Int J Public Health
DOI 10.1007/s00038-009-0105-z

HINTS & KINKS

Decomposing socioeconomic health inequalities

Niko Speybroeck · Peter Konings · John Lynch · Sam Harper · Dirk Berkvens · Vincent Lorant · Andrea Geckova · Ahmad Reza Hosseinpoor

Received: 28 May 2009 / Revised: 4 September 2009 / Accepted: 15 September 2009
© Birkhäuser Verlag, Basel/Switzerland 2010

Keywords Decomposition · Concentration index · Inequity · Health

This Hints & Kinks paper describes a technique enabling quantification of the contributions of determinants to socioeconomic inequality in health. This technique, differing from an analysis investigating the determinants of average health levels, has received considerable attention from health economists (van Doorslaer and Gerdtham 2003; van Doorslaer and Jones 2003; van Doorslaer et al. 2004; Wagstaff et al. 2003) but only more recently from epidemiologists as well (Harper and Lynch 2007; Lynch 2006; Hosseinpoor et al. 2006). This paper employs the relative concentration index (RCI), described in Konings et al. (2009), to summarize relative inequality across the entire socioeconomic distribution. The RCI of a continuous health outcome y results from a relative concentration curve, which graphs on the x-axis the cumulative percentage of the sample, ranked by an indicator of socioeconomic position such as education or income beginning with the poorest. The y-axis then indicates the cumulative percentage of the health outcome corresponding to each cumulative percentage of the distribution of the socioeconomic indicator. Figure 1 provides an example of a concentration curve, where the health variable is childhood malnutrition in Ghana in 2003. It shows that the level of malnutrition accumulates faster among the poor than among the better-off because the line is above the diagonal. The RCI is defined as twice the area between the concentration curve and the line of equality (the 45° diagonal from the bottom-left corner to the top-right). Details on how to compute the RCI can be found in Konings et al. (2009).

If y_i is linearly modeled, linking a health variable, y to a set of k health determinants, x_k, this can be expressed by:

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i \quad (1)$$

where ε is an error term.

Given the relationship between y_i and x_{ki} in Eq. 1, the RCI for y can be written as (Wagstaff et al. 2003):

The findings, interpretations and conclusions expressed in the paper by named authors are solely the responsibility of those authors.

N. Speybroeck · D. Berkvens
Institute of Tropical Medicine, Antwerp, Belgium

N. Speybroeck (✉) · V. Lorant
Institute of Health and Society,
Université Catholique de Louvain, Louvain, Belgium
e-mail: nspeybroeck@itg.be

