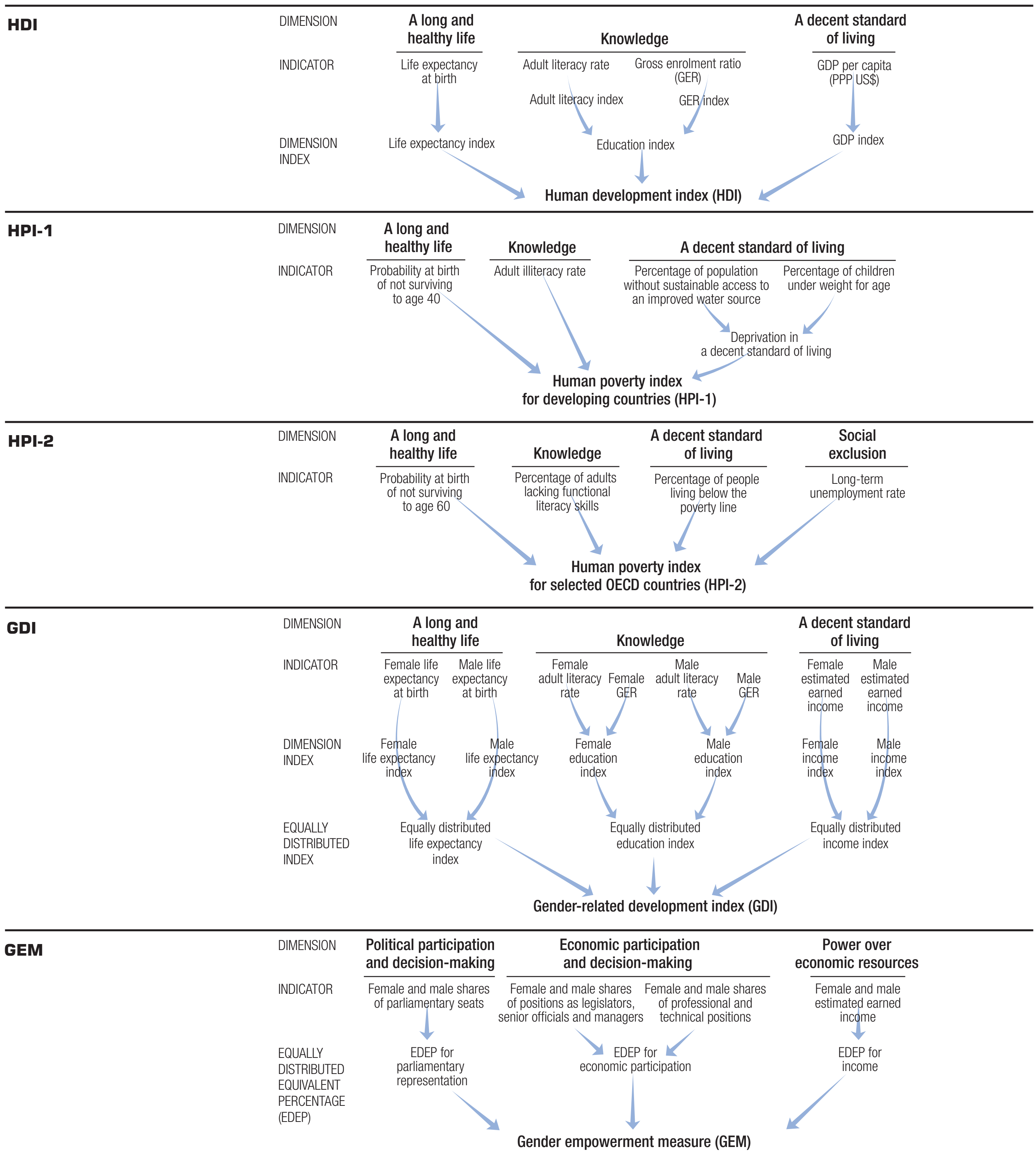
Calculating the human development indices

The diagrams here summarize how the five human development indices used in the *Human Development Report* are constructed, highlighting both their similarities and their differences. The text on the following pages provides a detailed explanation.

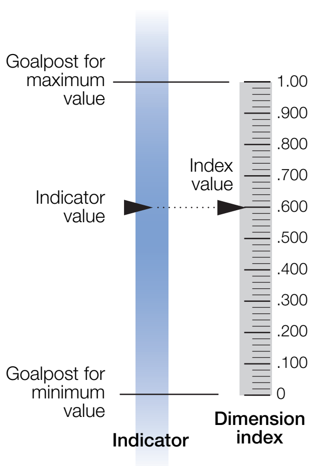


The human development index (HDI)

The HDI is a summary measure of human development. It measures the average achievements in a country in three basic dimensions of human development:

- A long and healthy life, as measured by life expectancy at birth.
- Knowledge, as measured by the adult literacy rate (with two-thirds weight) and the combined primary, secondary and tertiary gross enrolment ratio (with one-third weight).
- A decent standard of living, as measured by GDP per capita in purchasing power parity (PPP) terms in US dollars.

Before the HDI itself is calculated, an index needs to be created for each of these dimensions. To calculate these indices—the life expectancy, education and GDP indices—minimum and maximum values (goalposts) are chosen for each underlying indicator.



Performance in each dimension is expressed as a value between 0 and 1 by applying the following general formula:

$$\text{Dimension index} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}}$$

The HDI is then calculated as a simple average of the dimension indices. The box at right illustrates the calculation of the HDI for a sample country.

