

DECOMPOSITION OF CONCENTRATION INDEX USING GENERALISED LINEAR MODEL: ANALYSIS OF SOCIO-ECONOMIC DETERMINANTS OF HEALTH INEQUALITY IN THE NORTHERN TERRITORY OF AUSTRALIA

ZHAO, Yuejen*

Abstract

This paper aims to extend the existing linear decomposition techniques for the concentration index to generalised linear models. The second order Taylor expansion is used to derive a linear approximation of the partial concentration indices. This approach adheres closely to the standard decomposition method in the linear settings.

The method is illustrated using age-sex standardised mortality rate data from Australia to examine the socioeconomic determinants of Indigenous health inequality. The empirical results demonstrate the usefulness of this method. The new approach provides a more efficient and flexible way to quantify the contributions of the underlying determinants of health inequality in a nonlinear and multivariate context.

1. Introduction

The concentration index (*CI*) is commonly used as a summary measure for socioeconomic-related inequality in health or healthcare utilisation (Kakwani *et al.*, 1997; van Doorslaer *et al.*, 2004). It is a general form of the Gini index (*GI*) which is often employed to describe income disparity (Kakwani, 1980). The *CI* uses cumulative proportion of a health measure sorted by an underlying variable, plotted against the cumulative proportion of population to show socioeconomic health inequality (Wagstaff *et al.*, 1991). The *GI* and *CI* are both widely used to examine health inequality, its socioeconomic determinants and inequitable health care use (Wagstaff and Watanabe, 2003; Clark *et al.*, 2004; van Doorslaer *et al.*, 2004; Jones and Rice, 2004; Zhang and Wang, 2007). For a comprehensive review, see O'Donnell *et al.* (2008). The decomposition of *GI* and *CI* provides a useful means of analysing determinants of health inequality. Decomposition by groupings was the leading approach to quantifying how the socioeconomic determinants affect income inequality previously (Kakwani, 1980). Wagstaff and colleagues incorporated *CI* with a linear regression framework and demonstrated that the *CI* is additively decomposable into the contributions of individual determinants to socioeconomic health inequality (Wagstaff *et al.*, 2003). This method has been developed in linear models in which the response is assumed to be continuous and to follow a normal distribution.

However, statistical models for health research are intrinsically nonlinear. The distributions of health variables do not typically follow the normal distribution. Nonlinear regression models such as logistic, Poisson and negative binomial regressions are often used in health economics and epidemiology. It is desirable to extend the linear decomposition to nonlinear settings. Van Doorslaer and Masseria (2004) recommend the re-linearisation and use of a standard linear model to perform

* Yuejen Zhao, 1. Institute of Advanced Studies, Charles Darwin University, Darwin, NT, Australia, 2. Health Gains Planning Branch, Department of Health, Northern Territory, Australia. Email: yuejen.zhao@nt.gov.au

decomposition for nonlinear models. Van Doorslaer, Koolman and Jones (2004) noted that the decomposition of *CI* in nonlinear models can be implemented by analysing the marginal effects evaluated at the sample means. Huber (2008) proposed a micro-simulation model to decompose the *CI*. Nevertheless, micro-simulation models are often data intensive, difficult to develop and draw heavily on assumptions. Erreygers (2009) examined the decomposition of the corrected *CI*. Yiengprugsawan *et al.* (2010) argued that the marginal effect approach tends to yield inconsistent results depending on the choice of reference groups, and recommended the use of generalised linear models (GLM) with identity link. GLM is a class of statistical models involving a linear combination of parameters, probability distribution from the exponential family such as normal, binomial and Poisson, and a monotonic link function, notably identity, logit and log link (Dobson, 1990). The link function defines the numerical relationship between the mean of the response and the linear predictor of the explanatory variables. The response may be binary, discrete or continuous.

This study aims to extend the existing decomposition methods beyond linear additive models to all type of GLMs, and to develop a more general approach to regression based decomposition, which will help explore the underlying determinants of economic and health inequality. The mortality data are used to illustrate the method compared with the existing approaches. The paper is then concluded with a discussion on the merits and limitations of the innovative method.

