CI
Controlled Direct Effect (CDE)	0.024	0.002 to 0.046
Natural Direct Effect (NDE)	0.024	0.002 to 0.046
Natural Indirect Effect (NIE)	0.090	0.004 to 0.014
Total Effect	0.033	0.011 to 0.055
Women	Estimate	95% CI
Controlled Direct Effect (CDE)	0.033	0.011 to 0.055
Natural Direct Effect (NDE)	0.033	0.011 to 0.055
Natural Indirect Effect (NIE)	0.007	0.004 to 0.011
Total Effect	0.040	0.019 to 0.062

Conclusion

- Partnership status patterns are associated with biomarkers in mid adulthood
- The observed effects differed between men and women implying that the mechanisms that link partnership status and health may be gender specific
- In men, those that never married or cohabited had significantly higher levels on three haemostatic function biomarkers as well as worse respiratory function compared to men that were married and remained married for the duration of the observation period

Conclusion - II

- In women those that married in mid/late 20's or early 30's and remained married for the whole observation period had the best health
- Women that never married or cohabited had worse health compared to married women
- However, this effect was only manifested in fibrinogen levels, indicating that not marrying or cohabiting is less detrimental in women compared to men or that being married appears to be more beneficial to men

Conclusion - III

- We found that with the exception of worse respiratory function in men, non-marital cohabitation has similar effects to being married on mid-life health
- Not married cohabiters of both genders did not differ from married participants in the biomarkers used in our study
- For both genders transitions from and to marriage or non-marital cohabitation do not have a detrimental effect on mid-life health
- Smoking partly mediates the association between partnership status and fibrinogen

- Despite the wealth of the 1958 cohort, bias due to unknown unmeasured confounders cannot be ruled out, although sensitivity analysis where potential confounders were simulated supported our results
- The longitudinal typology captured the cumulative effect over 21 years of trajectories of partnership status in biomarkers in mid-life. Investigation of the short term effects of events such as marital dissolution was not possible with this approach
- Data on partnership status were based on self-reports. Although the latent variable specification of our longitudinal typology controls for measurement error, extreme bias (a participant misreporting in all nine indicators of our typology) may have influenced our results
- Our results can only be generalised to those born in 1958 and perhaps to other cohorts born close to this year

- Life course studies are methodologically challenging
- However, solutions with reasonable assumptions exist
- Directed Acyclic Graphs (DAGs) are very useful for life course studies

- DAGs should be used to express realistically complex hypotheses
- This will help us isolate the effect(s) of interest and choose the most appropriate model(s) and assumptions
- Isolate and reliably quantify policy modifiable effects
- Not always straightforward, but doable

Thank you for your attention