P. Konings
Department of Electrical Engineering,
SCD (SISTA), K.U. Leuven, Leuven, Belgium

J. Lynch
Sansom Institute, University of South Australia,
Adelaide, SA, Australia

J. Lynch
Department Social Medicine,
University of Bristol, Bristol, UK

S. Harper
Department of Epidemiology, Biostatistics,
and Occupational Health, McGill University,
Montreal, QC, Canada

A. Geckova
Kosice Institute for Society and Health,
Kosice, Slovak Republic

A. R. Hosseinpoor
World Health Organization, Geneva, Switzerland

Published online: 09 January 2010

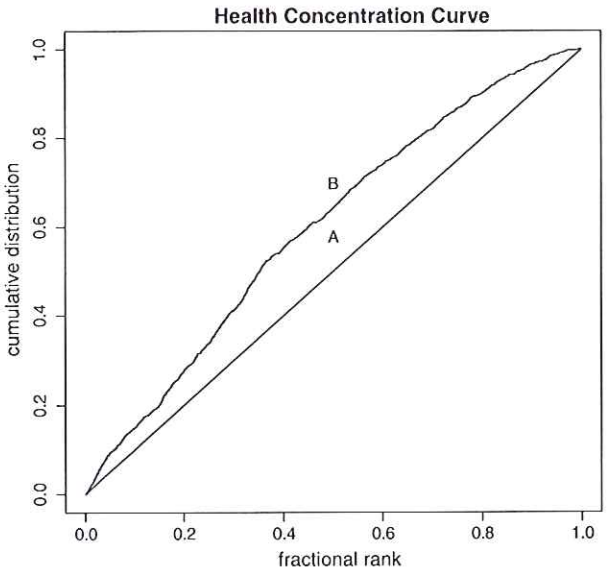


Fig. 1 The relative concentration curve: an example with malnutrition (Ghana, 2003; *Source* Demographic and Health Survey). The relative concentration index equals 2 × the area between the 45° line and the relative concentration curve = A/(A + B)

$$\mathrm{RCI} = \sum_k \left(\frac{\beta_k \bar{x}_k}{\mu}\right) C_k + \frac{\mathrm{GC}_\varepsilon}{\mu} = C_{\hat{y}} + \frac{\mathrm{GC}_\varepsilon}{\mu} \qquad (2)$$

with μ the mean of y, $\bar{x}_k$ the mean of x_k, and C_k the RCI for x_k (defined analogously to C). In the last term, GC_ε is the generalized concentration index for ε_i. It can be computed as a residual or using the formula $\mathrm{GC}_\varepsilon = \frac{2}{n}\sum_{i=1}^{n} \varepsilon_i R_i$, with R_i the fractional rank in the socioeconomic distribution, corresponding to ε_i (i.e. $R_i = 1/N$ for the poorest individual and $R_i = N/N$ for the richest).

Equation 2 shows that C is made up of two components. The first is a deterministic, or "explained" component (Wagstaff et al. 2003), equaling a sum of weighted concentration indices of the explanatory variables, where the weights are $(\beta_k \bar{x}_k/\mu)$. The second, a residual, or "unexplained", component reflects the inequality in health that is left unexplained by the determinants in the model.

A decomposition analysis can be conducted in a stepwise manner (Hosseinpoor et al. 2006). A first step is to estimate a regression model of the health variable to obtain the coefficients of the explanatory variables (β_k) in Eq. 1. The next step consists of calculating the average of the health variable and each of the determinants (μ and $\bar{x}_k$). A third step requires computing the RCIs for the health variable and for the determinants as well as the generalized concentration index of the error term (GC_ε). The RCI of each determinant is quantified using the computation explained in Konings et al. (2009). y_i and μ are now the value of the specific determinant for the ith individual, and the average of the determinant, respectively. These three steps provide all values required in Eq. 2. The last and remaining step quantifies the "pure" contribution of each determinant included in the model to the inequality in the health variable. This absolute contribution of each determinant is computed by multiplying the "contribution weight" related to a determinant and its RCI:

$$\left(\left(\frac{\beta_k \bar{x}_k}{\mu}\right) C_k\right).$$

The relative (proportional) contribution of each determinant is then obtained by dividing its absolute contribution by the RCI of the health outcome

$$\left(\left(\frac{\beta_k \bar{x}_k}{\mu}\right) C_k \Big/ C\right).$$

This sequence combines a regression analysis with distributional data (from the RCI) and thus allows estimating how determinants proportionally contribute to health inequality (e.g. the poor-rich gap). As an example the contribution of breastfeeding to childhood malnutrition inequalities in Ghana (Van de Poel et al. 2007) can be calculated by noting that the average "height for age" z score is $\mu = 1.58$ and the average duration of breastfeeding (in months) is $\bar{x}_k = 16.98$. The RCI for malnutrition was $C = -0.079$ and the RCI for breastfeeding was $C_k = -0.0042$. The coefficient in the regression for breastfeeding was $\beta_k = 0.01$. These parameters result in a pure contribution of $\left(\frac{\beta_k x_k}{\mu}\right) C_k = \left(\frac{0.01 \times 16.98}{1.58}\right) - 0.0042 = -0.00045$ and a relative contribution of breastfeeding $\left(\frac{\beta_k x_k}{\mu}\right) C_k \Big/ C = \frac{-0.00045}{-0.079} = 0.0057$ or 0.57%.