2. Method

To assess inequality, consider $\{y_i, z_i\}$ a non-negative shortfall measure y_i and a relevant underlying variable z_i for observation i , where $i=1,2,\dots,n$ and n is the number of observations. The mean $\bar{y} = \sum_i y_i / n$. Ranking the data by y_i yields

$$\{y_{(i)}, z_{(i)}\}, \text{ where } y_{(1)} \leq \dots \leq y_{(i)} \leq \dots \leq y_{(n)}.$$

Let $r_{(i)}$ be the cumulative proportion of observations. The total inequality can be calculated by (Kakwani, 1980)

$$GI = \frac{2}{\bar{y}n} \sum_{(i)} (y_{(i)} - \bar{y})(r_{(i)} - 0.5).$$

Here $0 \leq GI \leq 1$ where 0 and 1 represent perfect equality and perfect inequality respectively. Lorenz curve plots the cumulative proportion of $y_{(i)}$ against the cumulative proportion of $r_{(i)}$. If the data $\{y_i, z_i\}$ are re-ranked with respect to the underlying variable z_i so that $z_{[1]} \leq \dots \leq z_{[i]} \leq \dots \leq z_{[n]}$, the inequality with respect to z_i can be computed by

$$CI = \frac{2}{\bar{y}n} \sum_{[i]} (y_{[i]} - \bar{y})(r_{[i]} - 0.5),$$

where $r_{[i]}$ is the cumulative proportion of observations re-ranked by z_i (Kakwani, 1980). When there is no underlying inequality, the *CI* equals 0. If the variable $y_{[i]}$ is concentrated at a lower (or higher) level of the underlying variable $z_{[i]}$, the *CI* becomes negative (or positive). $|CI| \leq GI$. *GI* is a special case of the *CI* where $z_i = y_i$. The concentration curve plots the cumulative proportion of $y_{[i]}$ against the cumulative proportion of $r_{[i]}$.

Regression based decomposition of CI was initially developed in linear regression models (Oaxaca, 1973; Wagstaff *et al.*, 2003). Assume an estimate of y_i is $\hat{y}_i = \mathbf{x}_i \boldsymbol{\beta}$, where the covariates $\mathbf{x} = \{x_{ij}\}$ consist of p determinants and n observations, $\boldsymbol{\beta} = \{\hat{\beta}_1, \dots, \hat{\beta}_j, \dots, \hat{\beta}_p\}^T$ is a $p \times 1$ vector of parameter estimates. The CI can be decomposed into a deterministic component and a residual component (van Doorslaer *et al.*, 2004). In a GLM, the dependent variable y is assumed to follow a particular distribution from the exponential family and the estimate of y_i is $\hat{y}_i = f(\mathbf{x}_i \boldsymbol{\beta})$ where the link function is the inverse of $f(\cdot)$. The residual term is $e_i = y_i - \hat{y}_i$. Presumably both z_i and x_i are uncorrelated with e_i but affecting y_i . Without loss of generality, set the index $I \in \{GI, CI\}$ and $i \in \{(i), [i]\}$. It is straightforward to split the index

$$\begin{aligned} I &= \frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i \boldsymbol{\beta}) + e_i - \bar{y})(r_i - 0.5) \\ &= \frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i \boldsymbol{\beta}) - \bar{y})(r_i - 0.5) + \frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5). \end{aligned} \quad (1)$$

As inspired by van Doorslaer, Koolman and Jones (2004), the linear approximation is used to assess determinants of inequality in nonlinear settings. Provided $f(\mathbf{x}_i \boldsymbol{\beta})$ is differentiable, the Taylor series expansion can be used to estimate $f(\mathbf{x}_i \boldsymbol{\beta})$ approximately as a sum of limited terms of derivatives evaluated at a neighbourhood (Thomas and Finney, 1996). For notational simplicity, $f(\mathbf{x}_i \boldsymbol{\beta}) = f(\mathbf{x}_i)$. In a neighbourhood $\mathbf{x}_0 = \{x_{01}, \dots, x_{0p}\}$ where $f(\mathbf{x}_0)$ exists, the second order approximation of $f(\mathbf{x}_i \boldsymbol{\beta}) = f(\mathbf{x}_i)$ can be calculated using

$$\begin{aligned} f(\mathbf{x}_i \boldsymbol{\beta}) &= f(\mathbf{x}_0) + \sum_j (x_{ij} - x_{0j}) \frac{\partial f(x_j)}{\partial x_j} \Big|_{x_j=x_{0j}} + O(\|\mathbf{x}_i - \mathbf{x}_0\|^2) \\ &\approx f(\mathbf{x}_0) + \sum_j (x_{ij} - x_{0j}) \frac{df(\mathbf{x}_0 \boldsymbol{\beta})}{d(\mathbf{x}_0 \boldsymbol{\beta})} \beta_j \\ &= f^{(2)}(\mathbf{x}_i) \Big|_{x_j=x_{0j}} \end{aligned} \quad (2)$$

and $\bar{y} \approx f^{(2)}(\mathbf{x}_i) \Big|_{x_i=\bar{x}}$, given all the first order partial derivatives of $f(\mathbf{x}_0)$ are continuous.

