

**Who is healthier? Investigating
South African patterns of reporting
heterogeneity in self-rated health**

Ralitza Dobрева (UKZN)

ESSA 2015

Background

- Health is an important dimension of broader welfare concepts.
 - World Health Organization’s definition of health is
“a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (WHO, 1948).
- Health perceptions are key drivers of individuals’ decision-making about lifestyles and seeking treatment.
- In the South African context, heterogeneity in reporting may be a particularly important problem, given:
 - the large socio-economic differentials;
 - the diversity of cultural and language groups; and
 - the relative geographical isolation of people living in many rural areas.

Literature review

1. What does self-rated (or self-assessed) health status (SRH) measure? Is it a credible indicator of health?

Advantages and drawbacks of SRH

- *Advantages:*
 - practical advantage: **easy, cheap and fast to collect** in surveys using a single question;
 - conceptual advantages:
 - **all-encompassing in nature**: “while researchers are measuring the parts, respondents have access to the whole” (Idler and Benyamini, 1997: 28);
 - **dynamic, but includes elements of the respondents’ enduring self-concept** (Huisman and Deeg, 2010; Idler and Benyamini, 1997; Bailis et al, 2003).
 - **underpins decisions that have a real impact on the respondent’s wellbeing** (Bailis et al, 2003).
 - **encapsulates the availability of physical, emotional or social resources that can moderate a decline or aid a recovery** (Idler and Benyamini, 1997);
 - **independent predictor of mortality**, even when used *in addition to* a range of more objective health indicators (e.g. Idler and Benyamini, 1997; DeSalvo et al, 2006; Ardington and Gasealahwe, 2014).
 - In South Africa, SRH is found to be a significant predictor of two-year mortality, after socioeconomic status and other subjective and objective health indicators are taken into account (Ardington and Gasealahwe, 2014).

Literature review

1. What does self-rated (or self-assessed) health status (SRH) measure? Is it a credible indicator of health?

- *Drawbacks:*
 - **subjective in nature** and hence affected by measurement error (Crossley and Kennedy, 2002) ;
 - **reporting heterogeneity**, i.e. SRH may not be comparable between countries or even between rich and poor regions in the same country (Sen, 2002; Jürges, 2007) ;
 - **may be endogenous** to an important range of factors, making causal links difficult to identify (e.g. Currie and Madrian, 1999; Disney et al, 2006);
 - **the “justification hypothesis”** (e.g. Myers, 1982; Baker et al, 2004);
 - **may reflect certain aspects of health better than others** (Mavaddat et al, 2011);
 - **influenced by psychological factors and personality dispositions** (Goodwin and Engstrom, 2002; Huisman and Deeg, 2010).

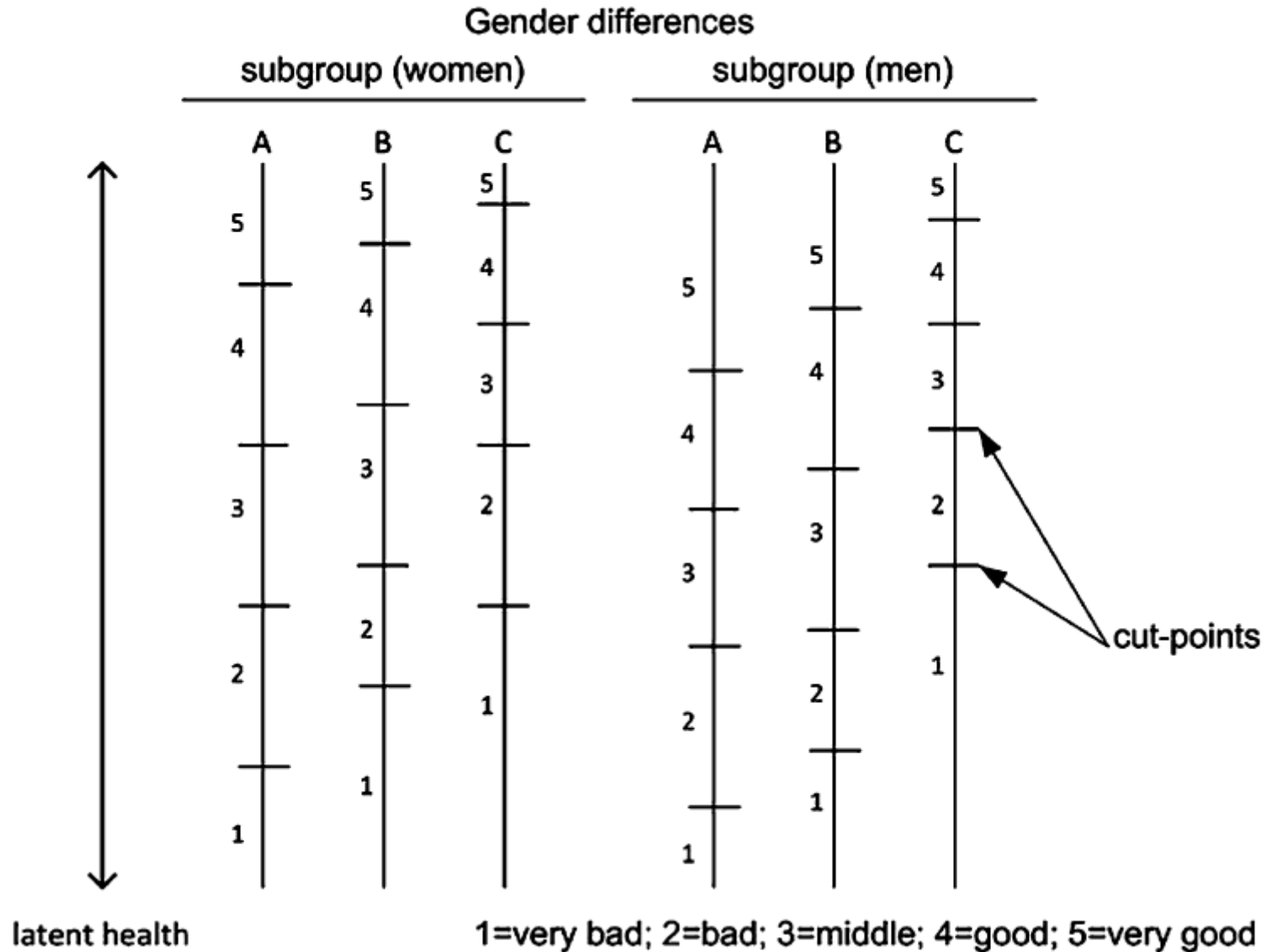
Literature review

1. What does self-rated (or self-assessed) health status (SRH) measure? Is it a credible indicator of health?

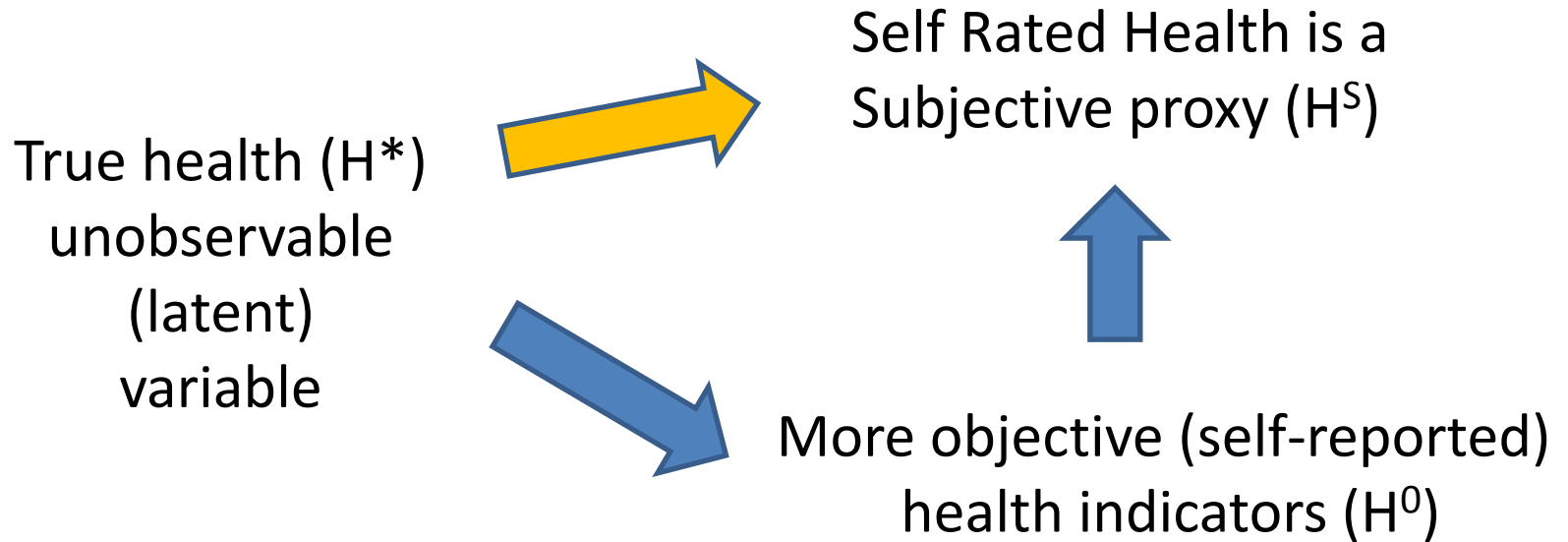
- Reporting heterogeneity:
 - Assuming an underlying unobservable (latent) and continuous “true” health variable, respondents project their perceptions into a reporting category, from “excellent” to “poor” (Jürges, 2007, Lindeboom and Van Doorslaer, 2004).
 - The categories are “verbal representations of different health states” (Jürges, 2007: 164).
 - Empirically, two types of reporting heterogeneity between groups (e.g. men and women):
 - *Cut-point shift*: differences in thresholds;
 - *Index shift*: parallel shift in all categories.
- Overall, SRH:
 - has a strong and reasonably consistent association with a wide range of more objective health measures.
 - remains a reliable independent predictor of mortality.
 - is determined by a cognitive process that is shaped by elements with various degrees of objectivity and subjectivity.

E.g. reporting heterogeneity between and within genders

(Fig. 2 in Schneider et al, 2012)



Methodology



Methodology

The ordered logit/probit model for panel data

- An ordered logit/probit technique built around the unobserved variable “true” health will be used.

$$\begin{aligned} Y_{it}^* &= X_{it}\beta + \varepsilon_{it} \\ i &= 1, \dots, N \\ t &= 1, \dots, T \end{aligned} \tag{1}$$

Y_{it}^* is the unobserved individual true health status

X_{it} is a vector of explanatory variables, incl. per capita hh income and inequality;

β is a parameter vector; ε_{it} is the error term.