Goalposts for calculating the HDI

Indicator	Maximum value	Minimum value
Life expectancy at birth (years)	85	25
Adult literacy rate (%)	100	0
Combined gross enrolment ratio (%)	100	0
GDP per capita (PPP US$)	40,000	100

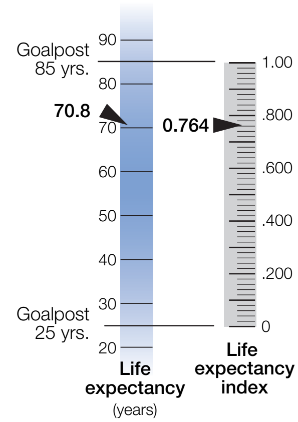
Calculating the HDI

This illustration of the calculation of the HDI uses data for Brazil.

1. Calculating the life expectancy index

The life expectancy index measures the relative achievement of a country in life expectancy at birth. For Brazil, with a life expectancy of 70.8 years in 2004, the life expectancy index is 0.764.

$$\text{Life expectancy index} = \frac{70.8 - 25}{85 - 25} = \mathbf{0.764}$$

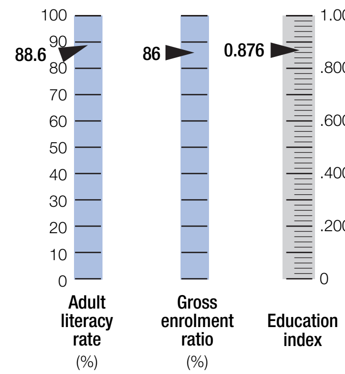


2. Calculating the education index

The education index measures a country's relative achievement in both adult literacy and combined primary, secondary and tertiary gross enrolment. First, an index for adult literacy and one for combined gross enrolment are calculated. Then these two indices are combined to create the education index, with two-thirds weight given to adult literacy and one-third weight to combined gross enrolment. For Brazil, with an adult literacy rate of 88.6% in 2004 and a combined gross enrolment ratio of 86% in 2004, the education index is 0.876.

$$\text{Adult literacy index} = \frac{88.6 - 0}{100 - 0} = 0.886$$

$$\text{Gross enrolment index} = \frac{86 - 0}{100 - 0} = 0.857$$

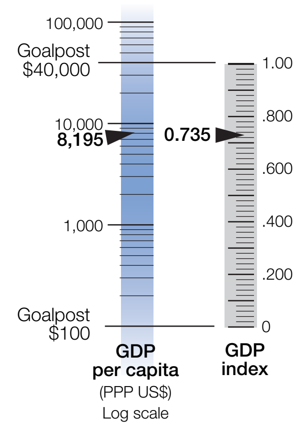


Education index = 2/3 (adult literacy index) + 1/3 (gross enrolment index)
= 2/3 (0.886) + 1/3 (0.857) = **0.876**

3. Calculating the GDP index

The GDP index is calculated using adjusted GDP per capita (PPP US$). In the HDI income serves as a surrogate for all the dimensions of human development not reflected in a long and healthy life and in knowledge. Income is adjusted because achieving a respectable level of human development does not require unlimited income. Accordingly, the logarithm of income is used. For Brazil, with a GDP per capita of $8,195 (PPP US$) in 2004, the GDP index is 0.735.

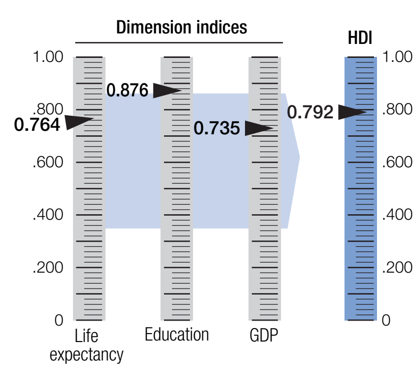
$$\text{GDP index} = \frac{\log(8{,}195) - \log(100)}{\log(40{,}000) - \log(100)} = \mathbf{0.735}$$



4. Calculating the HDI

Once the dimension indices have been calculated, determining the HDI is straightforward. It is a simple average of the three dimension indices.

HDI = 1/3 (life expectancy index) + 1/3 (education index)
+ 1/3 (GDP index)
= 1/3 (0.764) + 1/3 (0.876) + 1/3 (0.735) = **0.792**



The human poverty index for developing countries (HPI-1)

While the HDI measures average achievement, the HPI-1 measures *deprivations* in the three basic dimensions of human development captured in the HDI:

- A long and healthy life—vulnerability to death at a relatively early age, as measured by the probability at birth of not surviving to age 40.
- Knowledge—exclusion from the world of reading and communications, as measured by the adult illiteracy rate.
- A decent standard of living—lack of access to overall economic provisioning, as measured by the unweighted average of two indicators, the percentage of the population without sustainable access to an improved water source and the percentage of children under weight for age.

Calculating the HPI-1 is more straightforward than calculating the HDI. The indicators used to measure the deprivations are already normalized between 0 and 100 (because they are expressed as percentages), so there is no need to create dimension indices as for the HDI.

Originally, the measure of deprivation in a decent standard of living also included an indicator of access to health services. But because reliable data on access to health services are lacking for recent years, in this year's Report deprivation in a decent standard of living is measured by two rather than three indicators—the percentage of the population without sustainable access to an improved water source and the percentage of children under weight for age.