Replacing $f(\mathbf{x}_i \boldsymbol{\beta})$ and $\bar{y}$ in equation (1) by the linear approximations yields

$$\begin{aligned} f(\mathbf{x}_i \boldsymbol{\beta}) - \bar{y} &\approx f(\mathbf{x}_0) + \sum_j (x_{ij} - x_{0j}) \frac{df(\mathbf{x}_0 \boldsymbol{\beta})}{d(\mathbf{x}_0 \boldsymbol{\beta})} \beta_j - \left[f(\mathbf{x}_0) + \sum_j (\bar{x}_j - x_{0j}) \frac{df(\mathbf{x}_0 \boldsymbol{\beta})}{d(\mathbf{x}_{0i} \boldsymbol{\beta})} \beta_j \right] \\ &= \sum_j (x_{ij} - \bar{x}_j) \frac{df(\mathbf{x}_0 \boldsymbol{\beta})}{d(\mathbf{x}_0 \boldsymbol{\beta})} \beta_j \end{aligned}$$

Substituting this into equation (1) gives

$$I = \frac{2}{\bar{y}n} \sum_i \left(\sum_j (x_{ij} - \bar{x}_j) \frac{df(\mathbf{x}_0\boldsymbol{\beta})}{d(\mathbf{x}_0\boldsymbol{\beta})} \beta_j \right) (r_i - 0.5) + \frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5)$$

$$= \frac{2}{\bar{y}n} \sum_j \frac{df(\mathbf{x}_0\boldsymbol{\beta})}{d(\mathbf{x}_0\boldsymbol{\beta})} \beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5) + \frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5).$$

Assume $\sum_j \frac{df(\mathbf{x}_0\boldsymbol{\beta})}{d(\mathbf{x}_0\boldsymbol{\beta})} \beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5) \propto \sum_i (f(\mathbf{x}_i\boldsymbol{\beta}) - \bar{y})(r_i - 0.5)$ *ei.* the linear approximation $f^{(2)}(\mathbf{x}_i) \Big|_{\mathbf{x}_j = \mathbf{x}_{0j}}$ is a sufficiently good approximation to $f(\mathbf{x}_i\boldsymbol{\beta})$. The contributions of $\mathbf{x}_j$ can be estimated by

$$\frac{\frac{df(\mathbf{x}_0\boldsymbol{\beta})}{d(\mathbf{x}_0\boldsymbol{\beta})} \beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5)}{\sum_j \frac{df(\mathbf{x}_0\boldsymbol{\beta})}{d(\mathbf{x}_0\boldsymbol{\beta})} \beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5)} \left(1 - \frac{2}{\bar{y}nI} \sum_i e_i (r_i - 0.5) \right)$$

$$= \frac{\beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5)}{\sum_j \beta_j \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5)} \left(1 - \frac{2}{\bar{y}nI} \sum_i e_i (r_i - 0.5) \right). \quad (3)$$

Note that there is no $\mathbf{x}_0$ on the right side of this equation. In other words, regardless of what value $\mathbf{x}_0$ takes (zero, the sample means or any other value), the decomposition will produce unique results. Because e_i is separated in equation (1), the approximation errors due to equation (3) do not influence e_i . Because I and $\frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5)$ in equation (1) are known, the total contribution for $\mathbf{x}$ is

$$\frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i\boldsymbol{\beta}) - \bar{y})(r_i - 0.5) = I - \frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5).$$

Identity, log and inverse links will be considered in turn as follows, and logit link will be considered in the results section. In the case of identity link and ordinary least square (OLS), using the Taylor expansion,

$$f(\mathbf{x}_i\boldsymbol{\beta}) = \mathbf{x}_0\boldsymbol{\beta} + \sum_j (x_{ij} - x_{0j}) \beta_j = \mathbf{x}_i\boldsymbol{\beta}.$$

Replacing $f(\mathbf{x}_i\boldsymbol{\beta})$ in equation (1) by $\mathbf{x}_i\boldsymbol{\beta}$ leads to

$$I = \frac{2}{\bar{y}n} \sum_j \sum_i (x_{ij} \beta_j - \bar{x}_j \beta_j) (r_i - 0.5) + \frac{2}{\bar{y}n} \sum_i e_i (r_i - 0.5)$$

$$= \sum_j \frac{\beta_j \bar{x}_j C_j}{\bar{y}} + \frac{2}{\bar{y}n} \sum_i e_i r_i,$$

where $C_j = \frac{2}{\bar{x}_j n} \sum_i (x_{ij} - \bar{x}_j) (r_i - 0.5)$, which is identical to the results given in

