



2016

Health and development

An explicit approach to the design and assessment of country classification frameworks

Final report for the Equitable Access Initiative

January 7, 2016

Trygve Ottersen

Karen Grépin

Klara Henderson

Crossley Beth Pinkstaff

Charles Clift

Ole Frithjof Norheim

John-Arne Røttingen



Norwegian Institute of Public Health

Health and development

**An explicit approach to the design
and assessment of country classification
frameworks**

Final report for the Equitable Access Initiative

January 7, 2016

Trygve Ottersen^{a, b}

Karen Grépin^c

Klara Henderson^d

Crossley Beth Pinkstaff^c

Charles Clift^d

Ole Frithjof Norheim^b

John-Arne Røttingen^a

^a Norwegian Institute of Public Health, ^b University of Bergen, ^c New York University, ^d independent consultant

With research assistance from Usman Munir (Harvard University)

Developed by the Norwegian Institute of Public Health
Department of International Public Health
January 2016

Title:

Health and development
An explicit approach to the design
and assessment of country classification frameworks
Final report for the Equitable Access Initiative
January 7, 2016

Authors:

Trygve Ottersen^{a, b}
Karen Grépin^c
Klara Henderson^d
Crossley Beth Pinkstaff^c
Charles Clift^d
Ole Frithjof Norheim^b
John-Arne Røttingen^a

^a Norwegian Institute of Public Health, ^b University of Bergen,
^c New York University, ^d independent consultant

Ordering:

Only available in PDF format.

Cover layout:

Per Kristian Svendsen

Cover illustration:

Colourbox.com

Table of contents

Brief summary.....	4
Executive summary.....	6
1. The problem	12
2. NIPH's explicit approach	14
3. Findings from interviews and survey.....	23
4. The starting point: GNI per capita.....	33
5. Health needs	35
6. Domestic capacity	43
7. Expected impact.....	49
8. Other frameworks.....	53
9. Broad comparison of frameworks	58
10. Key choices for any institution.....	63
11. Use of frameworks.....	64
12. Limitations and next steps	66
13. Overall findings and recommendations	70
Appendix A: Disease-specific frameworks.....	72
Appendix B: Spearman correlation coefficients	74
Appendix C: Data sources and characteristics	76
Appendix D: List of supplementary material	80
References.....	81

Brief summary

Epidemiological, economic, and institutional transitions are challenging current country classifications in health and development. This has motivated the Equitable Access Initiative (EAI) to seek a new classification framework.

The Norwegian Institute of Public Health (NIPH) was commissioned by the EAI to provide input to the Initiative. The objective of the NIPH team has been to explore frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. These include decisions regarding eligibility, allocation, transition, and sustainability. This report summarizes the following six main outputs of our work.

Explicit step-by-step approach

We developed a step-by-step approach to the design and assessment of frameworks. This approach was intended to be intuitive and particularly explicit about inherent value choices. The approach integrates important concerns about health needs, the domestic capacity to address needs, and the expected impact of external financing.

Responses from the interviews and survey

We interviewed twenty stakeholders about their views on classification frameworks. Respondents believed that such frameworks have a potentially useful yet circumscribed role. Stakeholders considered needs and health needs the most important concern for a classification framework, followed by inequality, expected impact, and domestic capacity. This overlapped significantly with findings from an online survey of 285 stakeholders in which health inequality and burden of disease were seen as the most important country characteristics and income per capita as the least important characteristic.

Broad menu of frameworks

We designed twenty-eight frameworks through a uniform methodology. Most frameworks integrate gross national income per capita (GNIPC) and one other indicator into an index. Most of these indicators were closely related to either health needs, domestic capacity, or expected impact.

Head-on comparison of the implications

We modeled how the ranking of countries changes as one moves away from GNIPC alone to another framework. We did this both for individual countries and country groupings. We also calculated Spearman coefficients for correlations among frameworks.

List of key choices for any donor

Based on insights gained over the course of this study, we have articulated the key choices every donor will need to make when designing its classification framework.

Overall findings and recommendations

We found that addressing health needs is generally considered a key objective of external financing for health. We also found that including a health measure in the framework makes a significant difference for many countries and country categories and that it specifically improves the ranking of countries with large unmet health needs. Against this background, we recommend integrating GNIpc and one health-needs indicator for the initial classification of countries.

We found that which health-need indicator is chosen makes less of a difference than whether or not such an indicator is included at all. Moreover, no health-need indicator appears to be clearly superior to all others. We found that life expectancy and disability-adjusted life year rate (DALYR) are both good candidates for being included in a classification framework. In particular, we recommend that institutions at least consider using life expectancy as a generic health-need indicator.

We found that a range of other indicators was relevant for decisions regarding external financing for health, but none of these indicators was clearly suited for inclusion in the classification framework when judged against the following criteria: relevance, salience, validity, consistency, and timeliness and availability. Against this background, we recommend using a simple framework that integrates GNIpc and one health-need indicator and considering other indicators after this framework has been applied. In particular, we recommend supplementing the simple framework with a more discretionary assessment of within-country inequalities in health and income to the extent permitted by the data. We also recommend considering a range of indicators for domestic capacity and expected impact.

We found that nearly any move beyond GNIpc alone resulted in a lower rank for the group of low-income countries (LICs) and a higher rank for the group of middle-income countries (MICs). To the extent that this is undesirable, we recommend considering counteracting measures.

Although these general findings and recommendations can be informative, the optimal indicators and framework for any given institution eventually depend on organization-specific needs and objectives.

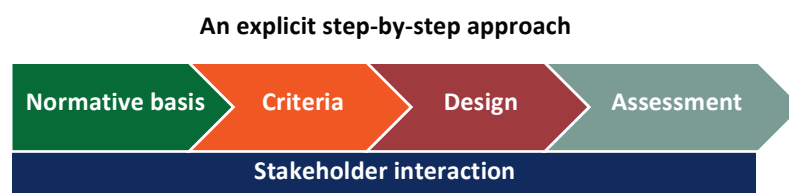
Executive summary

Epidemiological, economic, and institutional transitions are challenging current country classifications in health and development. This has motivated the Equitable Access Initiative (EAI) to seek a new classification framework.

The Norwegian Institute of Public Health (NIPH) was commissioned by the EAI to provide input to the Initiative. The objective of the NIPH team has been to explore frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. These include decisions regarding eligibility, allocation, transition, and sustainability. The six main outputs of our work are (1) an explicit step-by-step approach to the design and assessment of classification frameworks, (2) responses from interviews and a survey, (3) a broad menu of simple frameworks, (4) a head-on comparison of their implications, (5) a list of key choices for any donor, and (6) overall findings and recommendations for choice of indicators and use of frameworks.

Approach

We developed an explicit step-by-step approach to the design and assessment of frameworks. This approach is particularly explicit about the value choices inherent in framework design and assessment, and it emphasizes three concerns: health needs, domestic capacity to address needs, and the expected impact of external financing. The approach also seeks to be intuitive, inclusive in its design and assessment of a wide range of frameworks, and interactive in its use of stakeholder interviews and a stakeholder survey alongside consultations. Accordingly, we see this approach as an output in itself, and the EAI, the Convening Partners, and other institutions may use elements of the approach irrespective of whether they agree with the particular choices of the NIPH team.



Responses

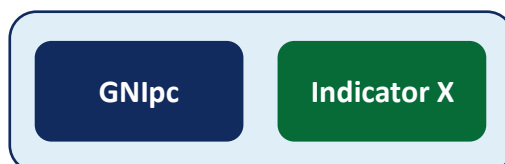
We interviewed twenty stakeholders about their views on classification frameworks devised to guide decisions on external financing for health. Such frameworks were considered potentially useful as an initial guide to decision making regarding external financing, provided they are simple and transparent. Respondents saw gross national income per capita (GNIpc) as relevant, but emphasized the need to go beyond this and consider additional factors. They identified measures of health needs and inequalities as the most important indicators for inclusion in a classification framework alongside GNIpc. Domestic capacity and expected impact were also considered relevant for external financing for health, but several stakeholders believed these concerns may better be taken into account at decision-making stages other than the initial classification of countries.

We also conducted an online survey with a discrete choice experiment, to which 285 stakeholders responded. Respondents were presented with a series of pairs of frameworks that differed in terms of how much importance each framework placed on four country characteristics: income per capita, burden of disease, strength of health system, and health inequalities. The survey findings suggested that most respondents want to prioritize countries with large health inequalities and burden of disease, which supported the interview findings. The survey further suggested that respondents do not prefer frameworks that put substantial weight on income per capita and that preferences depend on respondents' country of birth.

Menu

We generated a menu of twenty-eight frameworks. These were designed through a uniform methodology. Most of the frameworks integrate GNIpc and one other indicator into an index. There are six general health-need frameworks, based on indicators of under-five mortality rate, under-sixty mortality rate, life expectancy, healthy life expectancy, disability-adjusted life year rate (DALYR), and age-standardized DALYR. There are four domestic-capacity frameworks, based on debt service, tax ratio, total health expenditure per capita, and government health expenditure per capita. There are four expected-impact frameworks, based on absolute and relative improvement in under-five mortality rate and skilled birth attendance rate, respectively. There are six cross-cutting frameworks, based on inequality in life expectancy, Gini index for income, income share held by bottom 40%, skilled birth attendance rate, DTP3 immunization coverage, and out-of-pocket payments for health services. There are also three disease-specific frameworks, based on maternal mortality ratio, HIV prevalence, and tuberculosis prevalence. Finally, there are three three-criterion frameworks and two multicriterion frameworks integrating four indicators. One of these utilizes weights informed by the online survey.

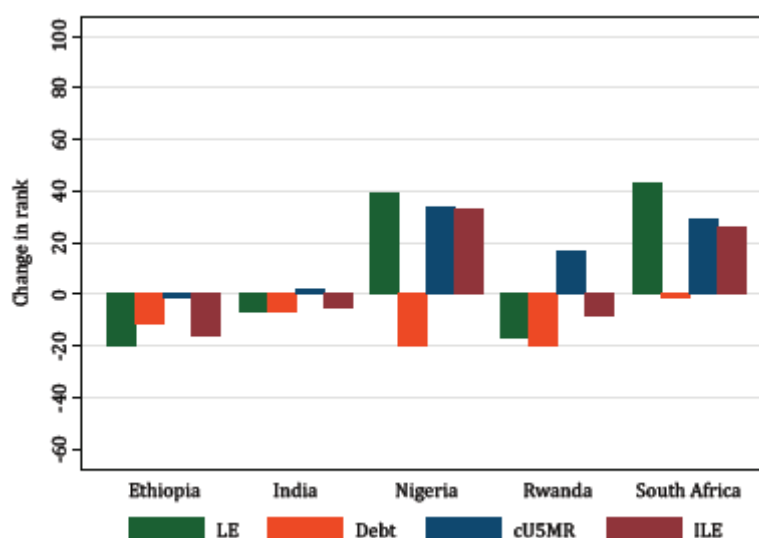
Simple frameworks with two indicators



Comparison

We assessed the frameworks through head-on comparisons of how they rank countries relative to a ranking based on GNIpc alone (our assumed baseline). First, we demonstrated the changes in rank for five focus countries, as illustrated below. These figures allowed us to determine the degree to which the inclusion of a particular indicator in a framework influenced the ranking of an individual country. Second, we identified average changes in rank for country groupings under different frameworks. These figures allowed us to determine the degree to which the inclusion of a particular indicator in a framework influenced the ranking for groups of countries. Third, we calculated Spearman coefficients for the correlations between GNIpc and other indicators or frameworks. Doing so helped clarify the implications of going beyond GNIpc, because a lower coefficient suggests greater changes as a result of such a move. Fourth, we generated two kinds of scatter plots. One type shows the changes in rank for a large set of countries, and the other shows the relationship between the normalized scores of GNIpc and the normalized scores of another indicator.

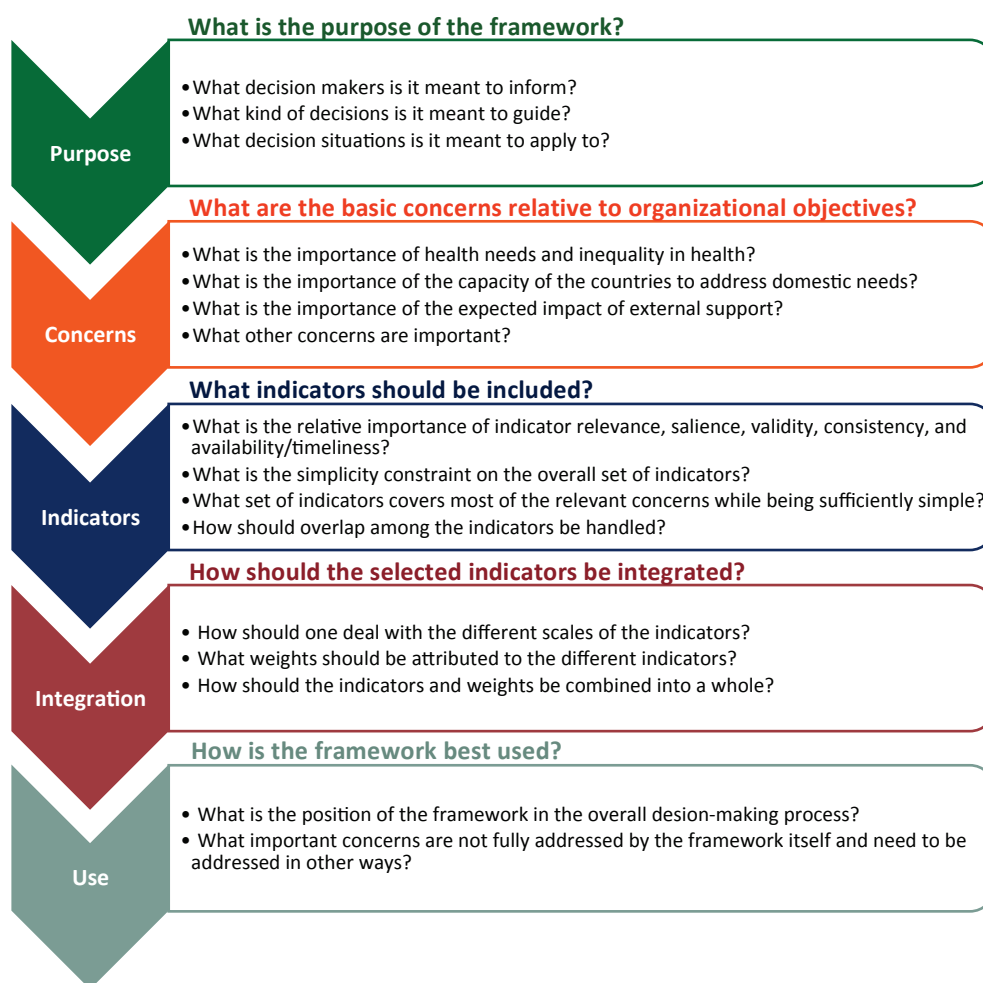
Selected changes for focus countries



Choices

The EAI Convening Partners and other institutions revise their classification frameworks from time to time. Based on insights gained over the course of the EAI project, we have articulated the key choices every donor will need to make when designing its classification framework. These choices are summarized below. The menu of frameworks and the implications described in this report can inform these choices.

Key choices in the design of frameworks



Overall findings and recommendations

With regard to health needs, we found that

1. addressing health needs is generally considered a key objective of external financing for health, and a general health-need indicator can be relevant for most institutions. We also found that including a health measure in the framework makes a significant difference for many countries and country categories and that it specifically improves the ranking of countries with large unmet health needs. Against this background, we recommend integrating GNIpc and one health-need indicator for the initial classification of countries.
2. which of the health-need indicators is chosen makes less of a difference than whether or not such an indicator is included at all. Moreover, no single indicator appears to be clearly superior to all others. We found that life expectancy and DALYR are both good candidates for being included in a classification framework. We recommend that institutions at least consider using life expectancy as a generic health-need indicator, but the optimal measure for any given actor will eventually depend on organization-specific needs and objectives.

3. although the implicit age weighting and the use of age standardization are linked to important value choices, they have received scarce attention in the context of external financing for health. We recommend critically examining the extent to which candidate health-need indicators give extra weight to deaths at a younger age and are sensitive to the country's age structure.
4. although inequality in health is considered an important concern for external financing for health, no robust measure of health inequality exists for this purpose. We recommend accelerating the development of such measures.

With regard to domestic capacity, we found that

5. domestic capacity is an important consideration for external financing for health. However, none of the examined indicators was clearly suitable for inclusion in the classification framework when judged against the following criteria: relevance, salience, validity, consistency, and timeliness and availability. We also found that the inclusion of domestic-capacity indicators might contribute to perverse incentives. Against this background, we suggest that indicators of domestic capacity other than GNIpc – such as debt service and tax ratio – are considered after the primary classification framework has been applied.

With regard to expected impact, we found that

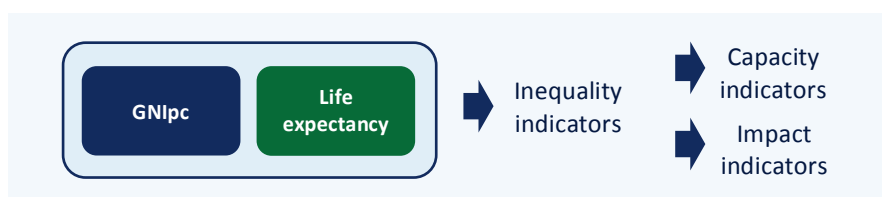
6. expected impact is an important consideration for external financing for health, but we conclude that none of the examined indicators is a good predictor of future impact. More generally, none of the indicators was clearly suitable for inclusion in the classification framework when judged against the following criteria: relevance, salience, validity, consistency, and timeliness and availability. We suggest that indicators of expected impact are considered after the primary classification framework has been applied and mainly at the program and project level, in the assessment of applications, and during the implementation phase. Potential country determinants of impact include level of conflict, institutional quality, and recent improvements in health-service coverage or health outcomes.

With regard to all examined frameworks, we found that

7. nearly any move beyond GNIpc alone resulted in a lower rank for the group of LICs and a higher rank for the group of MICs. To the extent that this is undesirable, we recommend considering counteracting measures.

Overall,

8. we recommend using a simple framework that integrates GNIpc and health needs for the initial classification of countries. We suggest supplementing the use of this framework with a more discretionary assessment of within-country inequalities in health and income to the extent permitted by the data.
9. we recommend considering a range of indicators for capacity and expected impact after the primary classification framework has been applied, including at the program and project level, in the assessment of applications, and during the implementation phase. This and the assessment of inequalities should be done in the context of country-level dialogue.
10. these general findings and recommendations may be informative, but the optimal indicators and framework for any given institution will eventually depend on organization-specific needs and objectives.

Decision-making stages

1. The problem

The world is changing, and so is the landscape of health and development. Transitions in health, economic development, and external financing for health have given rise to new opportunities as well as increasingly complex challenges.^{1, 2} These challenges, in turn, call for a new framework for classifying countries.

Health trends

There have been dramatic improvements in health outcomes over the past two decades.^{3, 4} However, there are vast inequalities between countries,⁵ and many countries have significant inequalities in health outcomes across gender, socioeconomic status, and place of residence.^{6, 7} These outcome inequalities are often accompanied by inequalities in access to health care.⁷⁻⁹ The bottom line is that, despite improvements, there remain vast unmet health needs across the world.

Economic trends

Economic changes over the last two decades have been no less profound. Economic growth rates have been impressive, and many countries have moved from low-income to middle-income status. The middle-income category now comprises 104 countries and 70% of the population, 32% of the gross domestic product (GDP), more than 75% of the poor,¹⁰ and almost 67% of the disease burden in the world (based on data from the World Bank and the Institute for Health Metrics and Evaluation (IHME)). At the same time, many middle-income countries (MICs) are characterized by substantial inequalities in income, health, and health services.

Trends in external assistance

Whereas domestic financing is the predominant source of financing for health in all but a few low-income countries (LICs), external funding plays a key role in many countries, by filling funding gaps, catalyzing domestic financing, or both. The last two decades have witnessed unprecedented growth in development assistance for health. It almost doubled from \$5.8 billion in 1990 to \$11.2 billion in 2001, and it nearly tripled to \$31.3 billion (expressed in 2011 US-dollar terms) by 2013.¹¹ However, development financing for health has plateaued in recent years, and there is a need to improve both the effectiveness and efficiency of development assistance.

