What is the shape of the relationship between socioeconomic status and health status?

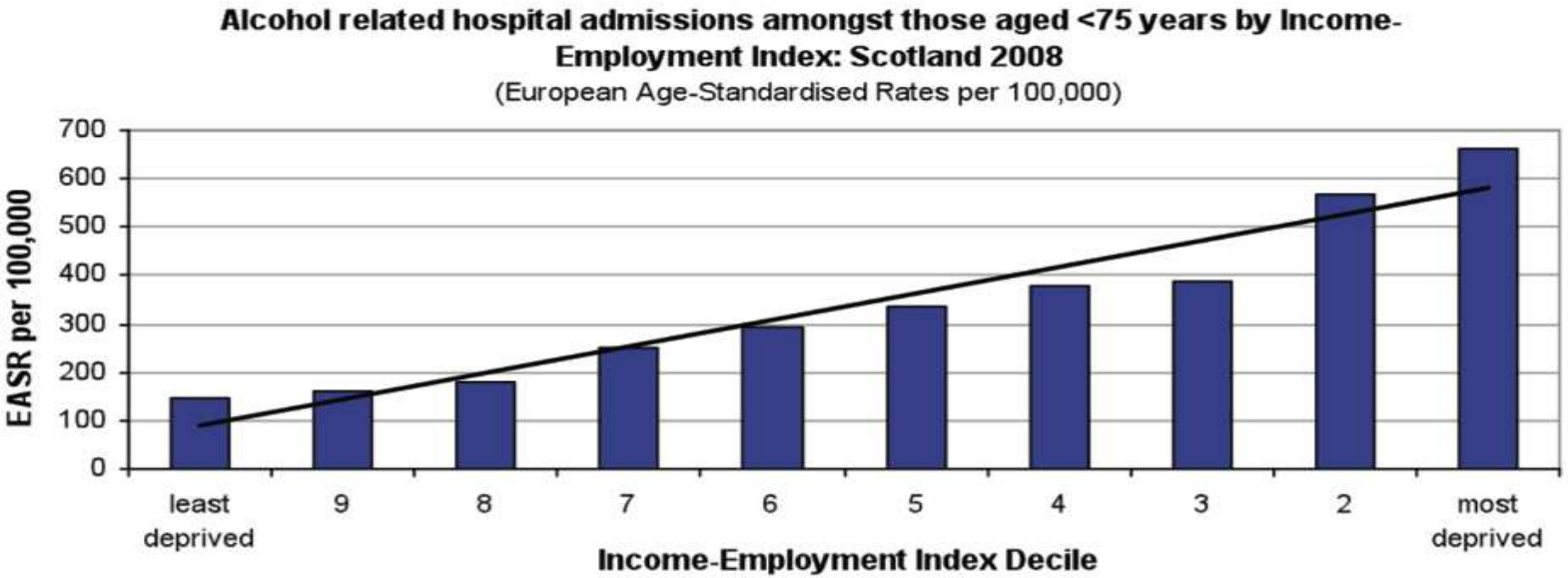
Rob Young & Frank Popham
Universities of Oxford & St Andrews