Wagstaff *et al.* (2003), and van Doorslaer and Gerdtham (2003). For log links,

$$f(\mathbf{x}_i\boldsymbol{\beta}) = \exp(\mathbf{x}_i\boldsymbol{\beta}) \approx 1 + \mathbf{x}_i\boldsymbol{\beta}.$$

Replacing $f(\mathbf{x}_i\boldsymbol{\beta})$ by the linear approximation leads to

$$\begin{aligned} & \frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i\boldsymbol{\beta}) - \bar{y})(r_i - 0.5) \\ & \approx \sum_j \frac{2\beta_j}{\bar{y}n} \sum_i (x_{ij} - \bar{x}_j)(r_i - 0.5). \end{aligned}$$

The contribution for $\mathbf{x}_j$ can then be calculated from equation (3). In the inverse link function,

$$f(\mathbf{x}_i\boldsymbol{\beta}) = 1/(x_i\boldsymbol{\beta}) \approx \frac{1}{\bar{x}\boldsymbol{\beta}} - \sum_j \frac{(x_{ij} - \bar{x}_j)\beta_j}{(\bar{x}\boldsymbol{\beta})^2}.$$

Replacing $f(\mathbf{x}_i\boldsymbol{\beta})$ by the linear approximation leads to

$$\begin{aligned} & \frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i\boldsymbol{\beta}) - \bar{y})(r_i - 0.5) \\ & \approx \sum_j \frac{2\beta_j}{\bar{y}n(\bar{x}\boldsymbol{\beta})^2} \sum_i (\bar{x}_j - x_{ij})(r_i - 0.5). \end{aligned}$$

3. Illustrative Example

This method is illustrated by using grouped data on geographic distribution of the age-sex standardised mortality rates in the Northern Territory, Australia (NT). Over the past decade Australia as a whole has sustained continuous improvements in population health (Australian Bureau of Statistics (ABS), 2008), but striking inequalities remain in health across geographical areas and between Indigenous and non-Indigenous populations (see key community profiles by Indigenous status in Table I). In this study, the *CI* decomposition is used to measure health inequality with respect to Indigenous status and examine the contribution of socioeconomic determinants to the Indigenous health inequality.

Table I. Indigenous Population Profile, the Northern Territory, Australia, 2006

					Indigenous	Non-Indigenous
Total population (Census count)					53,661	122,730
Median weekly individual income (AUD)					\$215	\$712
High school education (%)					8	46
Employed(%)					33	68
Life expectancy at birth (years)						
	Male				60	81
	Female				70	85

Source: ABS (2012). NT Government (2010).

The deaths registered between 2005 and 2007 were used to calculate the mortality rates for the 88 statistical local areas (SLA). To assess the horizontal health inequality, the age-sex directly standardised mortality rates are used as a shortfall response, using the 2006 Australian estimated resident population. A GLM with binomial distribution and logit link is applied for the covariates including above-high-school education

proportions, employment proportions and the index of relative socio-economic advantage and disadvantage (IRSAD) (ABS, 2006; ABS, 2010). The Indigenous percentage was taken as the underlying variable for calculation of the *CI* to quantify the Indigenous health inequality. The NT Indigenous mortality dataset is provided in the Appendix 1.

Figure 1. Lorenz and concentration curves for Indigenous mortality inequality

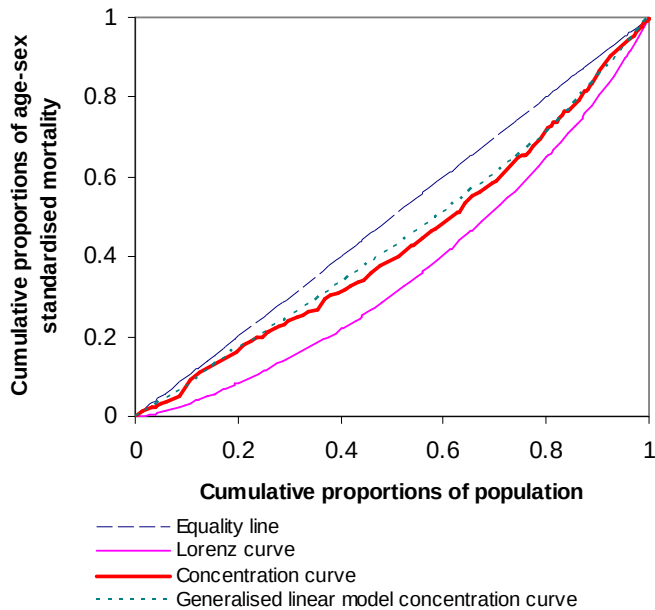


Table II. GLM – coefficient estimates, odds ratio and 95% confidence interval

	Coefficient	SE	OR	95% confidence interval	
IRSAD	-0.471	0.229	0.625	0.398	0.979
High school education	-0.974	0.377	0.378	0.180	0.791