Complex challenges

These powerful trends have brought about challenges as well as important opportunities. Prominent among these are two related challenges: those pertaining to the role of MICs and those pertaining to the transition away from eligibility for external financing for health.^{12, 13} Many have claimed that compared to LICs, MICs have a greater capacity to address unmet health needs without external financing. Nevertheless, these countries often have weak health systems, profound inequalities, and a large absolute degree of unmet health needs. These limitations have

led many researchers to claim that external financing for health is still highly relevant in MICs. At the same time, there is a widely shared concern that many countries risk losing eligibility too soon, too fast, and simultaneously from multiple donors, creating a financial “cliff” for these countries.¹³

The need for a new framework

These specific challenges point to the more fundamental challenge of how to classify countries when making decisions on external financing for health. In particular, many question the dominant role of gross national income per capita (GNIPC) in this context. One reason is that factors other than GNIPC are seen as relevant for countries’ capacity to address domestic needs.¹⁴⁻²⁰ Another reason is that GNIPC is not seen as adequately reflecting countries’ level of unmet human needs.^{10, 12, 21-26} It is therefore necessary to search for a new framework – one that goes beyond GNIPC and incorporates a broader set of indicators.

2. NIPH's explicit approach

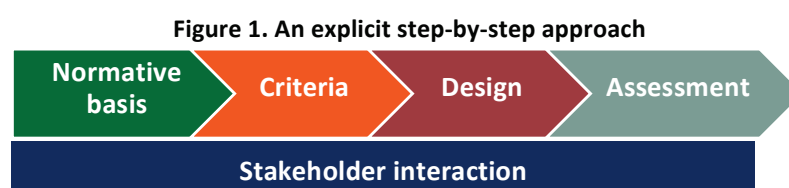
The Norwegian Institute of Public Health (NIPH) was commissioned by the Equitable Access Initiative (EAI) to provide input to the Initiative. The objective of the NIPH team has been to explore frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. Our report is explicit in its focus on external financing, because we believe the intended purpose of any framework is crucial to both its design and assessment. We do not seek to prescribe which frameworks specific institutions should adopt to guide their external financing for health, and our study represents a contribution only to the first, general steps toward a new classification of countries.

The six outputs of our work are (1) an explicit step-by-step approach to the design and assessment of classification frameworks, (2) responses from interviews and a survey, (3) a broad menu of simple frameworks, (4) a head-on comparison of their implications, (5) a list of key choices that any institution must make, and (6) overall findings and recommendations for choice of indicators and use of frameworks.

The NIPH team devised an explicit step-by-step approach to the design and assessment of frameworks. This approach is particularly explicit about the value choices inherent in the design and assessment of any framework. The approach also seeks to be intuitive, inclusive in its design and assessment of a wide range of frameworks, and interactive in its use of stakeholder interviews and a stakeholder survey alongside consultations. We see this approach also as an output in itself, and the EAI, the Convening Partners, and other institutions may use elements of the approach irrespective of whether they agree with the specific choices of the NIPH team.

As illustrated in Figure 1, the approach consists of 4 + 1 steps:

- Delineation of a normative basis
- Identification and assessment of indicators
- Design of candidate frameworks
- Assessment of candidate frameworks
- Cross-cutting interaction with stakeholders

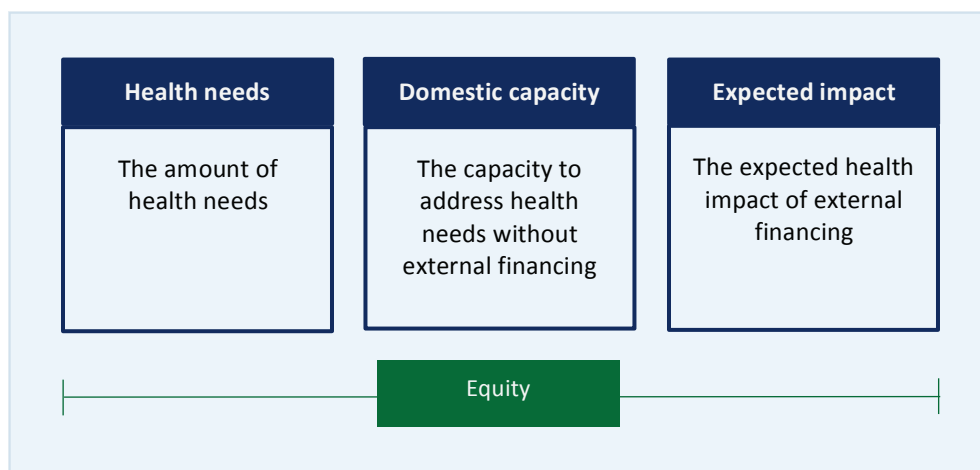


Delineation of normative basis

Decisions concerning external financing inherently involve value choices, although they are often left implicit. This is also the case for the rules and policies most donors use to guide such decisions. The first step in the design and assessment of classification frameworks is therefore to clearly delineate a normative basis for external financing for health.

The NIPH team delineated an explicit normative basis consisting of 3 + 1 concerns. As illustrated in Figure 2, this basis comprises health needs, domestic capacity, expected impact, and the cross-cutting concern for equity. The three concerns are clearly defined (see Box 1) and intended to be mutually exclusive to the greatest extent possible.

Figure 2. An explicit normative basis



The identification and definition of these concerns were based on a review of the literature on health financing, development assistance, and distributional justice, as well as on interaction with stakeholders.^{1, 15, 27-31} We believe that although all or most stakeholders will agree that each of these concerns is relevant for external financing for health, they will often disagree about their relative importance.

Box 1. Definitions of the basic concerns

Health needs: A population's or subpopulation's need for improvement in health or the determinants of health, including health services. The concern for health needs can incorporate the common concern for inequalities in health.

Domestic capacity: The capacity of countries to address domestic needs without external financing. Capacity so understood relates to ability to pay and fiscal space, as well as the constraints on these factors.

Expected impact: The expected impact of external financing for health in terms of changes in health or the determinants of health, including health-service coverage.

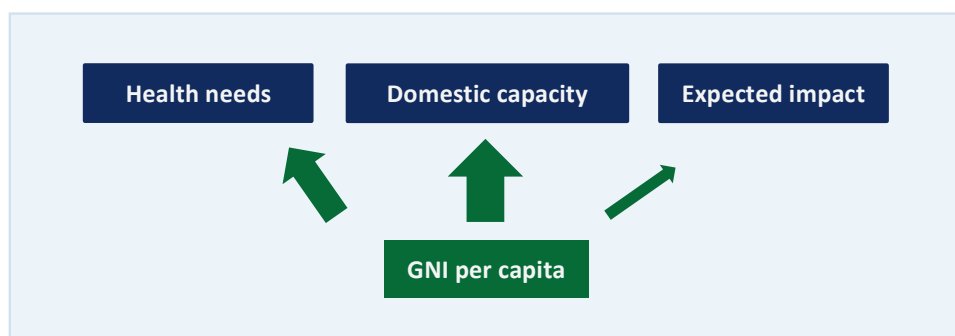
Equity: This concern is notoriously difficult to define. In the conceptual scheme delineated here, it is a concern with the distribution of resources, services, and outcomes across individuals, groups, and populations. So understood, equity is a cross-cutting concern sensitive to how well the three other concerns are addressed. Equity is often seen as closely related to inequalities of various kinds, but equity and equality are not the same and should not be conflated.

Identification and assessment of indicators

After delineating the normative basis, we identified the indicators that should be part of a classification framework. This was done through a number of smaller steps. We first developed a long list of candidate indicators, as shown in Supplementary File 1. This was based on a number of articles³¹⁻³³ as well as the list of candidate indicators generated at the EAI Expert Panel meeting in February 2014.

Based on this list, the team identified a set of indicators relevant for each of the concerns. When linking concerns and indicators, it is crucial to acknowledge that there is no one-to-one relationship between them, because the failure to do so has caused much confusion. In particular, many indicators are associated with more than one of the concerns delineated above. For example, in the context of external financing for health, GNIpc has probably most often been seen as an indicator of domestic capacity (through ability to pay), but it has also quite often been seen as an indirect indicator of health needs. Less commonly, it has been claimed that GNIpc, and income class based on GNIpc, is associated with the expected impact of external assistance.^{23, 24} These relationships are illustrated in Figure 3.

Figure 3. The lack of one-to-one correspondence between concerns and indicators



Inspired by criteria used elsewhere,³⁴⁻³⁶ we assessed these indicators on the basis of relevance, salience, validity, consistency, and availability and timeliness. Relevance was assessed relative to the normative basis and the specific purpose of designing a framework for the initial classification of countries that may be used across a wide range of actors and settings. Salience is particularly important in the present context, because it is crucial that the framework, as well as the resulting classification of countries, can be easily understood by policy makers and the general public. The criteria used to assess indicators are further described in Box 2.

Box 2. Criteria used to assess indicators

Relevance: The extent to which the indicator (if valid and reliable) provides meaningful information about country characteristics related to health needs, domestic capacity, expected impact, and/or the cross-cutting concern for equity, and the extent to which it does so for the specific purpose of classifying countries in a way that may be useful across a wide range of actors and settings.

Salience: The extent to which the indicator can be easily understood by policy makers and the general public.

Validity: The extent to which the indicator measures what it purports to measure.

Consistency: The degree to which data measurements are stable when repeated if the true situation has not changed.

Availability and timeliness: Availability depends on the number of missing values across countries, and timeliness is the recency and expected regularity of updates.

Design of candidate frameworks

The third step was the design of candidate frameworks, which we defined as the combination of two or more indicators into an index.

Key features

The design of these frameworks was characterized by simplicity, comprehensiveness, and uniformity.

Simplicity

When designing and assessing frameworks, we operated under a simplicity constraint. We required that each of the frameworks be relatively easy for policy makers and the general public to understand. Among other things, we took this to imply that the frameworks could not comprise a large potpourri of indicators and that the frameworks should preferably combine only two or three indicators. That frameworks should be simple was one of the most consistent messages we received in stakeholder interviews and consultations. Several stakeholders also specifically suggested that no more than three indicators should be integrated into a framework. Moreover, even if one were to seek multicriterion frameworks in the end, assessment of simple frameworks can help to clarify how different indicators will influence complex frameworks.

Comprehensiveness

We opted to design multiple frameworks. We believe it is useful for the EAI to be presented with a menu of options, including the implications of each option, before it makes recommendations. We also believe the optimal index differs across the Convening Partners and other institutions, as suggested in consultations and interviews. Moreover, before these partners and institutions reform the approach to country classification, it may be most useful for them to be presented with a menu of frameworks and their implications. These actors may choose to build off one of

the frameworks presented in this report or use the report's insights to construct a framework different from all these.

Uniformity

Each framework was designed and assessed using a uniform methodology in order to facilitate intuitive understanding and head-on comparison. To promote an intuitive understanding of the frameworks and their implications, the simplest methodological options were adopted whenever possible.

Steps in the design

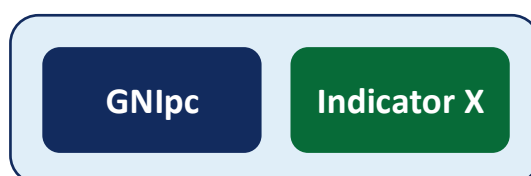
The design of each framework followed the same four substeps. Our starting point for each index was GNIPc. Because most of the frameworks comprised only two indicators, in what follows, we emphasized the design of two-criterion frameworks.

Selection of indicators

We first selected one indicator for integration with GNIPc into an index, as illustrated in Figure 4. Indicators were chosen from a preliminary list of about eighty indicators, identified based on literature reviews, reviews of existing donor and multilateral indicators, output from an EAI meeting, previous work by team members, and an initial assessment of data availability (see Supplementary File 1). Data sources and data characteristics for all indicators included in any framework are described in Appendix C.

The selection of indicators for each framework was based on the assessment of relevance, salience, validity, consistency, and availability and timeliness. Indicators was also selected with the aim of covering a wide range of concerns and more specific considerations.

Figure 4. Selection of one indicator to accompany GNI per capita



Determination of prioritization rule

We next specified the prioritization rule – a step that is often left implicit. The prioritization rule specifies what happens to the rank of a country when the level of a given indicator increases.²⁷ In our context, a higher rank means higher priority for external financing. For example, according to a common rule for GNIPc, rank and priority decrease monotonically with increases in GNIPc.

Normalization

The third substep was normalization. When indicators are combined, their scales matter. Many of the indicators have different natural scales, including different ranges. Some form of normalization is therefore likely to be better than none at all, although all normalization

procedures exhibit some problems. The choice of normalization procedure also affects the implicit weighting between criteria.

We used common min-max normalization.³⁴ This normalizes indicators, giving them an identical range [0, 1] by subtracting the minimum value and dividing by the range of the indicator values:

$$\frac{x - \min}{\max - \min}$$

where x is the value of the indicator before normalization, and \min and \max are respectively the lowest and highest observed values of the indicator among all LICs and MICs in our final sample of countries.

Application of weights and aggregation

When combining two or more indicators into a composite index, one needs to determine their relative weight.³⁴ We used equal weighting as the starting point throughout; that is, all indicators were given the same weight. Although equal weighting is by no means value neutral, it facilitates intuitive understanding of the frameworks and their implications. For a few frameworks, we applied weights informed by the online survey.

Each weight and normalized indicator constitute a product. To aggregate two or more products of this kind into a composite index, we employed linear aggregation, that is, straightforward summation of these products.³⁴ In the case of three indicators, the score of the composite index is then simply as follows:

$$w_A A + w_B B + w_C C$$

where A , B , and C are the indicator values, and w_A , w_B , and w_C are the weights assigned to each indicator. The sum of w_A , w_B , and w_C is always equal to 1 for each framework. Specifically, in the case of two indicators, w_A and w_B were both set at 0.5. For frameworks with three indicators, all three indicators were assigned an equal weight of 0.333.

Assessment of candidate frameworks

We assessed candidate frameworks through a direct comparison of their implications in terms of the ranking of countries. We also examined differences between the frameworks by generating scatter plots and Spearman correlation coefficients for the indicators and frameworks in question. As requested by the EAI, we refrained from going beyond rankings to examine eligibility thresholds and allocation of funds.

Baselines

All frameworks with only two indicators were compared to a framework of GNIPC alone; that is, we used a GNIPC baseline in most cases. Some frameworks with more than two indicators were also compared to a certain health-need framework, namely that of GNIPC combined with life expectancy.

Countries

For all the comparisons, all LICs and MICs that met the inclusion criteria identified above were included in at least part of our analyses. Rankings of all these countries are included in Supplementary File 5. For simplicity, comparisons reported in the main text concentrate on five focus countries: Ethiopia, India, Nigeria, Rwanda, and South Africa. This set of countries was selected because it illustrates the implications across a diverse set of countries and highlights the key challenges for any classification framework. Some of the key characteristics of these countries are shown in Table 1, which indicates that the countries differ in multiple dimensions. With the exception of income class, all of these characteristics were also integrated into at least one framework.

Table 1. Selected characteristics of focus countries

Country	Income class	GNIpc	LE	DALYR	HIVR	Debt	GHEpc	ILE	Gini	SBA
Ethiopia	LIC	470	64	48 475	1.2	1.4	15	30	34	23
India	LMIC	1 530	67	39 494	0.1	2.1	20	25	34	67
Nigeria	LMIC	2 700	53	73 320	3.3	0.1	32	41	43	38
Rwanda	LIC	670	64	43 972	2.9	0.6	41	30	51	94
South Africa	UMIC	7 410	57	55 894	18.8	2.5	287	26	65	95

Data for 2013, except SBA for South Africa, which is for 2008. GNIpc = gross national income per capita; LE = life expectancy; DALYR = disability-adjusted life year rate (per 100,000 population); HIVR = HIV prevalence rate; Debt = total debt service (% of GNI); GHEpc = government health expenditure per capita; ILE = inequality in life expectancy; Gini = Gini index for income; SBA = skilled birth attendance rate.

Country categories

Implications for five country categories were also examined. The first three are the World Bank's income classes. The other two are the 20% of countries with the lowest life expectancy and the 20% of countries with the lowest GNI-DALYR ratio, respectively. DALYR is the number of disability-adjusted life years (DALYs) per 100,000 population.

These categories were chosen because they speak to the current debate over classification frameworks. Income class is central, because much of the debate is framed in terms of the role of LICs compared to that of MICs in external financing for health. Although the 20% thresholds are somewhat arbitrary, the category based on life expectancy is pertinent, because concerns about countries with profound health needs figure prominently in the current debate. Similarly, the GNI-DALYR ratio points to a central relationship in the debate – that between economic capacity to address health needs and the current burden of disease in the country. Although this ratio may be relevant in its own right, it is worth mentioning that 96% of the 20% of countries with the lowest GNI-DALYR ratio in our sample are classified as LICs.

Relevance of comparisons

Comparisons of rankings do not by themselves indicate the optimal classification framework. However, they can assist in decisions concerning frameworks, as information about rankings can be used alongside information about other features of the frameworks to make an overall assessment. When assessing the implications of frameworks, we utilized input from consultations, interviews, and the survey.

Scatter plots

Whether an additional indicator is worth including in a framework will depend partly on the extent to which that indicator helps to distinguish between countries – and how it does so. We therefore generated two sets of scatter plots. The first plots the ranking following from one framework against the ranking following from another. One of these scatter plots is presented in the main text, and several others are shown in Supplementary File 4. The second set of scatter plots exhibits the normalized values of each indicator relative to the normalized value of GNIpc. These scatter plots are presented in Supplementary File 6.

Correlation coefficients

We also compared GNIpc with each of the candidate frameworks by calculating the Spearman coefficient of correlation between GNIpc and framework score. A Spearman coefficient is a measure of the rank correlation between two variables. The Spearman coefficient assumes only a monotonic association between variables, as opposed to the Pearson coefficient, which assumes a linear relationship. If the Spearman coefficient for the correlation between GNIpc and a framework score is lower, it suggests that moving from GNIpc alone to the framework in question is more likely to affect the overall ranking of countries. It should be noted, however, that the Spearman correlation coefficient only reflects the extent to which *ranks* deviate from a monotonic function. The correlation coefficient needs therefore not say much about how countries change relative to an eligibility threshold for external financing for health. We also calculated Spearman coefficients for the correlation between GNIpc and each of the indicators integrated into frameworks. All Spearman coefficients are presented in Appendix B.

Missing data

For indicators with only a few years of missing data by country, we employed linear interpolation to impute missing values for the preferred year, using available data for the years 2000–2014. This method assumes a linear relationship of values over years and utilizes nonmissing values for adjacent years to calculate an intermediate value. We used this method to calculate missing values for the following indicators: GNIpc, under-five mortality rate (U5MR), under-sixty mortality rate (U60MR), life expectancy (LE), debt service ratio (Debt), tax ratio (Tax), total health expenditure per capita (THEpc), government health expenditure per capita (GHEpc), absolute improvement in under-five mortality rate (cU5MR), relative improvement in under-five mortality rate (rcU5MR), absolute improvement in skilled birth attendance rate (cSBA), relative improvement in skilled birth attendance rate (rcSBA), income share held by bottom 40% (Income40), skilled birth attendance rate (SBA), DTP3 coverage rate (DTP3), out-of-pocket payments (OOPs), maternal mortality ratio (MMR), HIV prevalence rate (HIVR), and tuberculosis prevalence rate (TBR). SBA for South Africa received special treatment, because we could not impute the needed value and South Africa was one of the five focus countries. As a proxy for the 2013 estimate, we used a 2008 estimate provided by the World Health Organization (WHO) Global Health Observatory.

For many countries, there were too few observations on the Gini coefficient for the years 2000–2014. Interpolation was therefore not possible, and we instead used the most recent figure for each country in that period. Data on HALE and DALYR were unavailable for multiple consecutive years, and therefore we could not impute missing data for these indicators. Hence, for DALYR and HALE, as well as inequality in life expectancy (ILE, we used only data for 2013.

Our sample of countries included all countries classified as LICs and MICs in 2014 (N = 137) for which GNIpc was available (N = 131). Information about the data used for each of the indicators integrated into a framework is provided in Appendix C.