1. Introduction
2. Methodology
3. Results
4. Conclusions

1. Introduction

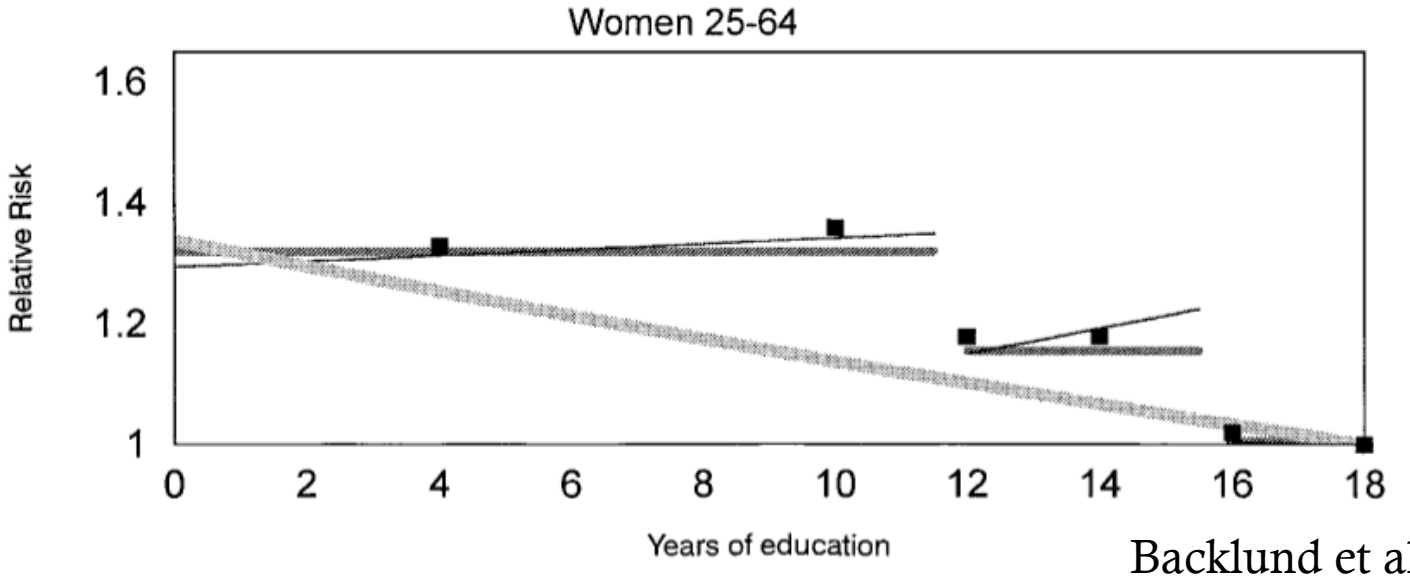
- Inequalities in health are observed across society, with those at the lower end of the social hierarchy generally suffering more.
- Health interventions to alleviate this, such as the Scottish Government's 'Equally Well' programme, depend on a robust and straightforward way to monitor these inequalities.
- A variety of measures of varying complexity can be found in the literature.

1. Introduction



- The Scottish Govt. models the dose-response relationship between SES and health as a linear one.
- The slope of the regression line is the relative difference between the top and bottom of the population distribution – the Relative Index of Inequality (RII).
- Is this assumption of linearity valid?

