Explorations of inequality: childhood immunization

Technical note and glossary

Suggested citation: Explorations of inequality: childhood immunization. Geneva: World Health Organization; 2018.

The statistics presented in the publication and its accompanying technical note are based on the re-analysis of Demographic and Health Surveys micro-data and are not necessarily the official statistics of WHO.

Contents

Survey sample sizes

The report contains data from standard Demographic and Health Surveys (DHS) Phase VI and Phase VII that were publicly available as of May 2018 *(1)*. (Note that, in India the National Family Health Survey (NFHS) follows the same protocol and methodology as the DHS.) Table 1 presents the list of countries included in this report, the year(s) of DHS or NFHS fieldwork, and survey sample sizes.

Table 1. Priority countries and survey sample sizes

Country, year(s) of DHS/NFHS fieldwork	Sample size of women aged 15–49 years	Sample size of children aged 12–23 months
Afghanistan, 2015	29 461	6 126
Chad, 2014–2015	17 719	3 131
Democratic Republic of the Congo, 2013–2014	17 651	3 485
Ethiopia, 2016	15 683	2 036
India, 2015–2016	699 686	51 544
Indonesia, 2012	45 607	3 615
Kenya, 2014	31 079	4 209
Nigeria, 2013	38 948	6 281
Pakistan, 2012–2013	13 558	2 199
Uganda, 2016	18 506	3 067

DTP3 immunization coverage indicator

The childhood immunization indicator was defined as the percentage of children aged 12–23 months who had received three doses of the combined diphtheria, tetanus toxoid and pertussis vaccine (DTP) vaccine at the time of the survey (i.e. DTP3 immunization coverage). A child was considered covered if any of the following situations applied:

- the child had a date recorded for the vaccination in the health card;
- the vaccination was recorded on the health card but without dates; or
- the vaccination was reported by the mother but the health card was not seen or did not exist.

For each child, coverage was calculated considering all DTP-containing vaccines recorded in the DHS/NFHS at the time of fieldwork, including pentavalent, tetravalent and/or DTP. Table 2 presents the names of the DTP-containing vaccines reported by the corresponding DHS/NFHS.

The WHO Collaborating Center for Health Equity Monitoring (International Center for Equity in Health, Federal University of Pelotas, Brazil) provided guidance on the computation of the childhood immunization indicator used in this report.

Table 2. DTP-containing vaccine considered for each country during the corresponding DHS/NFHS fieldwork period

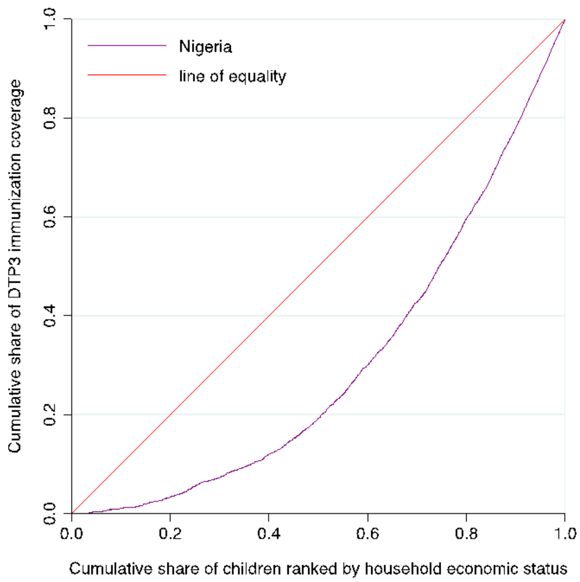
Country, year(s) of DHS/NFHS fieldwork	DTP-containing vaccine(s) reported in the DHS/NFHS
Afghanistan, 2015	Pentavalent (DTP+Hb+Hib)
Chad, 2014–2015	Pentavalent (DTP+Hb+Hib)
Democratic Republic of the Congo, 2013–2014	Pentavalent (DTP+Hb+Hib)
Ethiopia, 2016	Pentavalent (DTP+Hb+Hib)
India, 2015–2016	DTP
Indonesia, 2012	DTP
Kenya, 2014	Pentavalent (DTP+Hb+Hib)
Nigeria, 2013	DTP and Pentavalent (DTP+Hb+Hib)*
Pakistan, 2012–2013	DTP, Tetravalent (DTP+Hb) and Pentavalent (DTP+Hb+Hib)
Uganda, 2016	DTP

* In May 2012, Nigeria began the phased replacement of the diphtheria, pertussis, and tetanus (DPT) vaccine with the pentavalent vaccine (Nigeria DHS Report 2013).

