Modeling categorical variables 2:

Ordinal and multinomial models

Aleš Vomáčka 2023

Lesson Goals

- Learn how to work with ordinal binomial models
- Learn how to work with multinomial binomial models

Ordinal regression

Ordinal variables

Models for ordinal data

- Likert items
- Coarsed frequencies

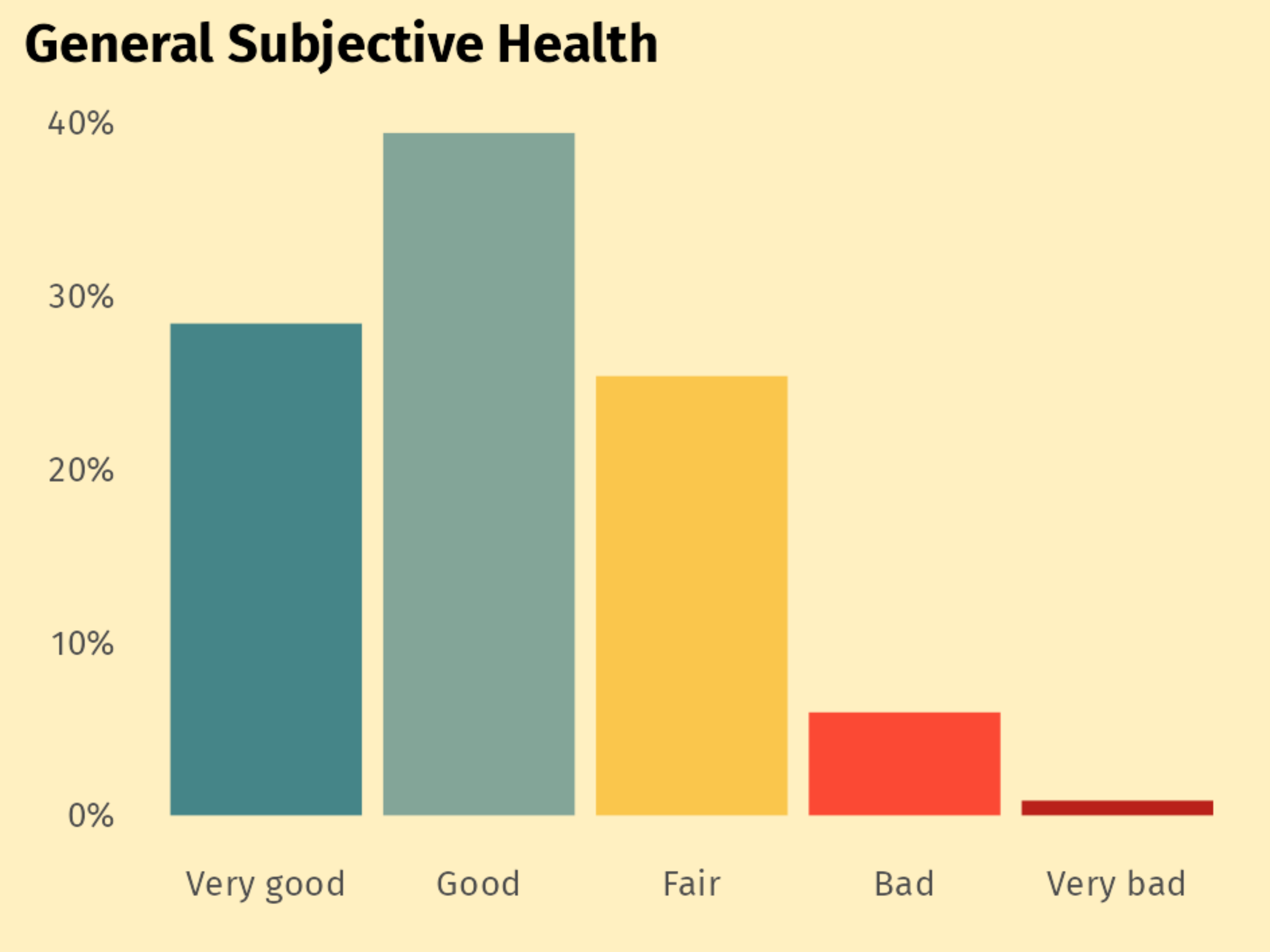
Extension of binary binomial models, most of what you've learned applies.

Sometimes called cumulative link models.

Subjective General health

How is your health in general?
Would you say it is ...