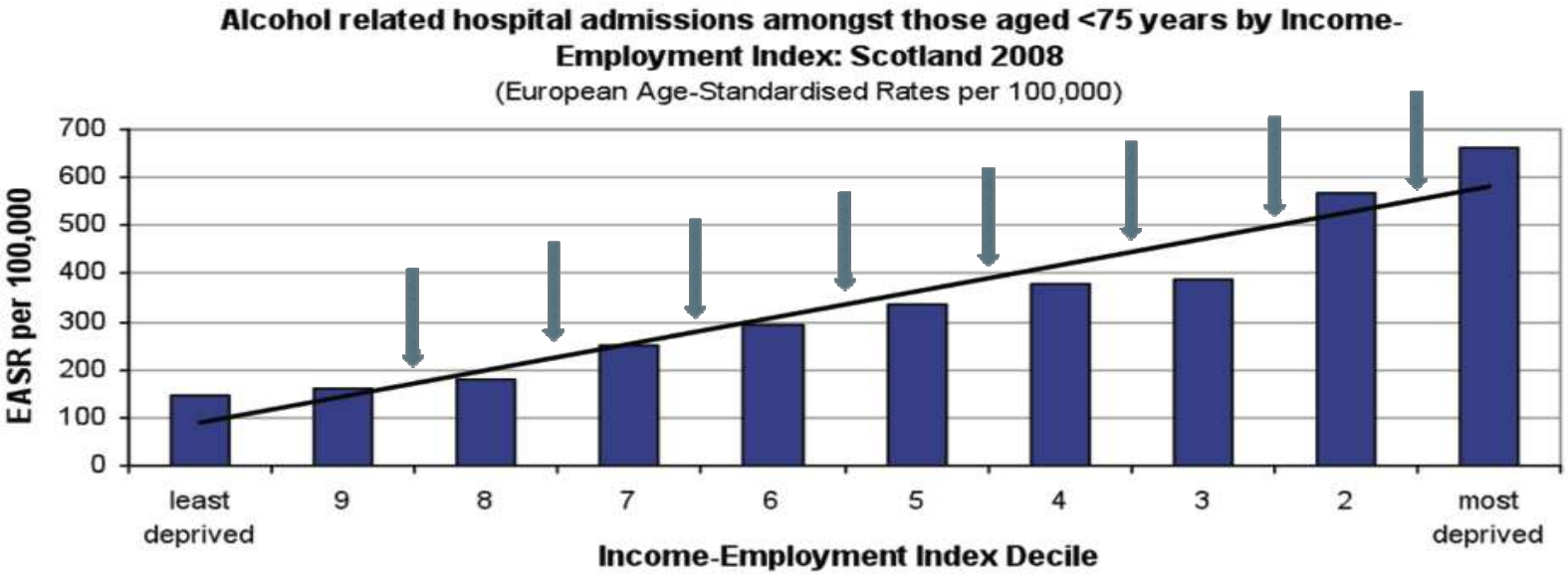
1. Introduction



Backlund et al., 1999

- The straight line is allowed to change slope and/or location only at previously well-defined education thresholds.
- There is no quantification of the overall cost of SES inequality.

2. Methodology



- Possible turning points ('knots') are placed at the end of every category, after the second.
- Population size can be accounted for where the groups need not be equally sized.

2. Methodology

- A model is fitted with a different slope between each knot, where the significance of each can be estimated.
- Non-significant knots ($p > 0.05$) are removed and the model refitted.
- This is repeated as many times as there are initial knots, although it frequently stabilises early on.
- A likelihood ratio test is used to confirm that any decreased parsimony is justified.

2. Methodology

- We first quantify the model fit using the Population Attributable Risk (PAR) which is suitable for all three model types:

$$PAR = \sum P_D \left(\frac{RR - 1}{RR} \right)$$

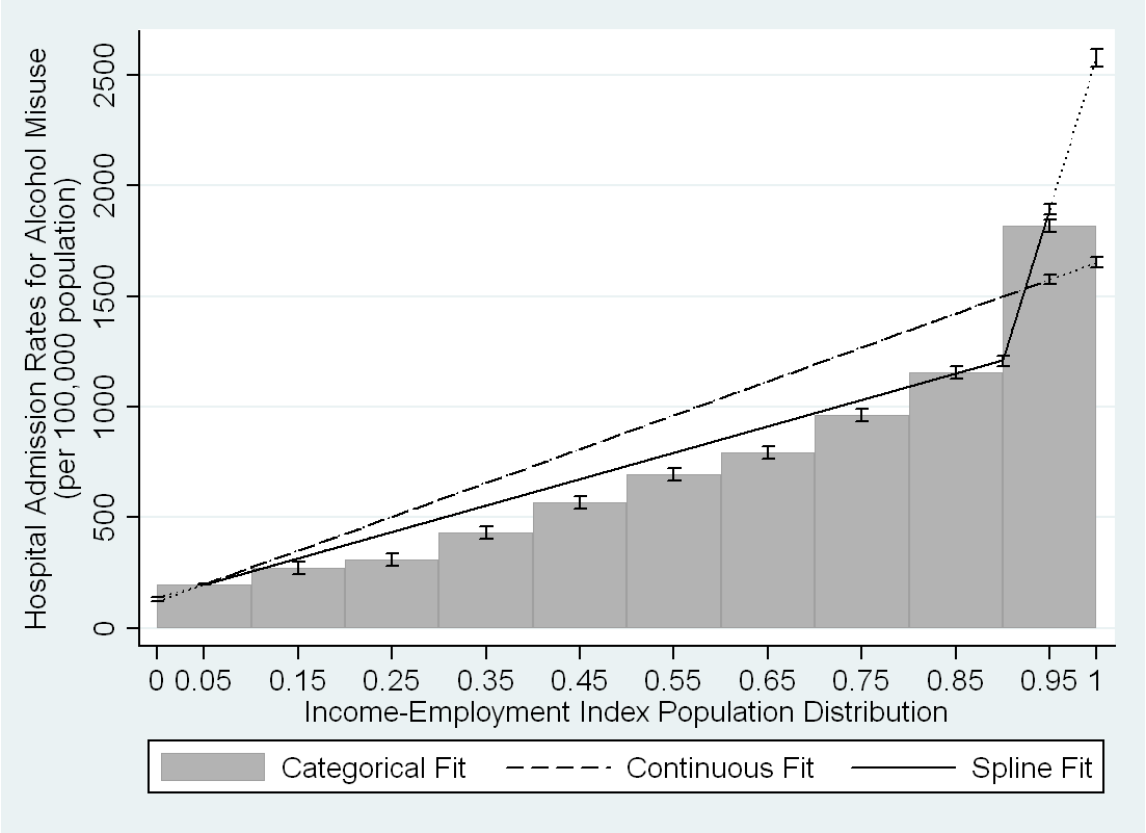
where P_D is the prevalence of the health outcome in that group and RR is the risk or incidence rate, relative to the highest SES group.