Concentration curve

The concentration curve is a graphical representation of how a health variable, in this case DTP3 immunization coverage, is distributed across the population, ordered according to a measure of socioeconomic status, such as household economic status. Therefore, there are two variables underlying the concentration curve: the health variable, the distribution of which is the subject of interest; and a variable of socioeconomic status, against which the distribution is to be assessed. As shown in Figure 1, the concentration curve plots the cumulative proportion of a health variable (y-axis) against the cumulative proportion of the population (or a representative sample), ranked by their socioeconomic status from the most disadvantaged to the least disadvantaged (x-axis). If the health variable is equally distributed across socioeconomic status, i.e. if there is no inequality, the concentration curve will be a 45° line (line of equality). If, by contrast, the health variable takes higher (lower) values among people with lower socioeconomic status, the concentration curve will lie above (below) the line of equality. The further the curve lies from the line of equality, the greater the degree of inequality in health.

Figure 1. Concentration curve for DTP3 immunization coverage in Nigeria (DHS 2013)



Concentration index

The concentration index provides a measure of the degree of inequality in a health variable over the distribution of another variable. In this report, concentration indices were constructed to measure inequality in DTP3 immunization coverage over the distribution of household wealth and mother´s education. Both indices were used as measures of socioeconomic inequality in DTP3 immunization coverage.

The standard version of the concentration index can be derived from the concentration curve and represents twice the area between the concentration curve and the 45° line of equality *(2,3)*. For a bounded health variable, including binary indicators such as DTP3 immunization coverage, Erreygers *(4)* proposed a modified version of the concentration index (i.e. the Erreygers Concentration Index, or ECI). The ECI satisfies the conditions that the absolute value of the index is the same regardless of whether inequality in health or in ill-health is being measured (mirror property), and that the value of the index is invariant to any feasible positive linear transformation of the health variable (scale and translation invariance) *(4,5)*. The ECI is defined as:

$$ECI(h) = \frac{8}{n^2(b_h - a_h)} \sum_{i=1}^{n} h_i R_i$$

where h_i is the health variable, R_i is the fractional rank of child i in the distribution of socioeconomic status, n is the number of observations and b_h and a_h are the variables upper bound and lower bound, respectively[1]. The equation shows that the concentration index can be

[1] For binary variables $(b_h - a_h)$ equals one.

interpreted as a sum of weighted health levels, with the weights being determined by the socioeconomic rank.

While the standard concentration index measures relative inequality for unbounded variables, the ECI is a measure of absolute inequality for bounded variables. ECI values have a possible range from -1 to +1. It has a negative value when the health indicator is concentrated among the poor (or less educated); and it has a positive value when the health indicator is concentrated among the rich (or more educated). When there is no inequality, the ECI value is 0. The command *conindex* with the *erreygers* option in STATA v14.2 was used to calculate wealth-related and education-related Erreygers concentration indices *(6)*.

The logistic regression

Given the binary nature of our outcome of interest —whether the child is vaccinated or not— a logit model was used to assess the adjusted associations between DTP3 immunization and a group of selected characteristics[2].

Using the logit transformation of the probability of DTP3 immunization, it is possible to remove its 0-1 range restriction, and model the logarithm of the odds of immunization as a linear function of the explanatory variables *(7,8)*:

$$\log(odds_{y_i}) = \alpha + \sum \beta_k x_{ki} + \varepsilon_i$$

where $odds_{y_i}$ are the odds of immunization for child i. The odds are given by the ratio of the probability of immunization p_i to its complement $(1 - p_i)$; that is, $\frac{p_i}{(1-p_i)}$. For example, if the probability of DTP3 immunization is 2/3 = 0.67, the odds are 0.67/0.33, or two to one. x_k is a set of observed k factors associated with child's DTP3 immunization including child, mother, household and geographic characteristics; and ε_i is the error term. Each estimated coefficient, $\hat{\beta}_k$, is the expected change in the log-odds of DTP3 immunization associated with a one unit increase in the value of the explanatory variable, holding the other variables constant. The exponential function of the regression coefficient ($e^{\hat{\beta}_k}$) gives the *odds ratio*, that is, the change in the odds of DTP3 immunization for a one unit increase in the explanatory variable.