1. Very good
2. Good
3. Fair
4. Bad
5. Very bad

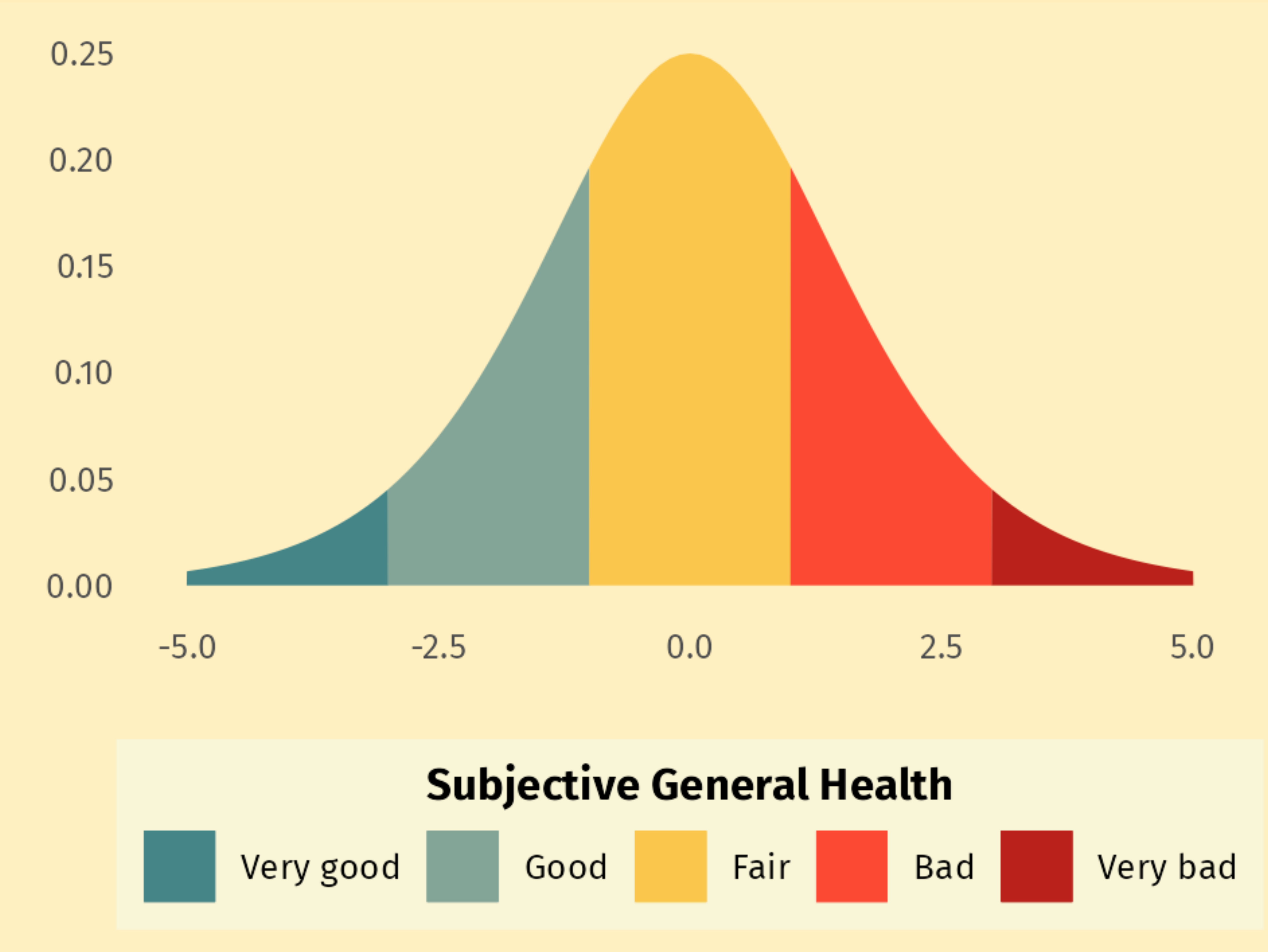


Ordinal model

$$logit(x <= X) \sim Binomial(X\beta)$$

Sequentially predicting the probability (or logit) of being in a category lower than x.

(Where x is are the response categories).



Predicting health by age and education

Being older increases the value of the latent scale

Predictor	Coefficient
Age	0.062
Education: Elementary	-
Education: High school (without diploma)	-0.248
Education: High school (with diploma)	-0.291
Education: University	-0.532
Intercepts:	
Very good|Good	0.610
Good|Fair	2.831
Fair|Bad	4.962
Bad|Very Bad	7.340

High value on the latent scale means you are in a bad shape

Interpreting ordinal regression

Same problems as with binary models:

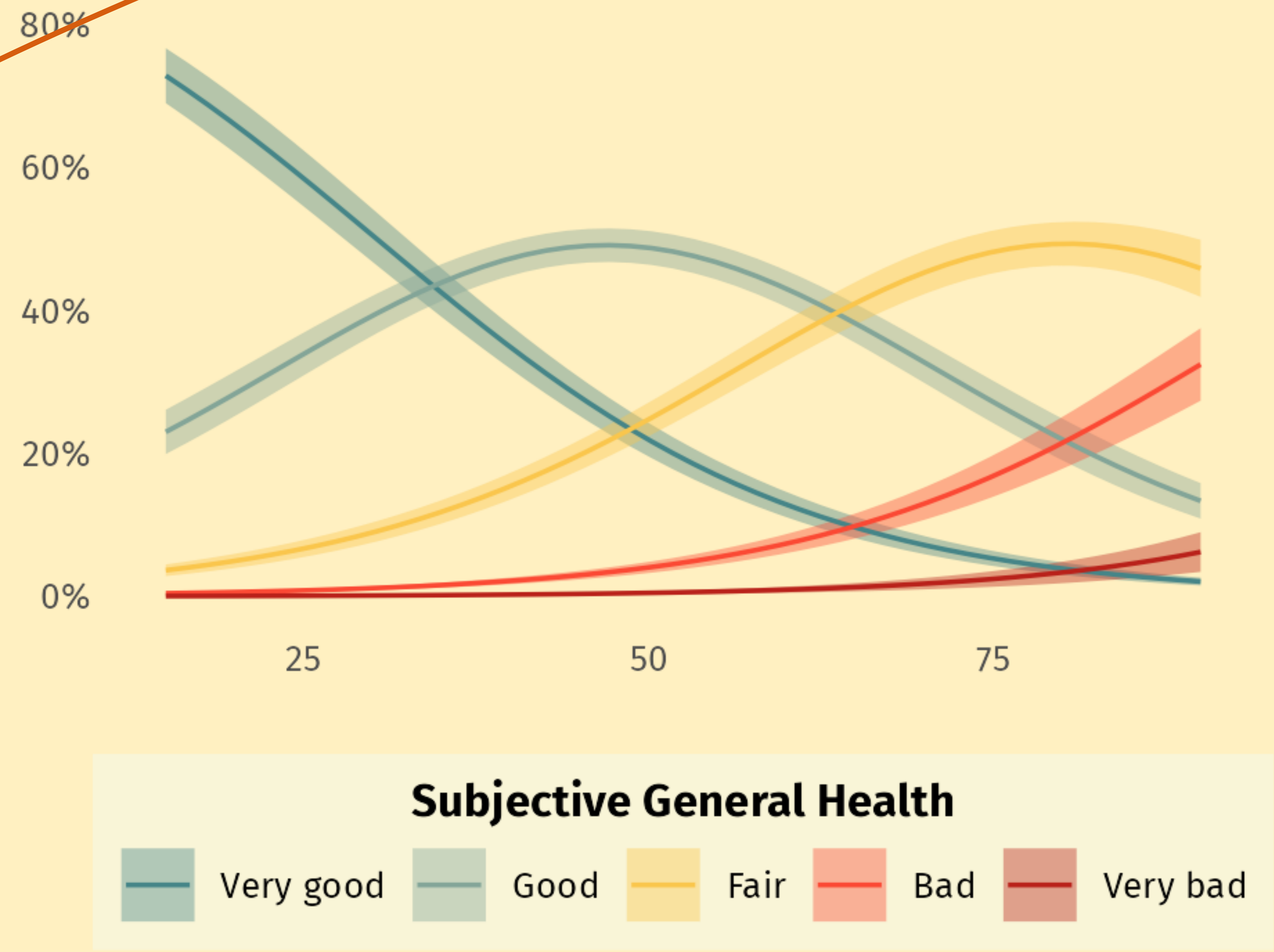
- Logits are an unintuitive scale
- Logits are non-collapsible