The human poverty index for selected OECD countries (HPI-2)

The HPI-2 measures deprivations in the same dimensions as the HPI-1 and also captures social exclusion. Thus it reflects deprivations in four dimensions:

- A long and healthy life—vulnerability to death at a relatively early age, as measured by the probability at birth of not surviving to age 60.
- Knowledge—exclusion from the world of reading and communications, as measured by the percentage of adults (ages 16–65) lacking functional literacy skills.
- A decent standard of living—as measured by the percentage of people living below the income poverty line (50% of the median adjusted household disposable income).
- Social exclusion—as measured by the rate of long-term unemployment (12 months or more).

Calculating the HPI-1

1. Measuring deprivation in a decent standard of living

An unweighted average of two indicators is used to measure deprivation in a decent standard of living.

Unweighted average = 1/2 (population without sustainable access to an improved water source)
+ 1/2 (children under weight for age)

A sample calculation: Namibia

Percentage of population without sustainable access to an improved water source = 13%
Percentage of children under weight for age = 24%

Unweighted average = 1/2 (13) + 1/2 (24) = 18.5%

2. Calculating the HPI-1

The formula for calculating the HPI-1 is as follows:

$$\text{HPI-1} = [1/3\,(P_1^{\alpha} + P_2^{\alpha} + P_3^{\alpha})]^{1/\alpha}$$

Where:
P_1 = Probability at birth of not surviving to age 40 (times 100)
P_2 = Adult illiteracy rate
P_3 = Unweighted average of population without sustainable access to an improved water source and children under weight for age
$\alpha = 3$

A sample calculation: Namibia

P_1 = 45.4%
P_2 = 15.0%
P_3 = 18.5%

$$\text{HPI-1} = [1/3\,(45.4^3 + 15.0^3 + 18.5^3)]^{1/3} = \mathbf{32.5}$$

Calculating the HPI-2

The formula for calculating the HPI-2 is as follows:

$$\text{HPI-2} = [1/4\,(P_1^{\alpha} + P_2^{\alpha} + P_3^{\alpha} + P_4^{\alpha})]^{1/\alpha}$$

Where:
P_1 = Probability at birth of not surviving to age 60 (times 100)
P_2 = Percentage of adults lacking functional literacy skills
P_3 = Percentage of population below income poverty line (50% of median adjusted household disposable income)
P_4 = Rate of long-term unemployment (lasting 12 months or more)
$\alpha = 3$

A sample calculation: Australia

P_1 = 7.7%
P_2 = 17.0%
P_3 = 14.3%
P_4 = 0.9%

$$\text{HPI-2} = [1/4\,(7.7^3 + 17.0^3 + 14.3^3 + 0.9^3)]^{1/3} = \mathbf{12.8}$$

Why $\alpha = 3$ in calculating the HPI-1 and HPI-2

The value of α has an important impact on the value of the HPI. If $\alpha = 1$, the HPI is the average of its dimensions. As α rises, greater weight is given to the dimension in which there is the most deprivation. Thus as α increases towards infinity, the HPI will tend towards the value of the dimension in which deprivation is greatest (for Namibia, the example used for calculating the HPI-1, it would be 45.4, equal to the probability at birth of not surviving to age 40).

In this Report the value 3 is used to give additional but not overwhelming weight to areas of more acute deprivation. For a detailed analysis of the HPI's mathematical formulation, see Sudhir Anand and Amartya Sen's "Concepts of Human Development and Poverty: A Multidimensional Perspective" and the technical note in *Human Development Report 1997* (see the list of selected readings at the end of this technical note).

The gender-related development index (GDI)

While the HDI measures average achievement, the GDI adjusts the average achievement to reflect the *inequalities* between men and women in the following dimensions:

- A long and healthy life, as measured by life expectancy at birth.
- Knowledge, as measured by the adult literacy rate and the combined primary, secondary and tertiary gross enrolment ratio.
- A decent standard of living, as measured by estimated earned income (PPP US$).

The calculation of the GDI involves three steps. First, female and male indices in each dimension are calculated according to this general formula:

$$\text{Dimension index} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}}$$

Second, the female and male indices in each dimension are combined in a way that penalizes differences in achievement between men and women. The resulting index, referred to as the equally distributed index, is calculated according to this general formula:

$$\text{Equally distributed index} = \{[\text{female population share (female index}^{1-\epsilon})] + [\text{male population share (male index}^{1-\epsilon})]\}^{1/1-\epsilon}$$

ϵ measures the aversion to inequality. In the GDI $\epsilon = 2$. Thus the general equation becomes:

$$\text{Equally distributed index} = \{[\text{female population share (female index}^{-1})] + [\text{male population share (male index}^{-1})]\}^{-1}$$

which gives the harmonic mean of the female and male indices.

Third, the GDI is calculated by combining the three equally distributed indices in an unweighted average.

Goalposts for calculating the GDI

Indicator	Maximum value	Minimum value
Female life expectancy at birth (years)	87.5	27.5
Male life expectancy at birth (years)	82.5	22.5
Adult literacy rate (%)	100	0
Combined gross enrolment ratio (%)	100	0
Estimated earned income (PPP US$)	40,000	100

Note: The maximum and minimum values (goalposts) for life expectancy are five years higher for women to take into account their longer life expectancy.

Calculating the GDI

This illustration of the calculation of the GDI uses data for Thailand.

1. Calculating the equally distributed life expectancy index

The first step is to calculate separate indices for female and male achievements in life expectancy, using the general formula for dimension indices.

