Scientific Institute
of Public Health

"Variance Estimation Methods for Health Expectancy by relative socio-economic Status"

Abatih E.

Van Oyen H., Bossuyt N. & Bruckers L.

Manila, 7th May, 2008

Contents

Introduction

Socio-economic inequalities in HE

- Uses educational level, income or occupation
- Inform health policy makers
- Definitions of SES
- Cohort effect
- Comparison between sub-groups

SE inequalities in HE on relative scale

- Neutral ground (Independent of definitions for SES)
- Accounts for the cohort effect (regression)
- Fosters international comparisons

Methodological issues

- Data
- Sampling design scheme of HIS
- Variance estimation
- Confidence intervals

Objectives:

- Estimation of variance of HE by Relative SES
- Incorporate survey design
- Confidence intervals
- Implications of methods

Methods

Measures of morbidity

- Limiting or extremely long-standing illness
- Perceived general health status
- Functional disabilities

Calculation of HE

- Sullivan's method
- Based on the life table

  Describes the survival experience of a real or hypothetica group of people followed from birth or other ages in thei time

- Abridged life table

  5 or 10 years age intervals

Sullivan's HE:

$$H\hat{E} = \frac{[\sum (1-\pi_i) * L_i]}{l_x}$$

Where; π_i are the prevalences of ill health conditions

L_i are the number of person years lived in the age inte

l_x number of persons surviving at the begining of the interval

Variance of HE:

$$S^2(HE_i) = \frac{1}{l_x^2} \sum_{i=x}^{w} L_i^2 * S^2(\pi)$$

Where;

$$S^2(\pi) = \frac{\pi * (1-\pi)}{N}$$

Measurement of inequalities in HE

- Higher socio-economic status versus lower socio-economic status
  - Sizes of two groups differ
  - Only extreme groups are compared
  - Association throughout levels of socio-economic status no taken into account
  - Cohort effect
- Regression based method
  - SES is operationalized as relative position on a SE scale (between 0 and 1)

Survey logistic regression

- Based on raw survey data
- Survey sampling design variables
  - Weights
  - Stratification variable
  - Clustering variable
- Survey logistic regression model:

$$\mathrm{logit}(\pi) = \beta_0 + \beta_j * x_j$$

$$\theta = \mathrm{logit}(\hat{\pi}) \quad \Longleftrightarrow \quad \hat{\pi} = \frac{\exp(\theta)}{1 + \exp(\theta)}$$

Variance of prevalence

Delta method

Letting $\rho = \text{var}(\theta)$

$$\text{var}(\hat{\pi}) = \left[\frac{\exp(\theta)}{[1+\exp(\theta)]^2}\right]^2 * \rho$$

Variance of HE

Sullivan's method

95% Confidence intervals

Normal approximation to the binomial:

$$H\hat{E} \pm s.e(H\hat{E}) * 1.96$$

Bootstrap weighted least squares regressio

- Aggregated data
- Survey design: weighted prevalences
- Assume relationship between prevalence(y) and relative position (x) on the social hierarchy are lin
- Regression model:

$$y=\alpha+\beta^{*}x+\varepsilon$$

- Weights=relative sizes of the educational levels fo each age group

- Generate n prevalences from a Bernoulli distribut
- Fit model n times for each age group

