15

Measuring and Explaining Inequity in Health Service Delivery

Equitable distribution of health care is a principle subscribed to in many countries, often explicitly in legislation or official policy documents (van Doorslaer, Wagstaff, and Rutten 1993). Egalitarian equity goals distinguish between horizontal equity—equal treatment of equals—and vertical equity—appropriate unequal treatment of unequals. In health care, most attention, both in policy and research, has been given to the horizontal equity principle, defined as "equal treatment for equal medical need, irrespective of other characteristics such as income, race, place of residence, etc." (van Doorslaer et al. 2000; Wagstaff and van Doorslaer 2000; Wagstaff, van Doorslaer, and Paci 1991). In this chapter, we discuss measurement and explanation of horizontal inequity in the delivery of health care.

In practice, it is not possible to examine the extent to which the horizontal equity principle is violated without simultaneously specifying a vertical equity norm. Researchers have usually assumed, implicitly or explicitly, that, on average, vertical equity is satisfied. That is, the observed differential utilization of health care resources across individuals in different states of need is appropriate. If that is accepted, then the measurement of horizontal inequity in health care use can proceed in much the same way as the standardization methods covered in chapter 5. For example, one seeks to establish whether there is differential utilization of health care by income after standardizing for differences in the need for health care in relation to income. In empirical analyses, expected utilization, given characteristics such as age, gender, and measures of health status, is used as a proxy for "need." Complications to the regression method of standardization arise because measures of health care utilization typically are nonnegative integer counts (e.g., numbers of visits, hospital days, etc.) with very skewed distributions. As discussed in chapter 11, nonlinear methods of estimation are then appropriate. But the standardization methods presented in chapter 5 do not immediately carry over to nonlinear models. They can be rescued only if relationships can be represented linearly. In this chapter, we therefore concentrate on standardization in nonlinear settings.

Once health care use has been standardized for need, inequity can be measured by the concentration index. Inequity could then be explained by decomposing the concentration index, as explained in chapter 13. In fact, with the decomposition approach, standardization for need and explanation of inequity can be done in one step. We describe this procedure in the final section of the chapter.

Measuring horizontal inequity

There will typically be inequality in the utilization of health care in relation to socioeconomic characteristics, such as income. Typically, in high-income countries poorer individuals consume more health care resources as a result of their lower health status and so greater need for health care. Obviously, such inequality in health care use cannot be interpreted as inequity. In low-income countries, the lack of health insurance and purchasing power among the poor typically mean that their utilization of health care is less than that of the better-off despite their greater need (Gwatkin et al. 2003; O'Donnell et al. forthcoming). In this case, the inequality in health care use does not fully reflect the inequity. To measure inequity, inequality in utilization of health care must be standardized for differences in need. After standardization, any residual inequality in utilization, by income for example, is interpreted as horizontal inequity, which could be pro-rich or pro-poor.

Standardization for differences in need could be done using either the direct or indirect method described in chapter 5. Although with demographic standardization the appropriate standardizing variables are immediately obvious, that is not true for need standardization. Need is a rather elusive concept that has been given a variety of interpretations in relation to the definition of equity in health care delivery (Culyer 1995; Culyer and Wagstaff 1993). By some definitions, measurement of need is not tractable, at least in the context of large-scale household surveys. In practice, researchers have relied on demographics plus health status and morbidity variables (e.g., self-assessed health, presence of chronic conditions, activity limitations, etc.) to proxy need.

Although both direct and indirect methods of standardization could be used, as we argued in chapter 5, when microdata are available there is little to commend the direct approach. Here, we restrict attention to the indirect method, which gives the difference between the actual distribution of use and the distribution that would be expected given the distribution of need. The latter is referred to as the *need-expected* distribution of health care.