FEMALE
Life expectancy: 74.0 years

$$\text{Life expectancy index} = \frac{74.0 - 27.5}{87.5 - 27.5} = 0.776$$

MALE
Life expectancy: 66.7 years

$$\text{Life expectancy index} = \frac{66.7 - 22.5}{82.5 - 22.5} = 0.737$$

Next, the female and male indices are combined to create the equally distributed life expectancy index, using the general formula for equally distributed indices.

FEMALE
Population share: 0.509
Life expectancy index: 0.776

MALE
Population share: 0.491
Life expectancy index: 0.737

$$\text{Equally distributed life expectancy index} = \{[0.509\ (0.776^{-1})] + [0.491\ (0.737^{-1})]\}^{-1} = \mathbf{0.756}$$

2. Calculating the equally distributed education index

First, indices for the adult literacy rate and the combined primary, secondary and tertiary gross enrolment ratio are calculated separately for females and males. Calculating these indices is straightforward, since the indicators used are already normalized between 0 and 100.

FEMALE
Adult literacy rate: 90.5%
Adult literacy index: 0.905
Gross enrolment ratio: 74.0%
Gross enrolment index: 0.740

MALE
Adult literacy rate: 94.9%
Adult literacy index: 0.949
Gross enrolment ratio: 73.4%
Gross enrolment index: 0.734

Second, the education index, which gives two-thirds weight to the adult literacy index and one-third weight to the gross enrolment index, is computed separately for females and males.

$$\text{Education index} = 2/3\ (\text{adult literacy index}) + 1/3\ (\text{gross enrolment index})$$

$$\text{Female education index} = 2/3\ (0.905) + 1/3\ (0.740) = 0.850$$

$$\text{Male education index} = 2/3\ (0.949) + 1/3\ (0.734) = 0.877$$

Finally, the female and male education indices are combined to create the equally distributed education index.

FEMALE
Population share: 0.509
Education index: 0.850

MALE
Population share: 0.491
Education index: 0.877

$$\text{Equally distributed education index} = \{[0.509\ (0.850^{-1})] + [0.491\ (0.877^{-1})]\}^{-1} = \mathbf{0.863}$$

3. Calculating the equally distributed income index

First, female and male earned income (PPP US$) are estimated (for details on this calculation, see the addendum to this technical note). Then the income index is calculated for each gender. As for the HDI, income is adjusted by taking the logarithm of estimated earned income (PPP US$):

$$\text{Income index} = \frac{\log(\text{actual value}) - \log(\text{minimum value})}{\log(\text{maximum value}) - \log(\text{minimum value})}$$

FEMALE
Estimated earned income (PPP US$): 6,036

$$\text{Income index} = \frac{\log(6{,}036) - \log(100)}{\log(40{,}000) - \log(100)} = 0.684$$

MALE
Estimated earned income (PPP US$): 10,214

$$\text{Income index} = \frac{\log(10{,}214) - \log(100)}{\log(40{,}000) - \log(100)} = 0.772$$

Calculating the GDI continues on next page

Calculating the GDI (continued)

Second, the female and male income indices are combined to create the equally distributed income index:

FEMALE	MALE
Population share: 0.509	Population share: 0.491
Income index: 0.684	Income index: 0.772

$$\text{Equally distributed income index} = \{[0.509\ (0.684^{-1})] + [0.491\ (0.772^{-1})]\}^{-1} = \mathbf{0.725}$$

4. Calculating the GDI

Calculating the GDI is straightforward. It is simply the unweighted average of the three component indices—the equally distributed life expectancy index, the equally distributed education index and the equally distributed income index.

$$\begin{aligned}\text{GDI} &= 1/3\ (\text{life expectancy index}) + 1/3\ (\text{education index}) + 1/3\ (\text{income index}) \\ &= 1/3\ (0.756) + 1/3\ (0.863) + 1/3\ (0.725) = \mathbf{0.781}\end{aligned}$$

Why $\epsilon = 2$ in calculating the GDI

The value of ϵ is the size of the penalty for gender inequality. The larger the value, the more heavily a society is penalized for having inequalities.

If $\epsilon = 0$, gender inequality is not penalized (in this case the GDI would have the same value as the HDI). As ϵ increases towards infinity, more and more weight is given to the lesser achieving group.

The value 2 is used in calculating the GDI (as well as the GEM). This value places a moderate penalty on gender inequality in achievement.

For a detailed analysis of the GDI's mathematical formulation, see Sudhir Anand and Amartya Sen's "Gender Inequality in Human Development: Theories and Measurement," Kalpana Bardhan and Stephan Klasen's "UNDP's Gender-Related Indices: A Critical Review" and the technical notes in *Human Development Report 1995* and *Human Development Report 1999* (see the list of selected readings at the end of this technical note).

The gender empowerment measure (GEM)

Focusing on women's opportunities rather than their capabilities, the GEM captures gender inequality in three key areas:

- Political participation and decision-making power, as measured by women's and men's percentage shares of parliamentary seats.
- Economic participation and decision-making power, as measured by two indicators—women's and men's percentage shares of positions as legislators, senior officials and managers and women's and men's percentage shares of professional and technical positions.
- Power over economic resources, as measured by women's and men's estimated earned income (PPP US$).