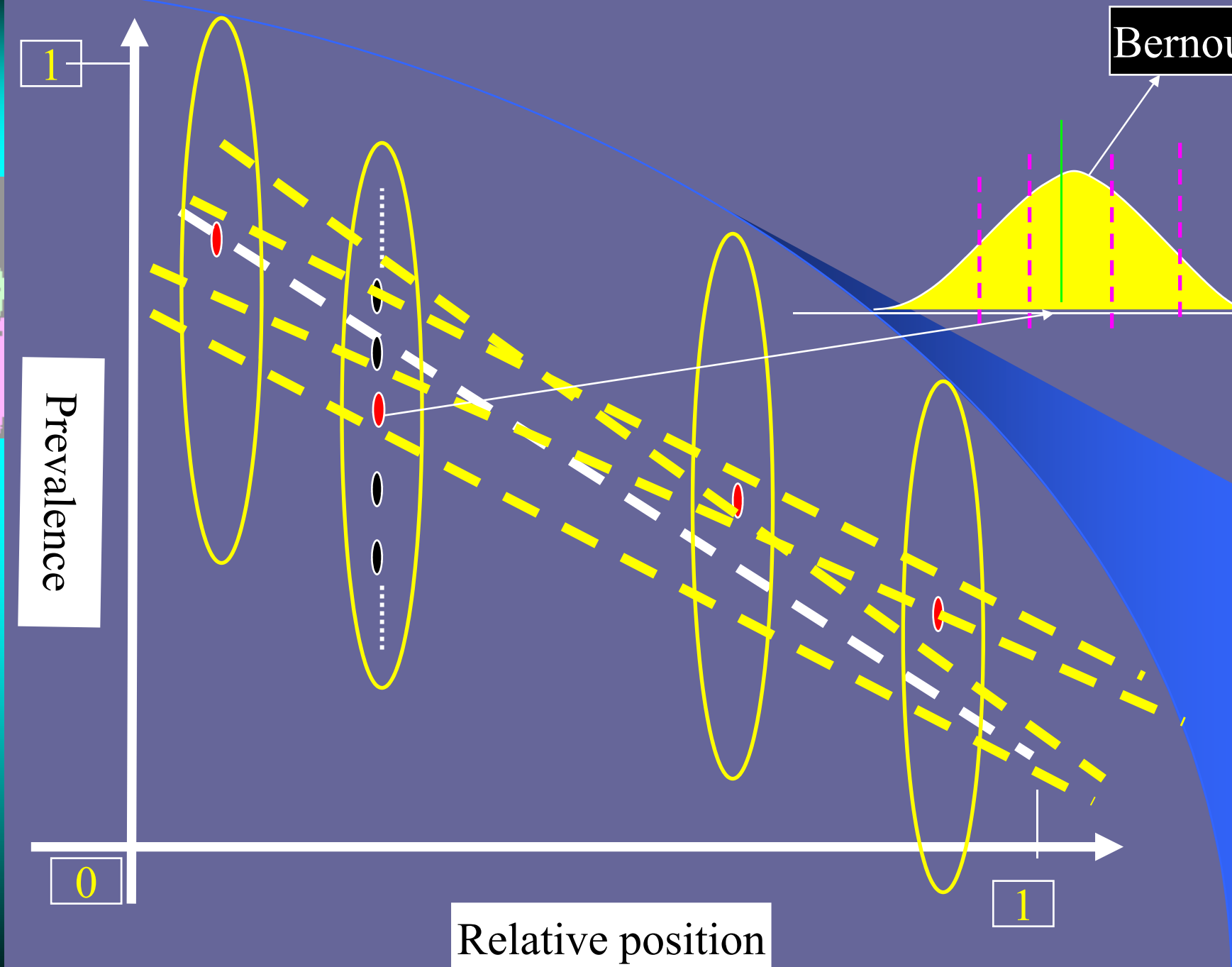
  Predict the prevalence of ill health condition for those
  the highest (x=1) and lowest (x=0) positions of the soc
  hierarchy
- Use Sullivan's method to estimate the HEs

HE and its Variance :

- Use distribution of the n generated HE

95% confidence intervals:

- Studentized confidence intervals
- Give better coverage


Bernou
1
Prevalence
0
1
Relative position

CI of inequalities in HE

– Cauchy Swartz inequality:

$$\mathrm{var}(H\hat{E}_1 - H\hat{E}_2) \leq [\sqrt{\mathrm{var}(H\hat{E}_1)} + \sqrt{\mathrm{var}(H\hat{E}_2)}]^2$$

$$(H\hat{E}_1 - H\hat{E}_2) \pm 1.96 * \sqrt{\mathrm{var}(H\hat{E}_1 - H\hat{E}_2)}$$

Case study

- Disability-free life expectancy (DFLE)

- Objectives
  - Estimate DFLE and their variances
    - Logistic regression
    - Bootstrap

- Testing for significance of differences in DFLE

- Compare survey logistic regression and bootstrap method

Mortality data

- Derived from the Belgian National Mortality Database

Morbidity data

- Health Interview Survey (HIS)(1997)
- Based on a complex sampling design scheme

Measure of morbidity

Functional disability

World Health Organisation (WHO) instrument including activities of daily living (ADL) e.g dressi hearing, seeing etc….