We compared six sets of frameworks. Four of these corresponded to concerns and indicator categories: a set of health-need frameworks, a set of domestic-capacity frameworks, a set of expected-impact frameworks, and a set of “other” frameworks. The fifth set included one framework from each of these four categories plus two multicriterion frameworks. The sixth set was similar to the previous one, except that the GNIpc-LE baseline was used. For each set, only the countries with available data (after imputation) for all the indicators pertaining to the set were included. Exactly this group of countries was also the basis for the baseline ranking. Accordingly, the countries ranked by each framework were consistent within each set, but could vary across sets. This strategy was pursued to reduce the effect of missing values concerning the changes in country rank. In particular, we sought to avoid a scenario in which the countries experienced a change in rank, when one moves from one framework to another, merely because the frameworks differ in number of countries with missing values.

Interaction with stakeholders

Throughout the project, the team interacted extensively with stakeholders. These stakeholders represented the EAI Convening Partners, multi- and bilateral donors, partner countries, civil society organizations, the private sector, and academia. Multiple meetings and consultations were arranged by the EAI. In addition, the NIPH sought to strengthen interaction with stakeholders through semistructured interviews and an online survey. These elicited stakeholders’ views on the most central questions pertaining to the design and assessment of classification frameworks. Findings from the interviews and the survey represent relevant input to the EAI in their own right and can be useful for the Convening Partners now and later. In addition, the findings informed our design and discussion of candidate frameworks.

3. Findings from interviews and survey

Interviews

During August and September 2015, the team conducted twenty interviews with stakeholders. Interview questions were approved by the EAI, and the interview guide is included in the background material for this report. Nineteen of the interviews took place over the phone, and one was conducted face-to-face. Nine of the twenty interviewees were members of either the EAI Conveners' Technical Working Group or the EAI Expert Panel. The stakeholders interviewed included six academics, four from civil society or the nongovernmental organization (NGO) sector, four from multilateral organizations, three from bilateral donors, two from recipient-country governments, and one from a philanthropic foundation. Several stakeholders held multiple roles, for example, representing both civil society and a recipient country. The recipient countries represented comprised India, Pakistan, South Africa, Thailand, and Zimbabwe.

The interviews concentrated on stakeholder views on alternative frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. Questions addressed the purpose of frameworks for classifying countries; the importance and role of the four concerns (needs and health needs, inequality, capacity, and expected impact) with respect to their inclusion in frameworks for classifying countries devised to guide decisions on external financing for health; and types of relevant indicators. Interview questions are further described in Supplementary file 3.

The role of frameworks

Stakeholders recognized the value of frameworks in facilitating informed decisions on external financing for health, including helping to determine eligibility, determining the terms of the contributions, and optimizing allocation of resources in relation to objectives. At the same time, stakeholders emphasized the circumscribed role of frameworks in the overall decision process, with respect to both decisions regarding external financing for health and those regarding external support more generally.

Stakeholders generally agreed that frameworks could be useful in providing an *initial* orientation for decision making that would then lead to consideration of a wider set of factors. These included qualitative factors related to a country's economic and health policies, social issues, tax systems, and performance, which respondents perceived as less able to be captured by a framework. Many felt that such issues necessitated a conversation with countries, going beyond a framework, in order to decide what assistance is appropriate. Many also felt that the different needs of countries required individual assessment and tailored responses to provide the right balance between need, capacity, and expected impact.

Along the same lines, many stakeholders expressed the concern that a new classification framework might be too simplistic to be useful in decisions concerning external financing for health. There was a perception that frameworks oversimplified the complexities of countries in diverse circumstances. At the same time, stakeholders generally maintained that any framework should be transparent and easy to understand – overly complex frameworks combining numerous indicators would not meet this criterion. This suggests that the tension between

simplicity and comprehensiveness should be carefully addressed in the design and choice of a framework.

Going beyond frameworks for financing decisions, many stakeholders emphasized the importance of support for the identification of the modality of assistance (such as technical assistance, capacity building, and advocacy) according to different country circumstances.

Relevant concerns

Of the four areas of concern examined directly (needs and health needs, capacity, expected impact, and inequality), *needs and health needs* was clearly identified as the most important by stakeholders, followed by *inequality*. One stakeholder suggested that needs and health needs was the starting point, and others that needs and health needs were the reason to allocate funds. It was also suggested that one should go beyond overall health needs and consider specific health needs, because there is not necessarily a close connection between overall disease burden and resource needs – some diseases are more expensive to address than others. At the same time, some stakeholders expressed the concern that donor responses were overly disease-focused, at the expense of strengthening health systems.

With regard to inequalities, there was a widespread view, particularly from recipient countries, that inequality or poverty – which were often discussed in tandem – was relevant to external financing for health. It was seen as an issue of shared responsibility among countries. Several claimed that we must avoid the “double penalty,” in which neither government nor external donors accept responsibility for poor people with unmet health needs in a country’s population. For some it was a matter of human rights. In line with this view, some stakeholders asserted that donors should not neglect countries just because they had passed a GNIpc threshold.

These considerations were not necessarily about amounts of assistance, but pertained equally to strategies for addressing unmet health needs, especially in countries with large inequalities. Such strategies could include a process of dialogue with government about how health policies could be improved and/or providing assistance through NGOs to disadvantaged groups. Notably, respondents from recipient countries tended to favor such approaches from donors as a means to persuade their governments to invest more and better in the health of their populations. This could include the use of conditionality. Moreover, some donors and other respondents were correspondingly keen that external assistance should provide the right incentives (and avoid disincentives) for governments to invest in health.

Several stakeholders suggested that considerations of *capacity* and *expected impact* are important – but often in ways external to the initial classification of countries and thus to the classification frameworks. For example, capacity and expected impact helped address questions about *how* to meet needs. This included questions regarding modality of assistance and time horizon for assistance.

In discussions of *expected impact*, one stakeholder stated that if a country performs well, it should have easier access to funding or more freedom in its use. If expected impact is likely to be low, the country should still be eligible for funds, but funders should ask, “Why is impact low?” and they should potentially engage with the country via another form of assistance or seek to maximize impact by focusing on a priority area rather than simply penalizing or rewarding poor performance. Some noted that it is important to differentiate between those countries that have control over the determinants for low expected impact and those that do not. Some donors

noted that there is now greater scrutiny on value for money and on results, such that donors have to justify investments in settings where the impacts are expected to be low, even if there is great need. Others noted the importance of unpacking the expected impact in order to understand the differences, for example, in time to impact and in impacts that are not measurable.

Capacity was generally considered the least important of the four concerns, and many emphasized aspects of capacity they perceived to be somewhat different from “ability to address health needs without external financing.” For example, several considered capacity to be primarily about the strength of the country’s health system and infrastructure, independently of its need for external resources. Some also referred to the issue of “absorptive capacity,” that is, the ability of a country to use external resources effectively. Others noted that capacity building is a longer-term objective than addressing immediate health needs. A number claimed that the Ebola crisis shed light on the weaknesses of health systems, and they noted that strengthening health systems was inherently a joint responsibility. Some emphasized that although capacity is relevant to decisions concerning external financing for health, it is not primarily relevant for the initial classification of countries. It was suggested, for example, that capacity was more relevant at the operational level.

Relevant indicators

One respondent described GNIpc as the *least worst indicator*, and others had similar sentiments. Stakeholders noted that it is updated every year, available for all countries, easily understood, correlated with other indicators, and transparent. Although most acknowledged its role as a principal indicator, there were a few caveats and concerns. A number of these were in relation to the classification of MICs. Here, stakeholders noted the need for deeper analysis of GNIpc and sensitivity to the following issues:

- dramatic and rapid shifts in GNIpc levels that may prematurely raise countries beyond their capacity
- the limitations of GNIpc in identifying the extensive health needs and poverty in MICs (subnational, at-risk groups)
- the need to be inclusive of broader criteria and indicators designed to bolster national investment in health (rather than displacing it) and to identify gaps in coverage and capacity

Stakeholders suggested that frameworks should include information about the following factors:

- Burden of disease or other health outcomes. This could include information about DALYs, U5MR, and other mortality and morbidity indicators. (Stakeholders noted the need to respect the time lag between increases in GNIpc and health-outcome indicators.)
- Health access. This could include information about public-health expenditure and level of access to publicly funded health services.
- Inequality. This could include information about the Gini coefficient and proportion of poor people in country.
- Capacity. This could include information about the density of the health workforce.
- Political commitment. This could include information about government health expenditure or government funds spent only on primary health care or a particular

health area. (This was suggested by stakeholders from multilateral organizations and recipient countries.)

Several stakeholders suggested that if possible, frameworks and associated decision making should take into account

- indicator trends, not just absolute levels
- time horizons of assistance given the nature of the disease, pace of country transition, and degree of inequality

Stakeholders were also conscious of the data-quality limitations and that they must be overcome if a significant number of these indicators are to be made operational.

Overall findings and use of inputs

Our findings suggest the following:

- Frameworks are potentially useful as an initial guide to decision making regarding external financing if they are simple and transparent.
- GNIpc is relevant, but there is a need to go beyond this and consider other indicators.
- Alongside GNIpc, measures of health needs and inequalities are the most important indicators to include in a classification framework.
- Domestic capacity and expected impact are relevant for decisions concerning external financing for health, but may best be taken into account at decision-making stages other than the initial classification of countries.
- Stakeholders are also concerned with the classification of countries for purposes other than determining eligibility and the level of external financing for health. These purposes include decision making pertaining to the modality of external support and pharmaceutical pricing.

The input from stakeholders has guided our analysis in phases II and III:

- Stakeholders' emphasis on keeping frameworks simple and transparent prompted us to concentrate on frameworks that integrate only two or three indicators.
- Stakeholders' emphasis on the circumscribed role of classification frameworks in the overall decision-making process prompted us to examine the limitations of such frameworks.
- Stakeholders' views on GNIpc prompted us to use that indicator as the basis for the design of new frameworks.
- Stakeholders' attribution of priority to health needs and inequality as areas of concern validated our emphasis on frameworks that integrate GNIpc and a health-need indicator. Stakeholder views also motivated us to examine more frameworks pertaining to inequality.
- Stakeholders' emphasis on aligning frameworks with organization-specific objectives motivated us to take a menu-based approach rather than an approach based on a few predetermined indicators. It also supported an approach that highlights the key value choices in the design and assessment of frameworks.

Survey

As part of our normative basis, we outlined a set of concerns that we believe most stakeholders take into consideration to at least some degree when making decisions about external financing for health. These include the concern for health needs, domestic capacity to address health needs without external financing for health, and the expected impact of such financing. A cross-cutting concern for equity is also part of the normative basis. Although inequity and inequality are not synonymous, we focused on inequalities in health in the context of the survey. Although we believe all stakeholders take all four concerns into consideration to some degree, it is unclear how much weight they would put against each, and we would expect that different stakeholders would have different preferences in this regard. To shed light on this question, we developed an online discrete choice experiment (DCE) to elicit the preferences of different stakeholders.

Methods

Respondents were presented with a series of pairs of frameworks. The frameworks differed in terms of how much importance each placed on four country characteristics corresponding to the four concerns. The characteristics were defined with the aim that they would be as mutually exclusive to one another as possible and easily understood by respondents. The four country characteristics were income per capita, burden of disease, strength of health system, and health inequalities.

For each country characteristic, we defined three possible levels of importance (low/no, some/medium, and high). Frameworks differed in terms of how much importance each placed on the four different country characteristics. Our choice of country characteristics and levels of importance produced an experimental design with eighty-one possible frameworks ($3 \times 3 \times 3 \times 3$). This led to a very large number of potential pairs of frameworks to present to each respondent, which was deemed impractical.

To simplify the choices, we developed an orthogonal main-effects plan that provided high levels of balance and minimal overlap in order to reduce the number of questions each respondent needed to answer to nine. One of the choices presented in the survey is shown in Figure 5.

Figure 5. Example of choice presented in survey

Among the two frameworks described below, which framework do you prefer to classify countries for decisions about external financing for health?

Country Characteristic	Framework A	Framework B
Income per capita	Low importance	High importance
Burden of disease	No importance	No importance
Strength of health system	No importance	Some importance
Level of health inequalities	No importance	Some importance

☐ Framework A
 ☐ Framework B

Survey Completion: 0% 100%

In addition, respondents were asked to answer a series of demographic questions on age, gender, country of birth, country of residence, highest level of education completed, and type of organization in which they worked. All respondents were asked to provide informed consent, and this research component received IRB human subjects exemption status from New York University.

Because the survey collected information on the characteristics of the respondents and we had experimentally defined differences between the frameworks, we estimated a conditional logit model, which is the preferred specification for estimating choice models of this nature.³⁷

The EAI Convening Partners assisted us with generating a sample of respondents. From the Global Fund, we received a list of e-mail addresses of people who had recently expressed interest in participating in a regional partnership forum or an e-Forum. For the other Convening Partners, EAI focal points were e-mailed, and these were asked to forward the survey link to relevant members of their organization with a request to complete the survey. We sent approximately 1,500 e-mail invitations.

Overview of respondents

A total of 285 people consented and completed the survey. Slightly less than half of the sample (45.3%) were female. The sample was highly educated, with more than half reporting a graduate degree (52.3%) and only 19.3% reporting an undergraduate degree or less. The sample had very wide geographic coverage, with responses coming from almost ninety countries. Respondents were most likely to have been born in a high-income country (HIC; 35.8%), followed by a lower-middle-income country (LMIC; 34.0%), a LIC (15.8%), and an upper-middle-income country (UMIC; 14.8%). Civil society organizations were very well represented in the sample – accounting for 44.6% of the responses. International organizations and academic/consultant institutions were also well represented (28.8% and 11.2%, respectively). Only 9.2% of respondents worked for aid-donor or aid-recipient governments. Table 2 provides a fuller description of the sample.

Table 2. Summary of sample demographics

Completed responses	n	%		
	285	100%		
Gender				
Female	129	45%		
Education				
Undergraduate or less	55	19%		
Graduate	149	52%		
Medical degree	33	12%		
PhD	36	13%		
Other or unknown	12	4%		
Organizational affiliation				
Civil society organization	127	45%		
International organization	82	29%		
Academic/commentator/consultant	32	11%		
Government in receipt of external assistance	13	5%		
Government providing external assistance	13	5%		
Industry	4	1%		
Other	14	5%		
Location of respondents				
	By country of birth		By country of residence	
Total number of countries	89		88	
High income	102	36%	103	36%
Upper-middle income	41	14%	40	14%
Lower-middle income	97	34%	97	34%
Low income	45	16%	45	16%

Findings

Our primary findings are summarized in Table 3, which presents the estimates of the importance of each of the country characteristics in the full sample of responses. Because the coefficients in this model are on a latent utility scale, they are not easily interpretable. However, the relative magnitudes of the coefficients allow us to comment on the relative preferences of respondents with regard to the country characteristics in the frameworks (e.g., we can rank the preferences of respondents with regard to country characteristics).

Table 3. Main effects

Attribute	Coefficient	Std. Err.	p-value
Country income			
Some/medium weight	0.21	0.10	0.04
High weight	0.27	0.10	0.01
Burden of disease			
Some/medium weight	0.84	0.12	0.00
High weight	1.23	0.10	0.00
Strength of health system			
Some/medium weight	0.43	0.15	0.00
High weight	0.83	0.09	0.00
Level of health inequality			
Some/medium weight	0.91	0.08	0.00
High weight	1.31	0.10	0.00
Number of observations		5130	
Number of responses		2565	
Number of respondents		285	
Number of responses per respondent		9	
Log likelihood		-1441	
Pseudo R2		0.1894	

Overall, the strongest predictor of framework choice was whether the frameworks in question assigned high importance to health inequalities. This suggests that when a framework had this attribute, it greatly increased the chances that it would be selected as the preferred framework in a set of framework choices. Almost as influential on framework choice was whether the framework in question assigned high importance to the level of disease burden. Also influencing framework choice was whether the frameworks in question assigned medium level of importance to either health inequalities or burden of disease, or high or medium importance to the strength of the health systems. Respondents placed very little weight on the importance that the framework attributed to income per capita. In our view, these findings suggest that with respect to external financing for health, most respondents wanted to prioritize countries with large health inequalities and burden of disease.

In Table 4, we show how these preferences varied by whether a person was born in a HIC or either a LIC or MIC (according to the World Bank classification for 2014). Respondents who were born in what is now a HIC were more likely to select frameworks assigning medium or high importance to health inequalities. These respondents also placed considerable weight on frameworks that assigned a high level of importance to the burden of disease or either a medium or high level of importance to the strength of the health system. Respondents born in HIC placed nearly no weight on the income level of the country.

Respondents born in what are now LICs or MICs had different preferences. The framework attribute that most heavily influenced their choice of framework was high importance assigned to level of burden of disease, followed by high importance assigned to level of health inequalities

and medium importance assigned to burden of disease. Using an interaction model, we also assessed whether there were overall differences between people born in HICs and those born in LICs or MICs, and many of these findings were confirmed. Respondents born in HICs were much more likely to choose frameworks that assigned medium or high importance to health inequalities. These respondents were also more likely to choose frameworks that assigned medium or high importance to the strength of the health system than were people born in LICs or MICs (results not shown).

Table 4. Main effects by country of birth

Attribute	High-income country			Low- and middle-income country		
	Coefficient	Std. Err.	p-value	Coefficient	Std. Err.	p-value
Country income						
Some/medium weight	0.11	0.17	0.50	0.27	0.13	0.04
High weight	0.07	0.16	0.67	0.39	0.13	0.00
Burden of disease						
Some/medium weight	0.74	0.20	0.00	0.91	0.15	0.00
High weight	1.03	0.16	0.00	1.36	0.13	0.00
Strength of health system						
Some/medium weight	0.92	0.26	0.00	0.19	0.18	0.29
High weight	0.94	0.16	0.00	0.79	0.11	0.00
Level of health inequality						
Some/medium weight	1.21	0.14	0.00	0.77	0.09	0.00
High weight	1.59	0.18	0.00	1.18	0.13	0.00
Number of observations		1836			3294	
Number of responses		918			1647	
Number of respondents		102			183	
Number of responses per respondent		9			9	
Log likelihood		-501			-932	
Pseudo R2		0.2126			0.1835	

Interpretation

Overall, the findings suggest that (1) most respondents wanted to prioritize countries with large health inequalities and burden of disease for external financing for health, (2) respondents were least concerned with income per capita, and (3) preferences varied across countries and especially across country income class.

The survey findings align with the interview findings. In the interviews, respondents were asked about the importance of needs and health needs, capacity, expected impact, and inequality. The respondents considered needs and health needs to be the most important, followed by inequality, and they considered capacity to be the least important.

The survey findings informed other parts of our approach, including the design and assessment of frameworks. Preferences regarding framework attributes can be used to suggest the importance people attribute to different indicators. This is the case even though the online survey focused on generic country characteristics rather than fully specified indicators. In addition, variations in preferences across respondents from different countries can inform

donors who want to be sensitive to such variation or want to focus on preferences specific only to certain countries.

The findings can also inform the specification of illustrative weights in the design of candidate frameworks. Although the coefficients from the conditional logit model are not directly applicable as weights in such frameworks, we believe the survey provides helpful information about the relative importance of indicators within frameworks. We believe that the survey suggests the following ranking of country characteristics in terms of their importance for external financing for health: health inequalities > burden of disease > strength of health system > income per capita. Based on this ranking, we assigned a weight of 0.4 to health inequalities, 0.3 to burden of disease, 0.2 to strength of the health system, and 0.1 to income per capita. We used these weights to design one candidate framework.

4. The starting point: GNI per capita

GNIpc is the natural starting point for the design of frameworks. Along with the related GDPpc, GNIpc is widely used across many settings, widely available, and frequently updated, and it is associated with many factors generally considered relevant for external financing for health.

GNIpc is the gross national income (GNI) divided by the midyear population. As defined by the World Bank, GNI is the sum of value added by all resident producers plus any product taxes (less subsidies) not included in the valuation of outputs plus net receipts of primary income (compensation of employees and property income) from abroad. GDP differs from GNI mainly because it is directly concerned with economic activity within the country's territory rather than the economic activity of the country's national citizens. However, the two measures overlap almost completely, and the differences between them will thus not be emphasized in this report.

GNIpc and GDPpc have been widely used as metrics of development and social progress.^{25, 26, 38, 39} Likewise, the two indicators have been widely used as a criterion for countries' eligibility for aid and for the allocation across countries deemed eligible.^{16, 32, 40-42} Among other uses, GNIpc and GDPpc have been used as indicators of unmet needs, lack of domestic capacity to address those needs, or both.^{40, 41} Moreover, GNIpc and to some extent GDPpc have played a similar role in the specific context of external financing for health.^{13, 32, 40, 43, 44} Here and elsewhere, one of the most widely used classifications of countries is the World Bank's income classification.⁴⁵ This classification system is based on GNIpc, and GNIpc thus exerts considerable influence through its use.