- Groups are defined by the initial categories, not the significant knots.

2. Methodology

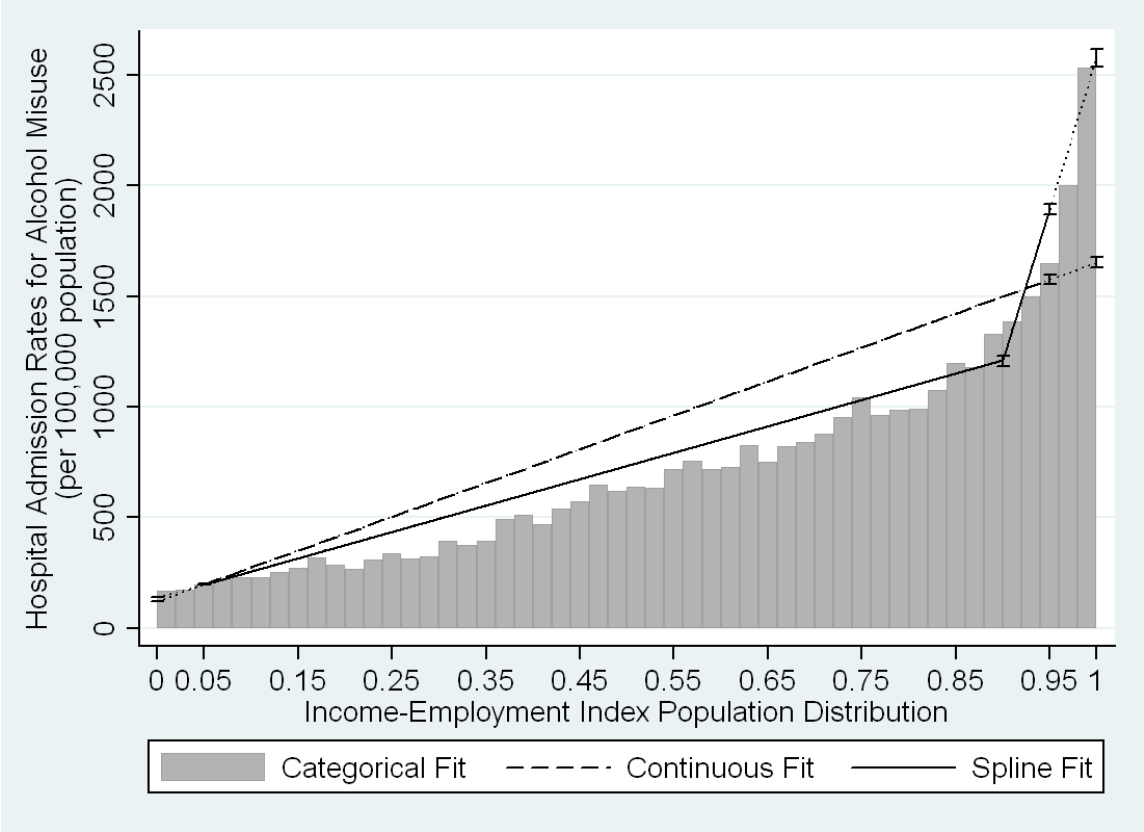
- We also estimate the RII as normal as the difference between the hypothetical individuals at the top and bottom of the population distribution.
- For a continuous model, this is simply the regression slope.
- For a spline model, the different slopes are weighted by the distance travelled by the spline.