Same solution - marginal effects on probability scale

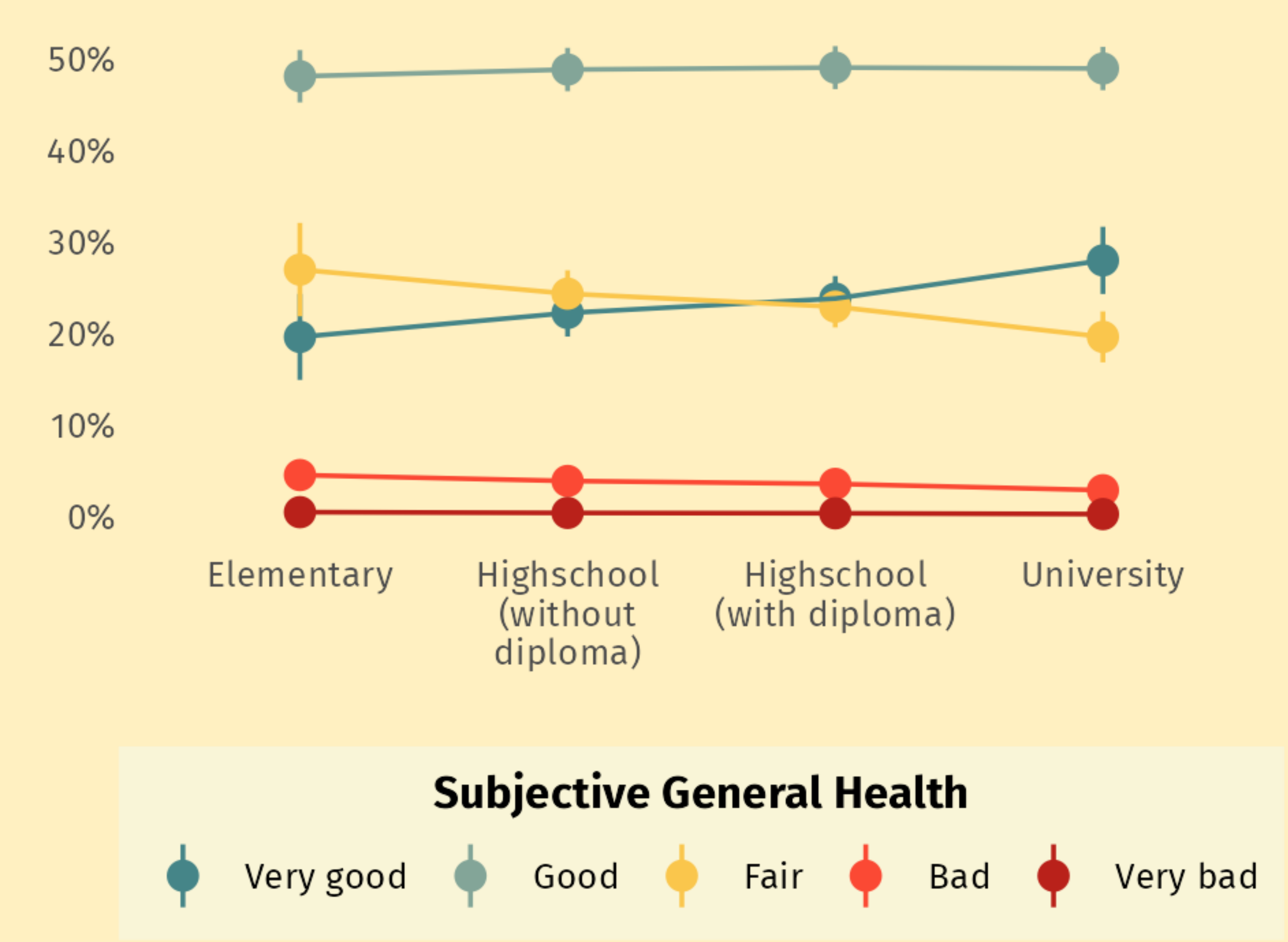
Marginal effects for Age

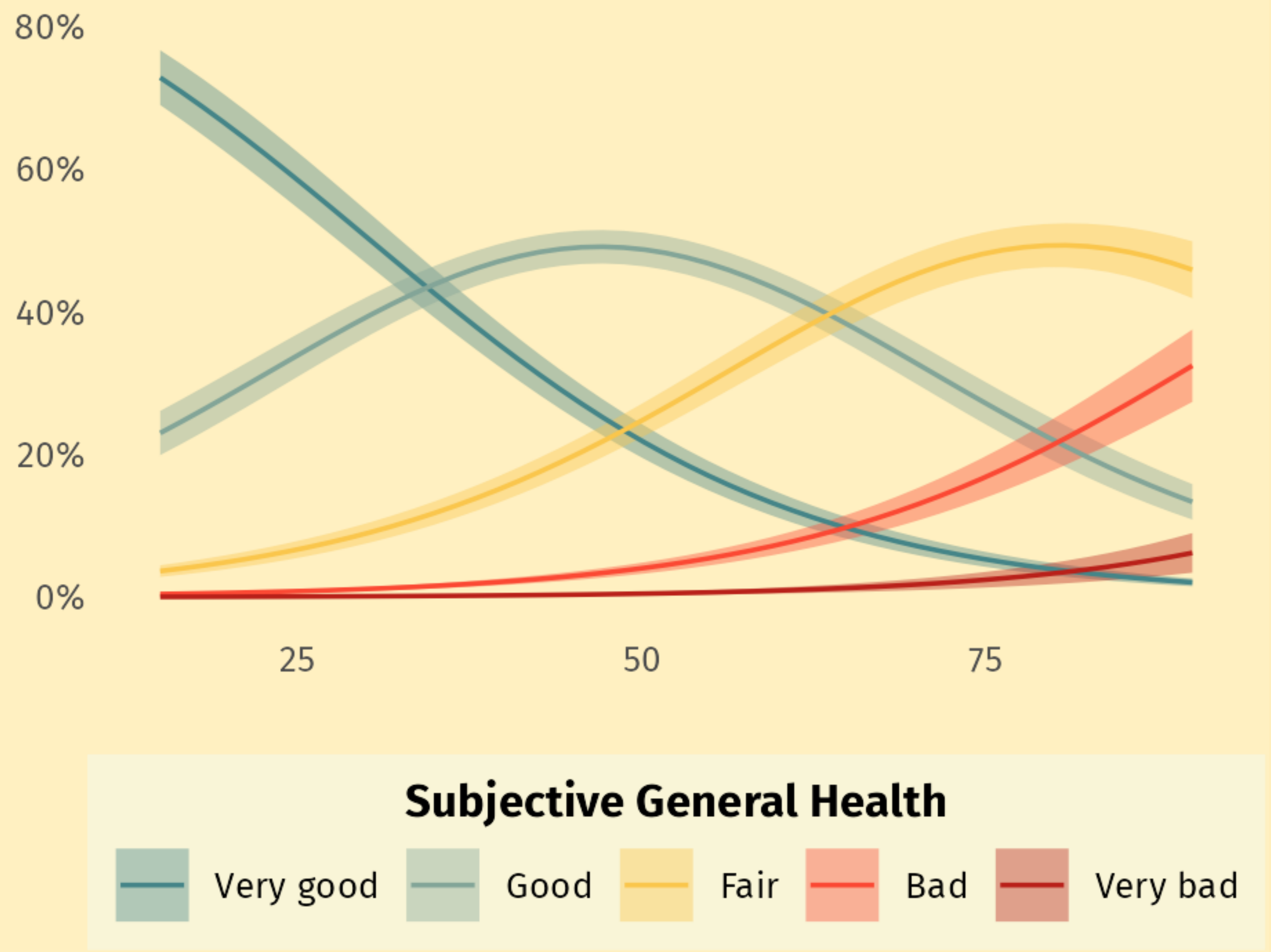
For each year of age, the probability of being in very good health drops by 1.06 percentage point.