Model specification

The model specification was primarily guided by a literature review and availability of data. The model was applied to all priority countries. Although the analysis does not aim to place any causal interpretation on the estimates, the inclusion of variables in the model was limited to those that could arguably be considered exogenous and not correlated with unobservable factors that affect immunization.

---

[2] Another standard model for binary response variables is the linear probability model that can be estimated using ordinary least squares. One problem with the linear probability model is that it can yield estimates of the probability of childhood immunization that are outside the 0-1 range; another problem is that the errors are heteroscedastic *(2,7)*.

Multicollinearity was also tested. That is, the condition where a very high correlation exists between two or more variables included in the model, resulting in unstable estimates (in our case, odds ratios). In Afghanistan and Ethiopia, for example, mother's ethnicity showed collinearity with subnational region. The model including ethnicity in both countries demonstrated that the variable was not statistically significant and the associations between DTP3 coverage and ethnicity was mainly explained through the regional variation in different ethnic subgroups. So, ethnicity was dropped from the logistic regressions in these countries, although ethnicity was used in the bivariate descriptive analysis.

Although the aim was to have a model as comparable as possible across countries, its implementation was restricted by data availability and, therefore, not all countries included all explanatory variables. Explanatory variables were included in the model as binary variables, with one indicator for each level of the categorical variable. Table 3 presents the list of variables that were included in the model for each priority country.

To explore whether the relationship between an explanatory variable and DTP3 immunization varied for different subgroups of the population, models including interaction terms were also tested. Evaluating interactions between explanatory variables was necessary to define appropriate models. For each country, interactions for each combination of the levels in key variables were added to the model. Specifically, we tested interactions between: mother's education and wealth quintile; mother's age at birth and wealth quintile; mother's age at birth and mother's education; and place of residence and wealth quintile. Each set of interactions between two variables was estimated individually and their overall significance was tested.

For almost all cases, interactions between each pair of variables were not statistically significant, except in three cases.

- In Afghanistan, the overall effect of the interaction between mother's age and education was significantly different from zero (p-value = 0.019), suggesting that education has a different effect on DTP3 immunization depending on mother's age.
- In Ethiopia, there was a significant interaction effect between mother's age and education (p-value = 0.001).
- In Ethiopia, the interaction between place of residence and wealth quintile was statistically significant (p-value = 0.021), indicating a differential effect of wealth on DTP3 immunization depending on place of residence.

Despite the interaction between mother's age and education being statistically significant in two cases, the more parsimonious model was preferred. This was because the sample sizes for some combination cells of mother's age and mother's education subgroups were too small, producing large uncertainty around the estimated odds ratios. Similarly, the interaction term between place of residence and wealth quintile in Ethiopia was not included in the final model, given the small sample sizes in the subsamples of quintiles 1 to 4 in the urban area. Furthermore, including interactions did not affect the estimated parameters for other factors and none of the statistically significant interactions were qualitative; that is, the directions of existing associations did not change after including interaction terms *(9)*.

Table 3. Model specification for DTP3 immunization coverage by country

Characteristic	Explanatory variables in the model	Afghanistan	Chad	Democratic Republic of the Congo	Ethiopia	India	Indonesia	Kenya	Nigeria	Pakistan	Uganda
Child's sex	Female	√	√	√	√	√	√	√	√	√	√
	(reference category: male)										
Birth order	1st born	√	√	√	√	√	√	√	√	√	√
	2nd-3rd born	√	√	√	√	√	√	√	√	√	√
	4th-5th born	√	√	√	√	√	√	√	√	√	√
	(reference category: 6th + born)										
Mother's age at birth	20-34 years	√	√	√	√	√	√	√	√	√	√
	35-49 years	√	√	√	√	√	√	√	√	√	√
	(reference category: 15-19 years)										
Mother's education[a]	Primary school	√	√	√	√	√	√	√	√	√	√
	Secondary school					√	√		√	√	√
	More than secondary school					√	√		√	√	√
	Secondary school or more	√	√	√	√			√			
	(reference category: no education)										
Mother's ethnicity or caste/tribe[b]	Country-specific categories			√		√		√			
	(reference category: country specific)										
Sex of household head[c]	Female		√	√	√	√	√	√	√	√	√
	(reference category: male)										
Household wealth	Quintile 2	√	√	√	√	√	√	√	√	√	√
	Quintile 3	√	√	√	√	√	√	√	√	√	√
	Quintile 4	√	√	√	√	√	√	√	√	√	√
	Quintile 5	√	√	√	√	√	√	√	√	√	√
	(omitted category: Quintile 1)										
Place of residence	Urban area	√	√	√	√	√	√	√	√	√	√
	(reference category: rural area)										
Subnational region	Country-specific categories	√	√	√	√	√	√	√	√	√	√
	(reference category: country specific)										