3. Results: Alcohol Misuse *vs.* Deprivation



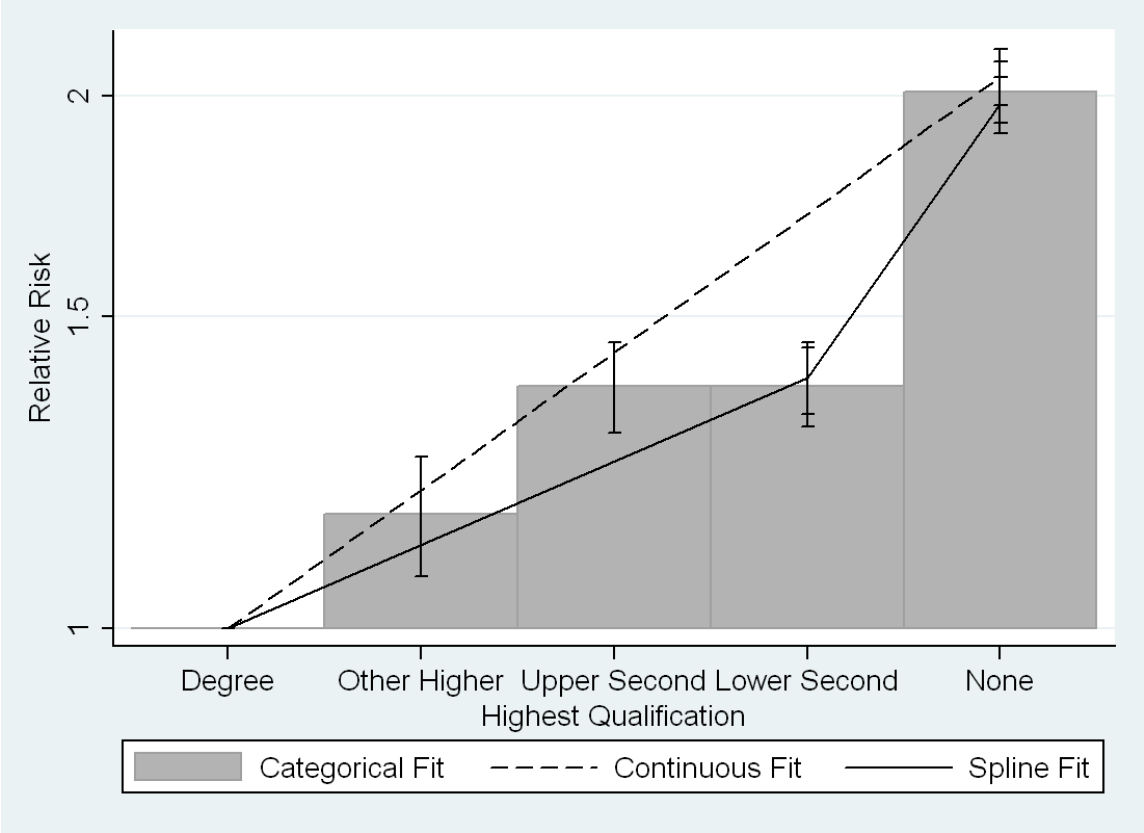
- The spline fit has statistical support – likelihood ratio test $p < 0.001$.
- There is a significant knot at 0.90 ($p < 0.001$).
- The spline PAR (0.82 ± 0.02) is closer to the categorical (0.94 ± 0.03) than the continuous (0.73 ± 0.03).
- The spline RII (4.4 ± 0.06) is much greater than the continuous (3.1 ± 0.03).

