Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

On the Measurement of Financial Protection: An Assessment of the Usefulness of the Catastrophic Health Expenditure Indicator to Monitor Progress Towards Universal Health Coverage

Karen A. Grépin ⬤, Bridget R. Irwin ⬤, and Benjamin Sas Trakinsky ⬤

Department of Health Sciences, Wilfrid Laurier University, Waterloo, Canada

ABSTRACT

Ensuring financial protection (FP) against health expenditures is a key component of Sustainable Development Goal (SDG) 3.8, which aims to achieve Universal Health Coverage (UHC). While the proportion of households with catastrophic health expenditures exceeding a proportion of their total income or consumption has been adopted as the official SDG indicator, other approaches exist and it is unclear how useful the official indicator is in tracking progress toward the FP sub-target across countries and across time. This paper evaluates the usefulness of the official SDG indicator to measure FP using the RACER framework and discusses how alternative indicators may improve upon the limitations of the official SDG indicator for global monitoring purposes. We find that while all FP indicators have some disadvantages, the official SDG indicator has some properties that severely limit its usefulness for global monitoring purposes. We recommend more research to understand how alternative indicators may enhance global monitoring, as well as improvements to the quality and quantity of underlying data to construct FP indicators in order to improve efforts to monitor progress toward UHC.

Introduction

The World Health Organization (WHO) defines Universal Health Coverage (UHC) as a state wherein "all people and communities can use the promotive, preventive, curative, rehabilitative and palliative health services they need, of sufficient quality to be effective, while also ensuring that the use of these services does not expose the user to financial hardship." Protecting households against financial hardship should be a key function of every health system and has been conceptualized as ensuring financial protection (FP), which has also been defined by the WHO as the state wherein "direct payments made to obtain health services do not expose people to financial hardship and do not threaten living standards."[1] Ensuring FP is thus a key component of achieving UHC, which has received increased political attention through the inclusion of a target to achieve UHC (target 3.8) among the Sustainable Development Goals (SDGs).

Illness is believed to be among the least predictable and largest economic shocks households can face.[2,3] Studies have shown that illness can lead to negative financial consequences through two primary channels: out-of-pocket health expenditures (OOPs) and the loss of income due to reduced labor supply or productivity.[4]

Paying for health services can have major financial consequences for households: survey data from 133 countries suggest that approximately 808 million people experienced catastrophic health expenditures in 2010,[5] while approximately 97 million people suffered impoverishment due to health spending.[6] Millions more also likely went without seeking healthcare altogether due to a lack of affordability and thus suffered from important unmet medical need.

To measure how well a country is doing with regards to ensuring FP, the most commonly used indicator is the proportion of the population experiencing catastrophic health expenditures (CHEs). While a full review of the indicators used to measure FP is beyond the scope of this paper, and reviews are available elsewhere,[7-9] the CHE indicator estimates the proportion of the population that live in households where out-of-pocket health spending (OOPs) exceeds a fixed threshold of a household's available resources, where available resources can be defined as either total household consumption or income (i.e. the budget share metric) or total resources after food or other essential expenditures have been subtracted from total household consumption or income (i.e. the capacity to pay metric). Complementary approaches to measure FP

CONTACT Karen A. Grépin ✉ kgrepin@wlu.ca 🖅 Department of Health Sciences, Wilfrid Laurier University, 75 University Avenue W., Waterloo, ON N2L3C5, Canada

also exist, including the widely used impoverishing health expenditure (IHE) indicator, which identifies households that are above a defined poverty line when OOPs are included in total household consumption or income but would be below it if OOPs were subtracted. Variations of both types of indicators also exist (e.g. different numerators, different adjustments to the denominators, or the use of different thresholds or poverty lines) and have seen widespread use in the literature.

Frameworks have been developed that include guidelines on which indicators and metrics should be used to monitor progress toward FP across countries and over time. In 2014, the WHO and the World Bank developed a joint monitoring framework that included a recommendation that both CHEs and IHEs indicators should be used to measure FP, and that indicators of FP should be further disaggregated for equity calculations, including along socioeconomic, geographic, and gender lines.[10] It also suggested that while countries should adapt a monitoring framework that is reflective of their unique epidemiological and demographic profile, countries should also strive to use internationally standardized indicators to allow for global monitoring. Regional frameworks have also been developed, for example the Regional Office of the WHO in Europe suggests using a variant of the CHE indicator that it developed and believes is better suited to monitor progress in the relatively higher-income countries of the Region.[11,12] However, the Interagency Expert Group on SDG Indicators (IAEG SDG), the institution that was created by the UN to develop and implement a global indicator monitoring framework for the SDGs, has recommended using only the CHE indicator, specifically defined as 10% (CHE10%) or 25% (CHE25%) of total household income or consumption as the official SDG indicator to monitor progress toward the FP sub-target (3.8.2) of UHC. The CHE10% metric, however, is more widely used in the literature.