When health care use is modelled by linear regression, the standardization procedure is exactly as presented in equations 5.1 through 5.3 of chapter 5. The need variables are included among the x's. The control, or z, variables should include nonneed correlates of health care utilization for which we do *not* want to standardize but which would bias the coefficients on the need variables if omitted from the regression (Gravelle 2003; Schokkaert and van de Voorde 2004). For example, suppose that some groups with poor health (an x variable), for example, people who are disabled or handicapped, receive more generous insurance coverage (a z variable) than the nondisabled. If we were to estimate the standardizing regression excluding a variable capturing the better coverage, then the coefficient on the poor health variable would—to some extent—pick up the effect of more generous cover, over and above the direct effect of greater need. That would overestimate the "appropriate vertical need difference" as embodied in the coefficient of poor health.¹

Once need-standardized utilization has been estimated, inequity can be tested by determining whether standardized use is unequally distributed by income, for example. Inequity could be measured by estimating the concentration index for need-standardized utilization, which has been referred to as the health inequity

¹The more generous cover to disabled persons may reflect society's concern that these individuals would receive less care than they need in the absence of such a subsidy. We assume here that—holding all other factors such as income and accessibility constant—it is the partial effect of poor health on health care use that, *on average*, reflects the appropriate vertical need difference between those who are in poor health and those who are not.

index (HI_{WV}) (Wagstaff and van Doorslaer 2000). Equivalently, this can be obtained as the difference between the concentration index for actual utilization and that for need-predicted utilization.

This procedure rests on the assumption that once observable need indicators have been controlled for, any residual variation in utilization is attributable to nonneed factors. Given that the data available on need indicators typically are limited, that is likely to be a strong assumption. It will result in biased measurement of horizontal inequity in the case that unobservable variation in need is correlated with income. Schokkaert, Dhaene, and Van de Voorde (1998) discuss this issue in the context of the related literature on risk adjustment.

Indirect standardization with nonlinear models

Measures of health care use are typically nonnegative integer counts, for example, number of visits to a doctor or days in a hospital. In a sample, there will typically be a large proportion of observations with no utilization and very few observations, corresponding to individuals falling severely ill, with utilization very much above the mean. Given this, it may be considered appropriate to model the determinants of the use/nonuse probability separately from the number of visits conditional on any use. Although the least squares regression method of indirect standardization could be used with such data, it would not guarantee that the predicted values from the standardizing regression (equation 5.2) lie in the permitted range of (0,1) for binary variables and at or above zero for nonnegative counts (see chapter 11). This can be avoided by using nonlinear estimators.

Let us write a nonlinear model of the relationship between a health care variable, y, which may be binary or a count, and need (x) and control (z) variables in terms of a general functional form G:

(15.1)
$$y_i = G\left(\alpha + \sum_j \beta_j x_{ji} + \sum_k \gamma_k z_{ki}\right) + \varepsilon_i,$$

where *G* will take particular forms for the probit, logit, Poisson, negative binomial, and so on models. If there were no *z* variables included in equation 15.1, then predicted values obtained from the model could be interpreted as need-expected utilization. Need-standardized utilization could then be defined as actual use minus need-expected utilization, as in equation 5.3, only in this case the mean of the prediction should be added, rather than the mean of the actual variable, to ensure that the mean of standardized utilization equals that of actual utilization.

However, as argued above, including *z* variables in the model is probably desired to avoid omitted-variables bias. Doing so in this nonlinear context leads to a problem because the effect of the *z* variables on need-standardized use can no longer be entirely neutralized by setting them equal to their means or indeed to any other vector of constants. As a result, the variance of the need-standardized use will depend on the values to which the *z* variables are set in the standardization procedure, and that will affect measures of income-related inequality, such as the concentration index. Accepting this, the analyst could define standardized use as follows:

(15.2)
$$\hat{y}_i^{IS} = y_i - G\left(\hat{\alpha} + \sum_j \hat{\beta}_j x_{ji} + \sum_k \hat{\gamma}_k \overline{z}_k\right) + \frac{1}{n} \sum_{i=1}^n G\left(\hat{\alpha} + \sum_j \hat{\beta}_j x_{ji} + \sum_k \hat{\gamma}_k \overline{z}_k\right),$$

where *n* is the sample size, and we have chosen to set the *z* variables to their means (\bar{z}_k) in obtaining the predictions. Note that the mean of \hat{y}_i^{IS} is equal to that of *y* but because *G* is not linearly additive, its variance would differ if the *z* variables were set to some other vector of values.

