Measuring inequity in health service delivery

Introduction
One of the main equity concerns in systems with egalitarian equity goals is to ensure that the horizontal equity standard is met such that “equals are treated equally” by the system, and the vertical norm that unequals are treated appropriately unequally. In health care, the horizontal equity principle is usually translated as “equal treatment for equal medical need, irrespective of other characteristics like income, race, place of residence, etc” [1-3]. In practice, it is impossible to examine the extent to which the horizontal equity principle is violated without simultaneously also specifying a vertical equity norm. Researchers have usually assumed, implicitly or explicitly, that the vertical equity norm is the relationship between use of health care and need for health care, which is observed on average in a given system.

If this implicit normalization is accepted, then the measurement and explanation of inequity in health care use and delivery can proceed in much the same way as is described in TN#4 for the analysis of age-sex standardized (adult) health inequalities. Complications only arise because of the often peculiar measurement level of health care utilisation variables which, at least in micro data settings, usually come in the form of integer counts (e.g. numbers of visits, hospital days, etc) with very skewed distributions. Then, appropriate estimation of the (partial) effects of ‘need’ variables on such use measures often requires intrinsically non-linear methods (see TN#11). By implication, the (direct and indirect) standardization results presented in TN#3 do not immediately carry over to this non-linear setting. They can be rescued only if the linearity of the relationships can somehow be recovered. In this note we therefore concentrate mainly on the computational issues in non-linear settings. To the extent that non-linear methods (e.g. probit, logit, etc.) are also required for estimation of health indicators, the methods presented here can be used in a similar fashion for the analysis of health inequality.

Measuring horizontal inequity in health service delivery
In principle, the problem of the analysis of horizontal (in)equity in health care use is analogous to the problem of demographic standardization of health distributions discussed in TN#3. The basic aim is to ascertain to what extent the actual, observed distribution of health care (e.g. by income) ‘matches’ the distribution of the need for such care. Again, the issue is one of standardization, in this case for need differences, and this could be achieved using either of the two basic methods, direct or indirect, when micro data are available. Unlike the variables to be used for demographic standardization, it is not a priori obvious which x variables would be good candidates to act as appropriate need proxies/predictors. In practice, researchers have relied on demographic and morbidity variables as recorded in health interview surveys (e.g. self-assessed health, presence of chronic conditions, activity limitations, etc.). Any variable which indicates a greater need for health care can, in principle, be used as a standardizing variable.

In principle, a linear regression-based framework could be used to standardize for need in both a direct and indirect way. However, as we argued and explained in TN#3, there appears to be little to commend the direct approach in this context. That is why we restrict attention here to the method of indirect standardization, which aims at comparing the actual distribution of use with the distribution that would be expected given the distribution of ‘need’. The latter is estimated as the need-expected distribution of health care.

Indirect standardization using linear regression models
The aim of any standardization is usually to see what the distribution of a variable of interest y would look like in the absence of differences in the distribution of the certain standardizing (need) variables. It requires estimates of (a) the distribution of the standardizing variables (x) and the variable of interest (y), and (b) the correlation between these variables.
The regression-based method proceeds by estimating an equation for health care \( (y_i) \) like

\[
y_i = \alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k z_{ik} + \epsilon_i
\]

where \( i \) denotes the individual, \( x_i \) are ‘need’ variables for which we want to standardize, and the \( z_k \) are non-need variables for which we do not want to standardize but for which we only want to control in the estimation of the \( \beta_j \). For example, suppose that some groups with poor health (an \( x \)-variable), e.g. the disabled or handicapped, receive more generous coverage (a \( z \) variable) of treatment than the non-disabled. If we were to estimate eq. (1) excluding a variable capturing the better coverage then the variable poor health would – to some extent – also reflect the more generous cover, over and above the ‘pure’ effect of greater need for care of these individuals “all else equal”.\(^1\) This would overestimate the ‘appropriate vertical need difference’ as embodied in the coefficient of that variable.