The recommendation to use the CHE indicator, however, was controversial,[13] and there is a lack of agreement in the literature about which indicators are best at measuring FP.[14–16] The use of CHE to measure FP long predates its adoption as the official SDG indicator, as do the debates about the strengths and limitations of it and alternative indicators.[17,18] As no indicator is ever perfectly able to capture all dimensions of an issue and given the known limitations of these indicators, it is surprising that there has been little discussion about the ability of the official SDG indicator to track and monitor progress toward FP, a gap this paper seeks to address.

Measurement is the quantification of abstract concepts using both theory and empirical evidence, both of which should inform the indicator selection process.[19]

However, the selection of indicators, especially those used for global monitoring, is rarely a purely technical exercise and is usually heavily influenced by political factors.[20] In addition, the quantification of concepts into specific numerical indicators can give power to ideas, which can greatly influence the extent to which an issue is prioritized as well as the way in which the issue is conceptualized and understood by policy makers.[20] Given the potential power of indicators to influence policy and practice, a more careful assessment of the official SDG indicator of FP is warranted.

The purpose of this paper is to assess the usefulness the official SDG indicator, to measure and monitor progress toward the FP sub-target of UHC across countries and over time. We do so by applying the RACER framework, a framework that was developed to evaluate indicators in terms of their relevance, acceptability, credibility, ease, and reliability,[21] and by critically bringing together evidence from the existing FP literature, including, where relevant, a discussion of how alternative indicators may address some of the limitations of the official FP indicator. In this article, we first discuss the RACER framework and its associated criteria, and then apply these criteria to the official SDG indicator, before attempting to draw some conclusions on the usefulness of the official FP indicator to measure and monitor UHC across countries. We hope the findings of this study will inform the ongoing discussion of global monitoring efforts and improve the assessment of progress in achieving FP and UHC for all.

Materials and Methods

Assessment Framework

The RACER framework was designed to assess the usefulness of indicators for informing policy decisions[21] and has been previously used to evaluate other SDG indicators, including those measuring social protection[22] and sustainable development.[23] The RACER criteria stipulate that in order for an indicator to be useful, it should be Relevant, in that it measures what it sets out to measure and reflects intended objectives, Acceptable, in that it is accepted by all key stakeholders, Credible, in that it is unambiguous, transparent, and easy to interpret, Easy, in that it is feasible to collect and analyze the necessary data, and Robust, in that it is sensitive, reliable, and complete, and generates high quality data. Gerdes et al.[24] have further extended this framework, developing a number of sub-criteria intended to specify and operationalize each criteria, which we have further adapted to assess the usefulness of the official FP indicator to monitor progress toward the FP sub-target of the UHC goal across countries and over time (Table 1). We

have dropped some sub-criteria, for example the modeling and forecasting sub-criteria as this is not relevant in the context of the SDGs. Using the sub-criteria to guide the discussion, we have used evidence from the literature to discuss the usefulness of the official FP indicators along each criterion.

Results

Relevant

Policy Support and Identification of Targets and Gaps

Given the definition provided by the WHO, to be relevant an FP indictor should be able to measure the extent to people are exposed to financial hardship through the use of health services and its impact on their living standards. The official SDG indicator, as well the widely used alternative indicators, use OOPs as the base of their calculation, which are then related to either some measure of a household's available resources or to a normative poverty line to become the measure of FP. These measures only capture the direct economic effects of using health services and fail to account for other indirect financial consequences of using health services, which may also reduce living standards.[2,4,25] While O'Donnell[15] has argued that only the direct effects of using health services and not of illness in general should in fact be the goal of health policy, some indirect financial effects may also result from the decision to seek health services, such as the opportunity cost of time spent seeking health services, and thus could underestimate the financial burden on households. More importantly, the official SDG indicator is blind to the way in which households finance OOPs through what are called coping mechanisms, many of

which can have additional short- or long-term financial implications for households. If households pay for health services through coping mechanisms, for example, CHE may overestimate the true financial effects on households, as actual consumption is likely to fall less than assumed.[26] Similarly, if households rely upon debt and other forms of borrowing to finance these expenditures, then CHE might actually underestimate the true financial burden on households in the long run.[26] If ensuring that households do not suffer financial hardship from using health services is the goal, then the current FP indicator may not fully capture the financial consequences of using health services on households.[27,28]