Box 15.1 Distribution of Preventive Health Care Utilization and Need in Jamaica

The table below shows the actual need-expected and need-standardized distributions for the probability of reporting at least one preventive visit to a doctor, nurse, or other health practitioner, by quintiles of equivalent expenditure in Jamaica derived from the 1989 Survey of Living Conditions (van Doorslaer and Wagstaff 1998). 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, and 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, for the poorest fifth of Jamaicans, the probability of reporting a preventive care contact is 6.5 percent lower than would be expected on average given their need, whereas the richest 20 percent of Jamaicans report a probability of such a contact that is 8.2 percent 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.

The figures reported in the table for need-predicted use and its difference from actual use are derived from a probit model including control variables, specifically (log) equivalent expenditure and health insurance status, which are set equal to their sample means to obtain the predictions. Need-standardized use is presented both with and without the inclusion of controls in the standardizing model and estimating this by both OLS and probit. In this example, results are not sensitive to either variation. It has been found elsewhere that concentration indices for standardized health care utilization are relatively insensitive to the use of OLS or nonlinear models for standardization (van Doorslaer, Masseria, and OECD Health Equity Research Group 2004; van Doorslaer et al. 2000; Wagstaff and van Doorslaer 2000). That is reassuring given the complications introduced by nonlinear models noted above. Insensitivity to the inclusion of control variables in the standardizing regression may be more specific to this example. Others have found more substantial differences (Gravelle 2003).

Quintile	Actual	Probit with controls		Need-standardized			
		Need- predicted	Difference = predicted – actual	With controls		Without controls	
				Probit	OLS	Probit	OLS
Poorest 20%	0.1717	0.2363	-0.0646	0.1450	0.1457	0.1483	0.148
2nd poorest 20%	0.2003	0.2158	-0.0155	0.1942	0.1943	0.1952	0.1950
Middle	0.2052	0.2119	-0.0067	0.2029	0.2030	0.2029	0.2029
2nd richest 20%	0.2157	0.1954	0.0203	0.2300	0.2297	0.2282	0.2285
Richest 20%	0.2706	0.1888	0.0817	0.2914	0.2908	0.2889	0.2893
Mean	0.2127	0.2097	0.0030	0.2127	0.2127	0.2127	0.2122
Concentration index/HI _{WV}	0.0928	-0.0452		0.1374	0.1362	0.1318	0.1322
Standard error	0.0122	0.0039		0.0117	0.0117	0.0117	0.0112
<i>t</i> -ratio	7.6249	-11.4721		11.7162	11.6182	11.2663	11.2968

Distributions of Actual Need-Predicted and Need-Standardized Preventive Visits to Doctor, Nurse, or Other Health Practitioner, Jamaica 1989

Computation

Stata computation for standardization by linear regression is provided in chapter 5. The general procedure is the same in the case of nonlinear models, with the replacement of the OLS command regr with the chosen estimator. So, in the case of a probit model, the need-predicted (yhat) and need-standardized (yst) probability of health care utilization would be generated as follows:

```
qui probit y $X $Z [pw=weight]
foreach z of global Z {
    gen copy_`z'=`z'
    qui sum `z' [aw=weight]
    replace `z' = r(mean)
}
predict yhat
foreach z of global Z {
    replace `z' = copy_`z'
    drop copy_`z'
}
sum m_yhat [aw=weight]
gen yst = y-yhat + r(mean)
```

where X and Z are globals containing lists of need and control variables, respectively. Note that the mean of predicted use and not the mean of actual use is added in generating standardized use. Obviously if control variables were not included, the predictions would be obtained immediately after the model is estimated and neither loop is required.

Quintile means can be estimated using tabstat and concentration indices computed as explained in chapter 8.

Explaining horizontal inequity

In chapter 13, we noted that if a health variable is specified as a linear function of determinants, then its concentration index can be decomposed into the contribution of each determinant, computed as the product of the health variable's elasticity with respect to the determinant and the latter's concentration index. This makes it possible to explain socioeconomic-related inequality in health care utilization. In fact, the decomposition method allows horizontal inequity in utilization to be both measured and explained in a very convenient way. The concentration index for need-standardized utilization is exactly equal to that which is obtained by subtracting the contributions of all need variables from the unstandardized concentration index (van Doorslaer, Koolman, and Jones 2004). Besides convenience, the advantage of this approach is that it allows the analyst to duck the potentially contentious division of determinants into need (x) and control (z) variables and so the determination of "justified" and "unjustified," or inequitable, inequality in health care utilization. The full decomposition results can be presented, and the user can choose which factors to treat as x variables and which to treat as z variables.