Category	Marginal effect
Very good	0.010633
Good	-0.000430
Fair	0.007292
Bad	0.003254
Very bad	0.000517



Marginal effects for Education





Questions?

InteRmezzo!

Ordinal regression assumptions

Proportional Odds Assumptions

Logistic ordinal regression assumes the effect of predictors is the same for all adjacent categories, the so-called proportional odds assumption.

Example:

> We assume that the effect of age on moving from "Very good health" to "good health" is the same as on moving from "good health" to "fair health".

Proportional odds assumptions

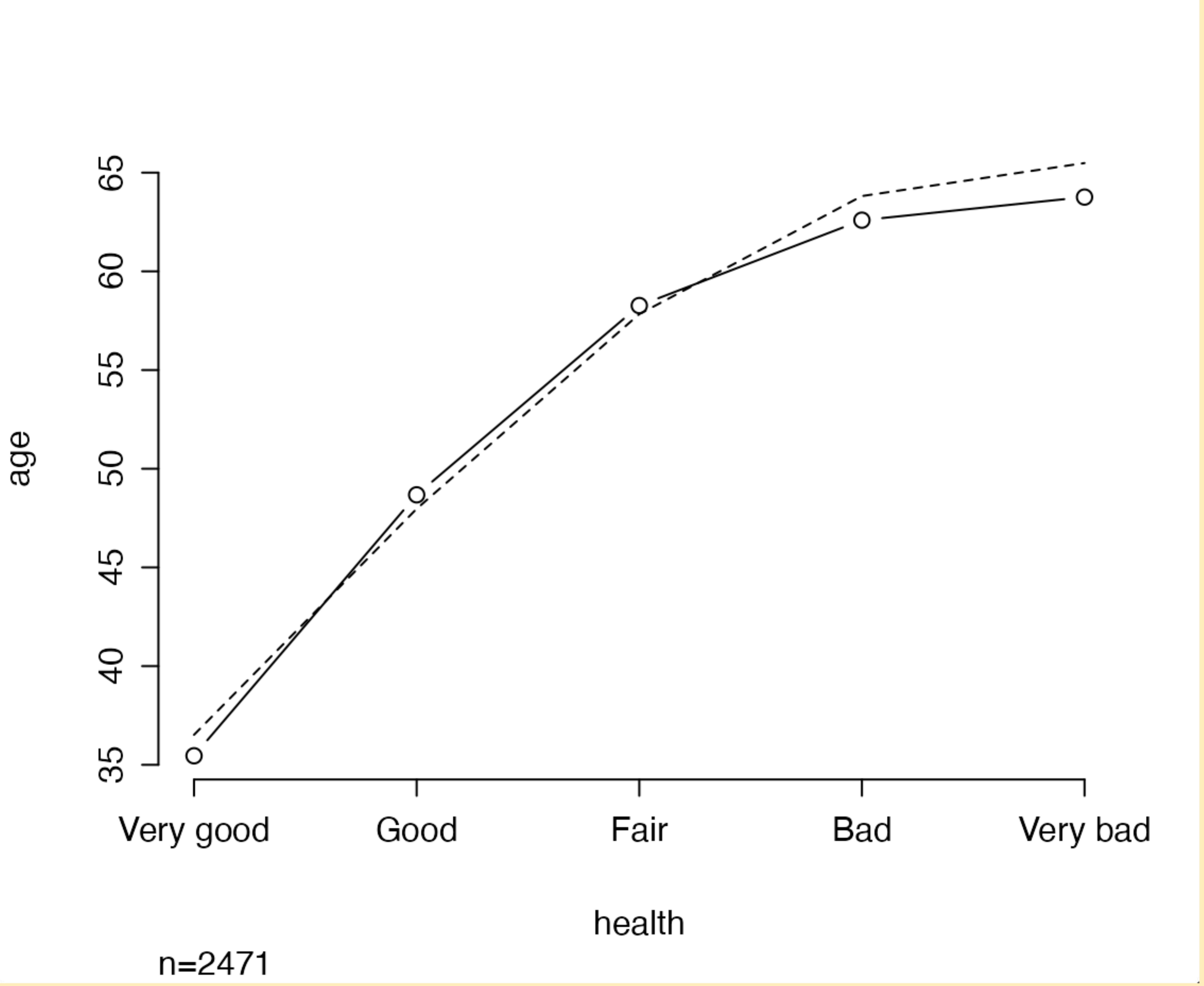
Predictor	Coefficient
Age	0.062
Education: Elementary	-
Education: High school (without diploma)	-0.248
Education: High school (with diploma)	-0.291
Education: University	-0.532
Intercepts:	
Very good|Good	0.610
Good|Fair	2.831
Fair|Bad	4.962
Bad|Very Bad	7.340

Checking proportional odds assumption

Looks good for age!

Plot average value of predictor against the dependent variable.

The full line should follow the dotted line.