Moderately limited: had difficulties performing one the activities

Severely limited: could only perform activities with help of others

Disability

Severely limited or moderately limited

Measure of socio-economic status

Educational level

Only available instrument from the NMD and the HIS

Variables:

- Prevalence (1 for disabled and 0 for non-disabled)
- Relative position (on a continuous scale from 0 to
- Region (1 for Flemish and 0 for Walloon)
- Gender (1 for men and 0 for women), and
- Agegrp (1 through 5)

25-74 by 10

Results

able 1: Comparison of results from the Bootstrap ar logistic regression methods for Flemish Women

	Bootstrap				Logistic regression			
	Lowest position		Highest position		Lowest position		Highest positio	
e	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Varian
	34.49	3.350	39.95	1.116	28.33	0.911	41.90	0.219
	24.81	3.250	30.97	1.073	20.16	0.829	32.36	0.215
	17.84	2.594	21.54	0.949	13.00	0.667	23.16	0.209
	11.01	1.953	12.85	0.750	6.88	0.458	14.44	0.192
	3.51	1.011	6.27	0.460	2.70	0.179	6.67	0.124

Results

able 1: Comparison of results from the Bootstrap ar logistic regression methods for Walloon Women

	Bootstrap				Logistic regression			
	Lowest position		Highest position		Lowest position		Highest positio	
e	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Variance	$DFLE_{25-74}$	Varianc
	25.06	2.349	41.93	1.070	25.57	0.883	40.25	0.259
	17.01	2.159	32.75	0.990	17.83	0.777	30.85	0.253
	11.15	1.558	23.63	0.909	11.25	0.585	21.87	0.242
	3.92	1.262	15.34	0.758	5.77	0.378	13.41	0.220
	1.05	0.467	7.16	0.436	2.21	0.132	6.10	0.131

e 3: Differences in $DFLE_{25-74}$ between Flemish women at the lowest and hig positions of the socio-economic hierarchy