In addition, by focusing only on actual financial outlays, the official SDG indictor fails to distinguish between households with zero health spending due to a lack of need versus those that do not use health services due to a lack of affordability.[9] This limitation has long been recognized in the literature and was part of the motivation for also including a health service coverage indicator alongside the FP indicator in the operationalization of the UHC monitoring framework. However, the current health coverage indicator only includes data on a limited number of key health services, many of which currently lack data to be adequately measured across countries and over time.[29] As a result, the current UHC measurement strategy is unlikely to be adequately distinguishing between improvements in UHC due to changes in FP without controlling for changes in the proportion of the population with unmet need. For example, if rates of unmet need were to increase in a country, we might estimate lower levels of CHE in that country. A proposed alternative indicator that integrates elements of both the CHE and the IHE indicators captures data on the proportion of the population with no health

Table 1. RACER criteria and sub-criteria, adapted for financial protection[24]

Relevant	
Policy Support & Identification of Targets and Gaps	Indicator is related to policy objectives, and allows for monitoring of progress and identification of gaps
Identification of Trends	Indicator can be used to track changes over time, identifying trends
Scope/Levels of Application	Indicator provides information relevant to the effective levels of application (e.g. allows for disaggregation if necessary)
Accepted	
Stakeholder Acceptance	Indicator is easily understood and accepted by stakeholders and is simple, both conceptually and in calculation
Credible	
Unambiguous	Indicator is unambiguous in its interpretation, both by policy makers and the general public, and allows for clear conclusions to be drawn which may guide political action
Transparency of Method	Data and calculation methods of the indicator are fully disclosed, interpretable, and reproducible
Easy	
Data Availability	Calculation of the indicator does not require data that is difficult or expensive to collect, or that is not measurable
Technical Feasibility	Calculation of the indicator is feasible using software and expertise available by those who would be using it
Complementarity & Integration	Indicator can be integrated with or complemented by the other indicators being assessed
Robust	
Defensible Theory	Indicator is based on a sound theory and assumptions are clearly stated and reasonable
Sensitivity	Indicator is sensitive to detecting policy-significant changes
Data Quality	Underlying data inputs are high quality and inaccuracies and errors are minimal
Completeness	Indicator is comprehensive and accounts for all aspects of financial protection
Reliability	Indicator is reliable in terms of its accuracy, repeatability, and how it is calculated

spending, however it also fails to fully capture unmet need.[30] Methods have been developed to measure unmet need for health services in other contexts,[31] usually based on self-reports of forgoing the use of health services due to unaffordability, but more research is needed to understand if such measures could be operationalized globally and could be integrated into global estimates of UHC.

Some conceptualizations of FP have also emphasized the idea that households should be protected not just against OOPs but also against the risk of needing to use health services,[8,15,32] which suggests that some believe that the measurement of FP should also include an ex-ante measure of risk and not simply an ex-post measure of OOPs.[8] Prior to the decision to use CHE as the official SDG indicator, the IAEG SDG had recommended using the proportion of the population with health insurance coverage as the official SDG indicator, partially motivated by these concerns as well concerns about data availability.[13] Vocal objections to this recommendation, however, eventually led to the replacement of the health insurance indictor, which only imperfectly measures FP as insurance coverage does not guarantee FP,[33] with the current FP indicator, but the focus on risk was lost in this replacement. While some health spending may actually be relatively predictable (e.g. chronic illnesses), other types of health spending (e.g. hospitalizations) may be much less so. Plus, conditional on being ill and seeking care, there is uncertainty about the magnitude of medical expenses. Insurance theory suggests that there are different welfare consequences to risk-averse households for certain versus uncertain health expenditures.[34] Yet the official SDG indictor fails to capture any of the financial consequences of uncertainty. While it is currently challenging to measure the financial consequences of uncertain health expenditures, indicators that capture elements of risk into estimates of financial protection against health expenditures have been proposed and should be further explored.[15,27]