[a] In Afghanistan, Chad, the Democratic Republic of the Congo, Ethiopia and Kenya, the combined category of "secondary school or more" was applied due to small sample sizes of women with higher levels of education.
[b] Mother's ethnicity was applied in two countries: Democratic Republic of the Congo and Kenya; caste/tribe was applied in India.
[c] Sex of household head was not included for Afghanistan, because nearly all households (99%) were headed by males.

Estimation and inference

Logistic regressions were estimated by maximum likelihood. Standard errors were corrected for within-cluster correlation, and sample weights were applied to all estimators. No additional adjustment was made for stratification[3]. All regression outputs were obtained using the *logit* command with the *or* option in the statistical package STATA v.14.2.

The statistical significance of odds ratios was determined by the 95% confidence intervals and the p-value of the test of statistical significance. If the p-value was below 0.05, the association was considered statistically significant, for a p-value between 0.05 and 0.10, the association was considered to be nearly significant, and for a p-value larger than 0.10 the association was considered not significant. The 95% confidence interval provided additional information about the precision of the estimate: a large confidence interval indicated a low level of precision, while a small confidence interval denoted a high level of precision. Joint hypothesis testing was also used to test whether the characteristics with multiple subgroups (e.g. regions, etc.), taken as a whole, were significantly associated with DTP3 immunization coverage.

Interpreting the odds ratio: Nigeria example

Each exponentiated coefficient in the logistic model is the ratio of two odds (*odds ratio*): the ratio of the odds of immunization for a certain subgroup of an explanatory variable to the odds of immunization for the reference subgroup of that variable. For example, the odds ratio for urban residence is the ratio of the odds of children living in urban areas being immunized to the odds of children living in rural areas being immunized. An odds ratio greater than 1 indicates that a certain level of the explanatory variable is associated with higher odds of immunization; an odds ratio less than 1 indicates lower odds of immunization; and an odds ratio equal to 1 means equal odds of immunization.

To illustrate how odds ratios are interpreted, Table 4 presents the estimated odds ratios for DTP3 immunization in Nigeria using DHS data for 2013. For example, an odds ratio for female household head of 1.43 means that, holding all other factors constant, the odds of DTP3 immunization for children living in female headed households (female = 1) are 1.43 times as high as the odds for children living in male headed households (female = 0). For ordinal variables with more than two subgroups, such as wealth quintiles or levels of schooling, the odds are interpreted with reference to a base level (reference subgroup) that is usually the lowest or highest order subgroup. In Nigeria, the estimated odds ratio of 7.27 for quintile 5 indicates that the odds of immunization among children in quintile 5 are more than seven times larger than for children in quintile 1 (the reference subgroup). Similarly, the odds ratio for more than secondary schooling demonstrates that the odds of immunization for children with mothers with more than secondary schooling was 6.64 times higher than for children with uneducated mothers. For categorical variables that do not follow a particular order, like subnational region,

---

[3] Following O'Donnell et al. *(2)*, the conservative strategy of not adjusting for stratification (in addition to correcting for clustering effects) was followed; in particular, considering that most DHS surveys would have stratified samples based on exogenous geographical factors, indicating no need for adjustment.

the subgroup with the lowest odds ratio was used as the reference subgroup; therefore, all other subgroups report always odds ratios larger than one.

The parameter of 0.03 for the constant corresponds to the baseline odds of immunization for children in all the reference subgroups, i.e. male children in the poorest quintile, with uneducated mothers, aged 15-19, living in male headed households, born sixth or more, and from rural areas in the North West region of Nigeria.