Note: GLM=generalised linear model with binomial distribution and logit link; IRSAD= index of relative socio-economic advantage and disadvantage / 1000. OR=odds ratio. SE=standard error.

For the purpose of illustration, the Lorenz curve for total mortality inequality and the concentration curve for Indigenous mortality inequality are shown in Figure 1 (see thick and thin solid lines respectively). Likelihood ratio tests are adopted for forward model selection. The optimal GLM model selected contains two variables: high school education and IRSAD. Table II displays the coefficient estimates and odds ratios with 95% confidence intervals for the optimal model, which shows that a higher mortality is associated with lower level of education and lower socioeconomic status. It can be argued that high school education may reduce mortality rates by over 60% and an increase in IRSAD may lead to a nearly 40% decline in mortality rates. The GLM concentration curve is the dotted line in Figure 1.

Table III. Decomposition of concentration index for Indigenous mortality inequality

	Socioeconomic determinant	New method		Contribution by existing methods	
		CI	Contribution	Marginal effect	Identity link
	High school education	0.063	47.5%	42.9%	42.9%
	IRSAD	0.049	36.5%	33.0%	46.0%
	GLM model	0.112	84.0%	84.0%	-
Indigenous mortality inequality		0.133	100.0%	100.0%	100.0%

Note: GLM=generalised linear model with binomial distribution and logit link; IRSAD= index of relative socio-economic advantage and disadvantage / 1000.

Table III shows the *CI*s decomposed by the high school education and IRSAD, and their contributions compared with those estimated using marginal effects (van Doorslaer *et al.*, 2004) and GLM with identity link (Yiengprugsawan *et al.*, 2010). The *CI* equals 0.133, suggesting that the Indigenous health inequality approximately accounts for half of the total mortality inequality ($GI=0.273$). The *CI* is positive, reflecting the fact that the mortality is more concentrated in the residential areas with higher Indigenous percentage. In logit link, using the Taylor expansion,

$$f(\mathbf{x}_i\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta})} \approx \frac{1}{2} + \frac{\mathbf{x}_i\boldsymbol{\beta}}{4}.$$

Replacing $f(\mathbf{x}_i\boldsymbol{\beta})$ by the linear approximation gives

$$\begin{aligned} & \frac{2}{\bar{y}n} \sum_i (f(\mathbf{x}_i\boldsymbol{\beta}) - \bar{y})(r_i - 0.5) \\ & \approx \sum_j \frac{\beta_j}{2\bar{y}n} \sum_i (x_{ij} - \bar{x}_j)(r_i - 0.5). \end{aligned}$$

Based on the optimal GLM, about 47.5% (0.063/0.133) of the Indigenous mortality inequality was attributable to poor education, and 36.5% (0.049/0.133) to low socioeconomic status. The corresponding contributions measured in the marginal effects were slightly lower, 42.9% for poor education and 33% for low socioeconomic status. The comparison indicates that the two methods are consistent, but the new Taylor approximation method appears to be more efficient and robust. The results obtained through GLM with identity link are different, indicating the contribution of IRSAD (46%) was greater than that of high school education (43%).

4. Discussion

This paper describes a more general method for exploring partial contributions of multiple sources of socioeconomic health inequality based on GLMs. The second order Taylor expansion is used for linear approximation. The inequality can be decomposed by groupings and by sources of explanatory variables. The method allows for the horizontal health inequality to be traced to covariates or determinants, and provides useful insights into the magnitude and direction of their impacts.

The *CI* is first decomposed into a deterministic part and a remainder. Then, the deterministic part is separated by the contributions of the covariates through linear approximation using equation (3). The example demonstrated the usefulness of the decomposition approach to revealing the sources of inequality, and decomposed Indigenous-related health inequalities in the age-sex standardised mortality by socio-economic determinants. The decomposition analysis demonstrated that the Indigenous health inequality can be largely explained by lower level of education and lower socioeconomic status.