The SDGs also include a number of cross-cutting goals, such as equity, thus a useful indicator of FP should be sensitive to the distributional aspects of FP. However, numerous studies in many international contexts have found the official SDG indicator to be the least sensitive to equity when tested alongside other FP indicators, and strong pro-rich gradients have frequently been found in the prevalence of CHE in many international contexts.[35–40] This pattern may partly be explained by the failure of the CHE indicator to capture lack of affordability among the poor, but could also be due to its inability to distinguish more discretionary forms of health spending,[16] including the fact that higher wealth households are more likely to use the private sector in most international contexts.[41] Indicators that adjust the denominator to account for the fact that poorer households require a larger proportion of their income to support subsistence spending or to account for more discretionary non-health spending among wealthier households, such as the capacity to pay indicator proposed by Xu et al.,[42] have consistently been shown to be more effective at identifying poorer households.[28] Given the potential of a pro-rich bias with the CHE indicator, caution should be placed on the interpretation of changes in this indicator in discussions relevant to equity and the adoption of more equity sensitive measures should be considered.

Identification of Trends

For an indicator to be useful for monitoring purposes, we argue that countries that have implemented successful reforms aimed at improving the financial burden households face in seeking health services should see an improvement in the FP indicator over time. Among rigorous evaluations of large-scale health system reforms conducted in many countries, some studies, such as those in Mexico,[43] Ghana,[44] and Thailand[45] have found reductions in the proportion of households experiencing CHEs. However, other studies have found opposite effects, such as those seen in China[46,47] and Peru.[48] These latter results have generally been explained by the fact that reforms may increase the utilization of health services or the types of services demanded, which could actually lead to higher levels of OOPs and thus CHE. Similarly, reforms aimed entirely at increasing utilization of health services, for example those that improve the quality of services delivered, all else equal, could actually lead to increases in OOPs. In both cases it would be hard to argue that households are not being made better off by such reforms; rather the contrary. However it does suggest that the official SDG indicator is limited in its ability to identify trends within countries, given that the way it is constructed, which makes it blind yet at the same time sensitive to changes in quantity or quality of health services utilization. An analysis of the trends in the FP indicator alone needs to be complimented with an analysis of the trends in a comprehensive measure of service coverage as well as potential policy changes in order for it to be useful. By ignoring these other factors and putting too much focus on the official FP indicator alone, we may actually disincentivize countries to undertake reforms that could lead to increased use of health services.

Scope/Levels of Application

While all of the FP indicators allow for some level of disaggregation (e.g. by geographic location or by wealth quintile), they are all calculated at the household level and thus cannot easily be disaggregated along other policy relevant lines, such as by gender or for other disadvantaged

groups that do not cluster well within households. Because health expenditures may differ for different groups of people and given the emphasis on equity enshrined in the SDGs, there is a need for such levels of disaggregation. Indeed, a mixed-methods study from Burkina Faso found that women faced additional pressures in paying for maternal health services and that such health spending can have important long-term economic and social consequences.[49] The official SDG indicator does not easily allow for such gendered inequities to be captured. Furthermore, only the indicators that use non-subsistence expenditures as the denominator attempt to adjust estimates for household size, which could also affect the ability of households to deal with a fixed level of health spending, all else equal. The lack of consideration of household size in the calculation of the official SDG indicator also likely affects the international comparability of this indicator across countries, since there are large differences in average household size globally, however, it is unknown how this may influence global monitoring.

Acceptable

Stakeholder Acceptance

Given the lack of consensus in the literature with regards to the best or most appropriate measure of FP, and the level of debate that has been associated with the indictor selection process,[13] there is reason to believe that all of the indicators suffer from issues related to acceptability regarding the choice of indicator. At the heart of the debate is the trade-off between identifying a measure that can be readily and easily constructed using data collected from routine, high coverage surveys, versus a measure that might better capture these concepts but that is not currently collected in as many countries. Ultimately, the choice to use CHE as the official SDG indicator was likely due to its longer history of use in the literature, as well as its ease of calculation and interpretation, all of which have made it relatively well accepted by stakeholders. That said, the CHE indicator was classified by IAEG SDG as a tier III indicator until recently, meaning that there was no internationally established methodology or standards for calculating it.[50] As such, the acceptability of the current indicator is not likely to be particularly strong among all stakeholders, although it is unclear if any of the other FP indicators are likely to be more acceptable to all stakeholders.