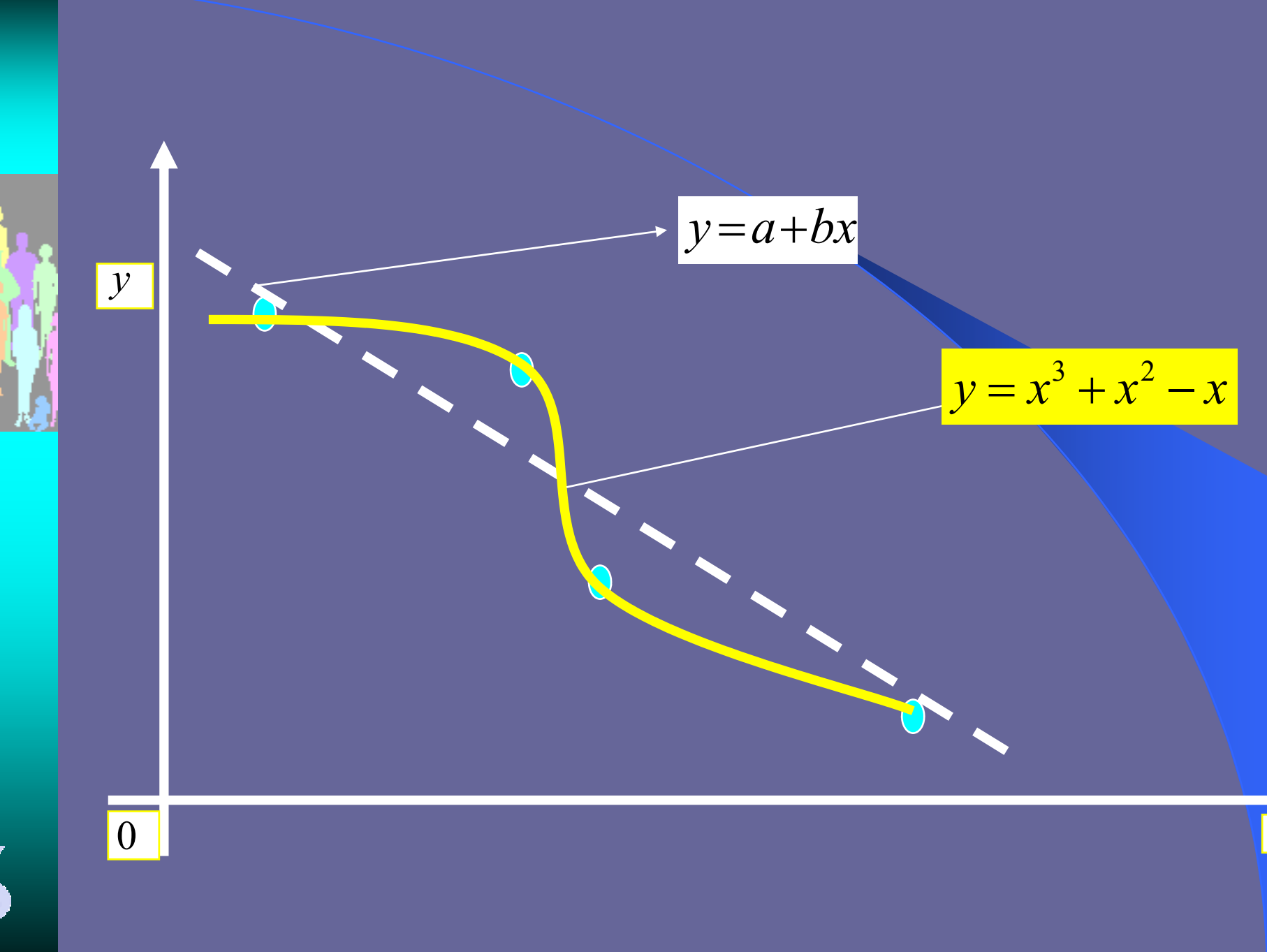
	BOOTSTRAP				LOGISTIC REGRESSION			
	Difference in $DFLE_{25-74}$	Approx SE of diff	Z-statistic	P-value	Difference in $DFLE_{25-74}$	Approx SE of diff	Z-statistic	P-va
	5.46	2.887	1.89	>0.05	13.57	1.423	9.54	<0.0
	7.16	2.839	2.52	<0.02	12.20	1.375	8.87	<0.0
	3.70	2.585	1.43	>0.10	10.16	1.273	7.98	<0.0
	1.84	2.264	0.81	>0.20	7.56	1.115	6.78	<0.0
	2.76	1.684	1.64	>0.05	3.97	0.775	5.12	<0.0

Discussion

- Sampling design of HIS should be taken into account
- Use raw survey data

Bootstrap method :

- Larger variance estimates
- Different conclusions
- Aggregated data (very small sample size)
- Partial account of Survey sampling design
- Linearity assumption


y
0
$y = a + bx$
$y = x^3 + x^2 - x$

Logistic regression

Raw survey data (larger sample size)

No linearity assumption

Full account of survey design

Correct variance estimates

Conclusions

- Use logistic regression method on raw survey data where available
- Use bootstrap when only aggregated data are available and sample size large enough
- Take survey design into account

Future perspectives

Assess performance of methods on incident rat
using Multi-state life table method

Develop strategies for taking full account of
survey design in the bootstrap method

Software

Stata and SPLUS

- Stata : Computes logit and its variance
- Splus: Aggregation destroys survey design

SAS and R

- SAS: proc surveylogistic, variance of p
- R : Survey design package

Thanks for your attention!