The Taylor approximation method provides an alternative approach to the *GI* and *CI* decomposition by covariates in GLMs. It appears to be superior to the existing method of marginal effects (van Doorslaer *et al.*, 2004) and more general than the method for GLM with identity link (Yiengprugsawan *et al.*, 2010). This study contributes to the existing literature in several ways. First, it derives the linear approximation for GLMs based on the Taylor expansion to decompose inequality into contributions of determinants. With the Taylor series, more accurate analysis can be achieved by including the high order terms. Second, it extends the method to cover both the *GI* and *CI* in nonlinear multivariate settings. The *GI* decomposition is best interpreted as quantifying the factors that explain the total observed inequality (based on the groupings, eg. SLA), whereas the *CI* decomposition explains how the health inequality (y_i) is related to an underlying variable (z_i), and how the socioeconomic determinants ($\mathbf{x}_i$) contribute to the underlying health inequality (say Indigenous health inequality). Third, the new decomposition method produces unique and consistent results, regardless of whether $\mathbf{x}_0$ is evaluated at the sample mean, zero or any other neighbourhood value. This is not the case for the marginal effect method, which needs to be evaluated at the sample means (van Doorslaer *et al.*, 2004; O'Donnell *et al.*, 2008). It is clear from equation (3) that the contribution of each determinant rests entirely on the magnitude and direction of β_j and the covariances between $\mathbf{x}$ and r , irrespective of initial values $\mathbf{x}_0$. The new method is applicable for all type link functions, whereas the marginal effect approach works only for the models with identity links (Yiengprugsawan *et al.*, 2010). It is not always plausible and appropriate, however, to assume identity link for GLMs (Dobson, 1990). By assuming identity link, the model results are likely to be similar to OLS. Fourth, the Taylor approximation method expands the decomposition method beyond the continuous health outcome to also cover binary, discrete and ordinal responses. Finally, the marginal effect approach seeks numerical solutions, by taking the marginal effects as the parameter estimates of linear model. The Taylor approximation method provides analytical solutions, and explicitly uses derivatives rather than marginal effects.

There are several limitations to this method. In this study, the degree of inequality for a nonlinear model is decomposed using the linear approximation. However, this is likely the inevitable price to be paid for nonlinear models (O'Donnell *et al.*, 2008). The inequality patterns found in the illustrative example are likely an underestimate of the true inequality since grouped data are used rather than individual data. The use of grouped data is driven by data availability and is necessary for direct standardisation to assess horizontal health inequality (Kakwani *et al.*, 1997). Further studies are warranted to investigate the use of high order terms of approximation and mixed partial derivatives.

In summary, the method described in this paper represents a new alternative and useful way to analyse the multiple sources of health inequality, based on GLMs, which are commonly used in economic research.

References

- ABS. 2006. SEIFA: *Socio-Economic Indexes for Areas*. ABS, Canberra. http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Seifa_entry_page. Accessed 20/12/2010 [2 February 2011].
- ABS. 2008. *Australian Social Trends 2008*. ABS, Canberra.
- ABS. 2010. *Population Statistics*. ABS, Canberra. <http://www.abs.gov.au/ausstats/abs@.nsf/web+pages/statistics?OpenDocument> [20 December 2010].
- ABS. 2012. *Community Profile*. ABS, Canberra. <http://www.abs.gov.au/websitedbs/censushome.nsf/home/communityprofiles> [5 June 2012].
- Clarke PM, Gerdtham UG, Connelly LB. 2003. A note on the decomposition of the health concentration index. *Health Economics* **12**:511-516.
- Dobson AJ. 1990. *An Introduction to Generalized Linear Models*. Chapman and Hall: Melbourne.
- Erreygers G. 2009. Correcting the concentration index. *Journal of Health Economics* **28**:504–515.
- Huber H. 2008. Decomposing the causes of inequalities in health care use: A microsimulations approach. *Journal of Health Economics* **27**: 1605–1613.
- Jones A, Rice N. 2004. Using longitudinal data to investigate socioeconomic inequality in health. In *Health Policy and Economics: Opportunities and Challenges*, Smith PC, Ginnelly L, Sculpher M (eds). Open University Press: Berkshire.
- Kakwani N. 1980. *Income Inequality and Poverty*. Oxford University Press: New York.
- Kakwani N, Wagstaff A, van Doorslaer E. 1997. Socioeconomic inequalities in health: Measurement, computation and statistical inference. *Journal of Econometrics* **77**: 87- 104.
- NT Government. 2010. *Health Gains Planning Information Sheet*. Department of Health and Families: Darwin. http://digitallibrary.health.nt.gov.au/dspace/bitstream/10137/480/1/Factsheet%202010_mortality_21Mar2011_final.pdf [5 June 2012].
- O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. 2008. *Analyzing Health Equity using Household Survey Data : a Guide to Techniques and their Implementation*. The World Bank: Washington, DC.
- Oaxaca R. 1973. Male-female wage differentials in urban labor markets. *International Economic Review* **14**: 693-709.
- Thomas GB, Finney RL. 1996. *Calculus and Analytic Geometry*. Addison-Wesley: Massachusetts.
- van Doorslaer E, Koolman X, Jones AM. 2004. Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics* **13**: 629–647.
- van Doorslaer E, Gerdtham UG. 2003. Does inequality in self-assessed health predict inequality in survival by income? evidence from Swedish data. *Social Science & Medicine* **57**: 1621-1629.