Notice that the 0.57 is slightly different from (Van de Poel et al. 2007) because of rounding errors. Notice that a negative contribution means that the effect of the explanatory variable on health (i.e. the regression coefficient) combined with the distribution of that variable over economic status is to "lower" socioeconomic inequality in health, favoring the poor. For example, age makes a negative contribution to malnutrition inequalities in Ghana (Van de Poel et al. 2007) because older children are both more likely to be malnourished and, probably due to higher child mortality rates among poorer households, more prevalent in the richer wealth quintiles.

The aforementioned methodology applies to continuous health outcomes. Non-normally distributed health outcomes require some modifications. A binary health outcome, death for example, requires a logit model which is non-linear in the probability of e.g. death, but linear in the natural logarithm of the odds of death. Moreover, since we are concerned with describing the inequality in predicted death, given the observed values of the determinants, attention will be limited to the first term in Eq. 2, i.e. the predicted inequality measured by $C\hat{y}$ then written as:

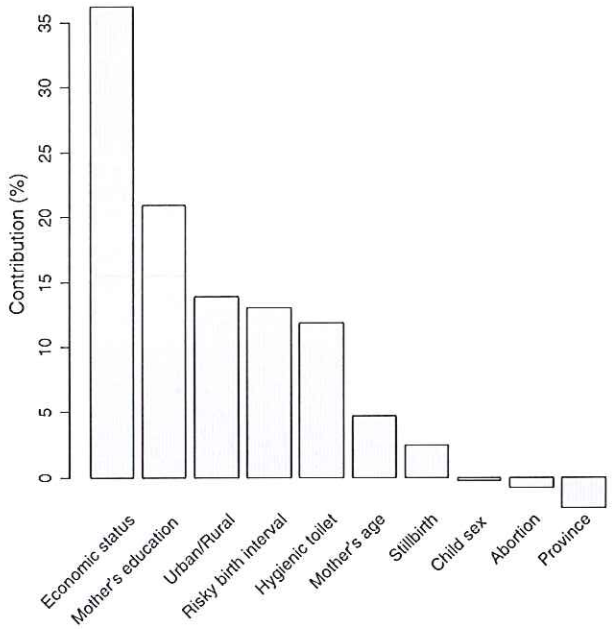


Fig. 2 The size and sign of the different contributions to the socioeconomic inequality of infant mortality in Iran (Hosseinpoor et al. 2006)

$$C_{\hat{y}} = \sum_k \left(\frac{\beta_k \bar{x}_k}{\mu} \right) C_k \tag{3}$$

For example, Hosseinpoor and colleagues (2006) decomposed the RCI of the Iranian infant mortality into its determining factors, as shown in Fig. 2. While 36% of income-related inequality in infant mortality in Iran was due to economic inequality per se, 64% was because important risk factors for infant mortality such as maternal illiteracy and access to a hygienic toilet were strongly correlated with income. The utility of the decomposition exercise is thus that it links monitoring health inequality with understanding its determinants (Harper and Lynch 2007).

Of course, the extent to which the selected determinants provide a well-specified model for the underlying data may affect the decomposition. The method presented here is only one analytical approach to decompose socioeconomic inequalities in health and other approaches, like the Oaxaca decomposition, can be used for analyzing inequalities (Van de Poel and Speybroeck 2009).

Most of today's software for inequality in health is written in Stata code (O'Donnell et al. 2008), but there is no implementation publicly available allowing for a decomposition of the RCI using generalized linear models. The methods described above are now implemented in an R package called decomp (available from the authors upon request). The R program is free of charge (http://www.r-project.org). We demonstrate code, including methods for bootstrapping confidence intervals in an "Appendix".