- We do not observe Y_{it}^* , but we observe:

$$\begin{aligned} Y_{it} &= 0 & \text{if } Y^* \leq 0 \\ Y_{it} &= 1 & \text{if } 0 < Y^* \leq \tau_1 \\ Y_{it} &= 2 & \text{if } \tau_1 < Y^* \leq \tau_2 \\ & \vdots \\ Y_{it} &= J & \text{if } \tau_{J-1} \leq Y^* \end{aligned} \tag{2}$$

Methodology

- The ordered logit model can be written as:

$$\Pr(Y_i > J) = g(X\beta) = \frac{\exp(\alpha_j + X_i\beta)}{1 + \exp(\alpha_j + X_i\beta)}$$

- It makes the “parallel lines assumption”.
- The generalised ordered logit model can be written as:

$$\Pr(Y_i > J) = g(X\beta_j) = \frac{\exp(\alpha_j + X_i\beta_j)}{1 + \exp(\alpha_j + X_i\beta_j)}$$

$$j = 1, \dots, M - 1$$

where M is the number of categories of the ordinal dependent variable.

- The parallel lines assumption is relaxed.
- A partial proportional odds logit model allows for the most flexibility making the assumption about some parameters.
- This allows for differential cut-points in SRH reporting categories to be identified wrt different demographic and socioeconomic indicators.

The Data: NIDS

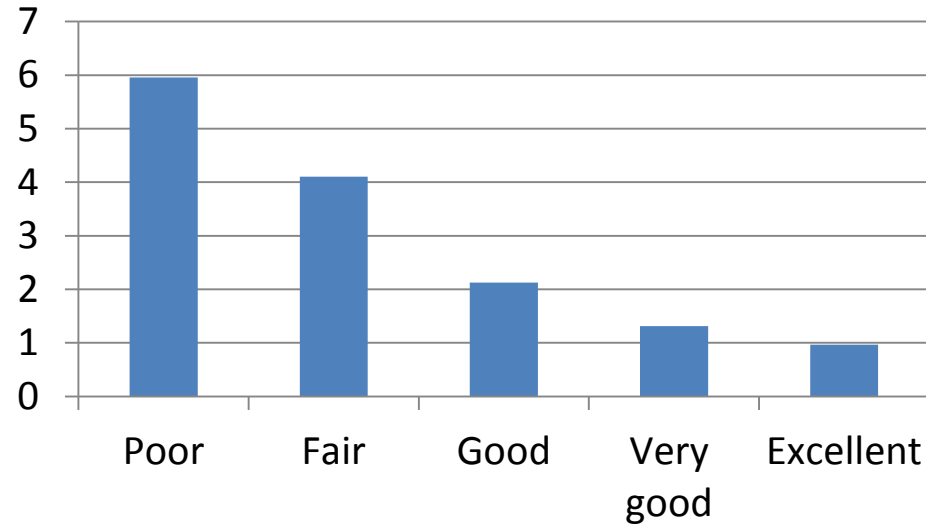
- Data: Waves 1 – 3 of NIDS
- Key variables for this study:
 - Self-rated health status (SRH):
 - “J1: How would you describe your health at present?
Would you say it is excellent, very good, good, fair, or poor?”
 - More objective self-reported measures of health:
 - A list of 24 symptoms experienced in the last 30 days
 - Chronic conditions (7 options) - TB, high blood pressure, diabetes, stroke, asthma, heart problems, and cancer.
 - Emotional health, reflected in a score according to the CES-D10 depression scale.
 - Objective health measures: BMI and blood pressure.

Socioeconomic variables

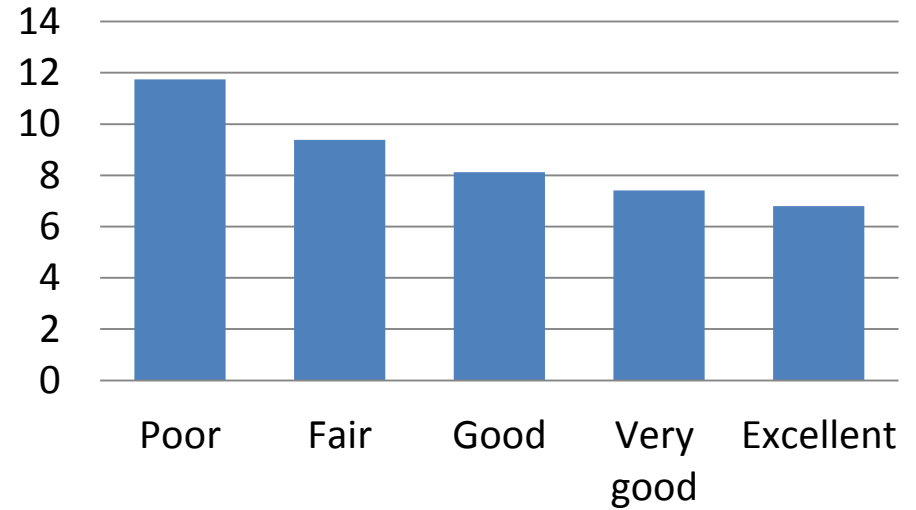
- Gender
- Age categories: 15 to 34; 35 to 45; 46 to 59; 60 and above.
- Marital status
- Dummy for presence of children in household
- Secondary and tertiary education dummies
- Employment status
- Quintiles of household expenditure per capita
- Geo-type: rural formal , tribal auth., urban formal, urban informal
- Medical aid membership