When interpreting the odds ratios, those that are nearly statistically significant or not statistically significant (p-value >=0.05) should also be taken into account. In Nigeria, while the odds ratio for birth order and place of residence were nearly significant, the odds ratio for child's sex was not statistically significant, suggesting that DTP3 immunization was not associated with child's sex.

Compounded vulnerability and advantage

By multiplying odds ratios of two or more associated factors, it is possible to provide results of compounded vulnerability/advantage. In our example, the odds of immunization for children of mothers aged 20-34 years and with more than secondary schooling in Nigeria were more than 12 times (6.64*1.89 = 12.6) higher than for children of teenaged mothers with no education. If, in addition, children in the advantaged group belonged to the richest 20% of the population and children in the disadvantaged group belonged to the poorest 20%, the odds were more than 90 times higher (6.64*1.89*7.27 = 91.4).

The STATA codes used in all analyses are available upon request (contact: equity_monitoring@who.int).

Table 4. Estimated odds ratios for DTP3 immunization in Nigeria (DHS 2013)

	Odds Ratio	p-value	95% Confidence Interval	
Wealth quintile *(base = quintile 1)*				
Quintile 2	1.93	0.000	1.42	2.63
Quintile 3	3.03	0.000	2.18	4.23
Quintile 4	4.82	0.000	3.35	6.93
Quintile 5	7.27	0.000	4.80	11.00
Mother's education *(base = none)*				
Primary	1.84	0.000	1.41	2.41
Secondary	3.73	0.000	2.73	5.10
More than secondary	6.64	0.000	4.09	10.81
Mother's age at birth *(base = 15-19)*				
20-34	1.89	0.000	1.37	2.61
35-49	2.38	0.000	1.60	3.54
Sex of household head *(base = male)*				
Female	1.43	0.009	1.09	1.86
Child's sex *(base = male)*				
Female	0.94	0.435	0.79	1.11
Child's birth order *(base = 6th+ born)*				
1	1.38	0.072	0.97	1.95
2-3	1.02	0.889	0.77	1.34
4-5	0.92	0.487	0.71	1.18
Place of residence *(base = rural)*				
Urban	1.26	0.063	0.99	1.61
Region *(base = North West)*				
North Central	2.49	0.000	1.75	3.53
North East	1.45	0.050	1.00	2.10
Sout East	6.71	0.000	4.39	10.26
Sout South	3.28	0.000	2.18	4.94
South West	2.13	0.000	1.47	3.09
Constant	0.03	0.000	0.02	0.04
N	*5,834*			
Pseudo R2	*0.3475*			
Wald Chi-square for regression (df=20)	*803.05*			
p-value (Chi-squared for wealth effects =0)	*0.000*			
p-value (Chi-squared for education effects =0)	*0.000*			
p-value (Chi-squared for age effects =0)	*0.000*			
p-value (Chi-squared for birth order effects =0)	*0.059*			
p-value (Chi-squared for regional effects =0)	*0.000*			

References

1. The DHS Program. Demographic and Health Surveys [Internet]. Washington (DC): US Agency for International Development; 2018 (dhsprogram.com, accessed 21 May 2018).
2. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. Analyzing health equity using household survey data. A guide to techniques and their implementation. Washington (DC): World Bank; 2008.
3. Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva: World Health Organization; 2013:126.
4. Erreygers G. Correcting the concentration index. J Health Econ. 2009 March;28(2):504–15.
5. Erreygers G, Van Ourti T. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. J Health Econ. 2011 July;30(4):685–94.
6. O'Donnell O, O'Neill S, Van Ourti T, Walsh B. conindex: estimation of concentration indices. Stata J. 2016;16(1):112.
7. Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge, MA: MIT Press; 2002.
8. Rodríguez G. Lecture notes on generalized linear models [Internet]; 2007 (http://data.princeton.edu/wws509/notes/, accessed 29 June 2018).
9. Szklo M, Nieto FJ. Epidemiology: beyond the basics. Sudbury, MA: Jones and Bartlett Publishers; 2007.

Glossary

Absolute inequality reflects the magnitude of difference in health between subgroups. Absolute measures of inequality retain the same unit of measure as the health indicator.