OLS estimates of the coefficients in (1) are then combined with actual values of the \( x \) variables and sample mean values of the \( z \) variables to obtained the predicted, or ‘\( x \)-expected’ values of utilisation, \( \hat{y}_i^x \):

\[
\hat{y}_i^x = \alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k z_{ik}
\]

Estimates of indirectly standardised utilisation, \( \hat{y}_i^\text{IS} \), are obtained as the difference between actual and \( x \)-expected utilisation, plus the sample mean \( \bar{y} \):

\[
\hat{y}_i^\text{IS} = y_i - \hat{y}_i^x + \bar{y}
\]

It has been shown (cf. [4, 5]) that when income is included alongside the \( x \) and \( z \) variables in (1), the decomposition method described in TN#14 provides a very flexible method to implement the IS procedure. The IS standardized concentration index of \( y \) then simply equals the unstandardized concentration index of \( y \) minus the contributions of each of the \( x \) variables, computed as the product of each \( x \)'s concentration index and its \( y \) elasticity. The decomposition, in effect, partitions the concentration index (or the area between the concentration curve and the diagonal) into the contributions of each \( x \) and \( z \) variable, and the reader or user of the information can select which contributions to “standardize for”.

In the case that there are no control, \( z \), variables, a shortcut method of obtaining an indirectly standardised concentration index is simply to include the standardising variables directly in a convenient regression equation for this index (see Notes 3 & 7). While this approach may provide a quick and straightforward means of generating need-standardized estimates, its usefulness in the case of health care utilization is very limited due to the inappropriateness of OLS in most contexts.

**Indirect standardization in the case of non-linear regressions**

TN#12 makes clear that \( y \) variables used to measure the use of health care often come in the form of discrete dependent variables with a probability or a count data interpretation. Moreover, differences in the effects of variables determining the probability of any health care utilisation versus the conditional use given at least one contact often require the use of so-called two-part models. Consistent estimation of the effects of \( x \) and \( z \) on \( y \) therefore often implies the estimation of intrinsically non-linear models like probit and logit for the use probabilities and count models or truncated count models for the conditional usage. As a consequence, some of the convenient standardization procedures feasible in the linear regression setting

---

\(^1\) In practice, the more generous cover may be perceived as a reflection of society’s concern that such individuals would receive less care than they need in the absence of such subsidy, but we assume here that – holding all other factors like income and accessibility etc constant – it is the partial (rather than the total) effect of poor health on health care use which, *on average*, reflects the appropriate vertical need difference between those who are in poor health and those who are not.
are often not available in this context. The method of obtaining \(x\)-expected values of \(y\) as described in eq. (3) breaks down because the equation is non-additive. When generating the \(x\)-expected values of \(y\), the \(z\) variables cannot be controlled for by setting them equal to their means. Other methods, based on appropriate linear approximations of the non-linear estimation results are then called for. One possibility is to use estimates of the marginal effects evaluated at the means (other options are possible, see e.g. [4]). Writing the equation in its linear approximation form will allow the methods described in the previous subsection to be used. This has the advantage of being a linear additive model of actual utilisation, but it is only an approximation. The inevitable price to be paid for this practice is that the standardization also only holds ‘approximately’, and it is not unique but depends on the values used for the evaluation.

If the general functional form \(G\) of such a non-linear model (like a probit, a logit, a negative binomial) can be written as:

\[
y_i = G(\alpha_i + \sum_j \beta_j x_{ij} + \sum_k \gamma_k x_{ik}) + \epsilon_i
\]

then a linear approximation of this function is given by:

\[
y_i = \alpha_i^m + \sum_j \beta_j^m x_{ij} + \sum_k \gamma_k^m x_{ik} + u_i
\]

where the \(\alpha_i^m, \beta_j^m\) and \(\gamma_k^m\) are defined as the average or marginal effects \(dy/dx\) of each \(x\) treated as fixed parameters and evaluated at the mean (or some other parameter), and \(u_i\) is the implied error term which includes approximation errors. Since eq. (5) represents a re-linearized equation derived from the non-linear model in eq. (4), it can be used in the same way as eq. (1) to generate need-expected distributions, \(\hat{y}_i^x\) and need-standardized distributions, \(\hat{y}_i^{IS}\). The only small difference is that in eq. (3), the actual sample mean \(\overline{y}\) has to be replaced by the mean of the predictions, \(\overline{\hat{y}}\), since the linear approximation does not guarantee that the predicted mean is identical to the actual mean.