- van Doorslaer E, Masseria C. 2004. *Income-Related Inequality in the Use of Medical Care in 21 OECD Countries*. OECD, Paris.
- Wagstaff A, Paci P, van Doorslaer E. 1991. On the measurement of inequalities in health. *Social Science & Medicine* **33**: 545-557.
- Wagstaff A, van Doorslaer E, Watanabe N. 2003. On decomposing the causes of health sector inequalities, with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics* **112**: 207-223.
- Wagstaff A, Watanabe N. 2003. What difference does the choice of ses make in health inequality measurement? *Health Economics* **12**: 885–890.
- Yiengprugsawan V, Lim LLY, Carmichael GA, Dear KBG, Sleigh AC. 2010. Decomposing socioeconomic inequality for binary health outcomes: an improved estimation that does not vary by choice of reference group. *BMC Research Notes* **3**: 57.
- Zhang Q, Wang Y. 2007. Using concentration index to study changes in socio-economic inequality of overweight among US adolescents between 1971 and 2002. *International Journal of Epidemiology* **36**: 916–925.

Appendix 1. The Northern Territory Indigenous Mortality Data

Aera Code	Indigenous	Income (\$AUD)	High school education	Employment	Standardised mortality rate	IRSAD
705051052	0.0000	404	0.2301	0.5870	0.00407	898.11
705102806	0.0282	828	0.4552	0.6885	0.00945	1116.56
705051044	0.0309	933	0.5555	0.6905	0.00519	1095.4
705051128	0.0347	565	0.2652	0.6024	0.00553	955.98
705051014	0.0385	681	0.6193	0.6850	0.00593	1082.16
705051108	0.0411	805	0.4807	0.6568	0.01156	1021.84
705051064	0.0483	737	0.4576	0.7063	0.00613	1045.24
705102811	0.0584	805	0.3952	0.6691	0.00070	1097.84
705051048	0.0599	704	0.4558	0.6876	0.00653	1044.49
710252409	0.0635	1121	0.3998	0.7094	0.01990	1063.75
705051088	0.0669	709	0.4600	0.6254	0.01146	1016.15
705202308	0.0671	641	0.2961	0.6275	0.00701	1025.82
705051018	0.0736	751	0.4419	0.5693	0.01326	1027.49
705051028	0.0744	726	0.4858	0.6106	0.00477	1030.47
705051078	0.0755	658	0.4824	0.6590	0.01076	1035.5
705051138	0.0765	730	0.4498	0.6142	0.00159	1079.79
705051094	0.0786	693	0.4530	0.6149	0.01432	996.6
710050770	0.0792	572	0.3256	0.5635	0.00873	983.91
705051098	0.0839	746	0.5280	0.7069	0.00608	1029.49
705202304	0.0846	831	0.5265	0.6589	0.00321	1074.59
705051124	0.0927	625	0.4361	0.6420	0.00458	1016.02
705051034	0.0961	656	0.4395	0.6162	0.01029	1019.5
705051104	0.0970	766	0.4865	0.6929	0.00579	1046.7
705051118	0.1032	540	0.3620	0.6021	0.00456	967.68
705051074	0.1098	630	0.3803	0.6013	0.00743	999.2
705102802	0.1158	727	0.4003	0.7422	0.00442	1013.56
705102824	0.1189	781	0.3608	0.6750	0.01790	1056.25
705051008	0.1227	679	0.4282	0.6454	0.00759	1013.77
705102804	0.1241	671	0.3563	0.6441	0.00436	991.1
705051134	0.1276	731	0.4126	0.6670	0.00793	1028.99
705051068	0.1298	620	0.4117	0.6233	0.00542	971.98
705051114	0.1389	649	0.4440	0.6235	0.01027	1006.96
705051004	0.1395	586	0.3905	0.5912	0.00386	981.24
705051024	0.1442	589	0.4018	0.5722	0.01229	909.59
705102818	0.1446	642	0.2893	0.5915	0.01097	957.88
710400207	0.1532	721	0.4413	0.6830	0.00623	1042.94
705051038	0.1653	576	0.3616	0.6110	0.01105	939.17
705051084	0.1699	473	0.3728	0.4538	0.00644	863.04
705051058	0.1718	613	0.3778	0.6154	0.01136	933.84
705102808	0.1850	530	0.2805	0.5004	0.00884	884.63
710152000	0.1957	962	0.3248	0.5783	0.00943	1023.24
710400205	0.1994	666	0.3871	0.6712	0.00970	992.61
710050759	0.2209	308	0.2087	0.3262	0.00538	815.67
710400208	0.2369	650	0.3891	0.6013	0.01952	933.12