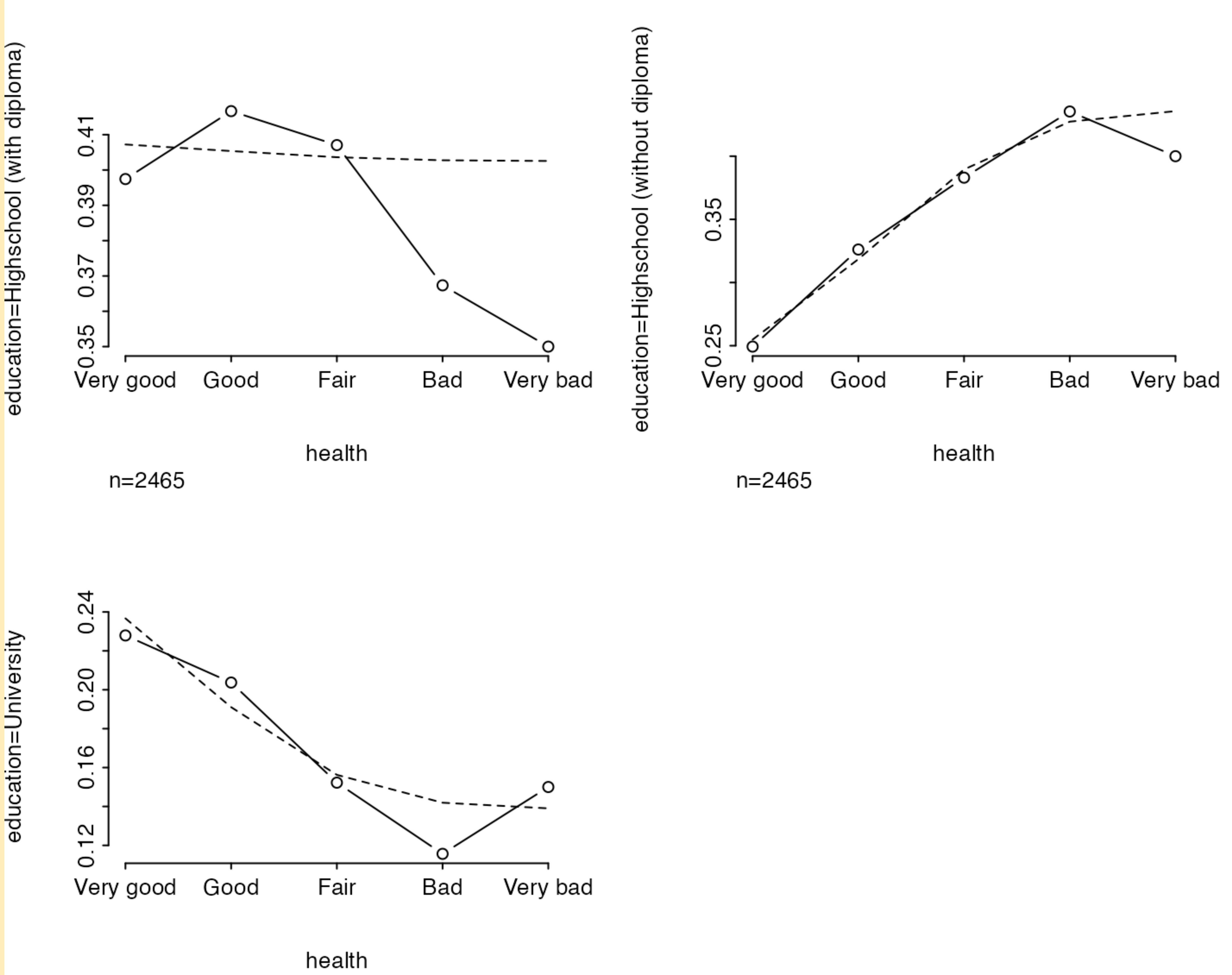


Checking proportional odds assumption

Less good for education

Plot average value of predictor against the dependent variable.

The full line should follow the dotted line.



Fixing proportional odds

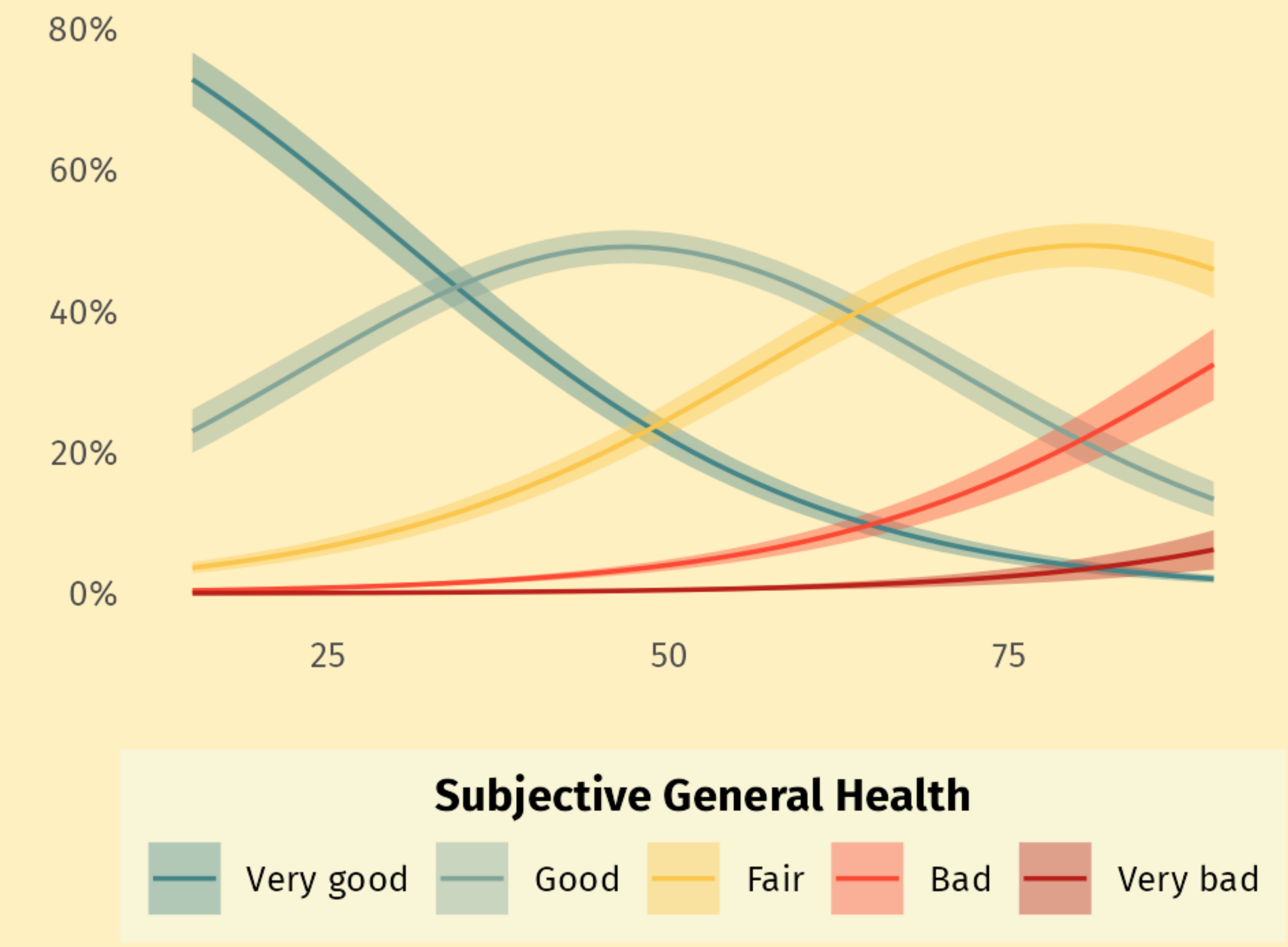
If the proportional odds assumption is broken for some predictor, point estimates will be biased.

We can fix this by adding more parameters to the model (partial cumulative model).