Linear approximation can be avoided as long as it is acceptable to include only \(x\) variables in the estimation of eqn. (1). In that case, predictions from the non-linear estimates can still be used to generate the distribution of ‘needed’ health care. Having obtained the predictions \(\left\{\hat{y}_i\right\}\), a need-standardized concentration index can be computed using the following convenient regression:

\[
2\sigma_R^2 \left[ \frac{y_i}{\overline{y}} - \frac{\hat{y}_i}{\overline{\hat{y}}} \right] = \delta_i + \delta_i R_i + v_i
\]

where \(R_i\) is the fractional rank in the income distribution and \(\sigma_R^2\) its variance. The OLS estimate of \(\delta_i\) will be equal to the need-standardized concentration index \(C^*\) (or \(HI_{WV}\)).

**Need standardization of a health care use distribution**

There is some evidence that despite the discrete nature of the dependent variables, in many cases the results obtained with non-linear methods are very close to those obtained by OLS (cf. e.g. [2, 3]). If researchers are prepared to rely on OLS regressions to analyze health care utilization, then all Stata commands presented in TN#4 to compute standardized distributions and indices can be used in this context, simply replacing health variables with use variables and demographic variables with need variables. Here, we describe some options to obtain need-standardized distributions and indices in the context of non-linear estimation models. The methods were used to generate the results presented in Box 1.
Box 1: The distribution of health care use and health care need in Jamaica

Table 1 shows the actual, the need-expected and the need-standardized distributions for the probability of reporting at least one (preventive) visit to a doctor, a nurse or other health practitioner, by quintiles of equivalent expenditure in Jamaica derived from the 1989 Survey of Living Conditions [6]. The indicators used in the prediction of needed health care are demographic variables (7 age-sex dummies), self-assessed health (4 dummies) and functional limitations of activities (7 dummies). It can be seen that the actual distribution observed is clearly pro-rich, while the need-expected distribution is pro-poor. This is a result of the fact that ‘need’ as proxied by demographic and morbidity characteristics is more concentrated among the lower income groups. As a result, poor Jamaicans’ probability of reporting a health system contact is 6.3% lower than would be expected on average given their need, while rich Jamaicans report a probability of such a contact which is 7.3% higher than expected. It is therefore no surprise that the need-standardized distribution shows an even more pro-rich distribution than the actual distribution. After need standardization, the richest quintile’s contact probability is twice that of the poorest.

It can be seen that in this example, controlling for other covariates in the estimation of the need effects does not have a great influence, despite the fact that both (the log of) equivalent expenditure and insurance status have a positive and significant impact on the use probability. When income level and having health insurance coverage are controlled for, the need-expected distribution is only very slightly more pro-poor (CI= -0.0485) than without such controls (CI= -0.0389).

Table 1: Quintile distributions of actual, need-predicted and need-standardized visits to doctor, nurse or other health practitioner, Jamaica 1989