GNIpc has a range of attractions as a criterion for external financing. Much work has been devoted to the development of the GNI and GDP, and, as mentioned above, these measures are widely available, frequently updated, and associated with many aspects generally considered relevant for external financing for health. Widespread use has also made the measures familiar to policy makers and the general public.

Despite these attractions, GNIpc has been heavily criticized as a criterion for external financing, and it is widely acknowledged that GNIpc has a number of shortcomings, especially when used in the absence of other criteria. This may be unsurprising given that this measure was not designed for this purpose. One limitation is that factors other than GNIpc and GDPpc are seen as relevant for countries' capacity to address domestic needs.¹⁴⁻²⁰ For example, factors such as institutional quality, level of corruption, and level of conflict can also be important. Another limitation is that GNIpc and GDPpc are not seen as adequately reflecting countries' level of unmet human needs.^{10, 12, 21-26} One reason for this is simply that human needs and human well-being also depend on other factors. The failure of GNIpc to fully capture health needs is a chief driver behind the many calls to go beyond this measure when making decisions for external financing for health. In addition, a country's GNIpc and GDPpc are average measures and thus do not directly reflect the distribution of income in the country.

These criticisms motivate an exploration of indicators to supplement or replace GNIpc. Given the current role of GNIpc, we believe supplementing it is more realistic and useful, at least in the short to medium term. We therefore used GNIpc as the starting point for our analyses and investigated ways to supplement GNIpc in order to devise a better framework for classifying

countries. This is a useful analytical approach even if one eventually places greater weight on indicators other than GNIPC.

In addition to the criticisms of GNIPC per se, there have also been criticisms of the thresholds used in GNIPC-based classifications of countries. Some have argued that these thresholds are too arbitrary to be useful, whereas others have argued that the thresholds are set at incorrect levels.^{10, 21, 23, 41, 42} We will not directly address the question of thresholds in this report. However, the general issue of GNIPC is obviously fundamental to the issue of GNIPC thresholds.

5. Health needs

Health needs is defined as a population's or subpopulation's need for improvement in health or the determinants of health, including health services.

Role of health needs

Improving health is typically considered the primary purpose of external financing for health, and most discussions of external financing for health have focused on addressing health needs and reducing disease burden.^{11, 29, 46} The concern for health needs suggests that external financing is more important in countries where health needs are greater. The necessity to go beyond GNIpc and make country classification more sensitive to health needs motivated the establishment of the EAI.⁴⁷

Some donors focus on specific types of health needs. For example, the Global Fund was established to address unmet health needs associated with HIV/AIDS, tuberculosis, and malaria. The mandate of Gavi, on the other hand, is to address unmet health needs associated with diseases preventable by immunization. As a result, donors vary in the weight they attribute to different types of health needs.

Selection of health-need indicators

From our identified set of health-need indicators, we selected the following for integration into a framework:

- Under-five mortality rate (U5MR)
- Under-sixty mortality rate (U60MR)
- Life expectancy (LE)
- Healthy life expectancy (HALE)
- Disability-adjusted life year rate (DALYR)
- Age-standardized disability-adjusted life year rate (DALYR_AS)

These are broad health-outcome measures. Each of the indicators were systematically assessed in terms of relevance, salience, validity, consistency, and availability and timeliness. We also examined three disease-specific indicators: maternal mortality ratio, HIV prevalence rate, and tuberculosis prevalence rate. The design and assessment of frameworks based on these indicators are described in the Appendix A.

Under-five mortality rate

U5MR is the probability per one thousand that a newborn baby will die before reaching age five, if subject to the age-specific mortality rates of the specified year. Estimates have been developed and are published by both the UN Inter-agency Group for Child Mortality Estimation and the IHME. As an outcome measure, U5MR is sensitive to a wide range of conditions, but it is restricted to those conditions that influence mortality before the age of five.

Under-sixty mortality rate

U60MR is here defined as the probability of dying between the ages of fifteen and sixty, if subject to age-specific mortality rates of the specified year. U60MR estimates are provided by the Population Division at the UN Department of Economic and Social Affairs (2015). U60MR as defined here focuses only on adult mortality and thus partly covers conditions other than U5MR. For example, this indicator would be more sensitive to premature mortality associated with HIV than the U5MR. In our context, a measure of U60MR that also included deaths below the age of fifteen is more relevant than one that captures adult mortality only. Such a measure would come closer to the under-seventy mortality rate, which has been suggested as a quantitative health target for the Sustainable Development Goals.³

Life expectancy

LE at birth indicates the number of years a newborn infant would be expected to live if prevailing patterns of age-specific mortality at the time of its birth were to apply throughout its life. LE estimates are readily available from several sources, including the UN Population Division. The LE indicator is widely used, and most people are familiar with it. LE covers a wider range of conditions than both U5MR and U60MR, because it is sensitive to deaths at all ages.

Healthy life expectancy

HALE combines mortality and nonfatal outcomes (i.e., morbidity) into a single measure of population health. HALE estimates are provided by both the IHME and the WHO. The methods for combining mortality and morbidity into one measure are controversial and have been altered over time.^{5, 48-50} Among the central issues are the quality of data used to generate estimates of morbidity and the methods employed to generate disability weights. HALE estimates have been provided irregularly up to 2015, but the IHME has announced that it will now be updating these annually.

Disability-adjusted life years

DALYR is the crude burden of DALYs per 100,000 individuals. DALYs is the sum of years of life lost due to premature mortality (YLLs) and years lived with disability (YLDs). DALY estimates are provided by the IHME. By representing disease burden, DALYR provides an intuitive estimate of total unmet health needs in a country. An important property of crude DALYR is that it is highly sensitive to the age structure of the population. This may be seen as an advantage by some and a disadvantage by others. DALYR estimates have been provided irregularly up to 2015, but the IHME has announced that it will now be updating these annually.

Age-standardized disability-adjusted life years

DALYR_AS is similar to crude DALYR, with one important exception. It uses a model-country age structure to generate age-standardized estimates that allow for comparisons across countries that are not sensitive to the age structure of the populations. To construct DALYR_AS, age-specific DALYR is first calculated for five-year age categories. Then, weights from a standardized population are applied to generate DALYR_AS. Age-standardized DALY estimates are provided by the IHME.

Prioritization rules used

The following simple prioritization rules were chosen for each of the indicators. Prioritization rules specify how country rank changes as the indicator increases.

- U5MR: Monotonically increasing. Higher levels of U5MR suggest countries are worse off, and they thus get a higher rank.
- U60MR: Monotonically increasing. Higher levels suggest countries are worse off, and they thus get a higher rank.
- LE: Monotonically decreasing. Higher levels suggest countries are better off, and they thus get a lower rank.
- HALE: Monotonically decreasing. Higher levels suggest countries are better off, and they thus get a lower rank.
- DALYR: Monotonically increasing. Higher levels suggest countries are worse off, and they thus get a higher rank.
- DALYR_AS: Monotonically increasing. Higher levels suggest countries are worse off, and they thus get a higher rank.

Implications for individual countries

As described above, we designed frameworks consisting of GNIpc and one health-need indicator. We subsequently assessed the implications of these frameworks.

General trends

General trends are reflected in the Spearman coefficients for correlations between GNIpc and each of the health-need frameworks. The Spearman correlation coefficients were all fairly high, ranging from 0.89 (U60MR), via 0.90 (LE) and 0.92 (DALYR), to 0.93 (U5MR, HALE, and DALYR_AS), as shown in Appendix B. It is important to note that these coefficients reflect the correlation between GNIpc and *frameworks*, not GNIpc and another indicator. The fairly high coefficients are of course partly due to the fact that GNIpc is itself part of the health-need frameworks, as well as the fact that richer countries tend to be healthier. Nevertheless, the coefficients are low enough to suggest that the difference between GNIpc alone and the health-need frameworks is material and potentially significant (in term of consequences for the countries involved). Appendix B also shows Spearman coefficients for the correlation between GNIpc and the health-need *indicators*.

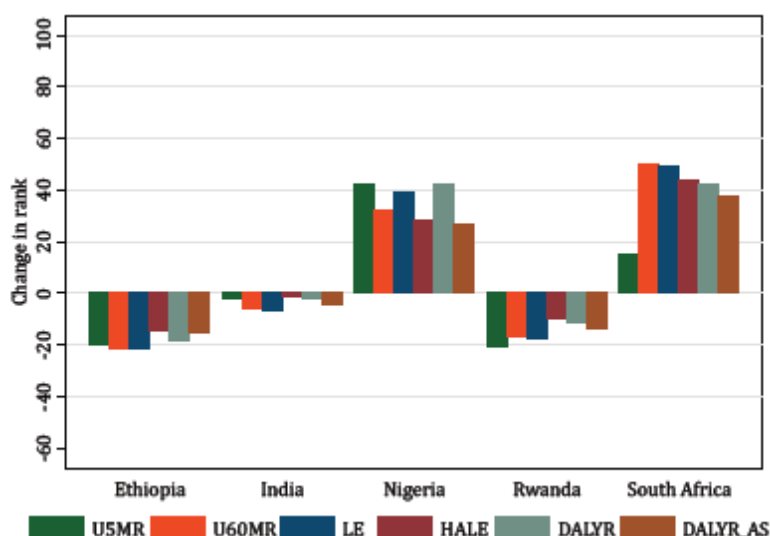
We also examined differences across the health-need frameworks themselves. The Spearman coefficients for the correlation between any two of these frameworks fell in the range of 0.93 (U5MR vs. U60MR) to 1.00 (HALE vs. DALYR_AS), as shown in Appendix B. These high scores suggest that which of the health-need indicators one chooses to integrate with GNIpc makes less difference than whether or not such an indicator is included in the framework at all.

Changes for specific countries

Although general trends are informative, it is also important to examine the underlying variation across countries. Figure 6 shows how the rank of the five focus countries changes as one moves

from GNIpc alone to health-need frameworks. Changes for all countries are presented in Supplementary File 5.

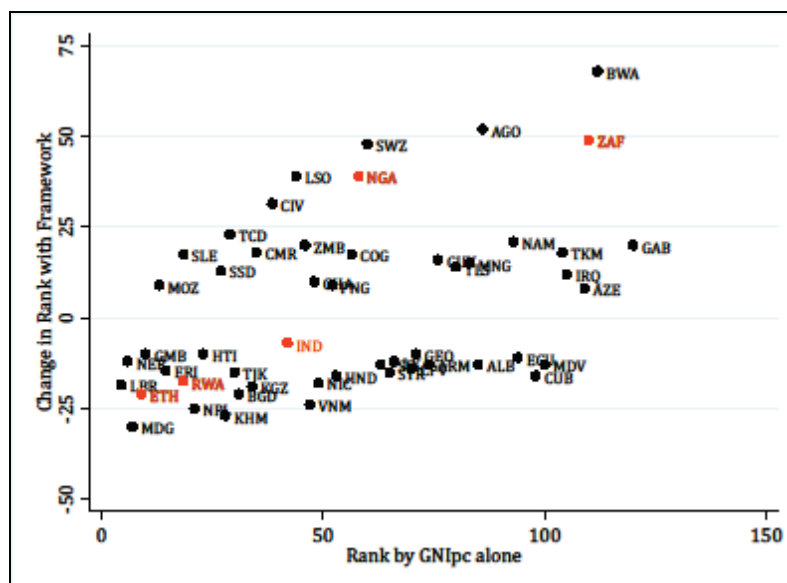
Figure 6. Changes in rank when moving from GNIpc alone to health-need frameworks (focus countries)



As the figure illustrates, Nigeria and South Africa get a substantially higher rank and thus priority, whereas Ethiopia and Rwanda get a significantly lower rank. Moreover, the differences across health-need frameworks are quite limited for the most part. The U5MR framework for South Africa is an exception, which is probably due to the country's high rate of adult mortality from HIV/AIDS.

Figure 7 shows the change in rank when moving from GNIpc to a LE framework for a larger set of countries. We focused on this framework for the reasons outlined below. The x-axis shows the country ranking based on GNIpc alone. The y-axis shows the countries' change in rank when moving from GNIpc to a LE framework. We show only the 20% of countries with largest increases in rank and the 20% of countries with the largest decreases in rank. The focus countries are highlighted in red. The three-digit country codes are presented in Supplementary File 2.

Figure 7. Changes in rank when moving from GNlpc alone to a life-expectancy framework (selected countries)

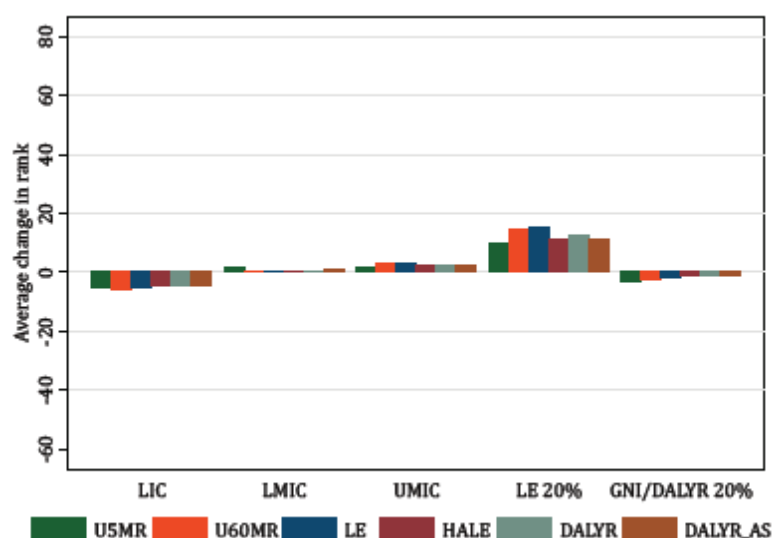


The figure shows that many countries experience substantial shifts in rank when moving from GNlpc to a LE framework. Botswana (BWA) gets the largest increase in rank due to its fairly high GNlpc (\$7,500) combined with a very low LE of 57 years. Madagascar (MDG) experiences the largest decrease in rank. This is due to a very low GNlpc (\$440) combined with a fairly high LE of 65 years. Highlighted in red are the five focus countries shown in Figure 6, Ethiopia (ETH), India (IND), Nigeria (NGA), Rwanda (RWA), and South Africa (ZAF).

Implications for country categories

It is also helpful to examine implications for country categories. Figure 8 shows the average changes in rank for countries in the five categories as one moves from GNlpc alone to health-need frameworks.

Figure 8. Average changes in rank for countries when moving from GNIpc alone to health-need frameworks



LICs on average get a somewhat lower rank irrespective of the health-need framework when moving from GNIpc alone. Conversely, LMIC and UMIC groups get a higher rank from any move away from GNIpc alone, but this increase is very small. At the same time, the variation across countries within the categories can be profound, as suggested by Figures 6 and 7. The countries with the lowest life expectancies exhibit a quite substantial increase in rank, again across all health-need indicators. This underscores the unsurprising but important point that if one integrates a health-need indicator, the countries with the poorest health outcomes will benefit. For the countries with the lowest GNI-DALYR ratios, however, the inclusion of a health-need indicator leads to a slight decrease in rank.

Assessment

The comparison of implications provides two broad findings. One pertains to health-need indicators in general. The inclusion of a health-need indicator in a framework does make a difference in country rankings, and it does so in a way that is likely to be considered relevant by most actors concerned with external financing for health. The other broad finding pertains to the difference among the health-need indicators examined. Put crudely, it doesn't matter much which one is chosen. However, some institutions may of course prefer one indicator over others based on their own needs and objectives.

Deciding whether to include a health-need indicator

One of the broad findings is that integrating a health-need indicator does significantly change the rank of many individual countries, compared to the ranking from GNIpc alone. This is the case even though the changes for country grouping are themselves quite limited. This suggests that the inclusion of a health-need indicator does allow differentiation between countries with similar levels of income but different levels of unmet health needs. The extent to which changes in rank represent changes in donors' decisions on external financing for health depends on several factors. For example, if a donor uses an aid-eligibility threshold, changes in the absolute

level of the indicator or index, and particularly such changes near the threshold level, are likely to be more important than rank per se.

The interviews, the survey, and other consultations all suggest that stakeholders generally consider health needs crucial for decisions on external financing for health and, more specifically, for a classification framework. All the health-need indicators examined are also available for almost all LICs and MICs, after imputations, as shown in Appendix C. These indicators generally scored high on the criteria of relevance, salience, validity, consistency, and availability and timeliness.

Overall, these findings make a strong case for including a health-need indicator in frameworks for initial classification of countries – provided this aligns with organization-specific needs and objectives. Some institutions may be more interested in disease-specific indicators (see Appendix A).

Choosing among specific health-need indicators

The second broad finding is that the differences in rankings across health-need frameworks are quite limited, as suggested by the high Spearman coefficients for correlation among health-need frameworks. Moreover, Figures 6 and 8 show for the most part small differences between health-need frameworks, compared to the difference between GNIpc and each of these frameworks. Again, there are some exceptions, as in the case of U5MR for South Africa.

The existing literature provides only limited discussions of the selection of generic health-need indicators for the specific purpose of external financing for health. We have emphasized the criteria of relevance, salience, validity, consistency, and availability and timeliness. Our discussions of health-need indicators, among ourselves and others, have highlighted four specific issues related the choice of health-need indicators: morbidity, decomposability, age weighting, and age standardization.

Morbidity

The first issue is whether an indicator directly accounts for nonfatal health outcomes. The main argument for including morbidity is that quality of life, not just length of life, is highly important to people. However, it could be argued that including morbidity makes the indicator unnecessarily complex; that nonfatal outcomes are not as well measured as mortality; that it involves controversial trade-offs between quality and length of life; that the purpose of external financing for health is primarily to save lives; that mortality and morbidity are closely correlated; and that emphasis on morbidity will unduly increase the priority assigned to UMICs.

Decomposability

The issue of decomposability pertains to whether the indicator can be decomposed with regard to factors found relevant to external financing for health. These factors may include cause of disease, risk factor, age, age structure, and geographical area. Decomposability can be useful if the donor primarily seeks to target only part of what is captured by the indicator – for example, only certain causes of death. It can be particularly useful if the donor's target varies across the settings in which it operates.

Age weighting

The issue of age weighting pertains to the weights attributed to loss of life years at different ages. Specifically, people may have different views about how loss of life years at different ages should influence a health-need indicator that is supposed to guide external financing for health. For example, postponing death from the age of thirty to the age of forty may be seen as a greater or otherwise more important reduction in unmet health needs than postponing death from the age of seventy to the age of eighty. Different health-need indicators account for such differences in various and sometimes subtle ways. The choice of indicator will thus determine how much weight is placed on health outcomes occurring at different ages.

Age standardization

The issue of age standardization pertains to the extent to which donors wish to take into consideration differences in age structures across countries. With regard to deaths, the overall, crude mortality rate in a population is influenced by both its age structure and the age-specific mortality rates. Some donors may want to take into account differences in age structure, because it is a key determinant of the actual burden of disease in a country. Other donors may want to disregard differences in age structure, either because they think these are not relevant per se or because they think the implications of doing so is favorable for other reasons. Depending on the circumstances, when the indicator is sensitive to actual differences in age structure, it may assign higher priority to countries with a high proportion of deaths at higher ages and with a high burden of certain kinds of diseases, such as noncommunicable diseases.

Overall

We believe that all the criteria and issues discussed here should be carefully considered when choosing a health-need indicator for external financing for health. Age weighting and age structure in particular seem to have received only limited consideration in this context.

Our initial assessment suggests that no single health-need indicator is clearly superior to all the others and that an institution's choice among health-need indicators will depend strongly on organization-specific needs and objectives. Although we cannot recommend any single indicator as appropriate for all institutions, we recommend that all institutions seeking to adopt a generic health-need indicator to at least consider LE. This indicator can be relevant for a wide range of institutions for multiple reasons. One is that LE is already a familiar and well-tested measure in numerous contexts, including as part of the Human Development Index. LE data are provided by several reliable institutions, available for all countries, and frequently updated. LE is also sensitive to all deaths and thus all causes of death. Overall, it can be regarded as the most dependable generic health-need indicator currently available. In addition, LE is sensitive to the age of death in a way that implicitly assigns greater weight to deaths at lower ages. For these reasons, we use LE whenever space only permits an emphasis on a single health-need indicator.