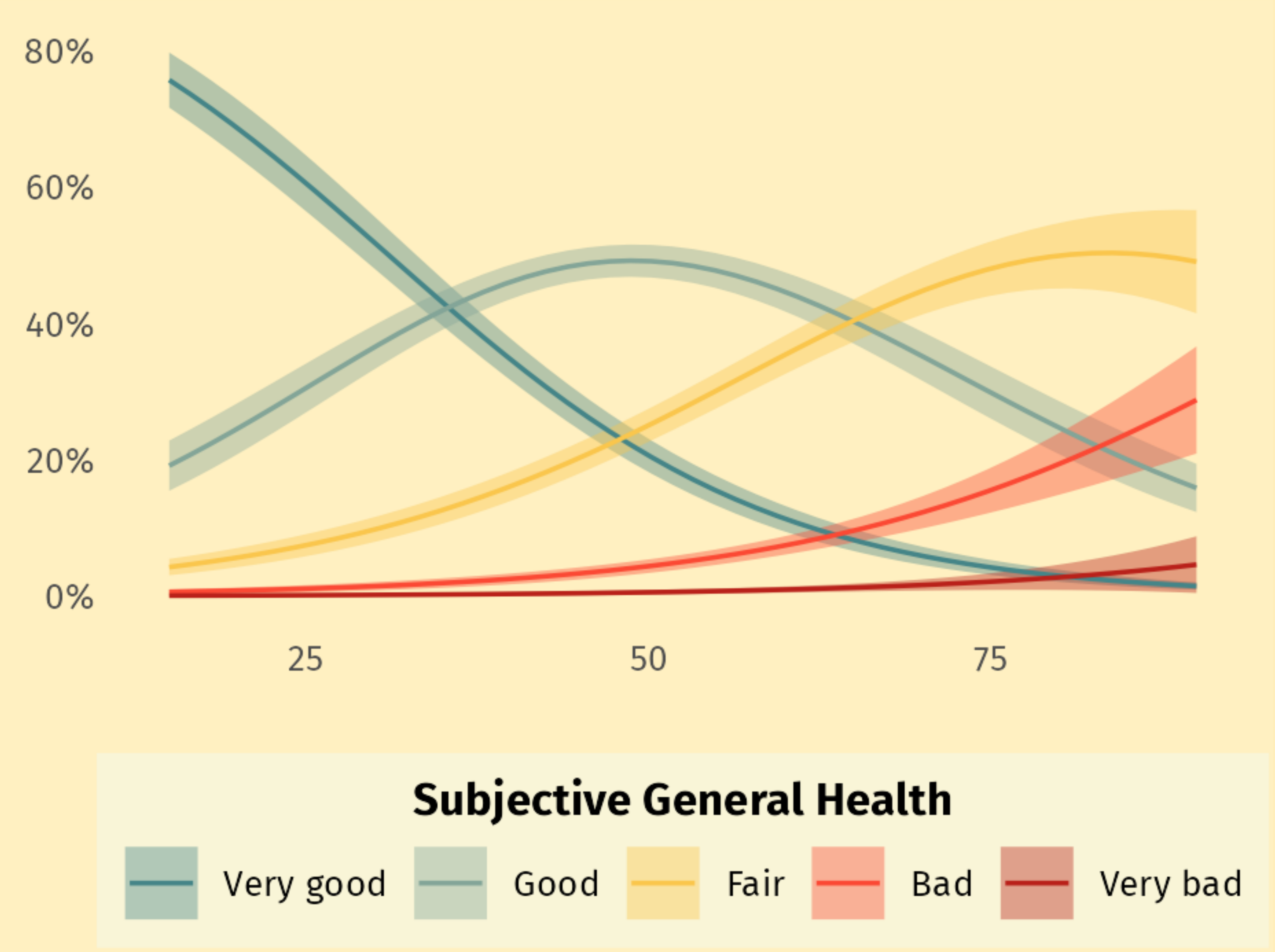
This fixes the bias, but is more demanding on sample size.

Age didn't need fixing...