Descriptives using Wave 1 of NIDS

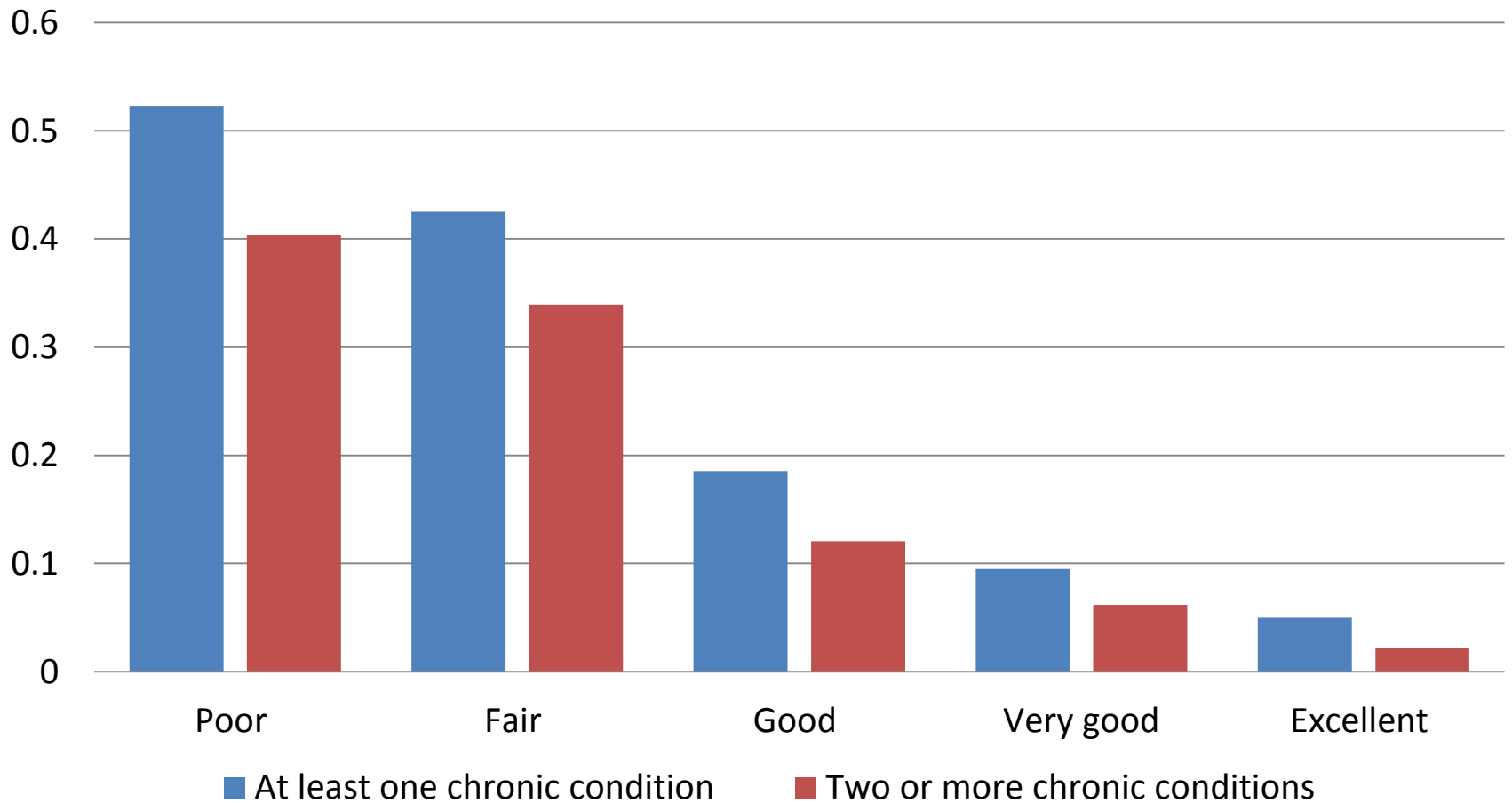
Mean number of symptoms vs SRH in W1



Mean Score on CES-D depression scale (out of 30) vs SRH in W1

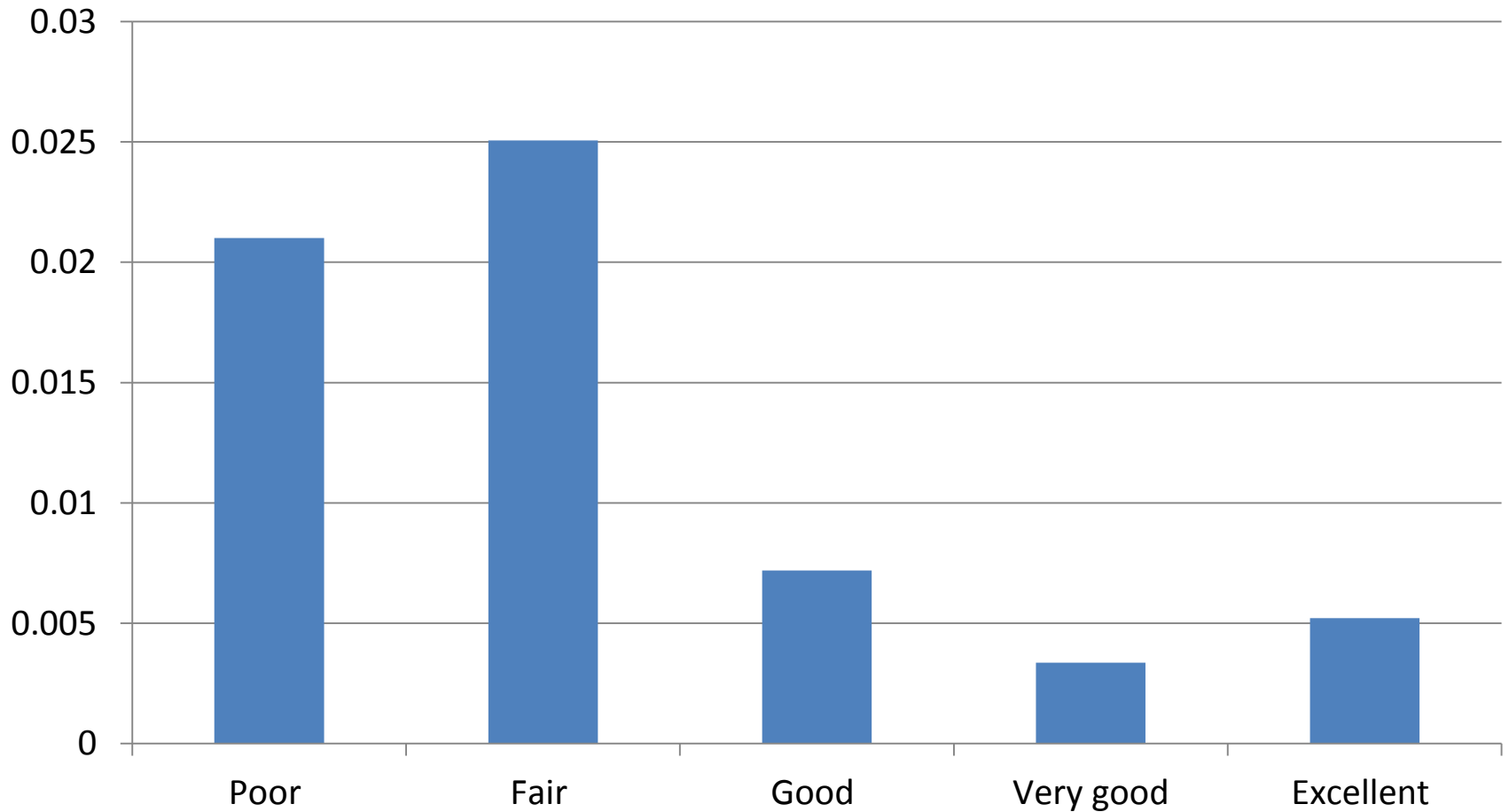


Chronic conditions vs SRH in W1 (proportions)

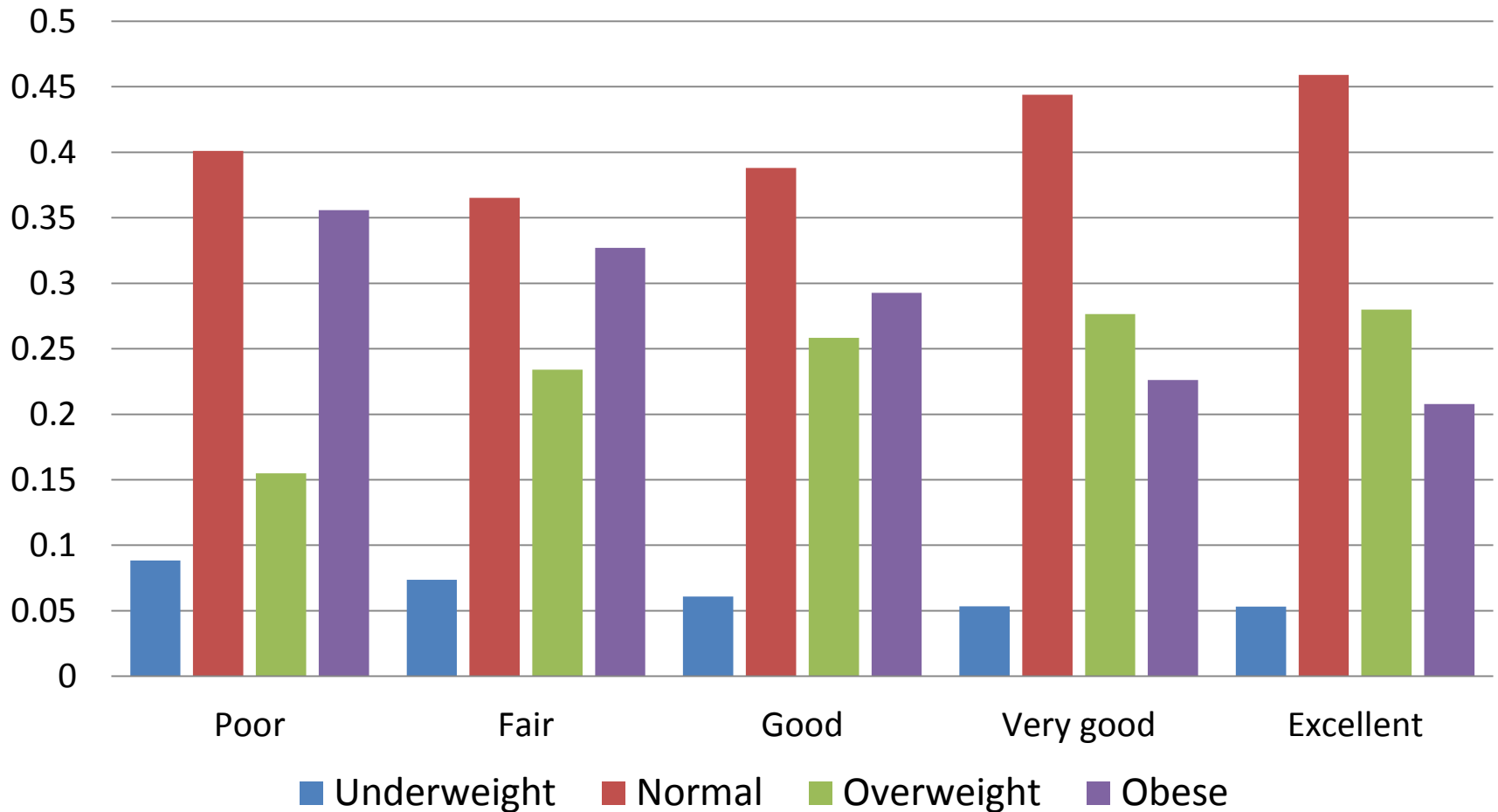


Hypertension vs SRH in W1 (proportions)