At the same time, we find that crude and age-standardized DALYR have several attractive properties. Although these indicators lack some of the features of LE just listed, they share with LE a sensitivity to all causes of death and a greater sensitivity to deaths at lower ages. At the same time, DALYR exhibits features that LE lacks. Among them are direct sensitivity to morbidity and the ability to further decompose health needs with regard to cause, risk factor, and age structure. Compared to LE, crude DALYR also has a greater sensitivity to differences in age structure across countries.

6. Domestic capacity

Domestic capacity is defined as the capacity of countries to address domestic needs without external financing. This is in line with many other definitions of capacity, and capacity so understood relates to ability to pay and lack of constraints.^{17, 18, 21, 51} ^a It also relates to the fiscal space for health, which can be understood as a country's ability to increase health spending without affecting other expenditures necessary for development and long-term solvency.¹⁹

GNlpc is commonly seen as an important indicator of domestic capacity, but also often seen as not being sufficiently sensitive to all relevant aspects of capacity.¹⁴⁻²⁰ We therefore examined indicators that can supplement GNlpc in order to devise frameworks that more fully capture capacity.

Role of capacity

In the context of external financing for health, the concern for capacity suggests that external financing is more relevant when a country's domestic capacity is lower. One reason for this is simply that countries with a lower capacity to make progress without external financing have a greater need for such funding – under a broader understanding of need than the one used above.^{22, 31, 32} Another reason is that it is inadvisable to weaken capable countries' incentives to address domestic needs through domestic means. If countries qualify for external financing for health irrespective of their capacity to address domestic health needs, they may not be motivated to invest sufficiently in health and health services.^{52, 53} It has been argued that reducing support to countries whose governments fail to use the country's capacity unfairly punishes needy individuals for factors beyond their control. At the same time, holding governments responsible in this way may be best for the world population overall if it helps preserve incentives. How to deal with domestic capacity is a central issue in the debate over the role of MICs in external financing for health.^{12, 21}

Selection of capacity indicators

The capacity of countries to address domestic health needs depends on a wide range of factors beyond GNlpc. We examined the following indicators:

^a *Capacity* is sometimes rather referring to absorptive capacity, the capacity to benefit, or the capacity of the country to use external financing effectively. However, to avoid conflating two different forms of capacity, we reserved the term *capacity* to describe the capacity of a country to meet needs without external financing.

- Debt service (Debt)
- Tax ratio (Tax)
- Total health expenditure per capita (THEpc)
- Government health expenditure per capita (GHEpc)

Measures of health expenditures were included because they may point to the countries' capacity to address health needs in the shorter term. The Country Policy and Institutional Assessment (CPIA) rating was not included due to lack of data.

Debt service

Total debt service is the sum of principal repayments and interest actually paid in currency, goods, or services on long-term debt, interest paid on short-term debt, and repayments (repurchases and charges) to the International Monetary Fund (IMF). We examined total debt service as a percentage of GNI.

A high debt-service level reduces domestic capacity to meet health needs, although the debt may have financed capacity-building activities. A variety of debt and debt-servicing ratios are available, and these have been variously employed to assess fiscal space for health.¹⁹ One measure with quite widely available data is total debt service as a percentage of GNI. This is a comprehensive measure in that it seeks to reflect all debt and to compare it with a broad measure of income, namely GNI.

Tax ratio

Tax revenue consists of compulsory transfers to the central government for public purposes. Certain compulsory transfers such as fines, penalties, and most social security contributions are excluded. Refunds and corrections of erroneously collected tax revenue are treated as negative revenue. We examined tax revenue as a percentage of GDP.

Total health expenditure per capita

Total health expenditure is the sum of public and private health expenditures. We examined total health expenditure per capita (THEpc).

Government health expenditure per capita

Government or public-health expenditure consists of recurrent and capital spending from government (central and local) budgets, external borrowings and grants (including donations from international agencies and nongovernmental organizations), and social (or compulsory) health insurance funds. We examined government health expenditure per capita (GHEpc).

Prioritization rules

The following simple prioritization rules were chosen for each of the indicators. Prioritization rules specify how country rank changes as the indicator increases.

- Debt: Monotonically increasing. Higher levels of debt service suggest that countries have less capacity to meet domestic health needs, and they thus get a higher rank.
- Tax: Monotonically decreasing. Higher tax ratio suggests that countries (governments) have greater capacity to meet domestic needs, and they thus get a lower rank.
- THEpc: Monotonically decreasing. Higher levels suggest that countries have greater capacity to meet domestic health needs in the shorter term, and they thus get a lower rank.
- PHEpc: Monotonically decreasing. Higher levels suggest that countries (the public-health systems) have greater capacity to meet domestic health needs in the shorter term, and they thus get a lower rank.

Implications for individual countries

We designed and subsequently assessed frameworks consisting of GNIpc and one capacity indicator.

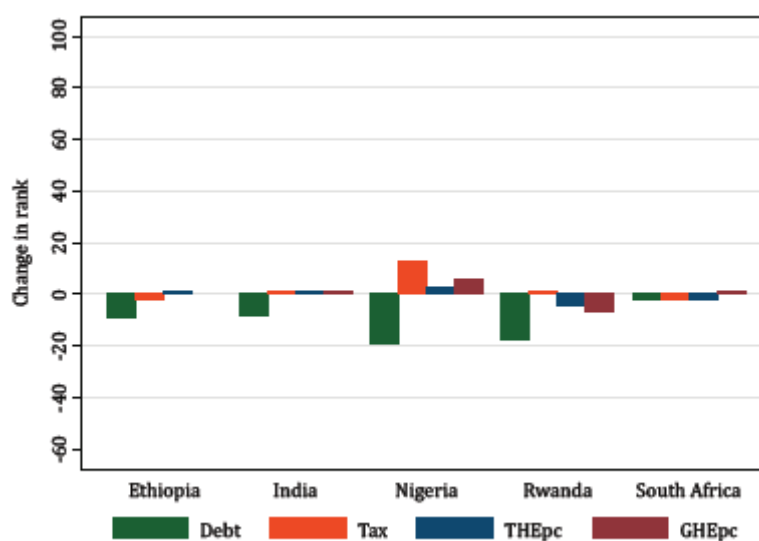
General trends

The Spearman coefficients for correlation between GNIpc and each of the capacity frameworks were for the most part high, with the exception of the debt framework. Specifically, the coefficients were 0.73 for the debt framework, 0.96 for the tax framework, and 0.99 for the THEpc and GHEpc frameworks. These are shown in Appendix B together with Spearman coefficients for the correlation between GNIpc and each of the indicators in question.

Changes for specific countries

We also examined the implications for individual countries. Figure 9 shows how the rank of the five focus countries changes as one moves from GNIpc alone to capacity frameworks. Changes for all countries are shown in Supplementary File 5.

Figure 9. Changes in rank when moving from GNIpc alone to capacity frameworks (focus countries)

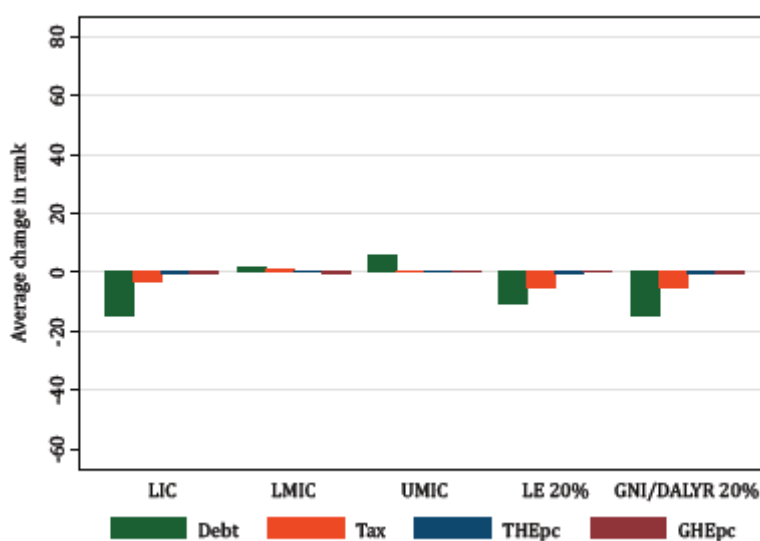


For capacity frameworks, there are more limited changes and more variation for specific countries than there were for the health-need frameworks. The most pronounced changes follow from the move from GNIpc alone to the debt framework – a move that results in all five countries receiving a lower rank. At the same time, Nigeria experiences a change in the opposite direction when moving to the tax framework. These implications are driven by the fact that Nigeria has a low level of debt service relative to GNI (0.1%), but also a very low tax ratio (1.6% in 2012).

Implications for country categories

Figure 10 shows the average change in rank of countries in the five categories as one moves from GNIpc alone to capacity frameworks.

Figure 10. Average changes in rank for countries when moving from GNIpc alone to capacity frameworks



Moreover, the figure shows that for the country categories, the main changes follow from the debt and, to a lesser extent, the tax framework. UMICs get on average a slight increase in rank from the debt framework, whereas the LICs and the countries with the lowest LE and GNI-DALYR ratio experience a decrease in rank.

Assessment

The Spearman correlation coefficients suggest that country rankings change quite substantially with the debt framework and not so much with the other frameworks. In any case, none of the capacity indicators (beyond GNIpc itself) is clearly suitable for inclusion in the framework for the initial classification of countries. This may be particularly evident if one is seeking a framework that can be applied across a wide range of situations and by a wide range of institutions. None of the capacity indicators perform consistently well against the criteria of relevance, salience, validity, consistency, and availability and timeliness. We discuss some challenges for specific indicators below, emphasizing the challenges to their relevance in the initial classification of countries.

Data on total debt service for 2013 were missing for 12% of LICs and MICs, after imputations, as shown in Appendix C. Another potential problem with including this indicator as a measure of capacity is that a high level may represent investments that improve the capacity to meet health needs, at least in the longer term. A third challenge is that debt level is at least partly a matter of government choice. If priority for external financing increases with debt service, there will be additional incentives to take on more debt. Fourth, one may want to focus on government debt rather than total debt. However, data on government debt service are currently even less readily available.

Data on tax ratio for 2013 were missing for as much as 23% of LICs and MICs, after imputations, as shown in Appendix C. For a given GNIpc or GDPpc, the tax ratio indicates something about the government's short- to medium-term capacity to address health needs. Other things being equal,

a lower tax ratio indicates a lower capacity on behalf of the government. However, a low tax ratio may be seen as partly or primarily the result of the government exerting insufficient effort in revenue mobilization. Especially in these cases, linking a low tax ratio to higher priority for external financing can create perverse incentives. From this perspective, tax ratio may be seen as having no more to do with external constraints on capacity – especially longer-term capacity – than how the government chooses to exert the capacity it has. At the same time, tax ratio is potentially problematic as an indicator of capacity from a short-term perspective. Even a fairly high GNIpc and tax ratio can be accompanied by a low short-term capacity to address health needs if only a small share of the government revenue is spent on improving health or if the spending is inefficient.

Data on THEpc and GHEpc are quite widely available. For 2013, data were missing for 6% of LICs and MICs, after imputations, as shown in Appendix C. By providing information on health spending, the two indicators shed light on the short-term capacity of countries and their governments to address health needs. One problem for comparison across countries is variation in the definition of health spending. Another relevant problem is analogous to that for tax ratio. THEpc and GHEpc may be seen as having more to do with what the country and government chooses to do with its capacity than true external constraints. THE and GHE normally represent quite small shares of government spending, and there are often less powerful actors advocating for sustained or increased spending for health than for many other areas. Especially against this background, it is easy to see how using THEpc and GHEpc as capacity indicators can create perverse incentives. If low levels of THEpc or GHEpc imply more external funding, governments may be discouraged to spend domestic funds on health. For some kinds of decisions, it can also be problematic that THEpc and GHEpc may partly include external financing for health.

None of these issues represent a decisive reason for not including the indicator in question in the primary classification framework. However, overall, we believe there is no clear case for integrating any of these indicators into a framework for initial classification of countries. For THEpc and GHEpc, this may be supported by the finding that rankings are quite unaffected by the move to these frameworks. This does not mean that any of the indicators are unimportant. It simply means they may better play other roles in the decision-making process than guiding the initial classification. This view of the role of capacity received some explicit support in the interviews. For example, one interviewee pointed out that capacity has greater significance at the operational level. Based on the preceding analysis, we suggest that capacity indicators other than GNIpc may have a role to play primarily at decision-making stages downstream from the initial classification of countries. The capacity indicators can inform copaying requirements, other preconditions, and *how* donors provide external financing for health, including modality and channel.

7. Expected impact

Expected impact is defined as the expected impact of external financing in terms of changes in health or the determinants of health, including health-service coverage. Expected impact so understood can overlap with effectiveness, efficiency, value for money, and performance.^{32, 40}

Role of expected impact

It is widely agreed that external financing for health should be directed to countries where it will have the greatest impact, other things being equal. Impact can be understood in terms of change in various health outcomes, including change in mortality rates and life expectancies. One may also consider the impact on health-service access and coverage.

Selection of impact indicators

There is limited agreement regarding which country characteristics are useful indicators of the impact of external financing in general or external financing for health.^{30, 54} One family of indicators consists of measures of past performance in terms of demonstrated improvements in health or health-service access. The basic idea is that greater improvements in the recent past indicate greater expected impact for the future. We examined the following indicators:

- Absolute improvement in under-five mortality rate (cU5MR)
- Relative improvement in under-five mortality rate (rcU5MR)
- Absolute improvement in skilled birth attendance rate (cSBA)
- Relative improvement in skilled birth attendance rate (rcSBA)

All these indicators pertained to improvements in the period 2003–2013. The Country Policy and Institutional Assessment (CPIA) rating would also have been a candidate indicator of expected impact, but it was not included due to lack of data. Some have suggested that fragility, conflict, and corruption are correlated with lower impact of external funds, at least in the shorter term.⁵⁵ Another suggestion put forward is that MICs may be more efficient targets of aid than LICs.^{21, 23, 24, 56}

Prioritization rules

For all four indicators, we used a monotonically increasing prioritization rule. The idea underlying this rule is that greater improvement in the past suggests higher impact from external financing for health in the future, and the countries in question thus get a higher rank.

Implications for individual countries

We designed and subsequently assessed frameworks consisting of GNIpc and one impact indicator.

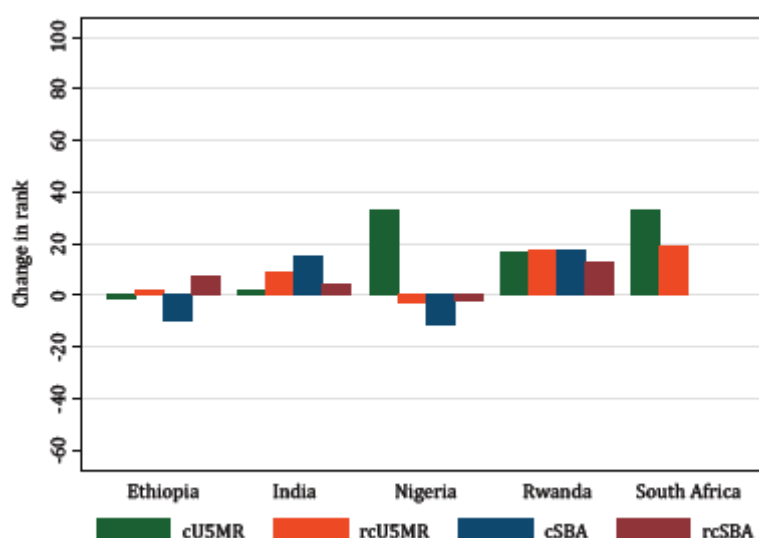
General trends

The Spearman coefficients for correlation between GNIpc and each of the impact frameworks were all high or fairly high, ranging from 0.91 (rcU5MR and cSBA), via 0.94 (cU5MR), to 0.98 (rcSBA), as shown in Appendix B. These scores suggest that at least with the former two frameworks, there are material and potentially significant changes in country ranking.

Changes for specific countries

We also examined the implications for specific countries. Figure 11 shows how the rank of the five focus countries changes as one moves from GNIpc alone to impact frameworks. Changes for all countries are shown in Supplementary File 5.

Figure 11. Changes in rank when moving from GNIpc alone to impact frameworks (focus countries)

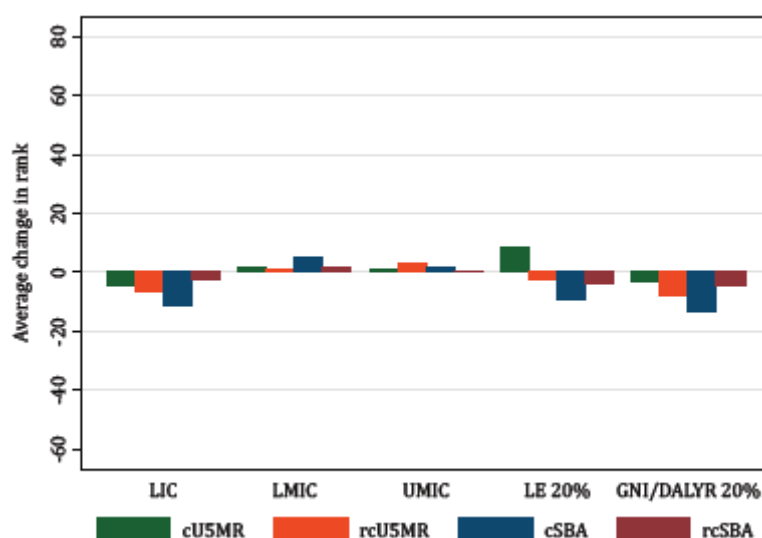


Rwanda gets a consistently higher rank because of substantial improvements in both U5MR and SBA. Specifically, from 2003 to 2013, U5MR was reduced from 139 to 48 deaths per one thousand live births, and the skilled birth attendance rate was improved from 36% to 94% of deliveries being aided by skilled staff. South Africa also gets an increase in rank when moving to any of the frameworks related to U5MR, but the country experiences no change in rank with regard to SBA.

Implications for country categories

Figure 12 shows the average changes in rank for countries in the five categories as one moves from GNIpc alone to impact frameworks.

Figure 12. Average changes in rank for countries when moving from GNIpc alone to impact frameworks



As the figure shows, the changes in rank are very moderate for the income-class categories. At the same time, the direction of change is consistent across indicators within each income class. LICs get a lower rank, and LMICs and UMICs get a slightly higher rank when moving from GNIpc alone. The countries with the lowest life expectancies and the countries with the lowest GNI-DALYR ratios experience decreases across the board, with only one exception.

Assessment

The Spearman correlation coefficients suggest that there is some room for the impact frameworks to make a significant difference, compared to GNIpc alone. Impact indicators can also influence incentives positively, potentially counteracting disincentives created by the use of some other indicators. Data on improvements in U5MR are quite readily available, as shown in Appendix C.

Despite these features, none of the indicators was clearly suitable for inclusion in a framework for the initial classification of countries when judged against the criteria of relevance, salience, validity, consistency, and timeliness and availability. Again, this may be particularly evident if one is seeking a framework that can be applied across a wide range of situations and by a wide range of institutions. However, not even for institutions focusing on U5MR or SBA are the indicators examined likely to provide trustworthy guidance about expected impact. There are two main reasons for this. One is that the determinants for major improvements in the recent past may have little or nothing to do with external financing in general or the specific target area being considered. Second, even if external financing – or factors that external financing is likely to influence – drove improvements in the past, it will not necessarily do so in the future. The context may have changed, or the indicator for which improvement is sought may have reached a level at which marginal changes are more difficult to achieve.

Against this background, it seems more reasonable to consider expected impact at decision-making stages downstream from the initial classification of countries. As with domestic capacity, this argument received some support in the interviews. Especially for donors funding programs

or projects, the expected impact may be assessed primarily with regard to these programs and projects. Here, one may consider certain country-level indicators, such as level of conflict, institutional quality, and level of corruption. However, program- or project-level indicators, such as those related to the technical merit of the program or project itself, are likely to be at least as important.

8. Other frameworks

We designed and assessed a range of frameworks based on indicators different from those discussed above. These indicators are treated separately because they cut across the three concerns to a particularly high degree. Specifically, we designed and assessed frameworks related to inequality, skilled birth attendance, child immunization, and out-of-pocket payments (OOPs) for health services.