705102814	0.2539	533	0.2379	0.4843	0.01182	870.27
710400203	0.2540	523	0.2637	0.5495	0.00972	987.86
710400201	0.2568	602	0.3118	0.6198	0.00821	992.84
705051054	0.2682	577	0.4129	0.6059	0.00673	910.4
710302200	0.2730	610	0.3166	0.5804	0.01372	957.72
710050700	0.3018	394	0.2292	0.4623	0.00682	894.95
710353409	0.3119	558	0.3027	0.5697	0.00122	901.49
710302360	0.3347	418	0.2249	0.5914	0.00802	844.5
710153309	0.3519	411	0.2282	0.5366	0.00279	766.03
710301409	0.3857	442	0.2451	0.5018	0.00847	843.72
710203030	0.4032	300	0.1633	0.4595	0.02109	835.56
710251609	0.4578	446	0.3127	0.4759	0.01492	1060.88
710403009	0.5100	433	0.2888	0.4836	0.01448	790.38
710353800	0.5601	355	0.1903	0.4834	0.01830	820.03
710304409	0.5837	334	0.1921	0.5952	0.01544	667.52
710200809	0.6804	244	0.1889	0.5346	0.01391	652.91
710304030	0.7016	384	0.1579	0.6585	0.01112	687.9
710301809	0.7457	230	0.1313	0.5435	0.00148	577.61
710300600	0.7918	263	0.1273	0.4661	0.01643	556.44
710351350	0.8682	185	0.1047	0.2290	0.02715	648.15
710405000	0.8745	227	0.0977	0.2593	0.01379	560.22
710354009	0.8789	214	0.0901	0.2230	0.00843	659.7
710401909	0.8868	218	0.0956	0.2271	0.00209	591.48
710400360	0.8871	223	0.0898	0.2528	0.01407	577.46
710403209	0.8916	222	0.0532	0.2245	0.01376	437.53
710403609	0.8919	215	0.0822	0.2709	0.01676	472.5
710202380	0.8959	279	0.1315	0.7181	0.01624	494.07
710300790	0.9036	221	0.1366	0.3352	0.01033	437.89
710302270	0.9163	207	0.1042	0.2365	0.02289	488.31
710252500	0.9199	209	0.1395	0.1296	0.01333	519.36
710104050	0.9201	210	0.1403	0.4284	0.02116	558.98
710152240	0.9216	214	0.1688	0.3387	0.01381	512.94
710154809	0.9280	216	0.1037	0.3014	0.02179	480.14
710251209	0.9304	201	0.0971	0.3028	0.01204	433.72
710204020	0.9318	188	0.0974	0.1746	0.00632	509.96
710302530	0.9341	191	0.0942	0.3188	0.02064	460.77
710305050	0.9375	219	0.1061	0.3365	0.01584	451.23
710400400	0.9489	210	0.1091	0.1494	0.00485	528.76
710402320	0.9558	207	0.1114	0.4066	0.02048	519.6
710403650	0.9581	210	0.0710	0.4302	0.00492	472.5
710250330	0.9665	211	0.0833	0.1154	0.01555	433.72
710350300	0.9789	181	0.0223	0.1825	0.00611	569.58
710304600	0.9801	244	0.0766	0.4931	0.01178	432.01
<u>710050540</u>	<u>1.0000</u>	<u>213</u>	<u>0.0857</u>	<u>0.4902</u>	<u>0.04426</u>	<u>435.28</u>

Note: IRSAD= index of relative socio-economic advantage and disadvantage