Decomposition results allow policy makers to move from tackling average health problems (the "level approach") to tackling inequalities of health (the "gap approach"). Results in Ghana (Van de Poel et al. 2007) suggested that factors strongly associated with average malnutrition are not necessarily contributing to relative socioeconomic inequality in malnutrition. Variables such as duration of breastfeeding can be quite strongly associated with a child's nutritional status, but do not contribute to socioeconomic inequality in malnutrition, because of a relatively equal distribution across socioeconomic groups.

A decomposition analysis allows a clear understanding of how factors affect inequality, i.e. through the more unequal distribution of determinants or through the greater association of determinants with health. Policies trying to reduce average bad health can be different from those aiming at lowering socioeconomic inequality in bad health. The latter are often determined not only by health system functions, but also by factors beyond the scope of health authorities, requiring a multisectoral approach to realize improvements of inequality in health across society.

Acknowledgments This paper resulted from cooperation initiated through meetings organized by the WHO European Office for Investment for Health and Development, Venice, Italy.

```
> plot(contribution(kenya.stunting.lm, kenya$wealth))
```

An analytical solution for the sampling variability and the standard errors for the contributions in the decomposition analysis is not available. However, statistical inference on the decomposition results can be obtained through bootstrapping. The following code shows how to conduct a bootstrap assuming a random sample, ingnoring survey design characteristics. For more advanced complex multi-stage sampling designs, the bootstrap needs to be adjusted, as was done by van Doorslaer (van Doorslaer and Koolman 2004). The example is therefore purely illustrative.

```
> model1 <- lm(stunting ~ meducY + sex + wealth + urban, data = kenya)

> nobs <- length(model1$fitted.values)
> B <- 1000
> results <- matrix(nrow = B, ncol = length(model1$coefficients))
> for (b in 1:B)
> {
>      bootstrapsample <- sample(nobs, nobs, replace = TRUE)
>      bootstrapdata <- model1$model[bootstrapsample,]
>      row.names(bootstrapdata) <- 1:nrow(bootstrapdata)
>      bootstrap.lm <- lm(stunting ~ meducY + sex + wealth + urban, data =
bootstrapdata)
>      results[b,] <- contribution(bootstrap.lm,
bootstrap.lm$model$wealth)$contribution
> }
> results <- data.frame(results)
> names(results) <- c('residual', names(model1$coefficients)[-1])
> resultstable <- apply(results, 2 , quantile, probs = c(0.05, 0.5, 0.95))
> # resultstable contains the contribution and associated confidence intervals
> resultstable

      residual    meducY  sexfemale    wealth    urban1
5%  -12.749012  4.188837 -2.1255662  74.09021 -6.232390
50%  -3.899108 14.522676 -0.8483391  87.96861  2.021826
95%   3.208132 24.800365  0.0992727 104.07006 10.934984
```

We can plot these results as follows

```
> library(gplots)
> resultstable <- resultstable[,order(resultstable[2,], decreasing = TRUE)]
> significant <- (resultstable[1,] / resultstable[3,]) > 0
> color <- ifelse(significant, 'dark gray', 'white')
> barplot2(resultstable[2,], col = color,      plot.grid = TRUE, plot.ci = TRUE,
           ci.l = resultstable[1,], ci.u = resultstable[3,], ylab =
'Contribution')
```

References

Harper S, Lynch J (2007) Commentary: using innovative inequality measures in epidemiology. Int J Epidemiol 36:926–928

Hosseinpoor AR, van Doorslaer E, Speybroeck N, Naghavi M, Mohammad K, Majdzadeh R, Delavar B, Jamshidi H, Vega J (2006) Decomposing socioeconomic inequality in infant mortality in Iran. Int J Epidemiol 35:1211–1219

Konings P, Harper S, Lynch J, Hosseinpoor AR, Berkvens D, Lorant V, Speybroeck N (2009) Analysis of socioeconomic health inequalities using the concentration index. Int J Public Health. doi:10.1007/s00038-009-0078-y