<table>
<thead>
<tr>
<th>Quint</th>
<th>Actual</th>
<th>Need</th>
<th>Difference</th>
<th>Need</th>
<th>Need-stand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>using Probit</td>
<td>predicted</td>
<td>minus actual</td>
<td>using Probit</td>
<td>and controls</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.173</td>
<td>0.236</td>
<td>-0.063</td>
<td>0.243</td>
<td>0.142</td>
</tr>
<tr>
<td>2</td>
<td>0.200</td>
<td>0.220</td>
<td>-0.020</td>
<td>0.221</td>
<td>0.192</td>
</tr>
<tr>
<td>3</td>
<td>0.204</td>
<td>0.215</td>
<td>-0.011</td>
<td>0.213</td>
<td>0.204</td>
</tr>
<tr>
<td>4</td>
<td>0.217</td>
<td>0.199</td>
<td>0.017</td>
<td>0.194</td>
<td>0.235</td>
</tr>
<tr>
<td>5</td>
<td>0.268</td>
<td>0.195</td>
<td>0.073</td>
<td>0.193</td>
<td>0.288</td>
</tr>
<tr>
<td>Mean</td>
<td>0.213</td>
<td>0.213</td>
<td>0.000</td>
<td>0.212</td>
<td>0.213</td>
</tr>
<tr>
<td>CI/95%</td>
<td>0.0929</td>
<td>-0.0389</td>
<td>0.1318</td>
<td>-0.0485</td>
<td>0.1412</td>
</tr>
<tr>
<td>St Err</td>
<td>0.0121</td>
<td>0.0036</td>
<td>0.0117</td>
<td>0.0038</td>
<td>0.0116</td>
</tr>
<tr>
<td>t</td>
<td>7.65</td>
<td>-10.95</td>
<td>11.31</td>
<td>-12.84</td>
<td>12.13</td>
</tr>
</tbody>
</table>
Stata syntax for need standardization

**Convenient standardization using OLS regression**

By far the easiest method of standardization is to use the convenient procedure and include ‘confounders’ (like need variables or dummies) in the convenient OLS regression procedure. Below are the commands for the Jamaican example (where no weighting was required). First we create the (fractional) ranking variable using

```
gen n=_N
egen rank= rank (deqexp), unique
gen frank = rank/n
```

Then we transform the relevant health care use indicator (e.g. vis) as follows

```
egen mvis = mean (vis)
egen sdrank = sd(frank)
gen varrank = sdrank * sdrank
/gen tvis = 2* varrank*vis/mvis
```

and run a Newey-West regression to obtain the estimate of the unstandardized concentration index and its standard error from

```
newey tstvis frank, lag(1) t(rank)
```

The conveniently need standardized estimate only requires the inclusion of a set of need indicator (dummy) variables. Define the vector of regressors as, for instance,

```
global x "var1 var2 ...."
```

Then estimate

```
newey tvis frank $x, lag(1) t(rank)
```

In most cases the above procedure will be sufficient to obtain reasonable estimates of a standardized concentration index and its standard error. While easy and straightforward, one disadvantage is that this method does not, for instance, directly generate, the quintile distribution of the standardized variable. The other is, obviously, that it does not appropriately deal with the non-continuous nature of the dependent variable.

**Indirect standardization using non-linear regressions**

A two-step procedure – which is only very slightly less convenient—requires the estimation of a ‘prediction’ equation (like eq. (5)) first to generate the values of need-expected health care use. This could be a probit or a logit in the case of a probability, or a count model in the case of a number of visits, or a combination of the two in the case of a two-part model.

```
dprobit vis $x
predict pvis
```

Then the indirectly standardized visit propbability (=stvis) can be computed using eq. (4) as follows
The quintile distributions of need-expected and standardized use can be generated using

tabstat vis pvis svis, by (quint)stats (mean)

Because the non-linear predictions do not necessarily have the same mean as the observed values, we need to compute

egen mpvis = mean (pvis)
egen mstvis = mean (stvis)

The estimates of the concentration index and its standard error of the need-expected probability can be obtained from the usual procedure:

gen tpvis = 2* varrank*pvis/mpvis
newey tpvis frank, lag(1) t(rank)

An estimate of the (indirectly standardized) concentration index (or $HI_{W}$) is obtained from

gen tdvis = tvis – tpvis
newey tdhui frank, lag(1) t(rank)

where the regression coefficient provides the estimate of the CI.