Benchmarking is the process of comparing data from similar areas or populations to get an idea of how one area/population performs in relation to others. Benchmarking provides context for a broader understanding of the state of inequality.

A **binary health indicator** takes on the value of zero or one, depending on the absence or presence of a specified condition.

A **bounded health indicator** has both finite lower and upper bounds.

Complex measures of inequality draw on data from all subgroups to produce a single number that is an expression of the level of inequality. For example, they can express inequality across all wealth quintiles, or among all regions in a country.

Concentration curve provides a graphical representation of how a variable health indicator (such as DTP3 immunization coverage) is distributed across subgroups of an ordered inequality dimension (such as household economic status). The concentration curve plots the cumulative proportions of a health variable (y-axis) against the cumulative percentage of the population (or a representative sample), ranked by a measure of socioeconomic status, beginning with the most disadvantaged, and ending with the least disadvantaged (x-axis).

Concentration index is a measure of the degree of inequality in a health indicator over the distribution of an ordered inequality dimension. In this report, concentration indices were constructed to measure inequality in DTP3 immunization over the distribution of household wealth and mother´s education

Disaggregated estimates are data that are broken down by population subgroup (as opposed to overall average).

An **explanatory variable** is introduced in a regression model to explain variation in the dependent health variable.

Health indicators (variables) are narrowly-defined measures that summarize information about a health topic.

Health inequalities are observable health differences between subgroups within a population. Health inequalities can be measured and monitored.

Health inequity is a normative concept that describes systematic differences in health between population subgroups that are deemed to be unjust, unfair and avoidable. Health inequity is linked to forms of disadvantage that are socially produced, such as poverty, discrimination and lack of access to services or goods.

An **inequality dimension** is the categorization upon which subgroups are formed for health inequality monitoring, such as wealth, education, region, sex, etc. The selection of dimensions of inequality typically reflects categories that are reasonably likely to reflect unfair differences between groups that could be corrected by changes to policies, programmes or practices.

A **logit model** is a nonlinear model for a binary response where the response probability can be transformed using the logit function and modelled as a linear function of the explanatory variables.

Monitoring is a process of repeatedly observing a situation to watch for changes over time. While monitoring can help to determine the impact of policies, programmes and practices, monitoring alone cannot typically explain the cause of troublesome trends. Rather, monitoring may be thought of as a warning system. Monitoring activities can both inform and direct research in a given area. Because monitoring tracks progress over time, it can be described as a continual cycle.

Multicollinearity is a condition where there is a very high correlation between two or more variables included in a regression model, resulting in unstable estimates.

Multiple regression analysis is a type of analysis where a dependent variable can be affected by more than one factor simultaneously.

Non-ordered inequality dimensions are not based on criteria that can be logically ranked. For example, region, ethnicity and religion dimensions of inequality typically contain subgroups that are non-ordered.

The **odds ratio** is the ratio of two odds: the odds that an outcome will occur given a particular level of an explanatory variable to the odds of the outcome occurring at the reference level of that variable. In logistic regression, an odds ratio can be summarized as the change in likelihood for a unit increase in the explanatory variable; the exponential function of the estimated coefficient represents the odds ratio associated with a one unit increase in the explanatory variable).

Ordered inequality dimensions have an inherent positioning and can be logically ranked. For example, wealth and education level are dimensions of inequality that typically contain subgroups that can be ordered.

Population share describes the percentage of the population that is represented by a given population subgroup. In cases where the health indicator does not affect the entire population, population share expresses the percentage of the *affected* population represented by a given population subgroup. For example, if looking at service coverage among pregnant women, then population share would express the percentage of pregnant women in a given subgroup out of all pregnant women in the population.

Population subgroups, in the context of health inequality monitoring, reflect ways of grouping a population based on a dimension of inequality. For example, population subgroups based on wealth are commonly grouped as quintiles, ranging from the poorest 20% to the richest 20%.

Relative inequality shows the proportional differences in health among subgroups. Relative measures of inequality are unit-less.

Summary measures of inequality yield a single number that reflects the level of inequality between two or more subgroups. Summary measures of inequality may indicate absolute or relative inequality, and may involve two subgroups (that is, simple pairwise measures) or more than two subgroups (that is, complex measures). Summary measures of inequality may be weighted or unweighted.

An **unbounded health indicator** has at least one infinite bound.