Hypertension: systolic > 179; diastolic > 109



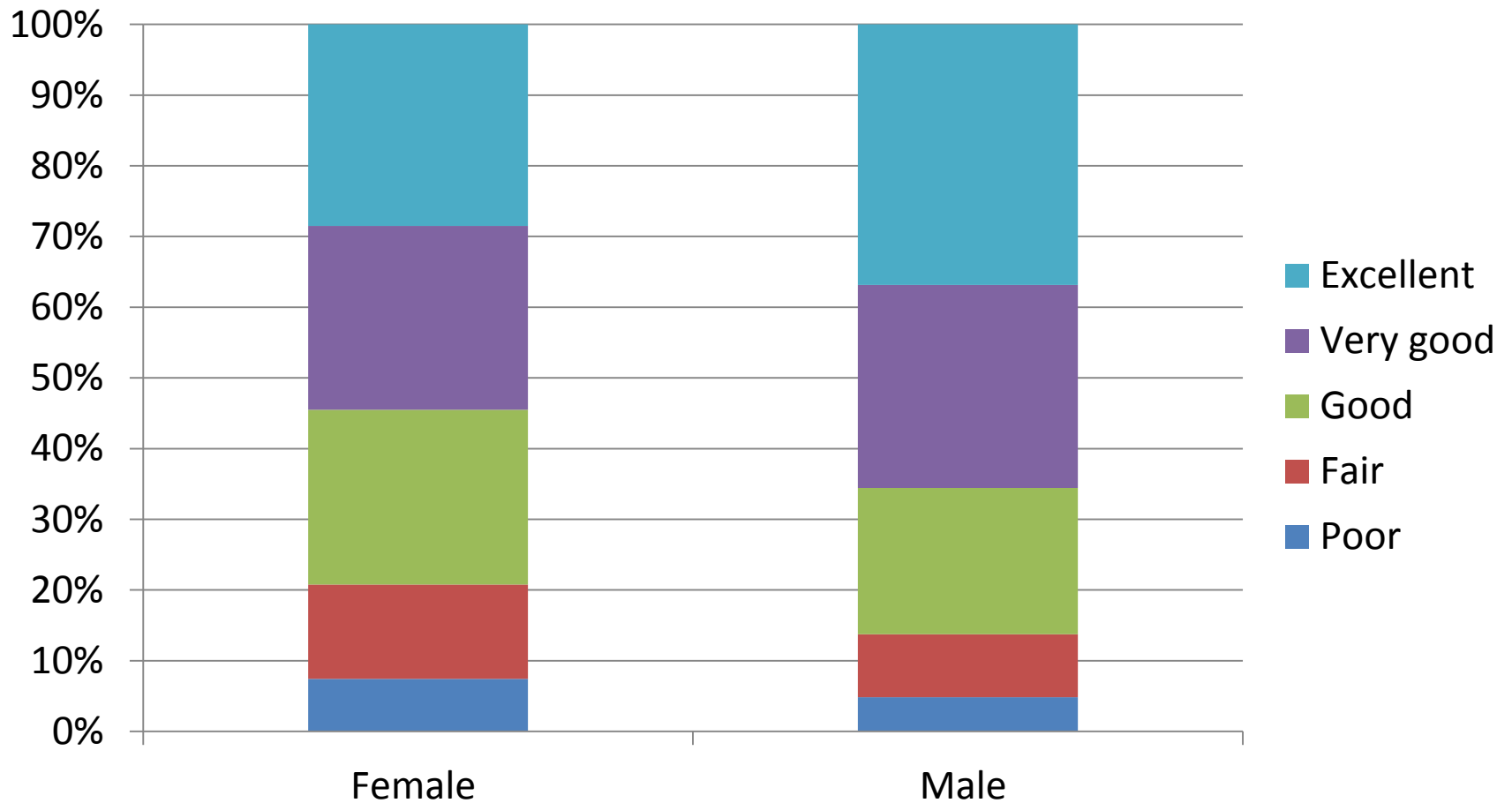
BMI categories vs SRH



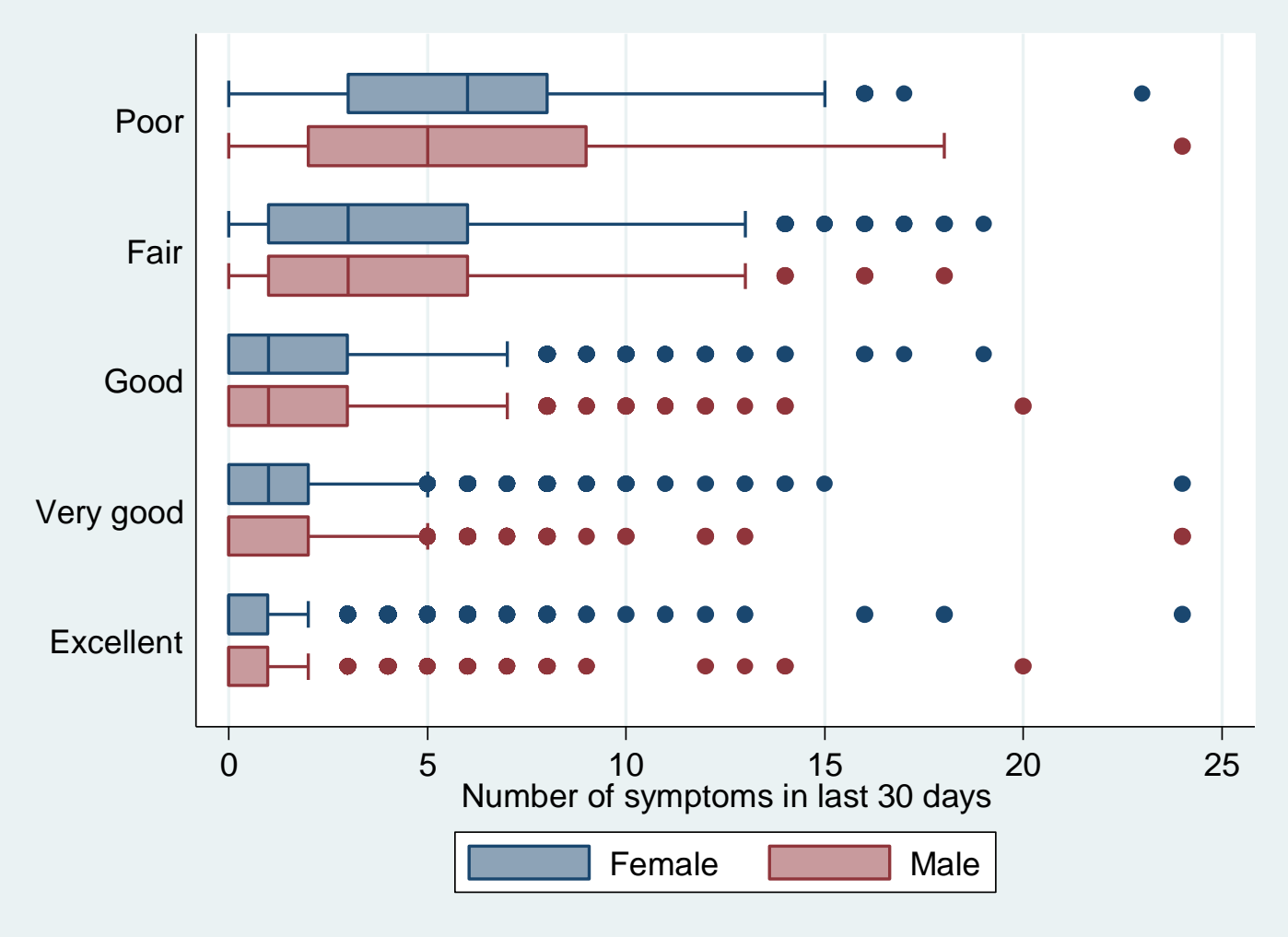
Who is healthier?

- What if we take everyone at their word?