For each of these three dimensions, an equally distributed equivalent percentage (EDEP) is calculated, as a population-weighted average, according to the following general formula:

$$\text{EDEP} = \{[\text{female population share (female index}^{1-\epsilon})] + [\text{male population share (male index}^{1-\epsilon})]\}^{1/1-\epsilon}$$

ϵ measures the aversion to inequality. In the GEM (as in the GDI) $\epsilon = 2$, which places a moderate penalty on inequality. The formula is thus:

$$\text{EDEP} = \{[\text{female population share (female index}^{-1})] + [\text{male population share (male index}^{-1})]\}^{-1}$$

For political and economic participation and decision-making, the EDEP is then indexed by dividing it by 50. The rationale for this indexation: in an ideal society, with equal empowerment of the sexes, the GEM variables would equal 50%—that is, women's share would equal men's share for each variable.

Where a male or female index value is zero, the EDEP according to the above formula is not defined. However, the limit of EDEP, when the index tends towards zero, is zero. Accordingly, in these cases the value of the EDEP is set to zero.

Finally, the GEM is calculated as a simple average of the three indexed EDEPs.

Calculating the GEM

This illustration of the calculation of the GEM uses data for Argentina.

1. Calculating the EDEP for parliamentary representation

The EDEP for parliamentary representation measures the relative empowerment of women in terms of their political participation. The EDEP is calculated using the female and male shares of the population and female and male percentage shares of parliamentary seats according to the general formula.

FEMALE	MALE
Population share: 0.511	Population share: 0.489
Parliamentary share: 36.5%	Parliamentary share: 63.5%

$$\text{EDEP for parliamentary representation} = \{[0.511\ (36.5^{-1})] + [0.489\ (63.5^{-1})]\}^{-1} = 46.07$$

Then this initial EDEP is indexed to an ideal value of 50%.

$$\text{Indexed EDEP for parliamentary representation} = \frac{46.07}{50} = \mathbf{0.921}$$

2. Calculating the EDEP for economic participation

Using the general formula, an EDEP is calculated for women's and men's percentage shares of positions as legislators, senior officials and managers, and another for women's and men's percentage shares of professional and technical positions. The simple average of the two measures gives the EDEP for economic participation.

FEMALE	MALE
Population share: 0.511	Population share: 0.489
Percentage share of positions as legislators, senior officials and managers: 25.4%	Percentage share of positions as legislators, senior officials and managers: 74.6%
Percentage share of professional and technical positions: 54.7%	Percentage share of professional and technical positions: 45.3%

$$\text{EDEP for positions as legislators, senior officials and managers} = \{[0.511\ (25.4^{-1})] + [0.489\ (74.6^{-1})]\}^{-1} = 37.46$$

$$\text{Indexed EDEP for positions as legislators, senior officials and managers} = \frac{37.46}{50} = 0.749$$

$$\text{EDEP for professional and technical positions} = \{[0.511\ (54.7^{-1})] + [0.489\ (45.3^{-1})]\}^{-1} = 49.67$$

$$\text{Indexed EDEP for professional and technical positions} = \frac{49.67}{50} = 0.993$$

The two indexed EDEPs are averaged to create the EDEP for economic participation:

$$\text{EDEP for economic participation} = \frac{0.749 + 0.993}{2} = \mathbf{0.871}$$

3. Calculating the EDEP for income

Earned income (PPP US$) is estimated for women and men separately and then indexed to goalposts as for the HDI and the GDI. For the GEM, however, the income index is based on unadjusted values, not the logarithm of estimated earned income. (For details on the estimation of earned income for men and women, see the addendum to this technical note.)

FEMALE	MALE
Population share: 0.511	Population share: 0.489
Estimated earned income (PPP US$): 9,258	Estimated earned income (PPP US$): 17,518
Income index = $\frac{9{,}258 - 100}{40{,}000 - 100} = 0.230$	Income index = $\frac{17{,}518 - 100}{40{,}000 - 100} = 0.437$

The female and male indices are then combined to create the equally distributed index:

$$\text{EDEP for income} = \{[0.511\ (0.230^{-1})] + [0.489\ (0.437^{-1})]\}^{-1} = \mathbf{0.299}$$

4. Calculating the GEM

Once the EDEP has been calculated for the three dimensions of the GEM, determining the GEM is straightforward. It is a simple average of the three EDEP indices.

$$\text{GEM} = \frac{0.921 + 0.871 + 0.299}{3} = \mathbf{0.697}$$

TECHNICAL NOTE 1 ADDENDUM

Female and male earned income

Despite the importance of having gender-disaggregated data on income, direct measures are unavailable. For this Report crude estimates of female and male earned income have therefore been derived.

Income can be seen in two ways: as a resource for consumption and as earnings by individuals. The use measure is difficult to disaggregate between men and women because they share resources within a family unit. By contrast, earnings are separable because different members of a family tend to have separate earned incomes.

The income measure used in the GDI and the GEM indicates a person's capacity to earn income. It is used in the GDI to capture the disparities between men and women in command over resources and in the GEM to capture women's economic independence. (For conceptual and methodological issues relating to this approach, see Sudhir Anand and Amartya Sen's "Gender Inequality in Human Development" and, in *Human Development Report 1995,* chapter 3 and technical notes 1 and 2; see the list of selected readings at the end of this technical note.)

Female and male earned income (PPP US$) are estimated using the following data:

- Ratio of the female nonagricultural wage to the male nonagricultural wage.
- Male and female shares of the economically active population.
- Total female and male population.
- GDP per capita (PPP US$).