Indicators and prioritization rules

Inequality

Especially from an equity perspective, it is not only countries' *aggregate* need that matters. From this perspective, we are equally concerned with how groups and individuals fare and the inequalities among these.^{6, 8, 57, 58} Against this background, inequality indicators may be useful for decisions concerning external financing for health.

We considered three inequality indicators:

- Inequality in life expectancy (ILE)
- Gini index for income (Gini)
- Income share held by bottom 40% (Income40)

Inequality in life expectancy

There are a few established measures of univariate inequalities in health, beyond simple ratios, for which data are available for a wide range of countries. One candidate measure is inequality in life expectancy (ILE), as prepared for the Human Development Reports.^{59, 60} This measure expresses inequality in distribution of expected length of life based on data from life tables estimated using the Atkinson inequality index. The measure has not been widely utilized beyond the Human Development Reports.

Gini index for income

The Gini index is the one of the most commonly used methods for measuring income inequality, and it has also been used extensively in public health. The Gini index for income measures the degree to which the distribution of income or consumption expenditure among individuals or households within a country deviates from an equal distribution. Although this index expresses inequality in the income dimension, it can be relevant for external financing for health. A greater Gini index for income suggests that a larger proportion of the population may have problems paying for and accessing health services. Just as important, to the extent that income is inversely correlated with health, a greater Gini index suggests larger health needs among parts of the population. Beyond the Gini index for income, there have been attempts to calculate a Gini index specifically for health outcomes.^{61, 62}

Income share held by bottom 40%

Percentage share of income or consumption is the share that accrues to subgroups of a population indicated by deciles or quintiles. The income share held by the bottom 40% is not an inequality measure per se, but it can convey important information about inequality when considered alongside country averages, such as GNIpc. Going beyond inequality per se, the income share held by the bottom 40% can also usefully express subgroup needs more directly.

Prioritization rules

The following prioritization rules were used for each of the indicators. Prioritization rules specify how country rank changes as the indicator increases.

- ILE: Monotonically increasing. Greater inequality suggests greater health needs in parts of the population, and the countries in question thus get a higher rank.
- Gini: Monotonically increasing. Greater inequality suggests lower income and greater needs in parts of the population, and the countries in question thus get a higher rank.
- Income40: Monotonically decreasing. Higher income held by this part of the population suggests lesser needs, and the countries in question thus get a lower rank.

Skilled birth attendance rate

SBA is the percentage of deliveries attended by personnel trained to give the necessary supervision, care, and advice to women during pregnancy, labor, and the postpartum period; to conduct deliveries on their own; and to care for newborns. Skilled birth attendance is an essential health service, and SBA can be seen as an important indicator of health-system strength.

The combining rule used was monotonically decreasing. Accordingly, higher SBA suggests a stronger health system and less need for support, and the countries in question thus get a lower rank.

DTP3 coverage

DTP3 coverage is the percentage of children aged twelve to twenty-three months who received three doses of vaccine against diphtheria, pertussis (or whooping cough), and tetanus before 12 months or at any time before the survey. These child vaccinations are also essential services, and DTP3 coverage can be seen as an important indicator of health-system strength.

The combining rule used was monotonically decreasing. Accordingly, higher vaccination rate suggests a stronger public-health system and less need for support, and the countries in question thus get a lower rank. It is worth noting that some organizations may use DTP3 coverage rates in a very different way, for example by requiring that countries reach a certain minimum level to receive support.

Out-of-pocket payments

OOPs are any direct outlays by households, including gratuities and in-kind payments, to health practitioners and suppliers of pharmaceuticals, therapeutic appliances, and other goods and services whose primary intent is to contribute to the restoration or enhancement of the health

status of individuals or population groups. It is a part of private health expenditure. The level of OOPPs is relevant because it reflects the extent to which the country has achieved universal health coverage (UHC).⁶³ A higher level of OOPP suggests that the country is farther away from providing financial risk protection and thus from achieving UHC.

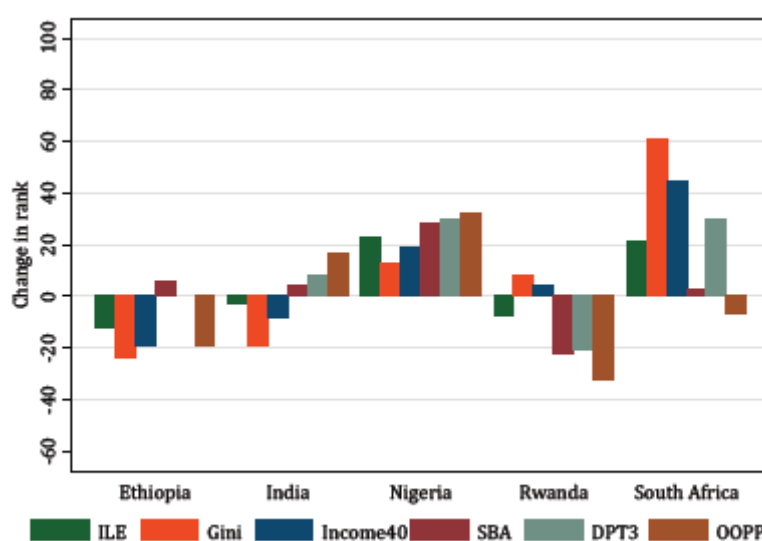
The combining rule used was monotonically increasing. Greater level of OOPPs suggests less financial risk protection and greater need for support, and the countries in question thus get a higher rank.

Implications for individual countries

Spearman coefficients for the correlation between GNlpc and each of the other frameworks varied considerably. For the inequality frameworks, it was 0.74 for Gini, 0.76 for Income40, and 0.93 for ILE. For the three other frameworks, it ranged from 0.82 for OOPP, via 0.93 for SBA, to 0.94 for DTP3. This suggests that the Gini, Income40, and OOPP frameworks are quite likely to generate significant changes in country rankings.

Figure 13 shows how the rank of the five focus countries changes as one moves from GNlpc alone to other frameworks.

Figure 13. Changes in rank when moving from GNlpc alone to other frameworks (focus countries)

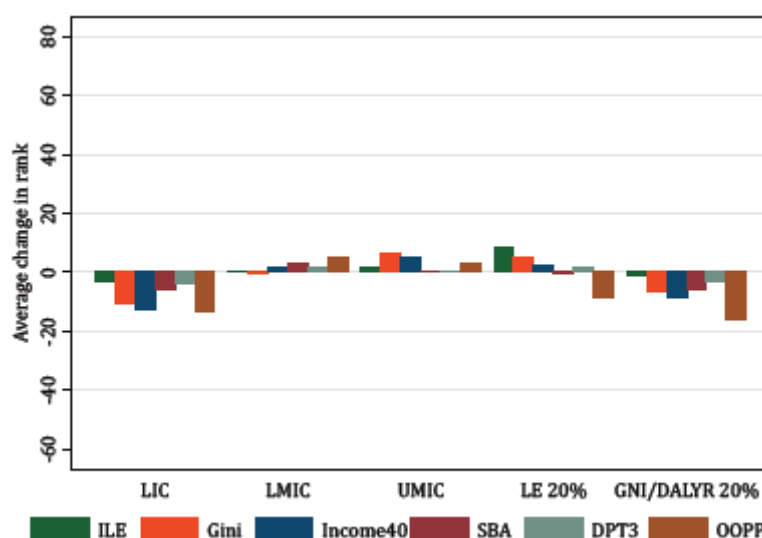


It is evident that South Africa gets a significantly higher rank due to large inequalities and that Nigeria experiences a higher rank as a result of the move to other frameworks. Ethiopia gets a lower rank for the most part, whereas the picture is more mixed for India and Rwanda.

Implications for country categories

Figure 14 shows the average changes in rank for countries in the five categories as one moves from GNlpc alone to other frameworks.

Figure 14. Average changes in rank for countries when moving from GNIpc alone to other frameworks



The figure shows that LICs on average get a lower rank when moving to any of the frameworks. LMICs and UMICs, to the contrary, experience for the most part a slight increase in rank. The countries with the lowest life expectancies get a slightly higher rank when moving to any of the inequality frameworks but a lower rank when moving to the OOPP framework.

Assessment

Both the correlation coefficients and the rank-change diagrams suggest that the Gini and Income40 frameworks in particular are accompanied by substantial changes for many countries. However, the indicators in question are associated with other challenges. Both are updated only irregularly and available for quite few countries in any given year. For 2013, data on Gini and Income40 were respectively unavailable for as much as 26% and 32% of the LICs and MICs, even after imputations (Appendix C). These two measures are also directly related to income and not to health, which is relevant even though income and health are closely correlated. ILE has the advantage of reflecting inequality in health directly and of being more readily available (Appendix C). At the same time, the ILE framework is more highly correlated with GNIpc, and the ILE framework is very highly correlated with the LE framework and the health-need frameworks more generally (Appendix B). This suggests that there may be little to gain from integrating ILE into the classification framework, especially if a health-need indicator has already been included. All inequality indicators are associated with challenges related to incentives. If a high degree of inequality implies higher priority for external funds, countries have a disincentive to promote equality.^{27, 33} For these reasons, it may be reasonable not to integrate any of the current inequality indicators into a framework for the initial classification for countries. Rather, the inequality measures may be used on a more discretionary basis, after the formal classification framework has been applied. One reason for proceeding in this way is to ensure that when large inequalities motivate external financing for health, funds are provided in ways that specifically target the reduction of those inequalities. In any case, it is important for the global-health community to develop more robust inequality measures. Such efforts can

support discretionary assessment, but they can also lead to a better candidate for integration in the classification framework itself.

The SBA or DTP3 frameworks seem to affect country rankings to a rather limited extent, but the changes for individual countries can be substantial, as shown in Figure 13. Both indicators are quite readily available for LICs and MICs (Appendix C). However, both SBA and DCP3 represent only a very limited – albeit a very important – part of the health system. This makes each less relevant as an indicator in a classification framework to be used across a wide range of settings and institutions. This limited scope also makes it quite easy for governments to improve skilled birth attendance without much improvement of the overall health system.

Finally, the OOPP framework results in rather substantial changes for both individual countries and country categories. It is also quite readily available (Appendix C). One challenge associated with OOPP is that it does not fully reflect cases in which people must forgo needed services because of user fees or other barriers. Another challenge is that OOPP may increase with the quality of services. There is also a disincentive for governments to reduce OOPP if it is a core criterion in country classification and if priority for external financing increases with OOPP.

Overall, SBA, DTP3, and OOPP seem relevant for many decisions concerning external financing for health. At the same time, none of the indicators was clearly suitable for inclusion in a framework for the initial classification of countries when judged against the criteria of relevance, salience, validity, consistency, and timeliness and availability. All three indicators may play their most important role at other stages of the decision-making process.

9. Broad comparison of frameworks

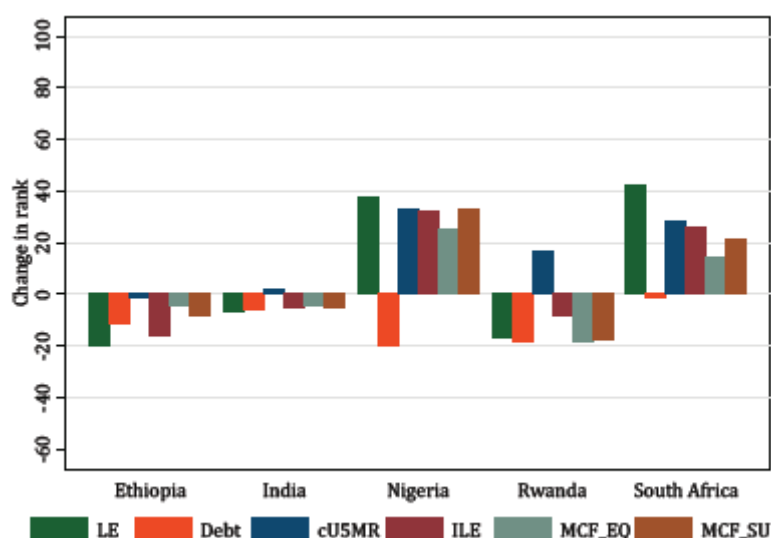
The preceding sections examined twenty frameworks across four categories, namely health needs, domestic capacity, external impact, and “other.” It can also be useful to juxtapose single frameworks from each of these categories. In this broad comparison of frameworks, we also included two multicriterion frameworks, comprising four criteria. Both of these frameworks integrated GNIpc, LE, ILE, and SBA. The indicators were selected in order to both cover several concerns and establish correspondence with the country characteristics in the online survey to the greatest extent possible. One of these multicriterion frameworks assigned equal weights to each indicator (MCF_EQ), whereas the other utilized weights informed by the online survey (MCF_SU). To integrate these weights, we interpreted the country characteristics of income per capita, burden of disease, health inequality, and health-system strength to be represented by GNIpc, LE, ILE, and SBA, respectively. Based on the findings from the survey and these relationships, we then assigned a weight of 0.1 to GNIpc, 0.3 to LE, 0.4 to ILE, and 0.2 to SBA. With these weights and these indicators, the multicriterion framework is very much a “health framework,” with little emphasis on national income or other factors.

Implications for individual countries

The Spearman correlation coefficients for the two-criterion frameworks were discussed above. They were 0.90 for the LE framework, 0.73 for the debt framework, and 0.93 for the ILE and SBA frameworks (when compared with the GNIpc baseline for their category). With regard to the multicriterion frameworks, the correlation coefficients were 0.97 for MCF_EQ and 0.98 for MCF_SU. These scores suggest that the latter multicriterion framework in particular is quite likely to generate significant changes in country ranking. At the same time, the correlation coefficient for this framework could have been even lower if we had chosen other indicators to represent the country characteristics addressed in the survey. For example, this would be expected if Gini replaced ILE as the indicator for health inequality.

Figure 15 shows how the rank of the five focus countries changes as one moves from GNIpc alone to various other frameworks. Most rank changes in this diagram differ slightly from the changes seen for the same country and same framework in the preceding diagrams. This is as expected, because the exact set of countries included varies somewhat across diagrams, depending on data availability. Changes for all countries are provided in Supplementary File 5.

Figure 15. Changes in rank when moving from GNIpc alone to various other frameworks (focus countries)

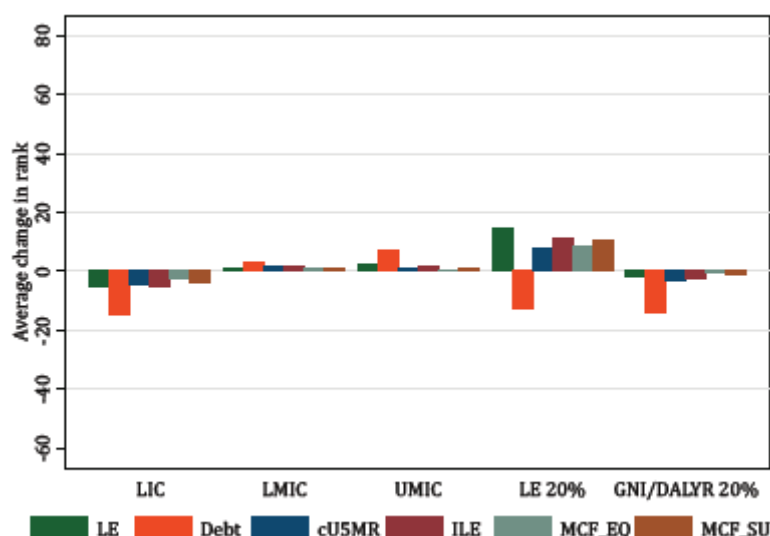


The implications of a move to any of the two-criterion frameworks were discussed above. When these implications are juxtaposed, we see that both Nigeria and South Africa get a substantially higher rank when moving from GNIpc to any other framework, except the debt framework. Conversely, Ethiopia, India, and Rwanda experience a lower rank when moving to any other two-criterion framework, except the cU5MR framework for India and Rwanda. Looking at the multicriterion frameworks, the changes in rank are pronounced for Nigeria, Rwanda, and South Africa. The difference between the multicriterion framework with equal weights and that with survey-based weights is modest.

Implications for country categories

Figure 16 shows the average change in rank for countries in the five categories as one moves from GNIpc alone to various other frameworks. As described above, these changes may differ slightly from those in preceding diagrams due to data availability.

Figure 16. Average changes in rank for countries when moving from GNIpc alone to various other frameworks



This figure clarifies some general trends. First, there are quite consistent changes across frameworks within categories, with the exception of the debt framework. Second, two broad trends across categories emerge. Among the income classes, all moves beyond GNIpc alone lead to a lower rank on average for LICs and a slightly higher rank for LMICs and UMICs. Among the two other categories, the 20% of countries with the lowest life expectancies mainly get a higher rank, whereas the countries with the lowest GNI-DALYR ratios consistently get a lower rank.

The basic pattern across income classes is not surprising, because income class is based on GNIpc and the changes in rank show what happens when one moves away from a ranking based on GNIpc alone. However, this pattern is a reminder of the potential effects of going beyond GNIpc. If one wants to transcend GNIpc but is concerned that LICs could be disadvantaged as a result, one may want to find ways to counteract this effect, for example through various exemptions.

Looking at the bottom 20% of countries in terms of life expectancy can be particularly relevant if one believes that external financing for health is primarily about assisting people with the poorest health outcomes. As Figure 16 shows, integrating either LE or ILE with GNIpc will improve the ranking of countries in this group. This is also the case for the multicriterion frameworks. However, among all the frameworks shown in Figure 16, the LE framework benefits these worst-off countries the most.

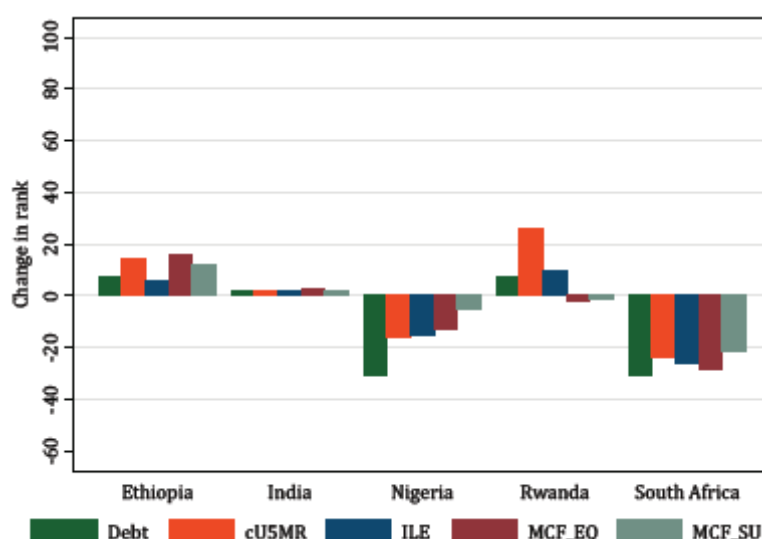
For the bottom 20% of countries in terms of GNI-DALYR ratio, the implications are different. As described, this ratio may be of particular interest for those who believe the most central consideration in external financing for health is the *relationship* between domestic capacity and health needs. For this group of countries, all departures from GNIpc alone, in Figure 16, result in a lower rank. However, with the exception of the debt framework, the changes are fairly small.

Broad comparison with different baseline

Finally, it can be useful to compare the frameworks while using a baseline other than GNIPC. We utilized the LE framework, that is, the framework integrating GNIPC and LE, as a baseline. This baseline may be highly relevant if one has decided to integrate a health-need indicator with GNIPC but wonders whether it would be worth adding further indicators to the index. It is important to note that in these comparisons, *all* frameworks include *both* GNIPC and LE, in order to show what happens when one adds other indicators to such an index. Accordingly, when compared to the GNIPC-LE baseline, what we call the debt framework comprises the GNIPC, LE, and debt criteria; the cU5MR framework comprises the GNIPC, LE, and cU5MR criteria; and the ILE framework comprises the GNIPC, LE, and ILE criteria. The multicriterion frameworks (MCF_EQ and MCF_SU) are designed as above; that is, they integrate the GNIPC, LE, ILE, and SBA criteria.