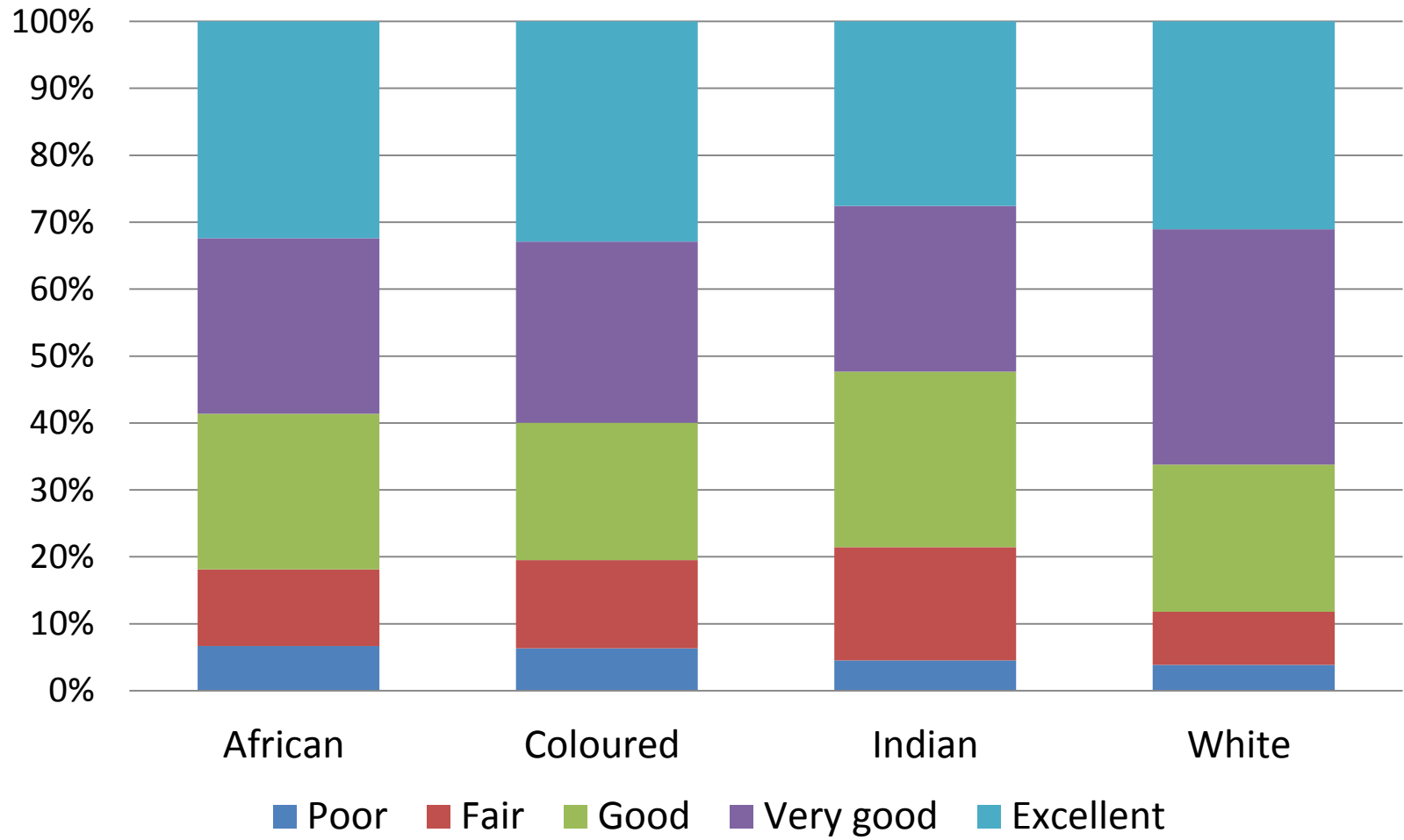
Self-rated health by gender



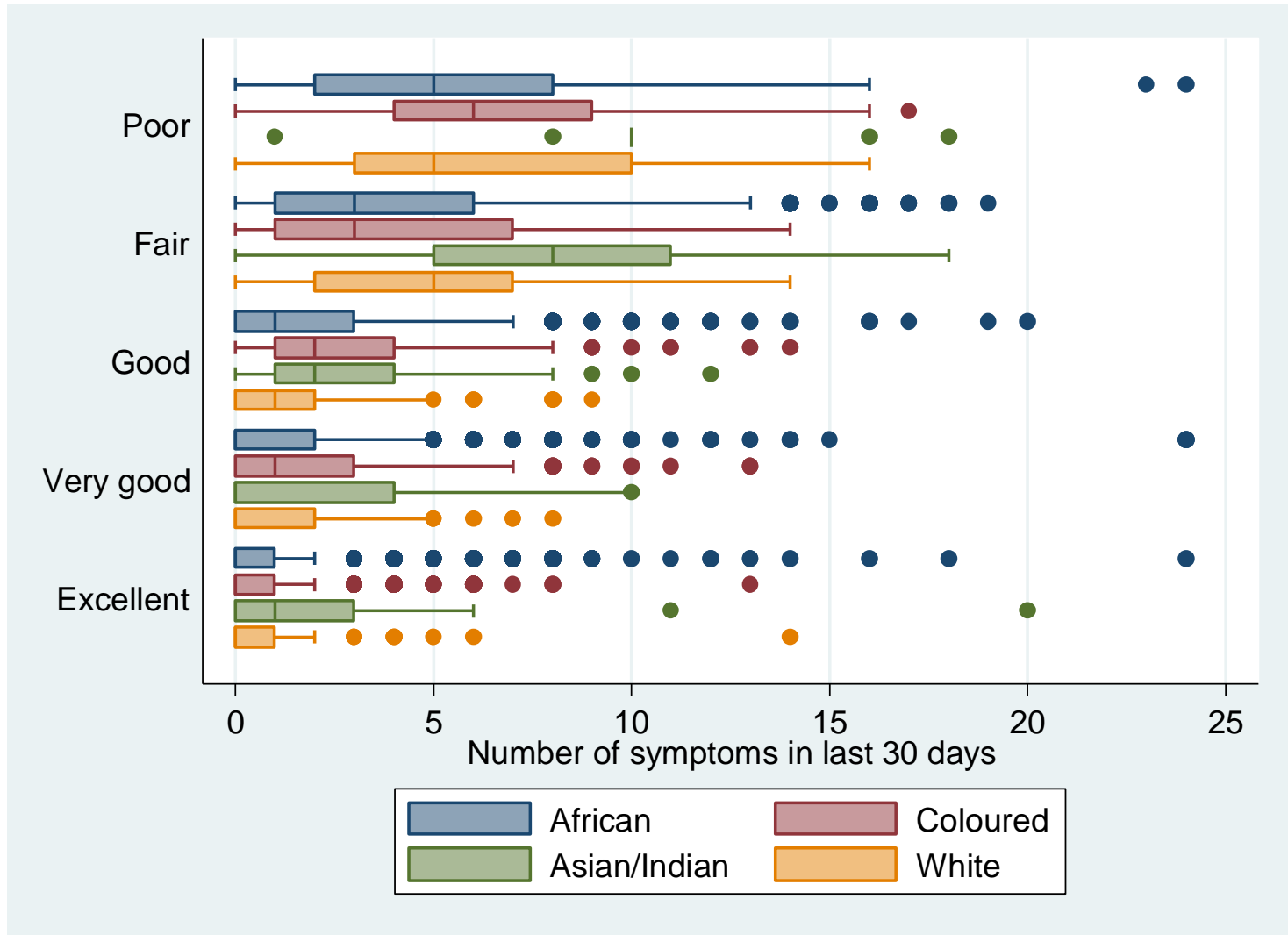
Symptoms and SRH by gender



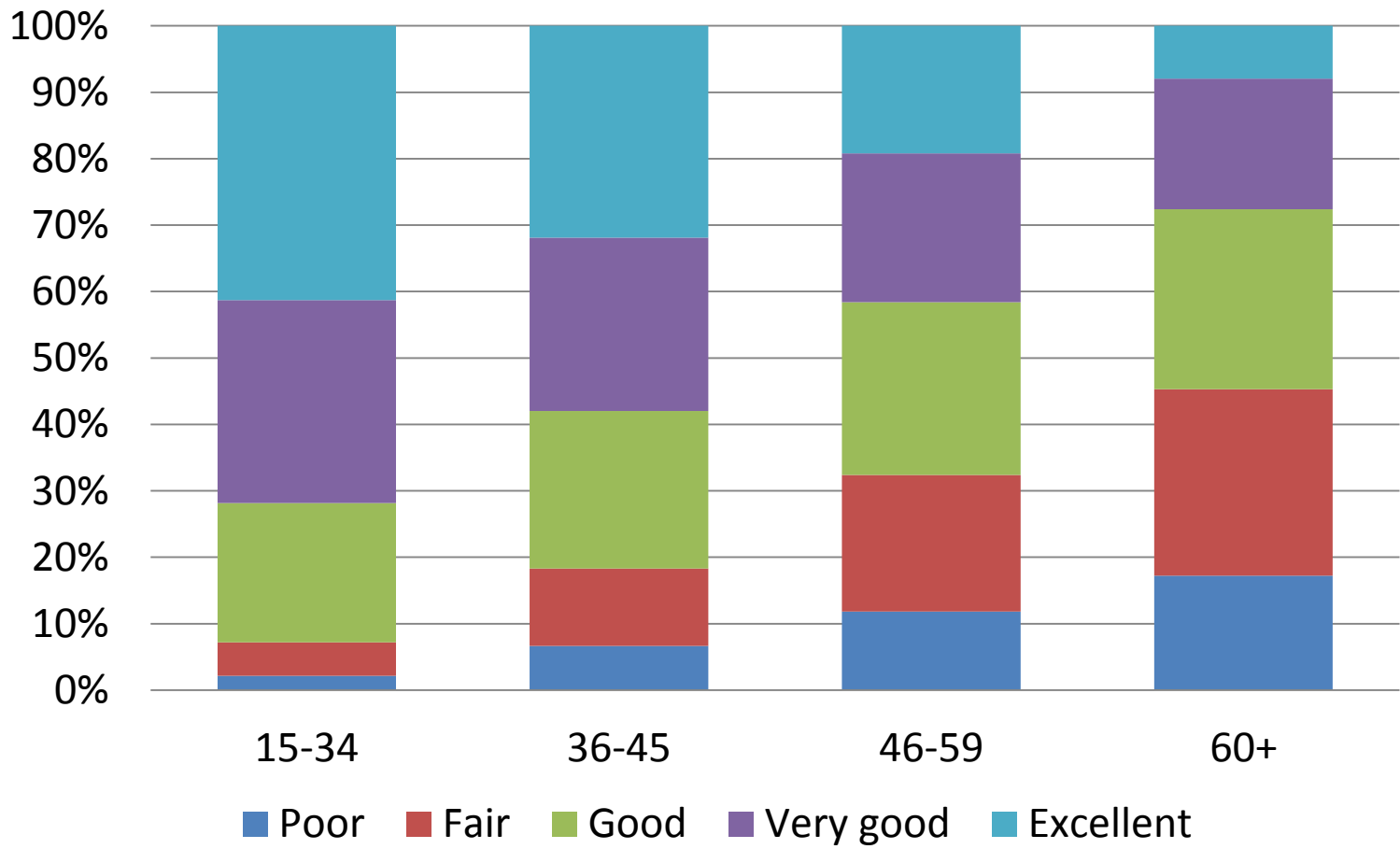
Self-rated health by race



Symptoms and SRH by race

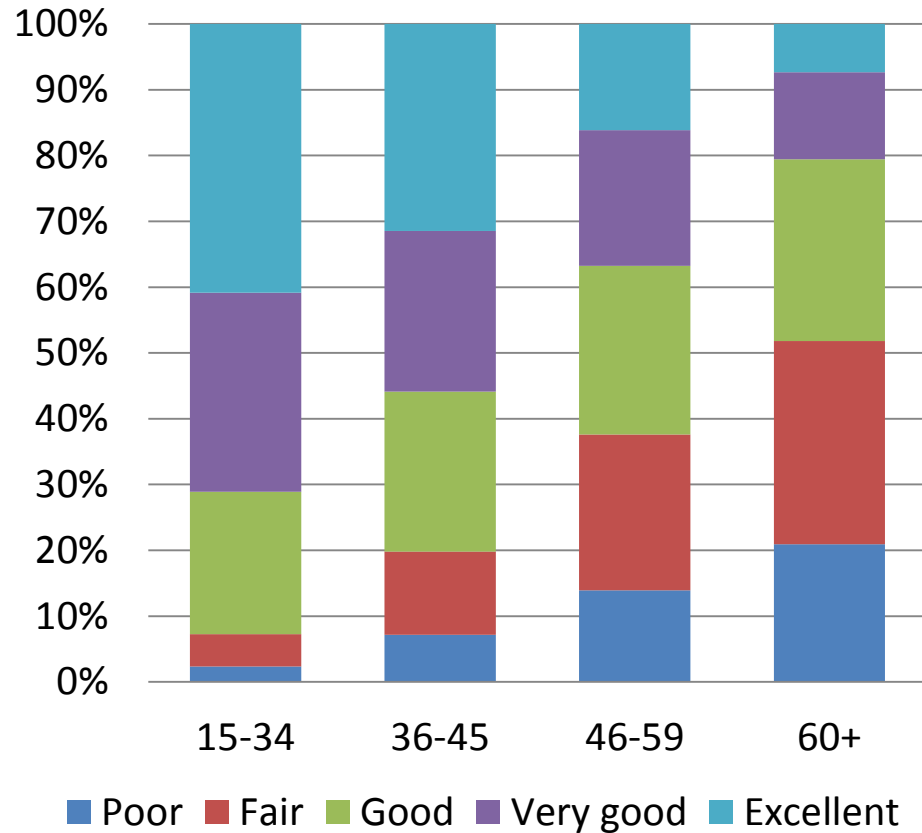


Self-rated health by age group