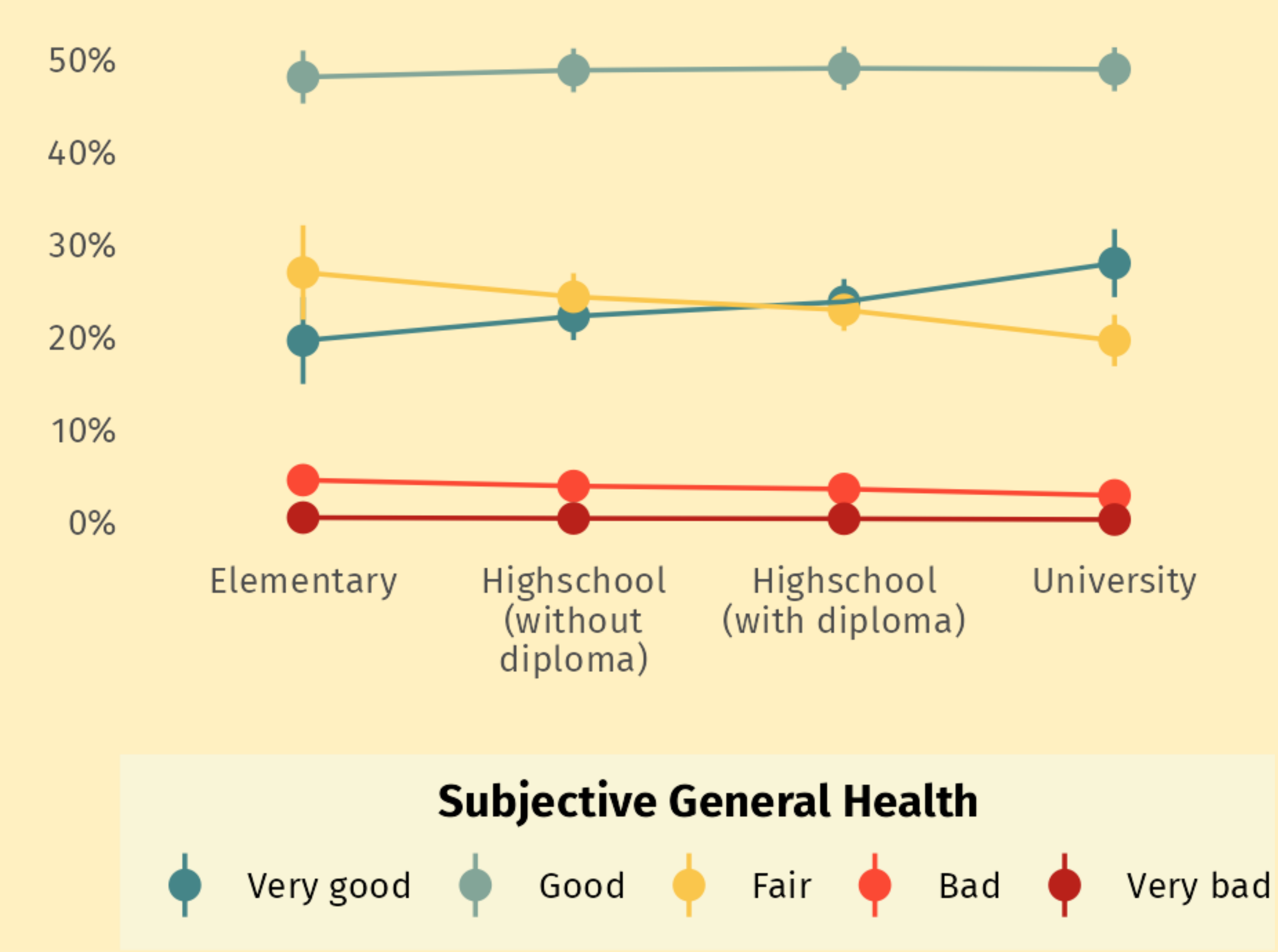
Assuming proportional odds

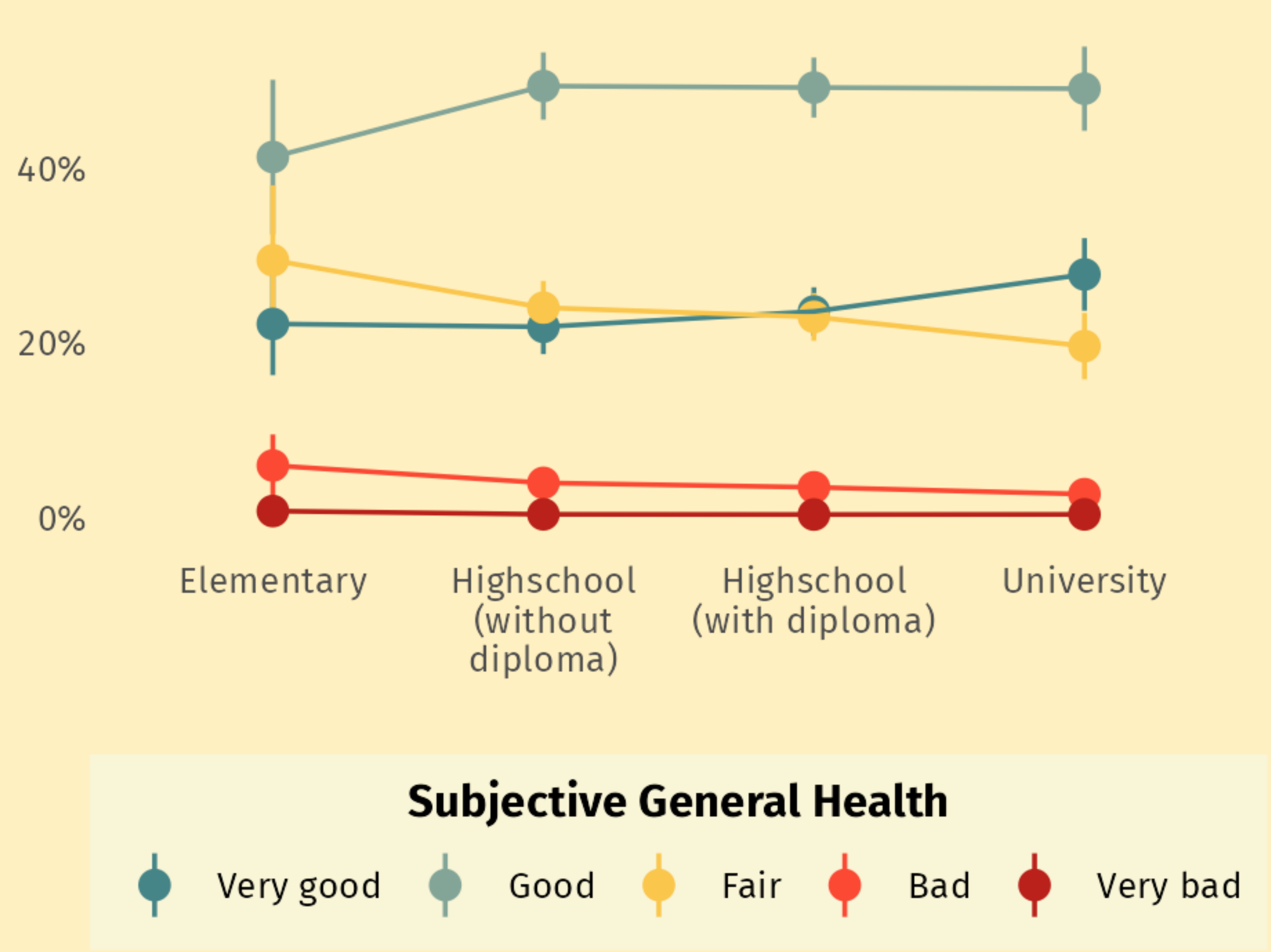


Not assuming proportional odds



...but there are (some) differences for Education





If you are not sure, just try the more flexible model

The diagnostic plots can't detect complex non proportional relationships, e.g. interactions, nonlinearity.

If you are not sure, just fit a more flexible model.

Questions?

InteRmezzo!

Item response theory (IRT)

Ordinal models very popular in measurement theory.

Allow to estimate properties of measurement tools, like difficulty, discrimination, measurement invariance.

Hanzlová, Radka. 2022. „An Item Response Theory Analysis and Psychometric Properties of the Czech Version of the Satisfaction with Life Scale.“ Survey Research Methods 16(3): 371–385. ISSN 1864-3361. Available at: https://doi.org/10.18148/srm/2022.v16i3.7940.