3. Results: Alcohol Misuse *vs.* Deprivation



- When the data is split into 50-quantiles, the spline fit based on 10 categories is surprisingly good.
- The large spline RII may be truly representative.

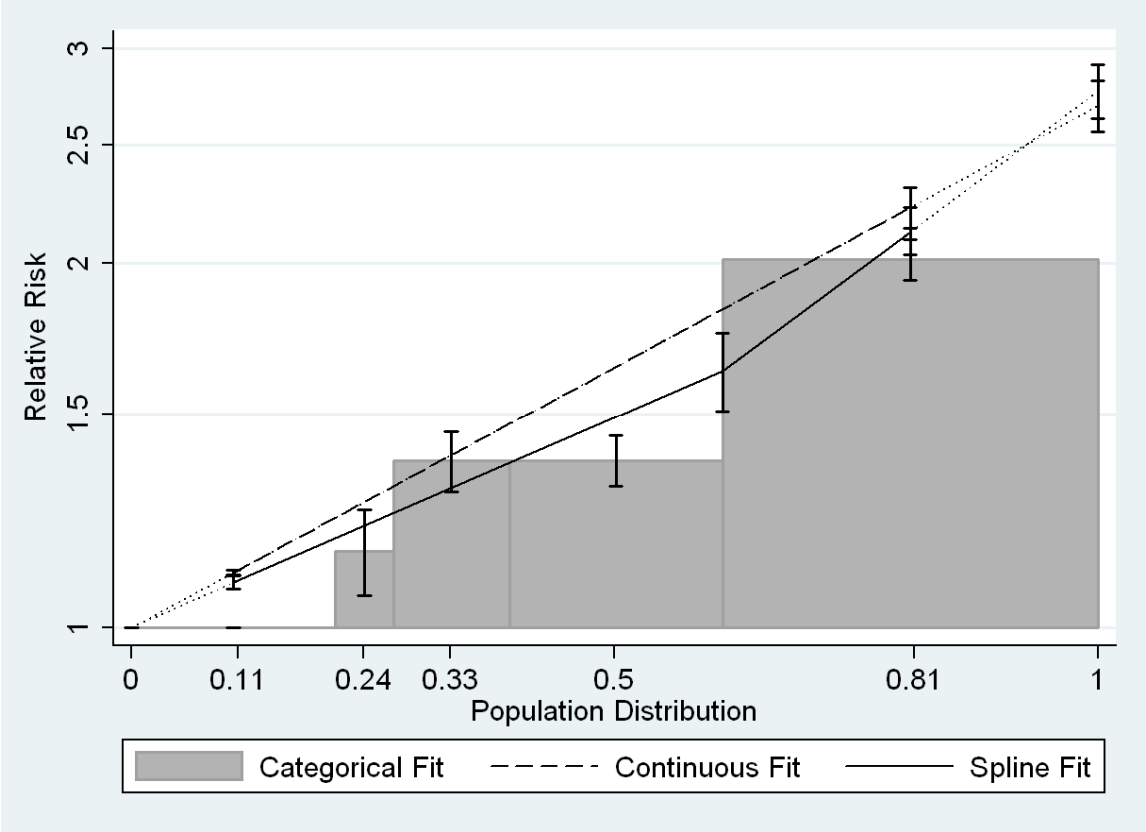
3. Results: Education status *vs.* Mortality



Source: Scottish Longitudinal Study

- The spline fit has statistical support.
- There is a significant knot after lower second ($p < 0.001$).
- The RII values converge on 2 – because the different sizes of the groups has not been accounted for?