The decomposition result holds for a linear model of health care. If a nonlinear model is used, then the decomposition is possible only if some linear approximation

to the nonlinear model is made. One possibility is to use estimates of the partial effects evaluated at the means (van Doorslaer, Koolman, and Jones 2004). That is, a linear approximation to equation 15.1 is given by

(15.3)
$$y_{i} = \alpha^{m} + \sum_{j} \beta_{j}^{m} x_{ji} + \sum_{k} \gamma_{k}^{m} z_{ki} + u_{i},$$

where the β_j^m and γ_k^m are the partial effects, dy/dx_j and dy/dz_k , of each variable treated as fixed parameters and evaluated at sample means; and u_i is the implied error term, which includes approximation errors. Because equation 15.3 is linearly additive, the decomposition result (Wagstaff, van Doorslaer, and Watanabe 2003) can be applied, such that the concentration index for *y* can be written as

(15.4)
$$C = \sum_{j} (\beta_{j}^{m} \overline{x}_{j} / \mu) C_{j} + \sum_{k} (\gamma_{k}^{m} \overline{z}_{k} / \mu) C_{k} + GC_{u} / \mu$$

Because the partial effects are evaluated at particular values of the variables, for example, the means, this decomposition is not unique. This is the inevitable price to be paid for the linear approximation. Also, unlike the truly linear case, the index of horizontal inequity, H_{WV} , obtained by subtracting the need contributions in equation 15.4 from the unstandardized concentration index will not equal the concentration index for need-standardized utilization calculated from the estimates of the nonlinear model parameters, as described in the previous section.

Note that equation 15.3 could itself be used to estimate need-standardized utilization and, unlike equation 15.2, its distribution would not depend on the values to which the control variables were set. Need-predicted utilization could be defined as

(15.5)
$$\hat{y}_i^X = \hat{\alpha}^m + \sum_j \hat{\beta}_j^m x_{ji} + \sum_k \hat{\gamma}_k^m \overline{z}_k$$

Then indirectly standardized use would be given by the following:

(15.6)
$$\tilde{y}_i^{IS} = y_i - \hat{y}_i^X + \overline{\hat{y}}$$

where $\overline{\hat{y}}$ is the mean of the predictions from equation 15.3 with all variables at actual values. Because equation 15.3 is linearly additive, the *z* variables cancel in the final two terms of equation 15.6, and the variance of \tilde{y}_i^{IS} , unlike that of \hat{y}_i^{IS} , does not depend on the values to which those variables are set in the need-prediction equation, 15.5. However, \tilde{y}_i^{IS} will depend on the values of both the *x* and *z* variables at which the partial effects are evaluated. There is no escaping the nonuniqueness of the standardization in the context of a nonlinear model including control variables.

Computation

Stata code for the concentration index decomposition based on linear regression is provided in chapter 13. For nonlinear estimators, the partial effects must be calculated from the parameter estimates and then the contributions calculated using these partial effects, as in equation 15.4. For the probit model, Stata has a programmed routine called dprobit that provides partial effects directly (see chapter 11):

```
dprobit y $X $Z [pw=weight]
matrix dfdx=e(dfdx)
```

The matrix command saves the partial effects into a matrix named dfdx. By default, partial effects are calculated at the sample means. They can be computed at another vector of values using the at (*matname*) option (see chapter 11). For other

Box 15.2 Decomposition of Inequality in Utilization of Preventive Care in Jamaica, 1989

We decompose the concentration index for any use of preventive health care in Jamaica. The probability of making any use of preventive care is estimated both by least squares, in which the decomposition is exactly as presented in chapter 13, and by probit, in which case we make a linear approximation to the model using the partial effects evaluated at sample means, as in equation 15.3, and then use the decomposition given by equation 15.4. Need and nonneed variables are as described in box 15.1, although, as pointed out above, the decomposition approach allows the user to choose which factors to consider as need proxies. We do not present the full decomposition results but provide, in the table below, the absolute and percentage contributions to the unstandardized concentration index for groups of "need" factors (age-sex dummies, self-assessed health dummies, and functional limitation dummies) and for the two "nonneed" factors. Results are not particularly sensitive to the estimation method. The residual difference between the unstandardized concentration index and the sum of the contributions of all need and nonneed factors is larger for the partial effects probit approach, largely because this gives a slightly larger estimate of the contribution of household expenditure.