Multinomial models

Multionomial models

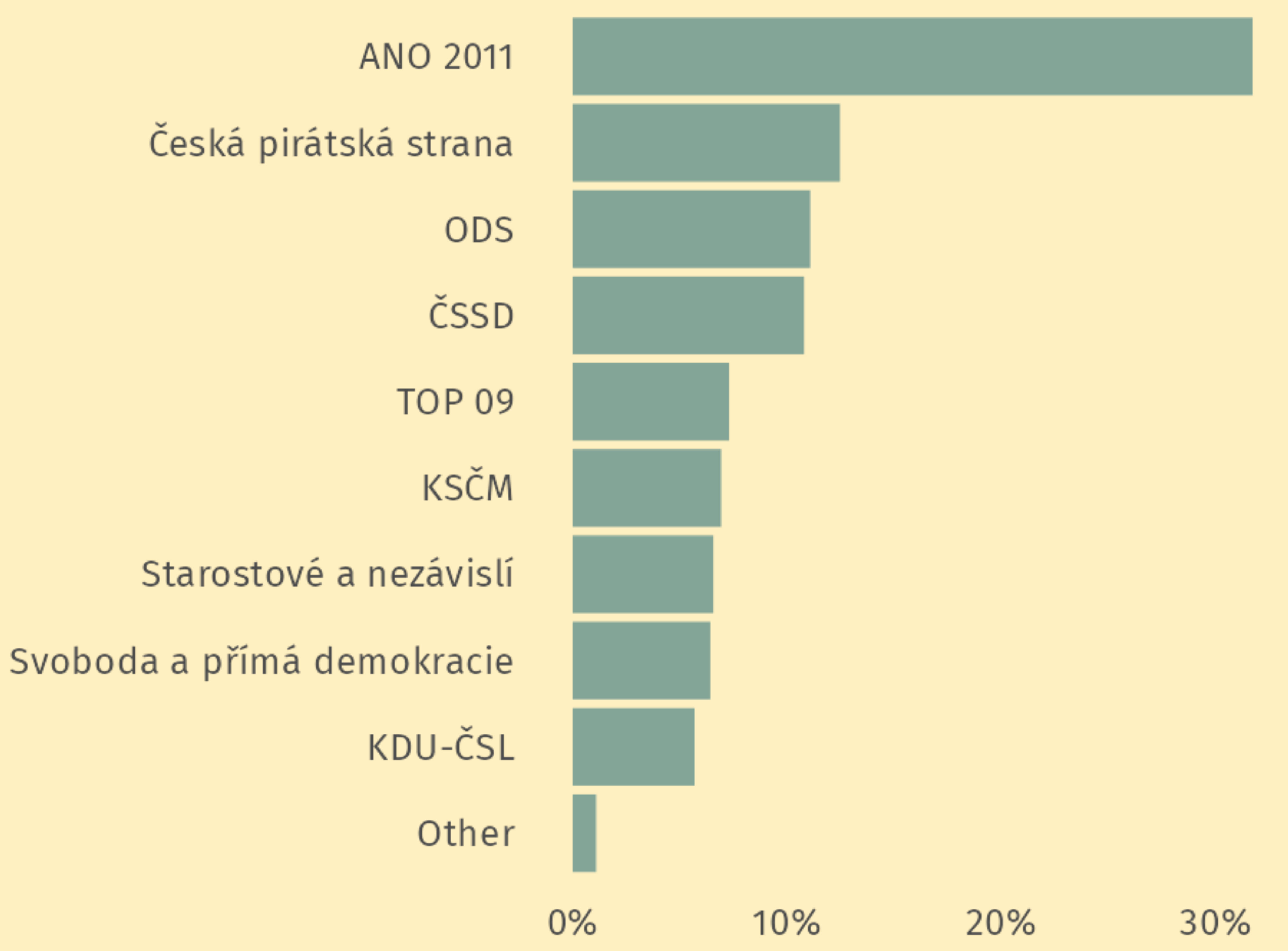
For nominal data

- Party preferences
- Product preferences
- University majors

Extension of binary model, just like the ordinal version before.

Party preferences in Czech Republic

What party did you vote for in the last parliamentary elections?



Multinomial logistic regression

Most flexible model, no assumption about order of response categories or distances between them.

Downside: most complex to estimate

- Computationally intensive
- Requires larger sample size (compared to binary/ordinal)
- Convergence problems are not uncommon

Multionomial regression

One response category is arbitrarily chosen as the reference group, similarly to dummy coding (it doesn't matter which).

Model estimates effect of predictors on change from reference group to each of the other ones.

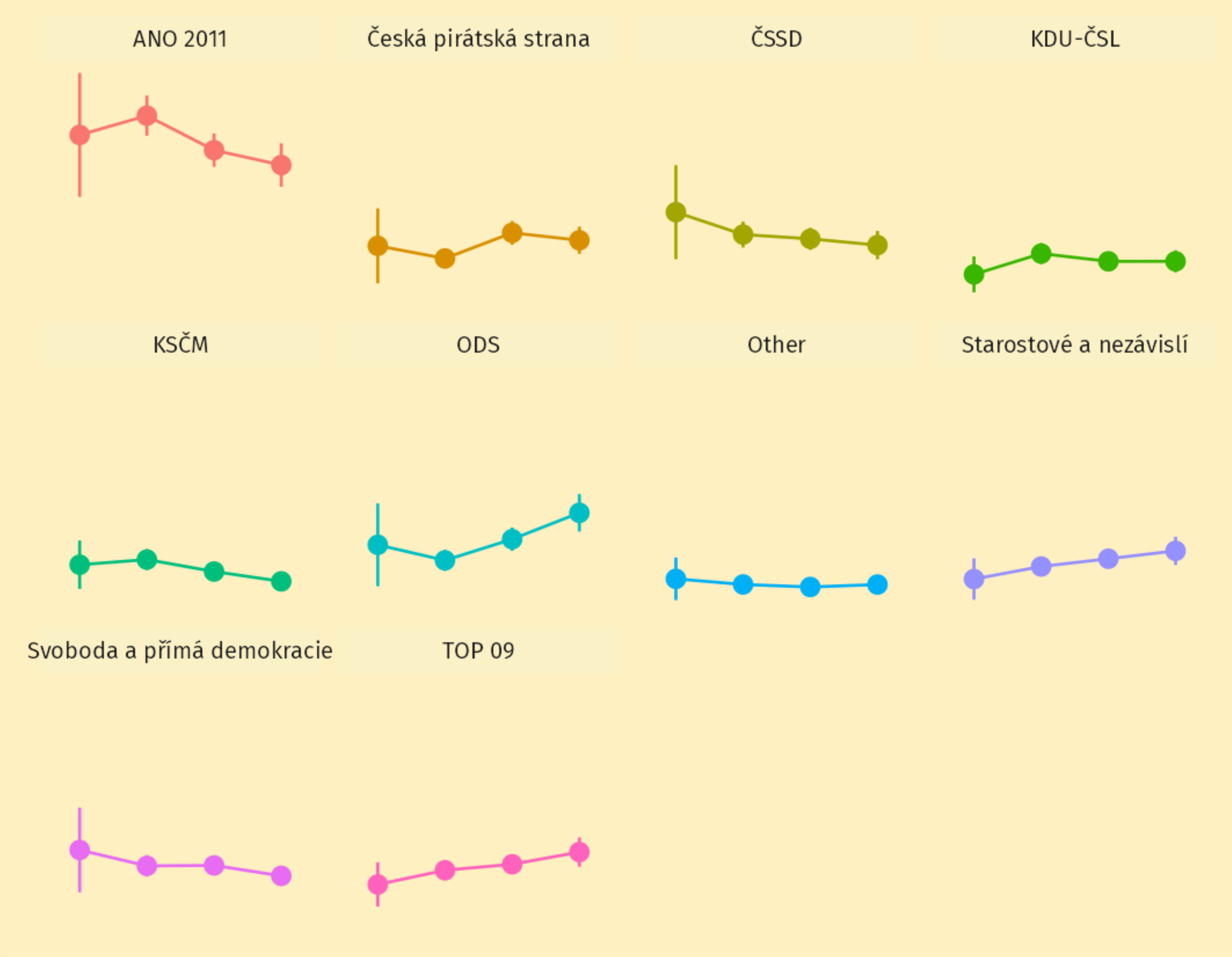
In practice, this means you'll get a ton of coefficients with little way to interpret them.

You know the answer and so do I: marginal effects on probability scale.

Party preferences and Age

Category	Marginal effect
Česká pirátská strana	-0.0055
Svoboda a přímá demokracie	-0.0001
9 TOP	-0.0008
ODS	-0.0003
Starostové a nezávislí	0.00030
KDU-ČSL	0.00010
ANO 2011	0.00111
ČSSD	0.00211
KSČM	0.00348
Other	-0.00034

Party preferences and Education



Questions?

InteRmezzo!

Assumptions

Independence of irrelevant alternatives assumption

The most important assumption is the independence of irrelevant alternatives.

We assume that adding/removing a category won't change relative popularity of the rest.

Example: In presidential elections between Babiš and Pavel, a third candidate entering the race shouldn't make some of Pavel's voters switch to Babiš (or vice versa).

Independence of irrelevant alternatives assumption

No easy way to check this - this is a problem of psychology, not statistics.

Matters less if we know the options are fixed (e.g. we know there wont be a third candidate).

Most problematic for for hypothetical predictions.

Can be solved by switching from logit link function to probit.

Questions?