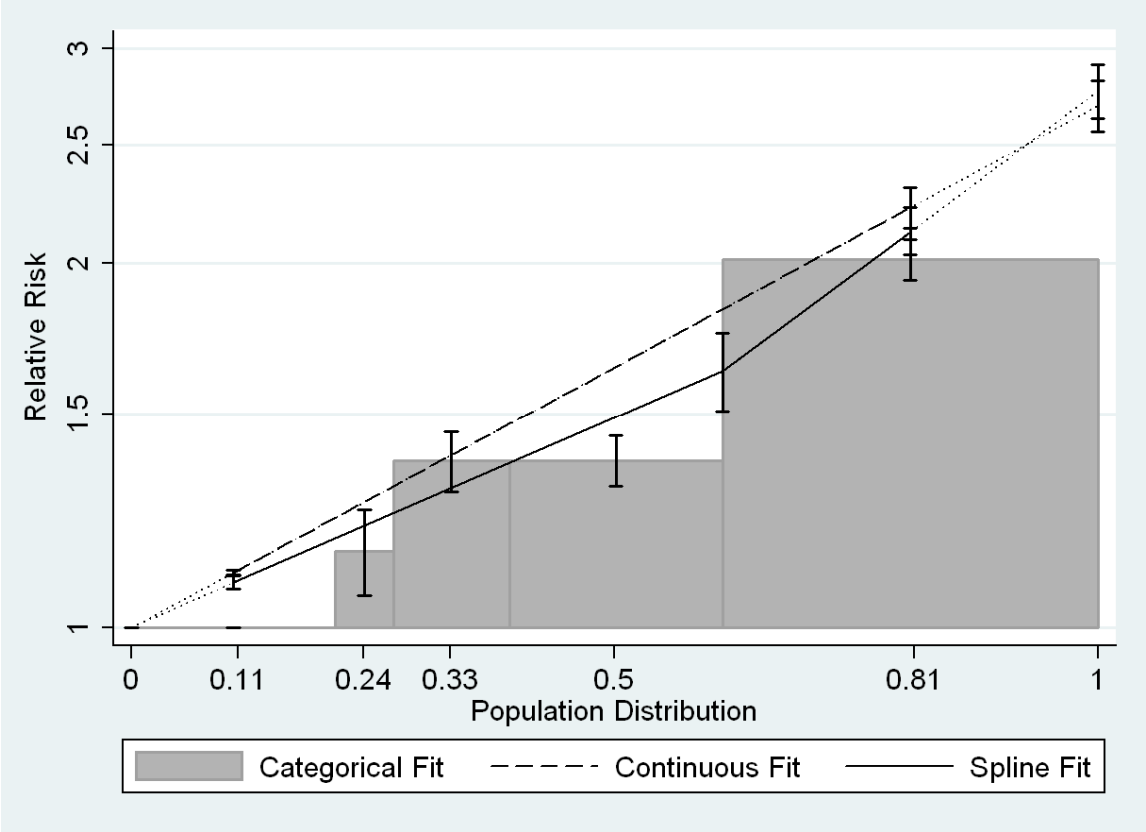
3. Results: Education status *vs.* Mortality



Source: Scottish Longitudinal Study

- Now including population size, as for the alcohol example.
- The knot at 'no qualifications' is only marginally significant ($p = 0.095$).
- The PAR values are all around 0.39, and the RII estimates for the continuous and spline models are similar.

3. Results: Education status *vs.* Mortality



Source: Scottish Longitudinal Study

- The RII estimates are based on hypothetical individuals at both ends of the population distribution.
- It may not be appropriate to extrapolate a data-led spline model beyond the data space (30% of the population, in this example).

4. Conclusions

- Individuals in the bottom 10% most deprived data zones have an increased rate of alcohol misuse disproportionate to their position in society.
- There is no non-linearity in the relationship between coronary heart disease and deprivation as measured above (data not shown).
- The increased risk of mortality for people with no qualifications is only significant when the population size of the groups is ignored.

4. Conclusions

- Our new, spline-based approach is simple yet powerful.
- It can simultaneously detect the number, location, and magnitude of knots in any relationships of the type described here.
- The use of two standard reporting statistics makes it straightforward to interpret the results.
- It can be implemented in any standard statistical package (we have used Stata, and have written this method as an ado file).

5. Acknowledgements

- John Frank and Sally Haw at the Scottish Collaboration for Public Health Research and Policy (SCPHRP).

- Alastair Leyland for helpful discussions.

- The help provided by staff of the Longitudinal Studies Centre - Scotland (LSCS) is acknowledged. The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office and the Scottish Government. The authors alone are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

- The MRC for funding this project.