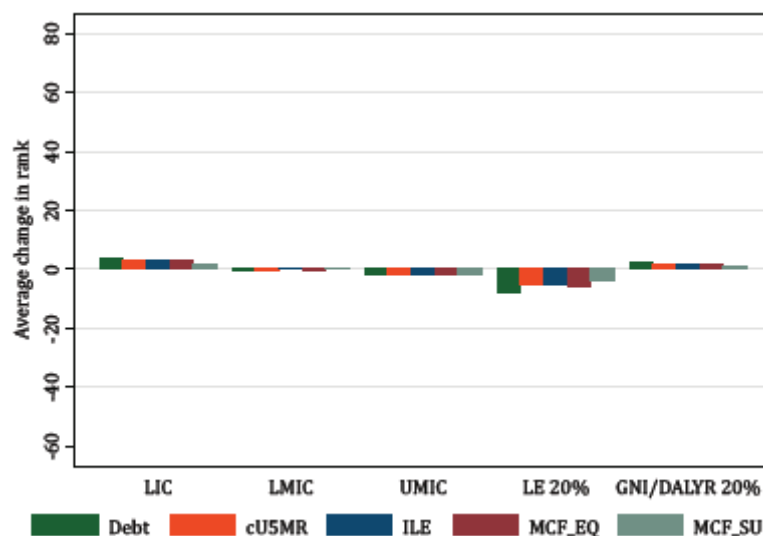
Figure 17 shows how the rank of the five focus countries changes as one moves from a LE framework to various other frameworks.

Figure 17. Changes in rank when moving from a life-expectancy framework to various other frameworks (focus countries)



The most clearly evident change here is South Africa's consistent and substantial decrease in rank when moving from an LE framework to any of the other frameworks. For some of the countries and some of the frameworks, there are significant changes in rank when departing from the LE baseline. Figure 18 shows the average change in rank for countries in the five categories as one moves from a LE framework to various other frameworks.

Figure 18. Average changes in rank for countries when moving from a life-expectancy framework to various other frameworks



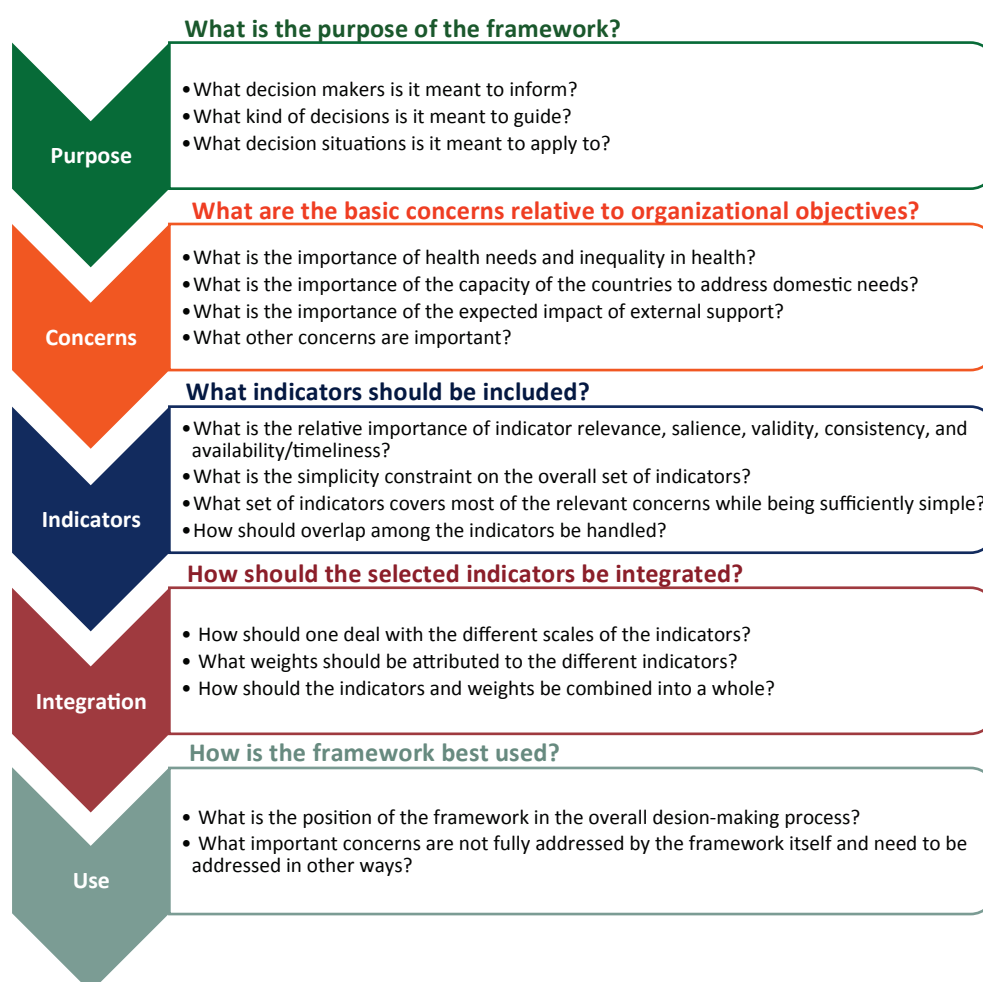
As the figure shows, the changes in rank are in general very small. The largest changes happen to the countries with the lowest life expectancies, but even those changes are modest. The diagram therefore suggests that if one already has a framework that integrates GNIpc and LE, adding any of the other indicators will make little difference to the country categories.

10. Key choices for any institution

The EAI Convening Partners and other institutions revise their classification frameworks from time to time. When doing so in the future, these institutions can capitalize on a wide range of insights from the EAI. An important outcome of the EAI process is that we, as a community, can better articulate the key issues and choices involved in the design of classification frameworks.

Based on insights gained over the course of the EAI project, we have sought to clearly articulate the key choices that every institution will need to make when designing its classification framework. These choices are summarized in Figure 19. In part, this line of reasoning maps onto the approach used in this project. The menu of frameworks and the implications described in this report can inform these choices.

Figure 19. Line of reasoning and key choices in the design of a classification framework



11. Use of frameworks

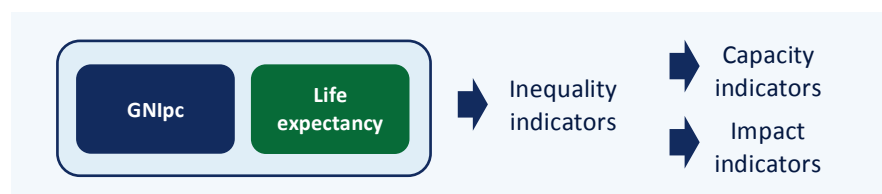
An indefinite number of frameworks exist for many different purposes. We have examined one kind of framework for one specific purpose, namely classifying countries in terms of characteristics relevant to decisions on external financing for health. We have examined frameworks in the form of indices rather than more loosely organized criteria, because we found that many stakeholders prefer such an index. It is important to acknowledge that GNIpc and a health-need indicator can be integrated into a common framework in ways different from those outlined here. In any case, it is important to acknowledge the roles this kind of framework can and cannot play and the cases in which one may need other kinds of frameworks.

Position in decision-making process

We have designed and assessed frameworks mainly for a certain position in the decision-making process pertaining to external financing for health. This position is the initial or primary classification of countries, which takes place early in the process. This means that the frameworks considered set the stage, but that other criteria and considerations may enter later in the process to support the final decision. We could have designed and assessed frameworks meant to cover most or the entire decision-making process, but this would have been highly complex and would have involved an in-depth discussion of the allocation of funds, which was an issue the analytical groups were asked not to address directly.

At the same time, we went beyond the initial classification of countries to suggest what positions in the decision-making process some indicators external to the primary classification framework should occupy. This is illustrated in Figure 20. Based on the analyses offered above, we suggest that institutions use only a simple framework that integrates GNIpc and health needs for the initial classification of countries. We suggest supplementing the use of this framework with a more discretionary assessment of within-country inequalities in health or income to the extent permitted by the data. Thereafter, we recommend the consideration of a range of capacity and impact indicators, including at the program and project levels, in the assessment of applications, and during the implementation phase. Because institutions have different decision-making processes, needs, and objectives, this structure will not be relevant to all of them.

Figure 20. Decision-making stages



The interviews supported this kind of role for classification frameworks. Stakeholders generally agreed that frameworks could be useful in providing an *initial* orientation for decision making that leads to a consideration of a wider set of factors, and they agreed that these factors could be addressed a dialogue between donor and recipient countries. A country-level dialogue could address more qualitative factors, including many factors closely related to domestic capacity and

expected impact, as well as many topics beyond classification, including the targeting of particular diseases or subpopulations.

Other decisions concerning health and development

We have considered frameworks designed for a specific purpose, namely for informing decisions concerning external financing for health. A simple framework that integrates GNIpc and health needs can be relevant also to other decisions regarding health and development. This includes decisions about forms of external support like advocacy and technical assistance, which are not primarily about financing, yet are often linked to financing from donors.

Even when a framework that integrates GNIpc and health needs is relevant for other decisions, the specific use is likely to differ. For example, if eligibility thresholds are used, these are likely to differ between external financing and technical support. There is also a range of decisions for which other indicators and other frameworks may be most relevant. Examples include decisions concerning (1) the speed and timing of a country's transition from eligibility status and the projected time horizon for external support, (2) which modality of external support to use, that is, whether financial support, technical support, advocacy, or a combination of these will be most effective, (3) the financing of global public goods for health, including research, infectious-disease surveillance, and regulation of harms to health, such as tobacco and arms, and (4) the pricing of drugs across countries.

12. Limitations and next steps

There are several limitations to the design and assessment of frameworks presented in this report. Some of these apply to most or all classification frameworks. Others are more specific to our approach.

Limitations of frameworks in general

We have designed and assessed frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. These frameworks were meant to be simple and to be applied early in the decision-making process.

All or most frameworks of this kind appear to share three challenges and limitations, related to comprehensiveness, incentives, and inequality and targeting, respectively. In addition, data availability and quality are limited for most candidate indicators.

Comprehensiveness

It is impossible to combine all relevant concerns and indicators into one simple framework or index. If one tried to include all relevant indicators in one or more contexts, numerous complications would arise, especially from overlap and interactions among indicators. Trade-offs between simplicity and comprehensiveness have to be made, and deciding which relevant indicators to include is a key priority-setting exercise that involves both technical considerations and value choices.

Incentives

Most or all frameworks struggle to deal adequately with incentives. The basic problem is that if poorer levels of an indicator imply higher priority for external support, countries' incentives to improve on this indicator may be reduced. This is a problem for health-need indicators and inequality indicators to the extent that greater health needs and greater inequality imply higher priority. This could also be a problem for capacity indicators that are highly dependent on government policies and effort. For example, government health expenditure per capita (GHEpc) can be an indicator of the health system's capacity to address health needs. Yet, GHEpc does not represent an external circumstance that is imposed on countries; it is very much a matter of choice.

The most obvious response to this problem is to adjust for the level of effort exerted in exercising the relevant capacities. The challenge with this kind of response is twofold. We currently lack a good measure of effort or true capacity to integrate into a classification framework. And even if we had one, we would still risk "punishing" unfortunate individuals for the behavior of their governments.

Against this background, we believe incentives need to be addressed primarily in ways external to the framework for classifying countries. There are many ways of doing this, but a discussion of these options falls outside the scope of this report.

Inequality and targeting

Within-country inequalities in health figure prominently in the current debate on external financing for health. The chief question is how we should address pockets of unmet health needs.

With regard to classification, an immediate response may be to apply the economic and health-needs indicators found useful at the national level to the subnational level. However, this does not directly help with the classification of *countries*. To that end, measures of national inequalities appear more useful. However, we have found no robust measure of that kind, that is, a measure that scores high on relevance, salience, validity, consistency, and timeliness and availability.

Even if a robust inequality measure were available, the link to country classification would not be straightforward. The reason for this is the intimate link between inequality and the appropriate response. It could be problematic if large inequalities implied higher priority for all kinds of external financing. It appears more reasonable that inequalities suggest higher priority only for support of policies and programs that primarily target those inequalities. The level of inequality may also influence the optimal modality of support. Especially for UMICs, policy and technical support are likely to be more relevant than financial support, although the latter may have an important catalytic role.

Limitations of our approach

Some of the limitations of our approach are the result of restrictions in the mandate, but most are the result of deliberate choices and trade-offs in the face of time constraints. We comment on nine limitations in particular:

1. We have focused on a certain kind of classification framework in health and development, that is, frameworks relevant to decisions on external financing for health that are to be applied early in the decision-making process. Such frameworks can be relevant also to other kinds of decisions, but this is not necessarily the case. Important decisions not specifically addressed by the frameworks in this report include decisions regarding modality of support and regarding eligibility for and level of nonfinancial support. The frameworks are also not directly sensitive to what diseases or kinds of functions are supported. In particular, the frameworks are not specifically designed to guide decisions concerning the financing of global public goods for health.
2. The frameworks designed and assessed in this report exhibit the general shortcomings discussed above, including those related to data availability and quality, comprehensiveness, incentives, and inequality and targeting. We have been quite strict in our assessment of data validity, consistency, and availability and timeliness.
3. We have examined implications mainly in terms of changes in rankings. This gives only an indication of the implications for most decisions concerning external financing for health. The final classification of countries depends not only on the dimension in which countries are assessed, but also on the thresholds applied within that dimension. These thresholds may or may not link directly to decisions regarding eligibility and transition. Among the eligible countries, the levels of financing are likely to be guided not primarily by rank, but by cardinal scores and absolute differences in the relevant indicators or indices.

4. We have used GNIpc as the starting point and as part of all frameworks. This does not preclude putting less weight on GNIpc than other indicators in the frameworks, including health-need indicators. However, the frameworks designed and assessed in this report are less relevant for those who would like to fully exclude GNIpc.
5. We have not been able to examine all relevant indicators, all relevant combinations of indicators, and all plausible prioritization rules. Doing so would have exceeded our time and space constraints and would have made the report highly complex. We believe, however, that the indicators and frameworks examined in this report cover a useful range of alternatives. Population size is one of the indicators that should have received more elaborate examination.
6. There is no agreement regarding the best method for framework or index construction, including the best methods for normalization, weighting, and aggregation. For the most part, we have used highly conventional and simple methods, with the aim of making the method as intuitive and transparent as possible. It would have been ideal to conduct a sensitivity analysis on some of these methodological choices.
7. We have not examined subnational applications of frameworks. However, the frameworks can be applied at this level – for example, to states or districts – to the extent that relevant data are available.
8. By using annual GNIpc, the examined frameworks do not help to smooth out sudden fluctuations in GNIpc. Moreover, the frameworks explored do not include projections. Such forecasting would be useful in several settings, but would also increase complexity and uncertainty.
9. The interviews and survey were based on a limited and nonrepresentative sample of respondents. We could not use the survey to derive weights directly applicable to the frameworks. In any case, the responses do not any way represent the “true” answers in the sense that they dictate what one *should* do. However, we believe the inputs from stakeholders have been very useful for the design and assessment of frameworks.

Key issues for further analysis

The design of approaches and frameworks for decisions on health and development is an evolving process. The global-health community and individual institutions continuously seek to improve on current practices while adapting to ongoing transitions. The analyses presented in this report can inform future reform efforts. These analyses have also revealed useful next steps for further inquiry. We find it particularly promising for analysts to

1. go beyond the pure ranking of countries in order to examine cardinal scores, absolute differences, and the implications of different classification thresholds. This is important, because any classification framework depends on both the choice of metric and the choice of threshold within that metric. These thresholds may or may not link directly to decisions concerning eligibility and transition.
2. go beyond the pure ranking of countries in order to examine the allocation of funds across countries. It is important to understand how much more or less funding countries will receive if one moves to a certain framework.

3. clarify the normative basis for health and development and for external financing for health.
4. thoroughly examine the most promising indicators for each of the concerns highlighted in this report, that is, health needs, capacity, and expected impact. In particular, it is important to examine which properties a health-need indicator should have in the specific context of external financing for health as well as which indicators exhibit those properties. GNIpc serves many purposes for which it was not specifically designed, and we should try to avoid that error when it comes to health-need indicators and external financing. It is also important to search for better inequality indicators and to examine the role of population size more systematically than has typically been the case in previous studies.
5. examine how the classification of countries can be linked to specific responses, such as targeting.
6. examine the desirability and implications of various methodological choices in index construction in the specific context of external financing for health.
7. conduct larger and more representative interview and survey studies.
8. examine how the general findings in this report and of the EAI can be tailored to fit organization-specific needs and objectives.
9. explore a wider set of strategies for smoothing countries' transition from external support.
10. assess and develop frameworks that inform decisions regarding optimal modality of support, levels of nonfinancial support, and the support of global public goods for health.

13. Overall findings and recommendations

Epidemiological, economic, and institutional transitions are challenging current approaches to classifying countries for decisions concerning health and development. This report has explored frameworks for classifying countries in terms of characteristics relevant to decisions on external financing for health. Specifically, the report has provided six kinds of outputs: (1) an explicit step-by-step approach to the design and assessment of classification frameworks, (2) responses from interviews and a survey, (3) a broad menu of simple frameworks, (4) a head-on comparison of their implications, (5) a list of key issues that any institution should address, and (6) overall findings and recommendations for choice of indicators and use of frameworks. We have also suggested next steps for further analysis.

With regard to **health needs**, we found that

1. addressing health needs is generally considered a key objective of external financing for health, and a general health-need indicator can be relevant for most institutions. We also found that including a health measure in the framework makes a significant difference for many countries and country categories and that it specifically improves the ranking of countries with large unmet health needs. Against this background, we recommend integrating GNIpc and one health-need indicator for the initial classification of countries.
2. which of the health-need indicators is chosen makes less of a difference than whether or not such an indicator is included at all. Moreover, no single indicator appears to be clearly superior to all others. We found that life expectancy and DALYR are both good candidates for being included in a classification framework. We recommend that institutions at least consider using life expectancy as a generic health-need indicator, but the optimal measure for any given actor will eventually depend on organization-specific needs and objectives.
3. although the implicit age weighting and the use of age standardization are linked to important value choices, they have received scarce attention in the context of external financing for health. We recommend critically examining the extent to which candidate health-need indicators give extra weight to deaths at a younger age and are sensitive to the country's age structure.
4. although inequality in health is considered an important concern for external financing for health, no robust measure of health inequality exists for this purpose. We recommend accelerating the development of such measures.

With regard to **domestic capacity**, we found that

5. domestic capacity is an important consideration for external financing for health. However, none of the examined indicators was clearly suitable for inclusion in the classification framework when judged against the following criteria: relevance, salience, validity, consistency, and timeliness and availability. We also found that the inclusion of domestic-capacity indicators might contribute to perverse incentives. Against this background, we suggest that indicators of domestic capacity other than GNIpc – such as debt service and tax ratio – are considered after the primary classification framework has been applied.

With regard to **expected impact**, we found that

6. expected impact is an important consideration for external financing for health, but we conclude that none of the examined indicators is a good predictor of future impact. More generally, none of the indicators was clearly suitable for inclusion in the classification framework when judged against the following criteria: relevance, salience, validity, consistency, and timeliness and availability. We suggest that indicators of expected impact are considered after the primary classification framework has been applied and mainly at the program and project level, in the assessment of applications, and during the implementation phase. Potential country determinants of impact include level of conflict, institutional quality, and recent improvements in health-service coverage or health outcomes.

With regard to **all examined frameworks**, we found that

7. nearly any move beyond GNIpc alone resulted in a lower rank for the group of LICs and a higher rank for the group of MICs. To the extent that this is undesirable, we recommend considering counteracting measures.

Overall,

8. we recommend using a simple framework that integrates GNIpc and health needs for the initial classification of countries. We suggest supplementing the use of this framework with a more discretionary assessment of within-country inequalities in health and income to the extent permitted by the data.
9. we recommend considering a range of indicators for capacity and expected impact after the primary classification framework has been applied, including at the program and project level, in the assessment of applications, and during the implementation phase. This and the assessment of inequalities should be done in the context of country-level dialogue.
10. these general findings and recommendations may be informative, but the optimal indicators and framework for any given institution will eventually depend on organization-specific needs and objectives.

We hope this report and these recommendations will provide useful input to the EAI and beyond.

Appendix A: Disease-specific frameworks

For some institutions it will be relevant to use disease-specific classification frameworks. For illustration, we designed and assessed three disease-specific health-need frameworks. There were based on the following indicators:

- Maternal mortality ratio (MMR)
- HIV prevalence rate (HIVR)
- Tuberculosis prevalence rate (TBR)

Maternal mortality ratio

Maternal mortality ratio (MMR) is the number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births. The data are estimated with a regression model using information on the proportion of maternal deaths among non-AIDS deaths in women ages 15-49, fertility, birth attendants, and GDP.