**Indirect standardization using the partial marginal effects from non-linear regressions**

If we want to include income (e.g. the logarithm of equivalent expenditure) or any other ‘control’ variable in the IS procedure, the procedure needs to be modified to allow for the use of the marginal effects. The aim is to obtain and to use ‘partial’ instead of ‘total’ need effects in the standardization. For instance, in order to control for two other covariates (e.g. $\text{logeqexp}$ and $\text{insur}$ in the Jamaican case) in the prediction regression, we re-estimate the probit model using the dprobit command in Stata. First, define a $z$ vector as

global z “logeqexp insur”
dprobit vis $z$ $x$, robust

This gives the marginal effects (dF/dx) estimates vector, which can be retrieved by

matrix DF = e(dfdx)

You can check this by displaying the relevant DF elements (in our case $z = [\text{logeqexp}, \text{insur}]$) as

estimates list
display DF[1,1] DF[1,2]

Then, generate the probit predicted probabilities using the ‘p’ option in the predict command

---

2 In principle, we only want to control for truly exogenous characteristics which may influence use and are associated with ‘need’. Income and insurance status are not very good examples of such variables to the extent that they could themselves be influenced by need variables. Availability of care in the region, for instance, would be a better example to the extent that it is not determined by (this individual’s) need.
predict probvis, p

When generating the indirectly standardized values of vis, make sure to set income and insurance at their population means by correcting the predictions as follows

\[\text{egen mlexp} = \text{mean(logeqexp)}\]
\[\text{egen minsur} = \text{mean(insur)}\]
\[\text{gen xpvis} = \text{probvis} + \text{DF}[1,1]*(\text{mlexp} - \text{logeqexp}) + \text{DF}[1,2]*(\text{minsur} - \text{insur})\]

where DF[1,1] and DF[1,2] denote the estimated marginal effect of the two z variables. If other ‘adjusters’ are to be included in the regression, then the same corrections have to be applied to the predictions for these other (z) variables.

Now, eq. (4) can be implemented again using

\[\text{gen stvis} = \text{mvis} + \text{pvis} - \text{xpvis}\]

and quintile tabulations and concentration indices estimated as above, by replacing the vis and pvis variables by stvis and xpvis.

When, instead of the probit model, other non-linear estimation techniques are being used to model health care utilization. It then requires to use the mfx command after the estimation command to generate the marginal effect estimates (cf TN#12). For instance, for a negative binomial regression use

\[\text{nbreg vis$z$}$x$, robust\]

to estimate the model and

\[\text{mfx compute, nose}\]

to generate the marginal fixed effects. Combining the model predictions obtained from

\[\text{predict npvis, n}\]

with the mfx estimates defined as

\[\text{matrix DFDX} = \text{e(Xmfx\_dydx)}\]
\[\text{di DFDX}[1,1] \quad \text{DFDX}[1,2] /* check */\]

need-predicted use defined as

\[\text{gen xnpvis} = \text{npvis} + \text{DFDX}[1,1]*(\text{mlexp} - \text{logeqexp}) + \text{DFDX}[1,2]*(\text{minsur} - \text{insur})\]

and need-standardized use determined using

\[\text{egen mxnpvis} = \text{mean(xnpvis)}\]
\[\text{drop stnpvis}\]
\[\text{gen stnpvis} = \text{mnpvis} + \text{npvis} - \text{xnpvis}\]

Quintile distributions and concentration indices can be obtained as usual for the various, newly generated nvis variables.

For two-part models, combining a dichotomous dependent variable in the first part with a (conditional or truncated) count model for the second part, the marginal effects-based approach using the linear
approximation does not work. The best approach in that case is to generate on the predictions of the two-part model and to compute the concentration index using a convenient regression of the (transformed) difference between actual and need-predicted use.

### Conclusion
The example presented in Box 1 for Jamaica shows that in many cases the convenient method will provide a reasonably good estimate of the need-standardized concentration index. Only when a distribution of standardized values is required, is the two-step procedure necessary. Linear approximations are necessary only when one wishes to estimate the correlation of ‘need’ variables conditional on other variables.

### Useful links
For further information on the use of other evaluations of marginal effects, and for testing for the importance of additional standardizing variables using bootstrap methods to obtain standard errors for the ‘contributions’ of variables to inequality, see e.g. [4].

### Bibliography