Lynch J (2006) It's not easy being interdisciplinary. Int J Epidemiol 35:1119–1122

O'Donnell O, van Doorslaer E, Lindelöw M, World B (2008) Analyzing health equity using household survey data: a guide to techniques and their implementation. World Bank, Washington

Van de Poel E, Speybroeck N (2009) Decomposing malnutrition inequalities between scheduled castes and tribes and the remaining Indian population. Ethn Health 14:271–287

Van de Poel E, Hosseinpoor AR, Jehu-Appiah C, Vega J, Speybroeck N (2007) Malnutrition and the disproportional burden on the poor: the case of Ghana. Int J Equity Health 6:21

van Doorslaer E, Gerdtham UG (2003) Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. Soc Sci Med 57:1621–1629

van Doorslaer E, Jones AM (2003) Inequalities in self-reported health: validation of a new approach to measurement. J Health Econ 22:61–87

van Doorslaer E, Koolman X (2004) Explaining the differences in income-related health inequalities across European countries. Health Economics 13:609–628

van Doorslaer E, Koolman X, Jones AM (2004) Explaining income-related inequalities in doctor utilisation in Europe. Health Econ 13:629–647

Wagstaff A, van Doorslaer E, Watanabe N (2003) On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. J Econ 112:207–223

Appendix

Appendix: R code to run the decomposition (Comments in grey).

We use a Demographic Health Survey dataset from Kenya (2003) as an illustration. Commands and output are in courier font. Commands are indicated by a prompt (>).

Start R, and load the decomp package containing the dataset :

```
> library(decomp)
```

We then fit a linear model for childhood stunting with sex, place of residence (urban/rural), household wealth and mother's education as covariates. Other generalized linear models (e.g. a logistic regression) are allowed.

Not all possible independent variables are included in the model, so the results shown are illustrative only.

```
> kenya.stunting.lm <- lm(stunting ~ meducY + sex + wealth + urban, weights
= popweight, data = kenya)
```

Next, we calculate the overall RCI and the contribution of the individual covariates with the contribution command

```
> contribution(kenya.stunting.lm, kenya$wealth)
Overall CI: -0.1164696

Decomposition:
          Contribution (%) Corrected
residual       -0.005950229        no
meducY         20.502829983        no
sexfemale      -0.764719310        no
wealth         87.971215576        no
urban1         -7.703376020        no
```

It is possible to display a barplot of the relative contributions as well: (see Fig. 3)

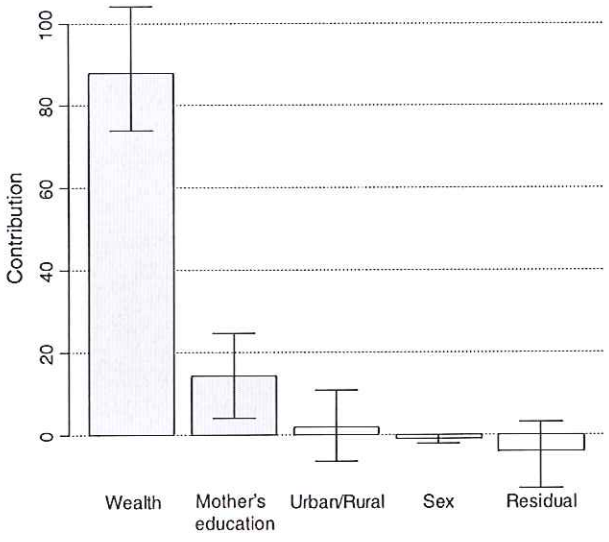


Fig. 3 The size and sign of the different contributions to the socio-economic inequality in childhood stunting in Kenya (Data Source: Kenya Demographic and Health Survey, 2003). Notice that not all possible variables are included in the model, so the results shown are illustrative only