HIV prevalence rate

HIV prevalence rate (HIVR) refers to the percentage of people ages 15-49 who are infected with HIV.

Tuberculosis prevalence rate

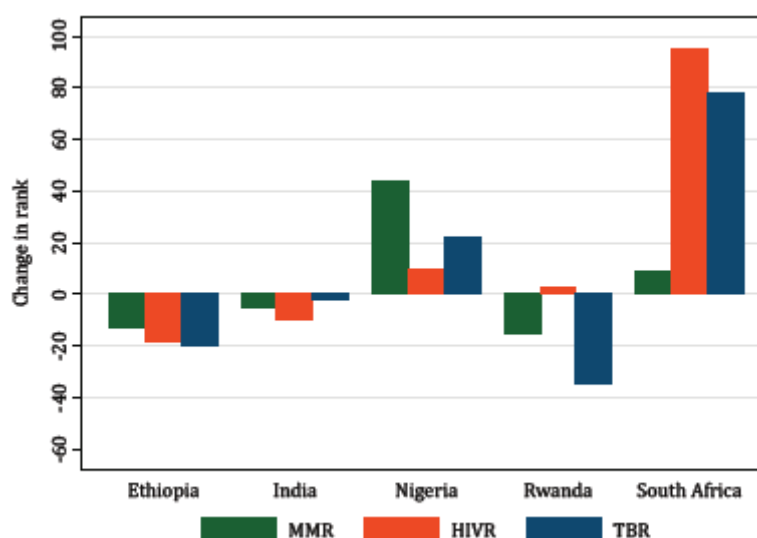
Tuberculosis prevalence rate (TBR) is the number of TB cases (all forms) per 100,000 population.

Prioritization rules used

For all indicators, the prioritization rule used was monotonically increasing. A higher level suggests countries are worse off and they therefore get a higher rank.

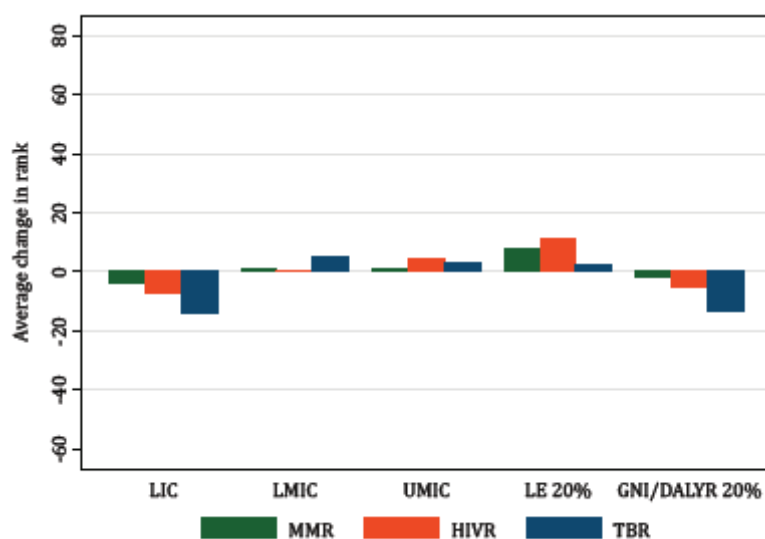
Implications for individual countries

Figure A1. Changes in rank when moving from GNIpc alone to disease-specific frameworks (focus countries)



Implications for country categories

Figure A2. Average changes in rank for countries when moving from GNIpc alone to disease-specific frameworks



Spearman coefficients for correlation between indicators

Health need (generic)		GNIpc (for health need set)	U5MR	U60MR	LE	HALE	DALYR	DALYR_AS
GNIpc (for health need set)		1						
U5MR		0,755	1					
U60MR		0,612	0,840	1				
LE		0,699	0,913	0,945	1			
HALE		0,705	0,889	0,917	0,936	1		
DALYR		0,660	0,803	0,881	0,894	0,924	1	
DALYR_AS		0,697	0,892	0,921	0,938	0,997	0,925	1
Health need (disease specific)		GNIpc (for disease specific set)	MMR	HIVR	TBR			
GNIpc (for disease specific set)		1						
MMR		0,761	1					
HIVR		0,414	0,623	1				
TBR		0,493	0,603	0,503	1			
Domestic capacity		GNIpc (for capacity set)	Debt	Tax	THEpc	GHEpc		
GNIpc (for capacity set)		1						
Debt		-0,541	1					
Tax		0,367	-0,339	1				
THEpc		0,950	-0,559	0,387	1			
GHEpc		0,930	-0,562	0,429	0,957	1		
Expected impact		GNIpc (for impact set)	cU5MR	rcU5MR	cSBA	rcSBA		
GNIpc (for impact set)		1						
cU5MR		0,691	1					
rcU5MR		0,158	0,563	1				
cSBA		0,407	0,408	0,153	1			
rcSBA		0,488	0,497	0,184	0,969	1		
Other		GNIpc (for other set)	ILE	Gini	Income40	SBA	DTP3	OOPP
GNIpc (for other set)		1						
ILE		0,748	1					
Gini		-0,062	0,213	1				
Income40		-0,100	0,123	0,914	1			
SBA		0,678	0,737	0,113	0,019	1		
DTP3		0,438	0,540	0,146	0,067	0,598	1	
OOPP		0,267	0,132	-0,390	-0,348	0,150	0,145	1

Appendix C: Data sources and characteristics

Indicator	Brief description	Long description	Year	Source	WDI indicator code	Countries with data (% of all LICs and MICs) ^a	Countries with missing data (% of study sample) ^b
GNIpc	Gross national income per capita, Atlas method (current US\$)	The gross national income, converted to U.S. dollars using the World Bank Atlas method, divided by the midyear population.	2013	World Bank	NY.GNP.PCAP.CD	4	0
U5MR	Under-five mortality rate (per 1,000 live births)	The probability per 1,000 that a newborn baby will die before reaching age five, if subject to age-specific mortality rates of the specified year.	2013	UN Inter-agency Group for Child Mortality Estimation	SH.DYN.MORT	3	1
U60MR	Mortality rate between ages 15 and 60 (per 1,000 adults)	The probability of dying between the ages of 15 and 60—that is, the probability of a 15-year-old dying before reaching age 60, if subject to current age-specific mortality rates between those ages.	2013	United Nations Population Division	SP.DYN.AMRT.FE and SP.DYN.AMRT.MA	6	4
LE	Life expectancy at birth (years)	The number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	2013	World Bank	SP.DYN.LE00.IN	4	2
HALE	Healthy life expectancy at birth (healthy life years)	The number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life, adjusted for the expected disability in those years.	2013	IHME	N/A	4	2

DALYR	Disability-adjusted life year rate (per 100,000 population)	The number of years of life lost due to premature mortality (YLLs) and years lived with disability (YLDs) per 100,000 population.	2013	IHME	N/A	4	2
DALYR_AS	Age-standardized disability-adjusted life year rate (per 100,000 population)	The number of years of life lost due to premature mortality (YLLs) and years lived with disability (YLDs) per 100,000 population, standardized by age.	2013	IHME	N/A	4	2
MMR	Maternal mortality ratio (per 100,000 live births)	The number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births.	2013	WHO	SH.STA.MMRT	6	4
HIVR	Prevalence of HIV (% of population ages 15-49)	The percentage of people aged 15-49 who are infected with HIV.	2013	UNAIDS	SH.DYN.AIDS.ZS	1	0
TBR	Tuberculosis prevalence rate (per 100,000 population)	The estimated number of TB cases (all forms) at a given point in time, expressed as the rate per 100,000 population.	2013	WHO	N/A	2	1
Debt	Total debt service (% of GNI)	The sum of principal repayments and interest actually paid in currency, goods, or services on long-term debt, interest paid on short-term debt, and repayments (repurchases and charges) to the IMF.	2013	World Bank	DT.TDS.DECT.GN.ZS	12	10
Tax	Tax revenue (% of GDP)	Compulsory transfers to the central government for public purposes. Certain compulsory transfers such as fines, penalties, and most social security contributions are excluded. Refunds and corrections of erroneously collected tax revenue are treated as negative revenue.	2013	IMF	GC.TAX.TOTL.GD.ZS	23	21

THEpc	Total health expenditure per capita (current US\$)	Sum of public and private health expenditures as a ratio of total population. It covers the provision of health services (preventive and curative), family planning activities, nutrition activities, and emergency aid designated for health, but does not include provision of water and sanitation.	2013	WHO	SH.XPD.PCAP	6	2
GHEpc	Government health expenditure per capita (current US\$)	Product of total health expenditure per capita (current US\$) and public health expenditure (% of total health expenditure).	2013	WHO	SH.XPD.PCAP and SH.XPD.PUBL	6	2
cU5MR	Absolute change in under-five mortality rate, 2003–2013	Absolute change in under-five mortality rate between 2003 and 2013, using 2003 as the base year.	2003–2013	UN Inter-agency Group for Child Mortality Estimation	SH.DYN.MORT (as input)	3	1
rcU5MR	Relative change in under-five mortality rate, 2003–2013	Relative change in under-five mortality rate between 2003 and 2013, using 2003 as the base year.	2003–2013	UN Inter-agency Group for Child Mortality Estimation	SH.DYN.MORT (as input)	3	1
cSBA	Absolute change in skilled birth attendance rate, 2003–2013	Absolute change in skilled birth attendance rate between 2003 and 2013, using 2003 as the base year.	2003–2013	UNICEF	SH.STA.BRTC.ZS (as input)	6	4
rcSBA	Relative change in skilled birth attendance rate, 2003–2013	Relative change in skilled birth attendance rate between 2003 and 2013, using 2003 as the base year.	2003–2013	UNICEF	SH.STA.BRTC.ZS (as input)	6	4
ILE	Inequality in life expectancy	Inequality in distribution of expected length of life based on data from life tables estimated using the Atkinson inequality index.	2013	UNDP	N/A	7	5
GINI	GINI index	The extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution.	Most recent (2003–2013)	World Bank	SI.POV.GINI	26	22

Appendix D: List of supplementary material

Supplementary Files:

1. Initial list of candidate indicators (pdf file)
2. 3-digit country codes (pdf file)
3. Interview guide (pdf file)
4. Rank-change scatter plots (pdf file)
5. Rank changes for all countries for all frameworks (Excel file)
6. Scatter plots of normalized indicators (pdf file)
7. Do files and data (zip file)

References

1. Røttingen JA, Ottersen T, Ablo A, Arhin-Tenkorang D, Benn C, Elovainio R, et al. Shared responsibilities for health: A coherent global framework for health financing. London: Chatham House; 2014.
2. Moon S, Omole O. Development assistance for health: Critiques and proposals for change. London: Chatham House; 2013.
3. Norheim OF, Jha P, Admasu K, Godal T, Hum RJ, Kruk ME, et al. Avoiding 40% of the premature deaths in each country, 2010–30: Review of national mortality trends to help quantify the UN Sustainable Development Goal for health. *Lancet*. 2015;385:239–52.
4. Verguet S, Norheim OF, Olson ZD, Yamey G, Jamison DT. Annual rates of decline in child, maternal, HIV, and tuberculosis mortality across 109 countries of low and middle income from 1990 to 2013: An assessment of the feasibility of post-2015 goals. *Lancet Global Health*. 2014;2:e698–709.
5. Salomon JA, Wang H, Freeman MK, Vos T, Flaxman AD, Lopez AD, et al. Healthy life expectancy for 187 countries, 1990–2010: A systematic analysis for the global burden disease study 2010. *Lancet*. 2012;380:2144–62.
6. Commission on Social Determinants of Health. Closing the gap in a generation: Health equity through action on the social determinants of health. Geneva: World Health Organization; 2008.
7. World Health Organization. World health statistics 2015. Geneva: World Health Organization; 2015.
8. World Health Organization. Making fair choices on the path to universal health coverage. Final report of the WHO Consultative Group on Equity and Universal Health Coverage. Geneva: World Health Organization; 2014.
9. WHO, World Bank. Tracking universal health coverage: First global monitoring report. Geneva: World Health Organization; 2015.
10. Sumner A. Where do the poor live? *World Development*. 2012;40:865–77.
11. Institute for Health Metrics and Evaluation. Financing global health 2013: Transition in an age of austerity. Seattle, WA: Institute for Health Metrics and Evaluation; 2014.
12. Ottersen T, Moon S, Røttingen J-A. The challenge of middle-income countries to development assistance for health: Recipients, donors, both, or none? Working paper: 2015.
13. Salvado RC, Walz J. Aid eligibility and income per capita: A sudden stop for MICs? DPAF Working Paper Series. Bill & Melinda Gates Foundation; 2013.
14. Guillaumont P. The concept of structural economic vulnerability and its relevance for the identification of the least developed countries and other purposes (nature, measurement, and evolution). CDP Background Paper No 12. New York: Department of Economic and Social Affairs, Development Policy and Analysis Division, United Nations; 2011.
15. Llavador HG, Roemer JE. An equal-opportunity approach to the allocation of international aid. *Journal of Development Economics*. 2001;64:147–71.
16. Bárcena A, Prado A, Titelman D, Pérez R. Middle-income countries: A structural-gap approach. Economic Commission for Latin America and the Caribbean (ECLAC). New York: United Nations; 2012.

17. Resch S, Ryckman T, Hecht R. Funding AIDS programmes in the era of shared responsibility: An analysis of domestic spending in 12 low-income and middle-income countries. *Lancet Global Health*. 2015;3:e52–61.
18. Knack S, Rogers FH, Heckelman JC. Crossing the threshold: A positive analysis of IBRD graduation policy. *Review of International Organizations*. 2012;7:145–76.
19. Gupta I, Mondal S. Health spending, macroeconomics and fiscal space in countries of the world health organization South-East Asia region. *WHO South-East Asia Journal of Public Health*. 2014;3:273–84.
20. Tandon A, Cashin C. Assessing public expenditure on health from a fiscal space perspective. Health, Nutrition and Population (HNP) Discussion Paper. Washington, DC: World Bank; 2010.
21. Glassman A, Duran D, Sumner A. Global health and the new bottom billion: What do shifts in global poverty and disease burden mean for donor agencies? *Global Policy*. 2013;4.
22. Results for Development Institute. Financing national AIDS responses for impact, fairness, and sustainability: A review of 12 PEPFAR countries in Africa. Washington, DC: Results for Development Institute; 2013.
23. Glennie J. The role of aid to middle-income countries: A contribution to evolving EU development policy. London: Overseas Development Institute; 2011.
24. Verbeke K, Renard R. Development cooperation with middle-income countries. Antwerpen: Institute of Development Policy and Management, University of Antwerp; 2011.
25. Sen A. Development as freedom. Oxford: Oxford University Press; 1999.
26. Stiglitz JE, Sen A, Fitoussi J-P. Mismeasuring our lives: Why GDP doesn't add up. New York: New Press; 2010.
27. Ottersen T, Moon S, Røttingen J-A. Development assistance for health: Simulating distributional implications of twelve allocation criteria. Working paper: 2015.
28. Cogneau D, Naudet J-D. Who deserves aid? Equality of opportunity, international aid, and poverty reduction. *World Development*. 2007;35:104–20.
29. Bell C, Fink C. Aid and health. Development aid: Why and how? Towards strategies for effectiveness. Paris: Agence Française de Développement; 2005.
30. Guillaumont P. Adapting aid allocation criteria to development goals. Essay for the 2008 Development Cooperation Forum. 2008.
31. Anderson E. Practices and implications of aid allocation. Background Study for the 2008 Development Cooperation Forum, UN Economic and Social Council (ECOSOC). 2008.
32. Ottersen T, Kamath A, Moon S, Røttingen J-A. Development assistance for health: Quantitative allocation criteria and contribution norms. London: Chatham House; 2014.
33. Basu S, Stuckler D, McKee M. An alternative mechanism for international health aid: Evaluating a global social protection fund. *Health Policy and Planning*. 2014;29:127–36.
34. Organization for Economic Co-operation and Development. Handbook on constructing composite indicators: Methodology and user guide. Paris: OECD Publications; 2008.
35. Bonita R, Beaglehole R, Kjellström T. Basic epidemiology. 2 ed. Geneva: World Health Organization; 2006.

36. United Nations Statistics Division. Discussion paper on principles of using quantification to operationalize the SDGs and criteria for indicator selection 2015.
37. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in econometrics*. New York: Academic Press; 1974. p. 105–42.
38. Nielsen L. How to classify countries based on their level of development. *Social Indicators Research*. 2013;114:1087-107.
39. Yang L. An inventory of composite measures of human progress. UNDP Human Development Report Office; 2014.
40. Ottersen T, Kamath A, Moon S, Røttingen J-A. Development assistance for health: Review of allocation criteria used by multi- and bilateral institutions. Working paper: 2015.
41. Kanbur R, Sumner A. Poor countries or poor people? Development assistance and the new geography of global poverty. *Journal of International Development*. 2012;24:686–95.
42. Alonso JA, Cortez AL, Klasen S. LDC and other country groupings: How useful are current approaches to classify countries in a more heterogeneous developing world? CDP Background Paper No 21 ST/ESA/2014/CDP/21. New York: United Nations,, Department of Economic and Social Affairs; 2014.
43. Dieleman JL, Graves CM, Templin T, Johnson E, Baral R, Leach-Kemon K, et al. Global health development assistance remained steady in 2013 but did not align with recipients' disease burden. *Health Affairs*. 2014.
44. Fan VY, Glassman A, Silverman RL. How a new funding model will shift allocations from the global fund to fight AIDS, tuberculosis, and malaria. *Health Affairs*. 2014.
45. Word Bank. Country and lending groups. Available from: <http://data.worldbank.org/about/country-and-lending-groups> [2 March 2015].
46. Gostin LO. Global health law. Cambridge, MA: Harvard University Press; 2014.
47. Equitable Access Initiative. Updated – terms of reference for the Equitable Access Initiative. 2015. Available from: http://www.theglobalfund.org/documents/eai/EAI_EquitableAccessInitiative_ToRs_en/ [28 June 2015].
48. Watts C, Cairncross S. Should the GBD risk factor rankings be used to guide policy? *Lancet*. 2012;380:2060–1.
49. Murray CJ, Ezzati M, Flaxman AD, Lim S, Lozano R, Michaud C, et al. Supplementary appendix to GBD 2010: Design, definitions, and metrics. *Lancet*. 2012;380:2063-6.
50. Hausman D. Measuring or valuing population health: Some conceptual problems. *Public Health Ethics*. 2012;5:229–39.
51. Reddy S, Heuty A. Achieving the MDGs: A critique and a strategy. Working paper. 2004.
52. Svensson J. When is foreign aid policy credible? Aid dependence and conditionality. *Journal of Development Economics*. 2000;61:61–84.
53. Gibson CC, Andersson K, Ostrom E, Shivakumar S. The samaritan's dilemma: The political economy of development aid. Oxford: Oxford University Press; 2005.

54. Cagé J. Improving upon the world bank's country policy and institutional assessment: A new performance indicator based on aid effectiveness. *Journal of Globalization and Development*. 2015;5.
55. McGillivray M. Aid allocation and fragile states. Naudé W, Santos-Paulino AU, McGillivray M, editors. *Fragile states: causes, costs, and responses*. New York: Oxford University Press; 2011. p. 166–85.
56. Pietschmann E. Assist those most in need, or most of those in need? The challenge of allocative efficiency for aid effectiveness. Bonn: German Development Institute; 2013.
57. Sen A. Why health equity? *Health Economics*. 2002;11:659-66.
58. Jamison DT, Summers LH, Alleyne G, Arrow KJ, Berkley S, Binagwaho A, et al. Global health 2035: A world converging within a generation. *Lancet*. 2013;382:1898-955.
59. UNDP. Human development report 2001. Making new technologies work for human development. New York: 2001.
60. UNDP. Human development report 2014. Sustaining human progress: Reducing vulnerabilities and building resilience. New York: 2014.
61. Norheim OF. Gini impact analysis: Measuring pure health inequity before and after interventions. *Public Health Ethics*. 2010;3:282–92.
62. Neumayer E, Plümper T. Inequalities of income and inequalities of longevity: A cross-country study. *American Journal of Public Health*. 2015;Published online ahead of print November 12.
63. WHO. The World Health Report 2010. Health systems financing: The path to universal coverage. Geneva: 2010.

www.fhi.no

Developed by the Norwegian Institute of Public Health
January 2016
Postboks 4404 Nydalen
NO-0403 Oslo
Tel: +47- 21 07 70 00
Only available in PDF format