Key

W_f / W_m = ratio of female nonagricultural wage to male nonagricultural wage
EA_f = female share of economically active population
EA_m = male share of economically active population
S_f = female share of wage bill
Y = total GDP (PPP US$)
N_f = total female population
N_m = total male population
Y_f = estimated female earned income (PPP US$)
Y_m = estimated male earned income (PPP US$)

Note

Calculations based on data in the technical note may yield results that differ from those in the indicator tables because of rounding.

Estimating female and male earned income

This illustration of the estimation of female and male earned income uses 2004 data for the Netherlands.

1. Calculating total GDP (PPP US$)

Total GDP (PPP US$) is calculated by multiplying the total population by GDP per capita (PPP US$).

Total population: 16,282 (thousand)
GDP per capita (PPP US$): 31,789
Total GDP (PPP US$) = 16,282 (31,789) = 517,586,944 (thousand)

2. Calculating the female share of the wage bill

Because data on wages in rural areas and in the informal sector are rare, the Report has used nonagricultural wages and assumed that the ratio of female wages to male wages in the nonagricultural sector applies to the rest of the economy. The female share of the wage bill is calculated using the ratio of the female nonagricultural wage to the male nonagricultural wage and the female and male percentage shares of the economically active population. Where data on the wage ratio are not available, a value of 75% is used.

Ratio of female to male nonagricultural wage (W_f/W_m) = 0.815
Female percentage share of economically active population (EA_f) = 44.0%
Male percentage share of economically active population (EA_m) = 56.0%

$$\text{Female share of wage bill } (S_f) = \frac{W_f/W_m\,(EA_f)}{[W_f/W_m\,(EA_f)] + EA_m} = \frac{0.815\,(44.0)}{[0.815\,(44.0)] + 56.0} = \mathbf{0.391}$$

3. Calculating female and male earned income (PPP US$)

An assumption has to be made that the female share of the wage bill is equal to the female share of GDP.

Female share of wage bill (S_f) = 0.391
Total GDP (PPP US$) ($Y$) = 517,586,944 (thousand)
Female population (N_f) = 8,202 (thousand)

$$\text{Estimated female earned income (PPP US\$) } (Y_f) = \frac{S_f\,(Y)}{N_f} = \frac{0.391\,(517{,}586{,}944)}{8{,}202} = \mathbf{24{,}652}$$

Male population (N_m) = 8,080 (thousand)

$$\text{Estimated male earned income (PPP US\$) } (Y_m) = \frac{Y - S_f\,(Y)}{N_m} = \frac{517{,}586{,}944 - [0.391\,(517{,}586{,}944)]}{8{,}080} = \mathbf{39{,}035}$$

Selected readings

Anand, Sudhir, and Amartya Sen. 1994. "Human Development Index: Methodology and Measurement." Occasional Paper 12. United Nations Development Programme, Human Development Report Office, New York. *(HDI)*

———. **1995.** "Gender Inequality in Human Development: Theories and Measurement." Occasional Paper 19. United Nations Development Programme, Human Development Report Office, New York. *(GDI, GEM)*

———. **1997.** "Concepts of Human Development and Poverty: A Multi-dimensional Perspective." In United Nations Development Programme, *Human Development Report 1997 Papers: Poverty and Human Development.* New York. *(HPI-1, HPI-2)*

Bardhan, Kalpana, and Stephan Klasen. 1999. "UNDP's Gender-Related Indices: A Critical Review." *World Development* 27 (6): 985–1010. *(GDI, GEM)*

United Nations Development Programme. 1995. *Human Development Report 1995.* New York: Oxford University Press. Technical notes 1 and 2 and chapter 3. *(GDI, GEM)*

———. **1997.** *Human Development Report 1997*. New York: Oxford University Press. Technical note 1 and chapter 1. *(HPI-1, HPI-2)*

———. **1999.** *Human Development Report 1999*. New York: Oxford University Press. Technical note. *(HDI, GDI)*

TECHNICAL NOTE 2

A human development index by income groups

The human development index (HDI) provides a composite snapshot of the national average of three important indicators of human well-being (see *Technical note 1*). But it does not capture variations around the average linked to inequality. This year's Report presents for the first time an HDI by income quintiles. The new measure, intended both to address a major human development issue and to stimulate discussion, points to large inequalities between rich and poor in many countries.

The HDI by income quintiles disaggregates performance by income quintile for 15 countries. Full details of the methodology used are in a background paper prepared for this year's Report (Grimm and others 2006). This technical note provides a brief summary.

Methodology

Construction of the HDI by income quintiles follows the same procedure as for the standard HDI. Life expectancy, school enrolment, literacy and income per capita data from household surveys are used to calculate the three dimension indices—health, education and income—by income quintile.

Data for the index are drawn from a variety of sources. For developing countries household income surveys are used to calculate the education and gross domestic product (GDP) indices for each quintile, and Demographic and Health Surveys are used to calculate the life expectancy index. Because the two data sets do not cover the same households, the information from the surveys is linked by approximating income for households in the Demographic and Health Surveys using variables that are available in both sets of surveys. The correlation between household income per capita and a set of household characteristics available in both surveys is estimated and used to generate a proxy for the income of households in the Demographic and Health Surveys. These characteristics include household structure, education and age of the household head, area of residence, housing characteristics and the like.

For the two developed countries in the study, Finland and the United States, GDP and education data are from the Luxembourg Income Study, and income and life expectancy data are from published empirical work.