SRH by age and race

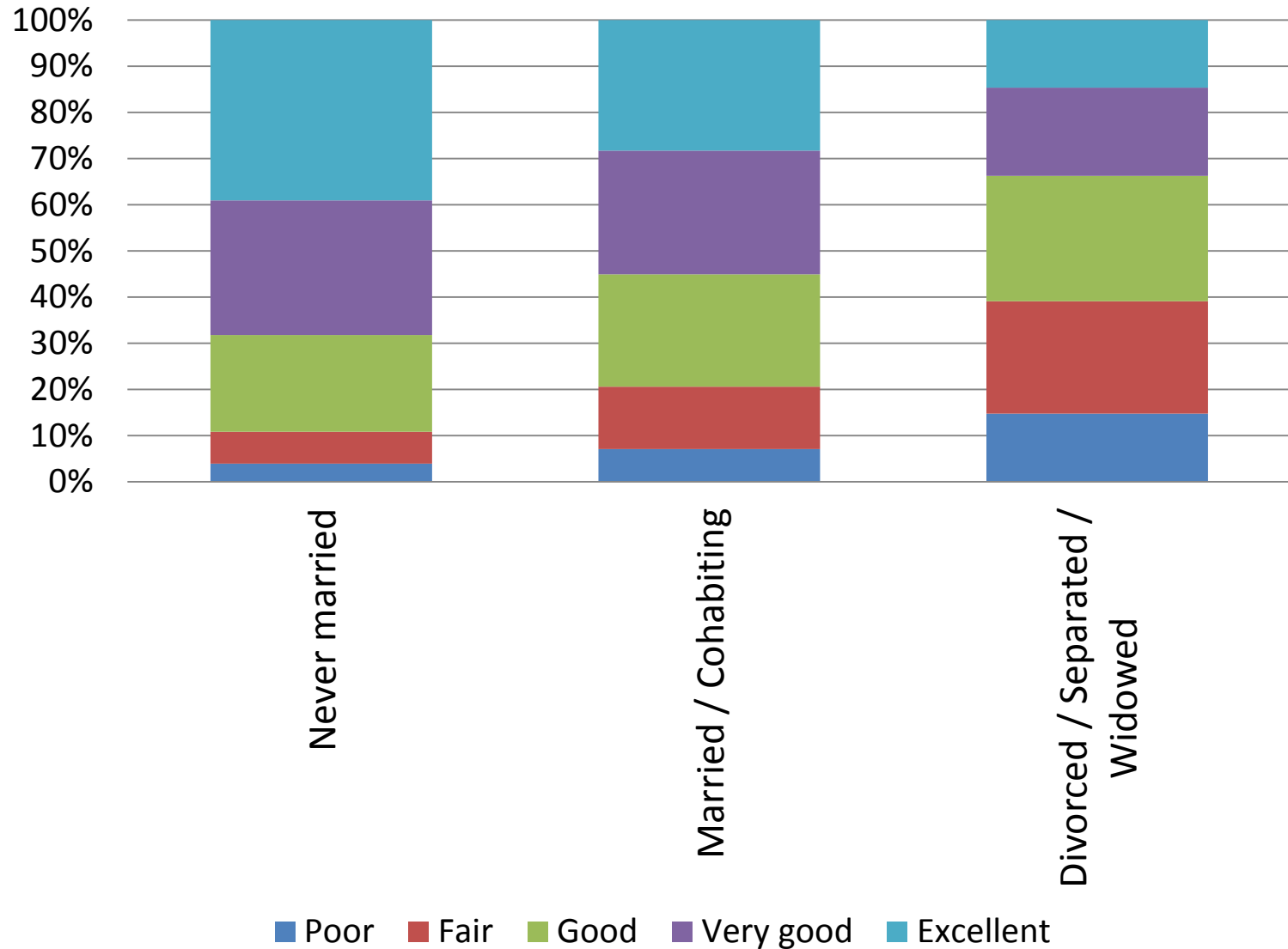
African



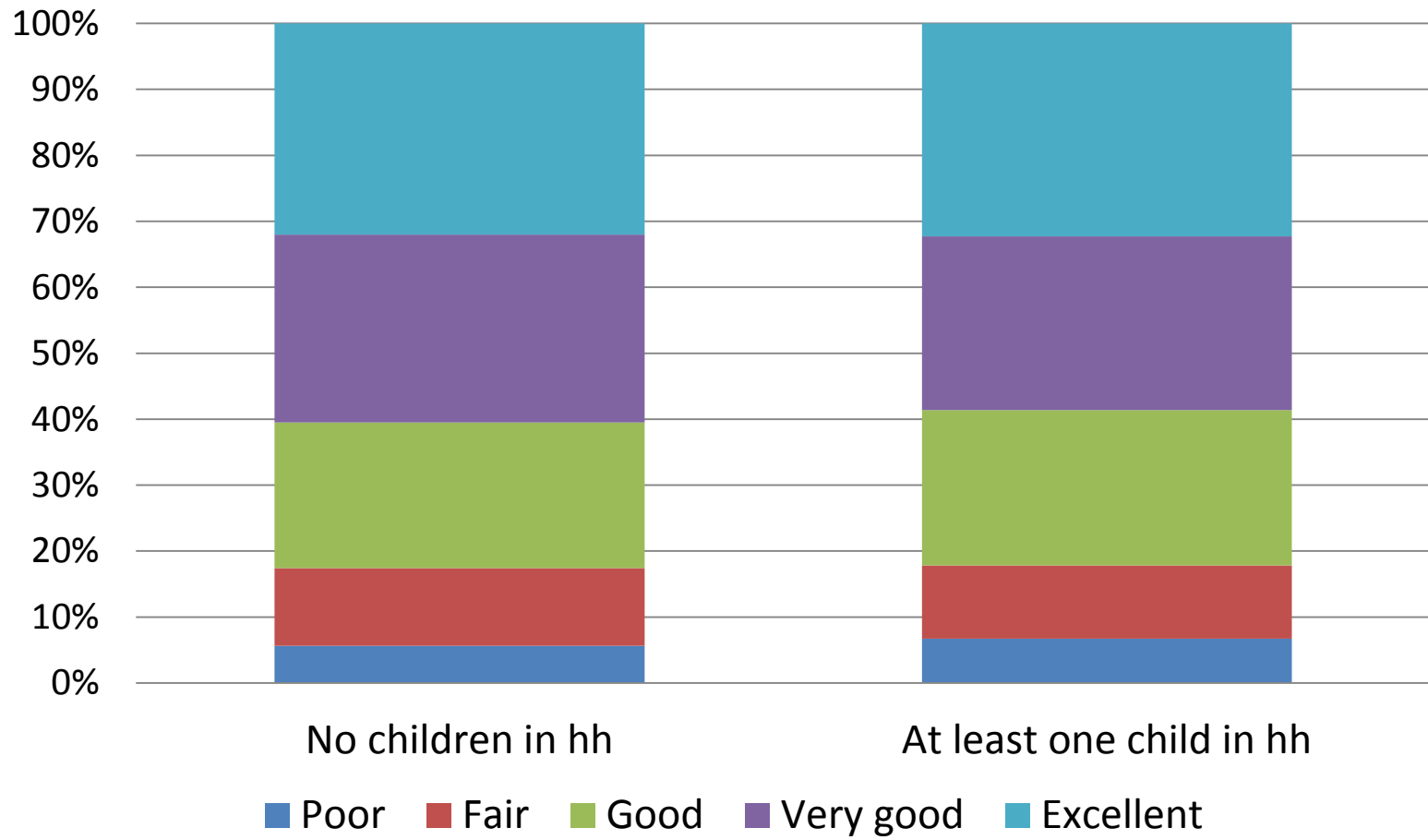
White



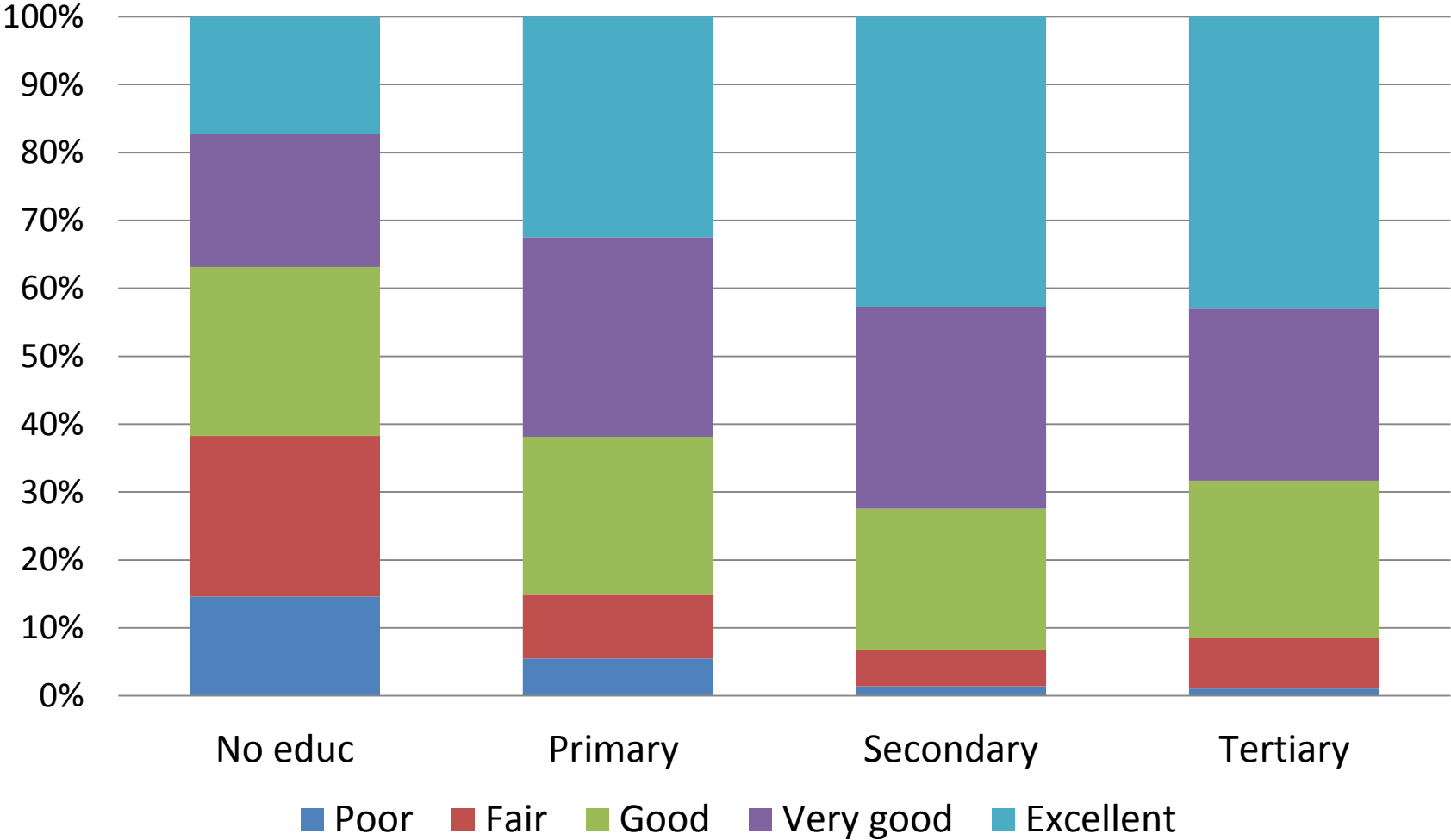
Self-rated health by marital status



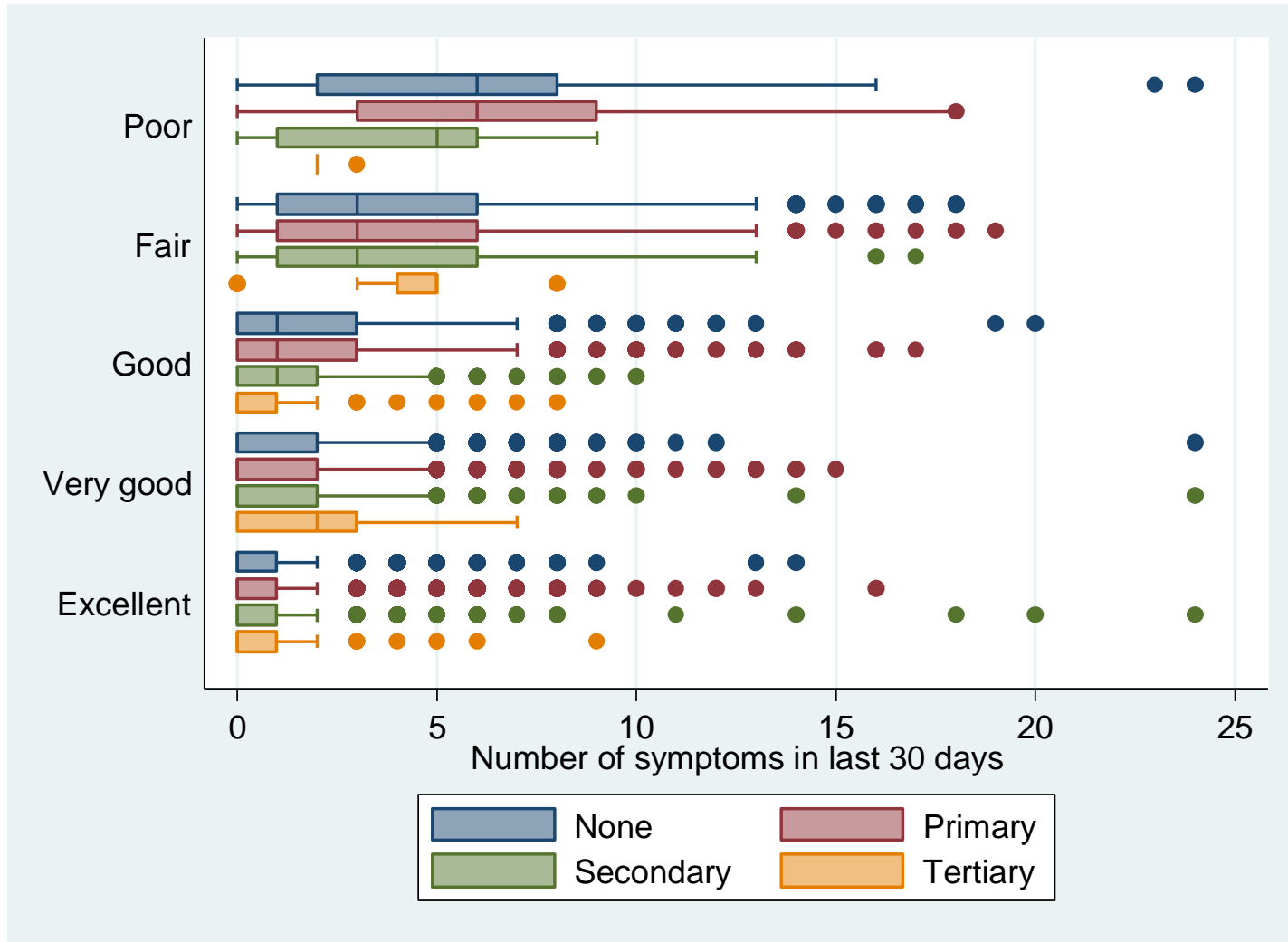
Self-rated health by presence of children in household