The contribution of all need factors is negative, indicating that if utilization were determined by need alone, it would be pro-poor. The aggregate contribution of all need factors is about 47 percent of the unstandardized index. Self-assessed health and functional limitations each contribute roughly twice as much as the age-sex groups. Although the distribution of need pushes utilization in a pro-poor direction, this is more than offset by the direct effect of household expenditure and of insurance coverage. If need were distributed equally, the direct effect of household expenditure on utilization would produce a concentration index 29 to 34 percent greater than that observed. There is also an indirect effect of household expenditure on utilization through health insurance coverage that raises the concentration index by 24 percent of its observed value.

The horizontal inequity index is positive, indicating that for given need, the betteroff make greater use of preventive care in Jamaica. The index is not particularly sensitive to the estimation method.

	0	LS	Probit partial effects		
-	Absolute	Percentage	Absolute	Percentage	
Need factors					
Age-sex groups	-0.0083	-8.9	-0.0110	-11.9	
Self-assessed health	-0.0169	-18.2	-0.0163	-17.5	
Functional limitations	-0.0182	-19.6	-0.0170	-18.3	
Subtotal	-0.0434	-46.7	-0.0443	-47.7	
Nonneed factors					
Log household expenditure	0.1196	128.8	0.1249	134.5	
Health insurance cover	0.0218	23.5	0.0221	23.8	
Subtotal	0.1414	152.3	0.1470	158.3	
Residual	-0.0052	-5.6	-0.0099	-10.6	
Total	0.0928		0.0928		
Horizontal inequity index	0.1362		0.1371		

Decomposition of Concentration Index for Access to Preventive Health Care in Jamaica, 1989 nonlinear models, the partial effects can be calculated using the mfx command after running the model (see chapter 11).

The contributions of need factors can then be computed with the following loop:

```
sca need=0
foreach x of global X {
     qui {
            mat b x' = dfdx[1, "x'"]
            sca b x' = b x' [1,1]
            corr r `x' [aw=weight], c
            sca cov x' = r(cov 12)
            sum `x' [aw=weight]
            sca m `x' = r(mean)
            sca elas x' = (b x' m x')/m y
            sca CI x' = 2*cov x'/m x'
            sca con_`x' = elas_`x'*CI_`x'
            sca prcnt `x' = con `x'/CI
            sca need=need+con `x'
      }
     di "`x' elasticity:", elas `x'
     di "`x' concentration index:", CI `x'
     di "`x' contribution:", con `x'
      di "`x' percentage contribution:", prcnt `x'
}
```

where CI is a scalar equal to the unstandardized concentration index computed as in chapter 8. The scalar need will contain the sum of the contributions of all the need factors. The contributions of the nonneed factors can be computed by running the same loop over the global Z containing the nonneed factors and renaming the scalar need to nonneed. The total contributions of all need factors and of nonneed factors and the horizontal inequity index (HI_{WV}) can then be displayed as follows:

```
di "Inequality due to need factors:", need
di "Inequality due to non-need factors:", nonneed
sca HI = CI - need
di "Horizontal Inequity Index:", HI
```

Further reading

Detailed discussion of the issues touched on in this chapter can be found in Wagstaff and van Doorslaer (2000); Gravelle (2003); Schokkaert and van de Voorde (2004); and van Doorslaer, Koolman, and Jones (2004). Standard errors for the contributions to the concentration index decomposition can be obtained by bootstrapping (van Doorslaer, Koolman, and Jones 2004). Gravelle, Morris, and Sutton (2006) make a valuable contribution in placing the empirical study of equity in health care in the context of a social welfare maximization model. That helps to make explicit the links between normative and positive analysis of the distribution of health care, a point that has also been emphasized by Schokkaert and van de Voorde (2004). It also helps clarify the conditions required for the identification of horizontal and vertical equity and to distinguish between the two.