Data for the construction of the index are derived as follows.

Life expectancy

Calculations are based on infant mortality data from Demographic and Health Surveys. Infant mortality has proven a reliable proxy for overall mortality patterns and thus for life expectancy. Infant mortality rates for each income quintile are applied to Ledermann model life tables (a tool for estimating life expectancy based on the historical relationship between life expectancy and infant mortality).

The education index

The education index is based on adult literacy and school enrolment data. Adult literacy data are available directly from the household income surveys for each income quintile. To calculate the quintile-specific gross enrolment index, the combined gross enrolment ratio for each quintile is calculated. Each individual ages

The work on the human development index by income group was undertaken by Michael Grimm, Kenneth Harttgen, Stephan Klasen and Mark Misselhorn, with inputs from Teresa Munzi and Tim Smeeding from the Luxembourg Income Study team.

5–23 attending school or university, whether general or vocational, is considered enrolled. The quintile-specific gross enrolment index is then calculated using the same minimum and maximum values that are used in calculating the standard HDI.

GDP index

The GDP index is calculated using the income variable from the household income survey. For conceptual reasons and because of measurement errors, mean income per capita calculated from the household income surveys can be very different from GDP per capita from national accounts data, which are used to calculate the GDP index in the standard HDI. To eliminate differences in national price levels, household income per capita calculated from the household income surveys is expressed in US dollars in purchasing power parity (PPP) terms using conversion factors based on price data from the latest International Comparison Program surveys provided by the World Bank. This income per capita is then rescaled using the ratio between the household income variable and GDP per capita expressed in PPP (taken from the standard HDI).

Finally, these data are rescaled to the same average as that of the standard HDI for the relevant year. The HDI by income quintiles is then calculated according to the standard formula (see *Technical note 1*):

$$\frac{\text{Life expectancy index} + \text{education index} + \text{GDP index}}{3} = \text{Human development index}$$

This calculation is carried out for each quintile.

Issues for discussion

The HDI by income quintiles exercise provides a simple, intuitive and transparent approach for measuring important human development disparities within countries. It provides a useful composite indicator for tracking inequalities in income and wider inequalities in opportunity linked to health and education. However, the use of the HDI model to examine national inequalities raises a number of conceptual and methodological problems.

Consider first the relationship between income and the other indicators. The HDI by income quintiles measures annual incomes, which fluctuate considerably due to shocks and to lifecycle developments. Taking an annual average snapshot of the income of a household in, say, the poorest quintile can obscure very large dynamic changes over time. This produces additional methodological problems, not least because linking more stable health and education outcomes to fluctuating incomes can bias the results.

Data quality in the household surveys presents another set of problems. These problems are addressed here by the simplifying assumptions outlined above and explained in more detail in Grimm and others (2006). But aligning demographic and health survey and household income survey data is inherently problematic, and other approaches are possible. For developed countries, data quality is a less immediate problem. But cross-country comparisons remain difficult. In the case of Finland and the United States the assessment of life expectancy by income groups is based on data for the early 1990s linked to current incomes. However, data constraints mean that the income measure differs from that used for the other two components. In addition, Luxembourg Income Study data do not contain enrolment data, which must then be proxied by attainment data.

One final concern relates to the scale of inequality. In proportionate terms, differences between the rich and poor are much larger in the income dimension than in the health and education dimension. Arguably, smaller differences in health and education might, however, be just as important from a human development point of view and should therefore attract a greater weight in the HDI by income quintiles than they currently have. These are broader methodological issues inherent in such composite indices that will be investigated in future Reports.

TECHNICAL NOTE 3

Measuring risk in lack of access to water and sanitation

Access to water and sanitation is a matter of life and death. But what are the parameters of risk associated with not having access? Given the scale of illness and death associated with the problem, that question has received surprisingly little attention.

Chapter 1 sets out the results of a research exercise looking at the risks associated with deprivation in access to water and sanitation. The approach borrows from analytical techniques used in medical and economic research to examine the relationship between behaviour or treatment and health outcomes. It focuses on the association between access to specific types of water and sanitation infrastructure and changes in the risk of illness or premature death. More specifically, the exercise captures how access to water and sanitation affects the risk of neonatal (0–1 months) and post-neonatal (1–12 months) mortality, as well as the risk of diarrhoea, the leading water-related cause of death in children.

Data

Data for the research are derived from Demographic and Health Surveys, which collect information on a wide set of socioeconomic variables at the individual, household and community levels and are usually conducted every five years to allow comparison over time. Each survey sample consists of 5,000–30,000 households. The samples are not longitudinal by design, but they are representative at the national, urban and rural levels. Although Demographic and Health Surveys' primary focus is women ages 15–49, they also collect information on several demographic indicators for all members of the household, including children.

Some 22 surveys from 18 countries were used to construct the data set (table 1). Surveys conducted in or since 2000 were used in most cases to include the most recent information available. For the analysis here, children were the primary unit of analysis.