Credible

Unambiguous

As discussed above, a key limitation of the FP indicators is that it is difficult to interpret the underlying causes behind any changes in the measure of OOPs, or the numerator, but it is less widely appreciated that the CHE indicators are also highly sensitive to changes in the denominator. If systematic changes in the levels of income or consumption in a country occur, most likely due to influences outside the health sector, it would also affect the measures of FP. For example, we may see a reduction in the incidence of CHEs if average household income or consumption rises without any efforts to improve coverage of FP. Therefore, the current official FP indicator performs poorly at providing an unambiguous measure of changes in FP due to the influence of general economic conditions of households on its denominator. The IHE indicator overcomes this limitation to some extent when health spending is compared to a fixed, potentially internationally comparable, poverty line rather than household level income or consumption, which provides some advantages for global monitoring purposes. More research is needed to understand how general economic conditions influence estimates of FP and whether fixed poverty lines represent an improvement upon the current indicator to measure FP in the long-run, such as the time period over which the SDGs should be achieved.

Transparency of Methods

The official SDG indicator as well as all of the FP indicators rely upon similar methods to calculate health expenditures and are thus similarly transparent in the methods used to aggregate the numerator. However, because there are no internationally standardized survey instruments used to measure OOPs, and since countries tend to use different survey instruments in different years, it is unclear how comparable estimates of OOPs are for cross-country comparisons over time. Several studies have demonstrated that the use of different survey instruments can have a substantial effect on overall estimates, with, for example, variations in the length of the questionnaire, the recall period, and the unit of analysis all significantly affecting the resulting estimates.[51–55] Similarly, although there are widely recommended approaches to calculating total household consumption or income, again due to diversity of survey instruments, there is also likely substantial variation across countries in terms of how these are calculated. This lack of transparency may limit the comparability of estimates of FP across countries. Standardized categories to estimate household expenditures, such as the classification of individual consumption according to purpose (COICOP), have been developed and should be used to increase estimates of OOPs across countries and to improve the transparency of the calculation of FP estimates.

Easy

Data availability

The official SDG indicator is currently categorized as a tier II indicator by the IAEG SDG, which is reserved for indicators for which there are established international methodologies and standards but for which "data are not regularly produced by all countries."[50] Data to calculate FP indicators are typically sourced from household budget and expenditure surveys (HBES), which are conducted irregularly in most countries, ranging from every year to up to every five years, and many countries do not conduct surveys at all. A recent effort to aggregate all available data to calculate FP indictors across countries was only able to identify data from 122 countries, and among those with data, only 93 had data from more than one time point in recent decades.[6] Furthermore, the median year that data was calculated was 2005, which may be too old to provide a meaningful benchmark to begin to monitor progress toward the SDGs. Although a more updated version of the same database was recently released,[56] which was able to identify data from at least a dozen or so more countries, the overall picture, especially over time, remains highly incomplete. Additional data collection efforts are likely necessary in order to provide a more complete picture of FP and to be able to adequately track progress going forward.

Technical Feasibility

One of the key features of the CHE indictor that likely influenced its selection as the official SDG is the relatively easy calculation that is required to estimate it: researchers need only to have data on household OOPs and total household income or consumption and then divide one by the other.[15] It is technically feasible to calculate all of the FP indicators discussed in this paper using the same underlying household survey data, provided that the survey utilized acceptable methods to measure and report health expenditures, food and non-food spending, and consumption or income. While many HBES provide household welfare aggregates, since these aggregates are not disaggregated along the lines needed, it is technically more challenging to calculate the capacity to pay metric relative to the budget share metric. Although a number of resources have been developed to help researchers calculate FP indicators,[57,58] unless the needed aggregates are made available to the user, calculating other indicators or metrics may be technically more demanding than calculating the official SDG indicator. Indeed, a recent study calculating CHEs across countries relied primarily upon welfare aggregates due to the computational effort that would have been required to individually calculate these budget estimates for each country.[6]

Complementarity and Integration

Another limitation of the methods used to calculate the FP indicators is that they rely upon data sourced from non-health household surveys, such as HBES, which rarely collect detailed and comprehensive data on health or health service utilization. Therefore, it is not currently possible to measure whether the same households are seeing improvements in both FP as well as service coverage, the other important component of UHC. Similarly, it is not currently possible to develop indicators that are adjusted for nonuse of health services or the types of service providers used. There is a need to implement surveys or that will allow for critical health, health service, and health expenditure data to be collected along with household income or consumption data to allow for the monitoring of progress in similar populations. A number of countries, such as Kenya, Tunisia, and the Philippines, have recently launched national health expenditure and utilization surveys which allow for detailed data on both health and expenditures, potentially providing more complete estimates of FP through an analysis of underlying drivers of health expenditures.