References

- Culyer, A. J. 1995. "Need: The Idea Won't Do—But We Still Need It" [editorial]. Soc Sci Med 40(6): 727–30.
- Culyer, A. J., and A. Wagstaff. 1993. "Equity and Equality in Health and Health Care." *Journal of Health Economics* 12(4): 431–57.
- Gravelle, H. 2003. "Measuring Income Related Inequality in Health: Standardisation and the Partial Concentration Index." *Health Economics* 12(10): 803–19.
- Gravelle, H., S. Morris, and M. Sutton. 2006. "Economic Studies of Equity in the Consumption of Health Care." In *The Elgar Companion to Health Economics*, ed. A. M. Jones, 193–204. Cheltenham, England: Edward Elgar.
- Gwatkin, D. R., S. Rustein, K. Johnson, R. Pande, and A. Wagstaff. 2003. Initial Country-Level Information about Socio-Economic Differentials in Health, Nutrition and Population, Volumes I and II. Washington, DC: World Bank, Health, Population and Nutrition.
- O'Donnell, O., E. van Doorslaer, R. P. Rannan-Eliya, A. Somanathan, S. R. Adhikari, D. Harbianto, C. G. Garg, P. Hanvoravongchai, M. N. Huq, A. Karan, G. M. Leung, C-W Ng, B. R. Pande, K. Tin, L. Trisnantoro, C. Vasavid, Y. Zhang, and Y. Zhao. 2007. "The Incidence of Public Spending on Health Care: Comparative Evidence from Asia." *World Bank Economic Review* 21(1): 93–123.
- Schokkaert, E., G. Dhaene, and C. Van de Voorde. 1998. "Risk Adjustment and the Trade-off between Efficiency and Risk Selection: An Application of the Theory of Fair Compensation." *Health Economics* 7(5): 465–80.
- Schokkaert, E., and C. van de Voorde. 2004. "Risk Selection and the Specification of the Conventional Risk Adjustment Formula." *Journal of Health Economics* 23(6): 1237–59.
- van Doorslaer, E., X. Koolman, and A. M. Jones. 2004. "Explaining Income-Related Inequalities in Doctor Utilization in Europe." *Health Economics* 13(7): 629–47.
- van Doorslaer, E., C. Masseria, and OECD Health Equity Research Group. 2004. "Income-Related Inequality in the Use of Medical Care in 21 OECD Countries." In *Towards High-Performing Health Systems: Policy Studies*, ed. OECD Health Project, 109–66. Paris: OECD.
- van Doorslaer, E., and A. Wagstaff. 1998. "Inequity in the Delivery of Health Care: Methods and Results for Jamaica." Paper prepared for the HDN Network, LAC Region, Equi-LAC—WP#3. World Bank, Washington, DC.
- van Doorslaer, E., A. Wagstaff, and F. Rutten. 1993. *Equity in the Finance and Delivery of Health Care: An International Perspective*. Oxford, England: Oxford University Press.
- van Doorslaer, E., A. Wagstaff, H. van der Burg, T. Christiansen, D. De Graeve, U.-G. Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Hakkinen, and J. John. 2000. "Equity in the Delivery of Health Care in Europe and the U.S." *Journal of Health Economics* 19(5): 553–84.
- Wagstaff, A., and E. van Doorslaer. 2000. "Measuring and Testing for Inequity in the Delivery of Health Care." *Journal of Human Resources* 35(4): 716–33.
- Wagstaff, A., E. van Doorslaer, and P. Paci. 1991. "On the Measurement of Horizontal Inequity in the Delivery of Health Care." *Journal of Health Economics* 10(2): 169–205.
- Wagstaff, A., E. van Doorslaer, and N. Watanabe. 2003. "On Decomposing the Causes of Health Sector Inequalities, with an Application to Malnutrition Inequalities in Vietnam." *Journal of Econometrics* 112(1): 219–27.