Methodology

The methodology follows a two-step approach. First, the elements that affect the chance of survival in different stages of life were identified, disentangling the effects of individual, household and community characteristics that contribute to mortality and illness. For neonatal mortality the main variable was defined as a discrete indicator with two values: zero if the child is alive and one if the child died during the first month of life. For diarrhoea a discrete outcome approach was used, with a one indicating a diarrhoeal episode within the two weeks

Table 1 Country coverage

Country	Year	Sample size
Bangladesh	1999–2000	6,368
Benin	2001	5,349
Cameroon	2004	8,125
Egypt	1995 2000	12,135 11,467
Ethiopia	2000	10,873
Gabon	2000	4,405
Ghana	2003	3,844
Guatemala	1998–99	4,943
Haiti	2000	6,685
Indonesia	2002-03	16,206
Mali	2001	13,097
Morocco	2003–04	6,180
Nepal	2001	6,931
Nicaragua	2001	6,986
Peru	1996 2000	17,549 13,697
Uganda	2000–01	7,113
Viet Nam	1997 2002	1,775 1,317
Zambia	2001–02	6,877
Zimbabwe	1999	3,643

prior to the interview. A logit model was then estimated in both cases (box 1).

A different model and different outcome variable were used to estimate the impact of specific elements on post-neonatal survival. All children older than one month were included, with the outcome variable indicating the occurrence of death between the 2nd and 11th months of life. A Cox proportional hazard model was then used to estimate the chances of survival.

At each step a set of control variables was used to identify the effects of specific characteristics. The control variables include individual variables (such as the sex of the child, birth intervals and whether the child was breastfed), household variables (such as type of dwelling, education of the mother and wealth of the household as measured by an asset index) and community-level variables (such as urban or rural, region of residence and so on). A regression analysis was then conducted to isolate the specific risks associated with each type of sanitation and water facility, using the absence of water and sanitation infrastructure as the reference scenario.

Typically, the wealth of households is measured by a standard asset index, which measures possessions such as vehicles and televisions as well as access to water and sanitation. Since the main interest of the study is the effect of water and sanitation infrastructure on health outcomes, an asset index that excludes these variables was constructed. Following standard procedures, eight household assets were included to calculate the first principal component, which was then used to construct a standardized index. This index was then used to divide households into wealth quintiles.

Finally, the robustness of the research was further tested. In particular, the mortality study was expanded using propensity score matching to check for endogeneity of the outcome variable or unobserved characteristics that may be correlated with access to water and sanitation.

Most of the results are shown and discussed in chapter 1. For further details, refer to the background papers prepared for this year's Report by Fuentes, Pfütze and Seck.[1]

Note

1 Fuentes, Pfütze and Seck 2006a, 2006b.

Box 1 Technical model for measuring risk

Two basic statistical methods were used to capture the risk underlying access to water and sanitation.

For neonatal mortality and incidence of diarrhoea, a standard logit model was used. Logit estimations are used when the outcome variable has two possible values (thus logits are often referred to as binary models). The two possible outcomes are labelled as failure ($Y = 0$) or success ($Y = 1$).

Parameters in logit estimations can be interpreted as the change in probability associated with a unit increase in the independent variables. The resulting parameters thus show the change in probability of the event conditional on the individual, household and community characteristics.

Formally, in the logit model the dependant variable Y_i is assumed to follow a Bernoulli distribution conditional on the vector of explanatory variable $\boldsymbol{X}_i$. The probability of success is written as

$$P(Y_i = 1|x_i) = \Lambda(x_i\beta) \text{ and } P(Y_i = 0|x_i) = 1 - \Lambda(x_i\beta)$$

with $\Lambda(z) = (1 + \exp^{-z})^{-1}$ being the cumulative distribution function of the logistic model.

The conditional density can be written as

$$f(y_i|x_i) = \Lambda(x_i\beta)^{y_i}[1 - \Lambda(x_i\beta)]^{1-y_i}.$$

The log likelihood function becomes

$$l(\beta) = \sum_{i=1}^{n} \log f(y_i|x_i) = \sum_{y_i=1} \log \Lambda(x_i\beta) + \sum_{y_i=0} \log[1 - \Lambda(x_i\beta)].$$

The maximum likelihood estimate $\hat{\beta}$ of β is the value that maximizes the log likelihood function $l(\beta)$.

For the determinant factors in post-neonatal mortality a more elaborate estimation framework is needed because of the problem of censored observations. The data used do not contain observations for the entire period of analysis for all children. For example, a child who is four months old at the time of the interview and dies at the age of five months will not be recorded by the survey as a death; this characteristic creates a bias that needs to be corrected. One way to address this problem is to restrict the sample to children who were at least 12 months old at the time of the interview. However, this would eliminate a considerable number of observations. Instead, a hazard model is used to account for censoring issues. Based on the extensive literature on mortality, a Cox proportional hazard model is applied. The model is a semi-parametric estimation, given that the underlying hazard rate is not modelled by some functional form. This model has only one requisite structural assumption: the effect of the covariates on the relative hazard rate must be constant over the period under consideration.

Formally, the (conditional) hazard function of the Cox model given a k-dimensional vector of covariates ($\boldsymbol{X}$) can be written as

$$\lambda(t|X) = \lambda_0(t)\exp(\beta' X) ,$$

where $\beta' = (\beta_1, \beta_2, ..., \beta_k)'$ is the vector of parameters (proportional change in the hazard function) and $\lambda_0(t)$ is the baseline hazard function.

The parameters β' can be estimated without estimating $\lambda_0(t)$ using maximum likelihood. If i denotes the index of ordered failure times t_i $i = (1, 2, ..., N)$, d_i the number of observations that fail at t_i, D_i the set of observations at t_i and R_i the risk set, the partial log likelihood function can be written as

$$l(\beta) = \sum_{i=1}^{N} d_i [\beta' X_i - \ln \sum_{j \in R_i} \exp(\beta' X_j)].$$