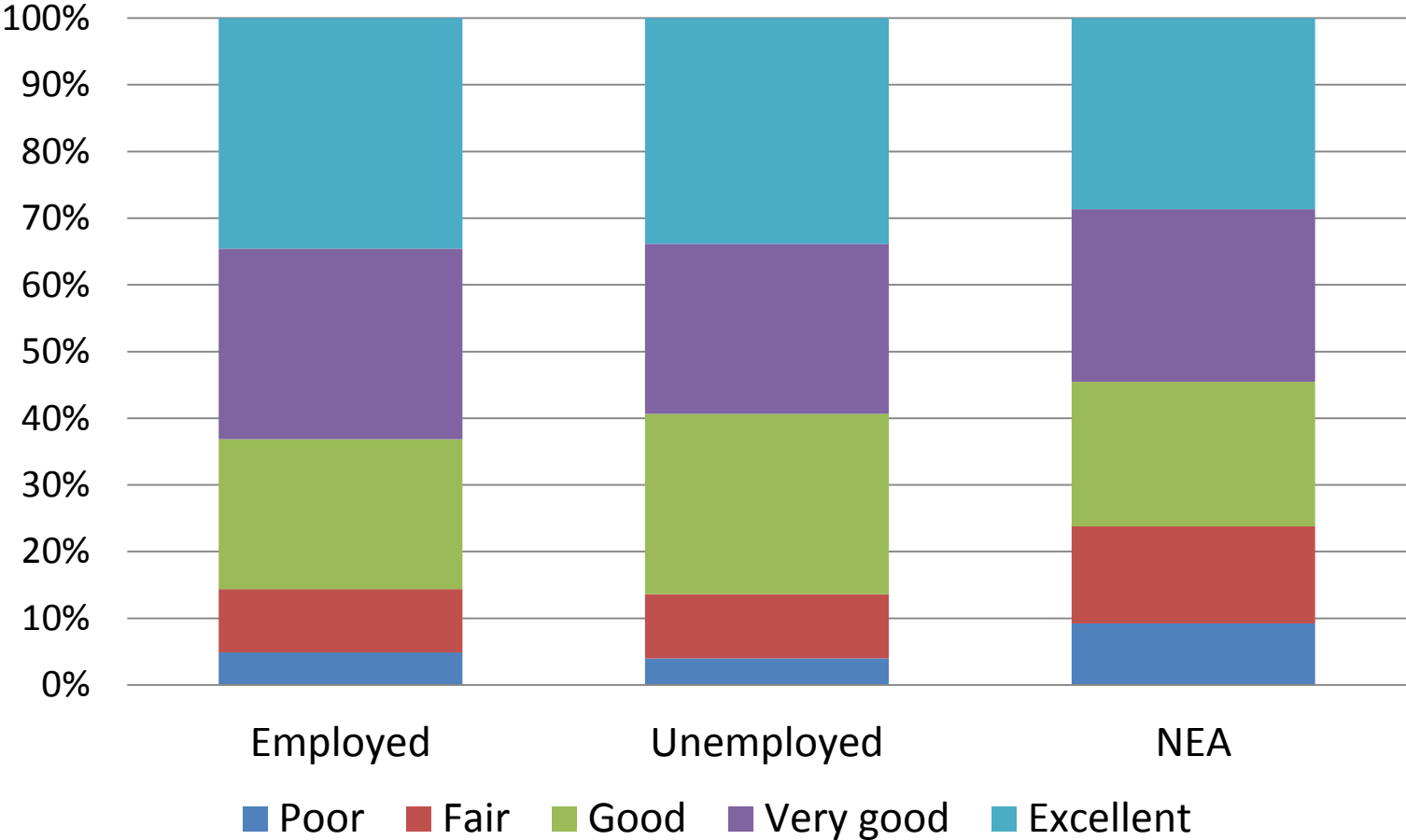
Self-rated health by education



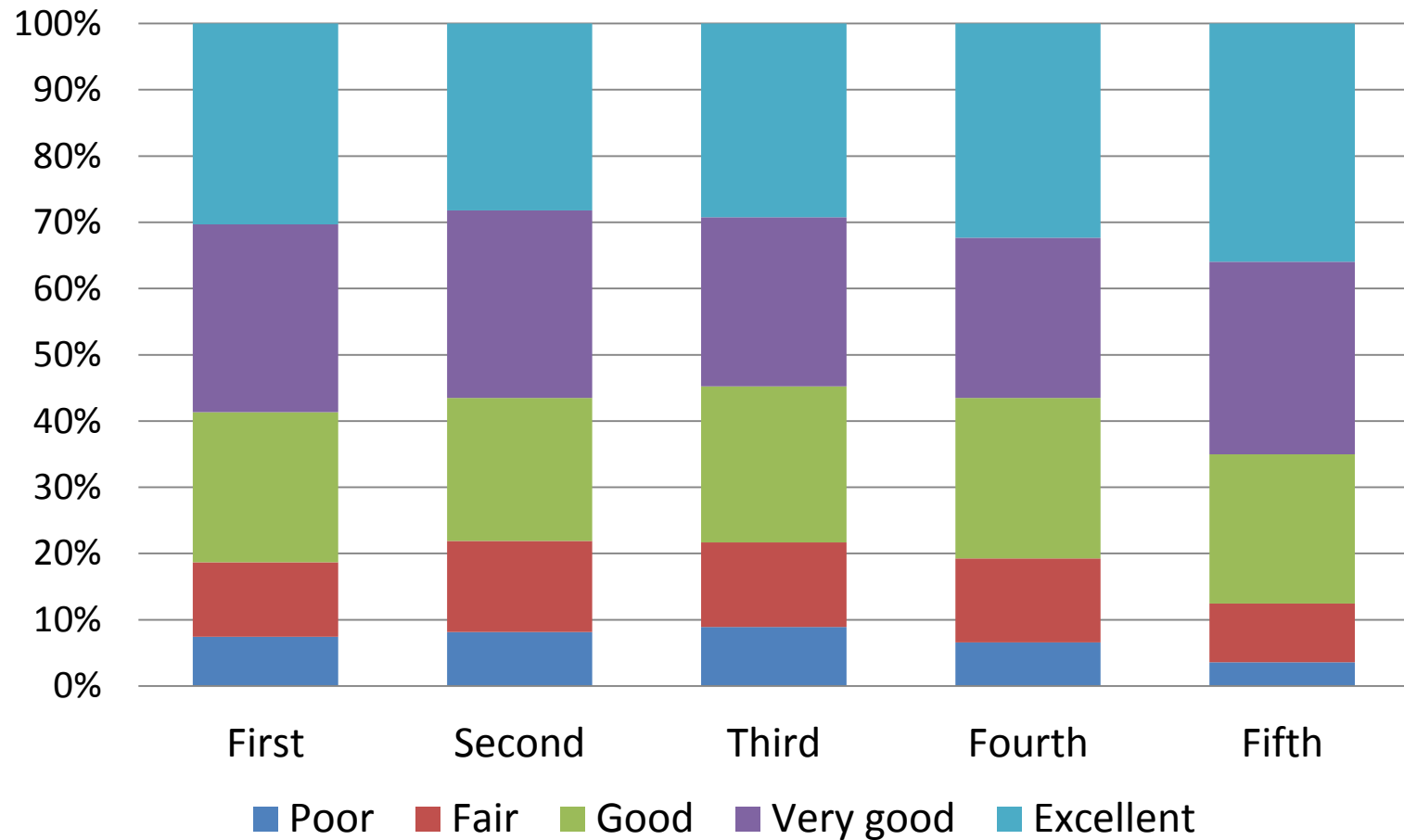
Symptoms and SRH by education category



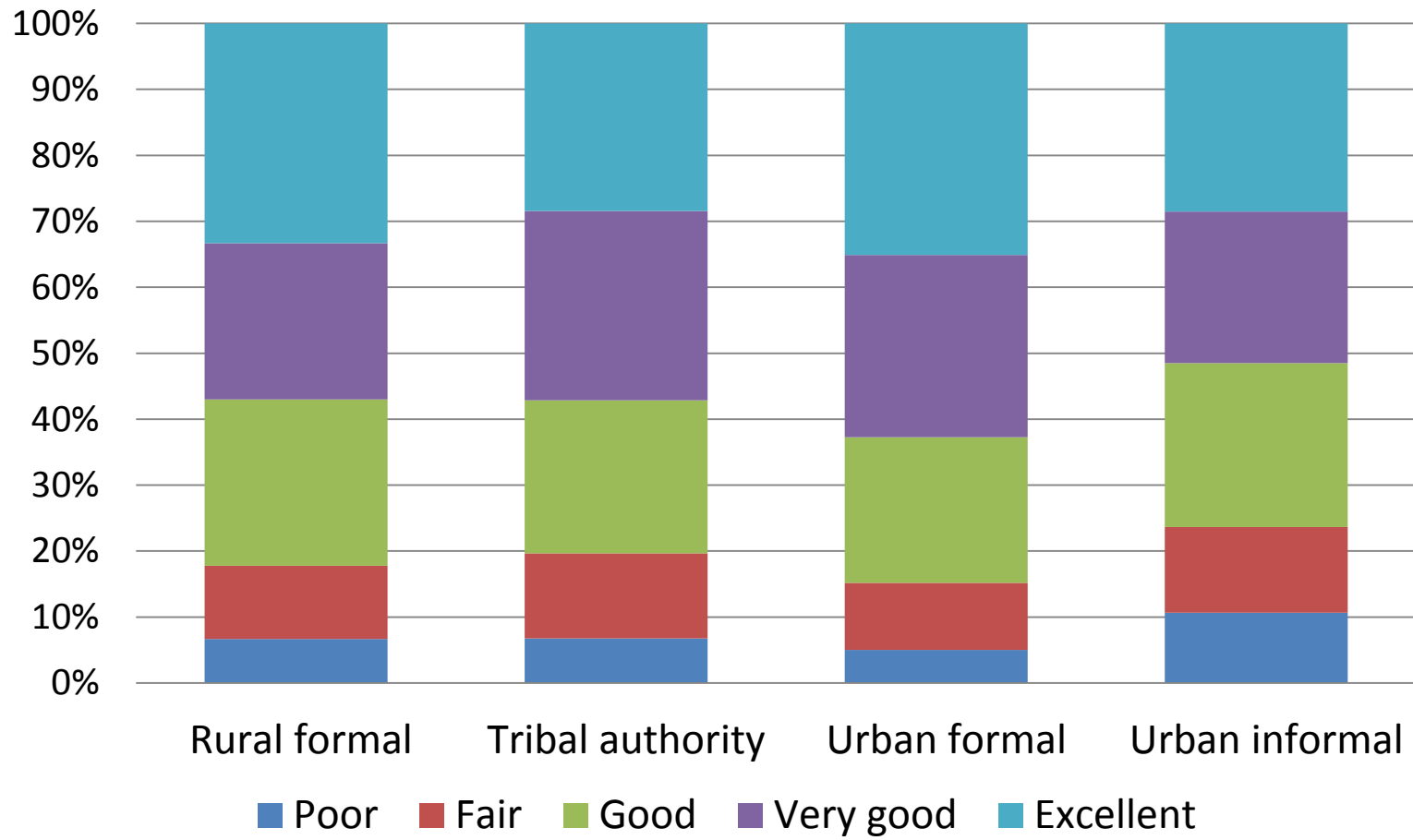
Self-rated health by employment status



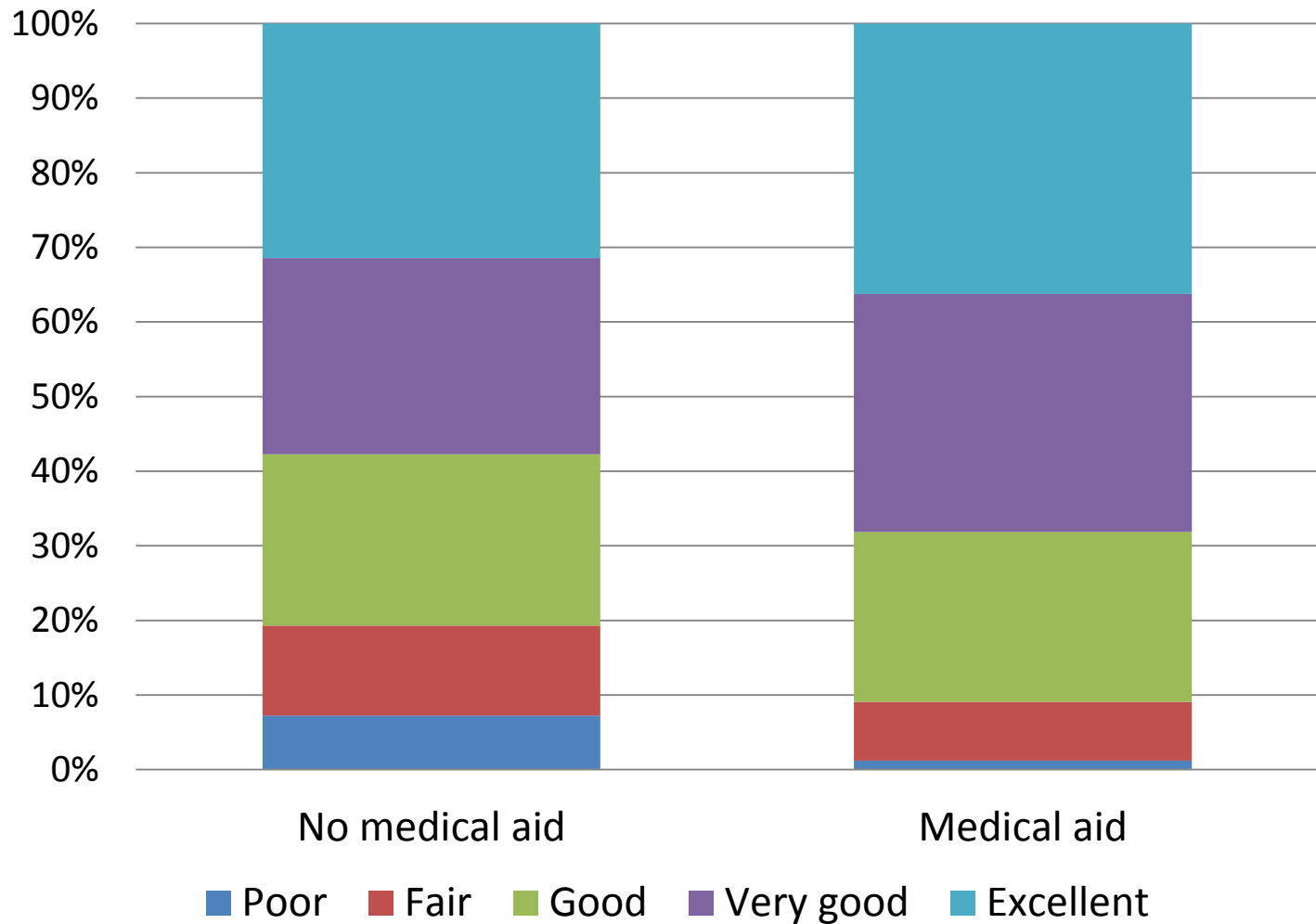
Self-rated health by per capita expenditure quintile



Self-rated health by geo type



Self-rated health by medical aid membership in W1



Reporting heterogeneity

Generalised ordered logit using CSMs from w1-w3 of NIDS

(pooled, adjusting for clusters in pid,

Dep var: SRH)

	Poor	Fair	Good	Very good
Male	0.122***	0.122***	0.122***	0.122***
Age cohort 35 - 45	-0.666***	-0.793***	-0.422***	-0.354***
Age cohort 46 - 59	-0.836***	-1.174***	-0.821***	-0.719***
Age cohort 60+	-1.100***	-1.576***	-1.357***	-1.247***
Coloured	-0.015	0.003	0.123*	-0.084
Indian	-0.008	-0.008	-0.008	-0.008
White	-0.257	0.098	0.200**	-0.037
Unemployed	-0.099**	-0.099**	-0.099**	-0.099**
Not economically active	-0.429***	-0.279***	-0.085**	-0.002
Secondary school	0.843***	0.270***	0.222***	0.210***
Tertiary educ	0.378***	0.378***	0.378***	0.378***
Exp quint 2	0.002	-0.127*	0.073	0.105*
Exp quint 3	0.084	-0.073	0.071	0.257***
Exp quint 4	0.159	-0.078	0.054	0.269***
Exp quint 5	0.22	-0.08	0.129**	0.344***

Reporting heterogeneity (generalised ordered logit)

Continued...

	Poor	Fair	Good	Very good
Married/Cohabiting	-0.064	-0.064	-0.064	-0.064
Divorced / Sep / Wid	-0.127**	-0.127**	-0.127**	-0.127**
Children in hh	-0.024	-0.046	-0.079*	0.048
Tribal authority	0.312***	0.102	-0.013	-0.138**
Urban formal	0.112**	0.112**	0.112**	0.112**
Urban informal	0.021	0.021	0.021	0.021
Medical aid	0.583**	0.221**	-0.029	-0.016
Symptom index	-0.058***	-0.065***	-0.050***	-0.046***
Smoker	-0.149***	-0.149***	-0.149***	-0.149***
Regular drinker	-0.037	-0.037	-0.037	-0.037
Chronic (1 or more)	-0.556***	-0.556***	-0.556***	-0.556***
Chronic (2 or more)	-0.390***	-0.390***	-0.390***	-0.390***
CES-D Depression ind	-0.095***	-0.071***	-0.051***	-0.032***
Intercept	5.438***	4.176***	1.732***	-0.051

Reporting heterogeneity (generalised ordered logit)

- Test parallel lines assumption using generalised ordered logit (with autofit routine)
- Parallel lines assumption is NOT met for (highlighted vars):
 - Age cohort dummies;
 - Race: Coloured and White (African is base category);
 - Employment status: Unemployed and NEA (employed is base category);
 - Secondary education (but not tertiary; primary is base category);
 - Expenditure quintiles;
 - Children;
 - Geo type: Tribal authority area;
 - Medical aid;
 - Symptom index;
 - CES-D scale for depression.

Conclusion

- Reporting heterogeneity
 - Cut point shifts exist with regard to:
 - Age cohorts
 - Race: Coloured and White
 - NEA
 - Secondary education
 - Expenditure quintiles
 - Children in household
 - Geo type: Tribal authority area
 - Medical aid
 - Symptom index
 - CES-D depression scale
 - Index shifts exist with regard to:
 - Gender: Male > 0
 - Unemployed < 0
 - Tertiary education > 0
 - Divorced / separated / widowed < 0
 - Geo type: Urban formal > 0

Way forward

- Sensitivity analysis
- Logit or probit?
- Weights? CSMs only? Balanced panel?
- Correction for reporting heterogeneity (e.g. for NEA index shift) when using SRH as dependent variable?

Thank you!