Robust

Defensible Theory

The use of consumption rather than income as a measure of available resources draws upon the permanent-income theory, which postulates that households may have unpredictable streams of income, but that they will smooth their consumption over time to maximize their lifetime utility.[59] This is particularly important in countries where a large portion of the population works in the agricultural sector, as income varies a lot by season, whereas consumption tends to be more stable. Studies have shown that health spending may also vary according to season.[60] Current surveys used to source FP indicators rarely take this type of seasonal variation into consideration and estimates of FP may vary depending on when the survey is conducted.[40] Additionally, the use of a 10% threshold (or other thresholds used in the calculation of other FP indicators) has no theoretical backing and it unknown if there is any association between crossing this threshold and the likelihood that families face a policy-relevant level of financial hardship or reduction in their living standards.

Sensitivity

Hsu et al.[7] found that country rankings are highly sensitive to both the choice of threshold used to construct the CHE indicator and the choice of household resources used in the denominator in other FP indicators, and

noted that varying the thresholds and denominators also leads to very different rankings among countries, which poses an important challenge to global monitoring. In addition, other studies have shown that the type of survey used to calculate FP can greatly influence estimates of OOPs[30] and thus cross-country comparisons are likely highly sensitive to the type of survey used to calculate FP. Countries also vary with regards to whether they use data from the general expenditure module or from a health module to calculate OOPs. A study assessing the incidence of CHEs in India using different household surveys found that CHE estimates varied dramatically from survey to survey depending on the number of items used to capture health and total expenditures.[61] Overall sensitivity of FP measures to survey features is likely underappreciated and to date there is limited evidence available to fully understand how sensitive current estimates of FP are to these types of changes.

Data Quality

The underlying quality of the data needed for the calculation of the official SDG indicator is unclear across countries. Consumption modules between the different surveys vary greatly in terms of depth and detail, and issues such as inconsistent recall periods lead to biases including telescoping errors, rule of thumb errors, recall biases, and personal leave out biases.[52,62] Furthermore, although standards such as COPIP have been developed to measure standardized OOPs across countries, there is substantial variation across countries in terms of which health items are included, the specific wording of questions used, and the recall periods used. Therefore, the underlying data used to calculate the official SDG indicator are likely to suffer from important data quality issues.

Completeness

While the CHE indicator alone may provide an incomplete estimate of FP in a country, as suggested by the WHO monitoring framework, by complimenting the official SDG indicator with another indicator, such as with an IHE indicator, it may provide a more complete picture of FP in a country.[30] As previously mentioned, however, several dimensions of FP are not covered by all of the indicators, such as the use of coping mechanism, thus using two indicators may not provide a complete picture of FP among households.

Reliability

The reliability of the different indicators has been assessed in a number of studies, all of which have demonstrated a high degree of variation in FP estimates both between and within indicators depending on the context in which they are used. Similarly, Lu et al.[53] evaluated the sensitivity of estimates of CHEs to survey design over time, and observed not only a significant difference in incidence depending on the survey used, but also a lack of consistency from year to year in whether or not a particular survey method produced estimates higher or lower than the other. Limited additional evidence is available to assess the reliability of the current FP indicator.

Discussion

This paper assessed the usefulness of the official SDG indicator to measure and monitor the FP component of UHC across counties and over time. To our knowledge, this is the first study to attempt to draw conclusions regarding the usefulness of indicators for monitoring of UHC. Our assessment, which used the RACER evaluation framework, revealed three main findings.

First, while all FP indicators have shortcomings, the official SDG indicator has a number of important limitations. Importantly, the CHE indicator has been shown to exhibit a pro-rich prevalence in many countries and thus likely does a very poor job at tracking the distribution of FP across households. It also lacks a robust theoretical basis to support its prominence as the official measure of FP. While the official SDG indicator has some important merits, such as that it is easy to calculate and is among the most widely used of the indicators, the limitations of the CHE indicator may mean that it is necessary to rethink the choice of this indicator to monitor progress achieving SDG 3.8.2.

Second, our assessment reveals that, due to their reliance on unreliable, incomparable, and potentially poor-quality survey data, all of the current indicators are subject to important measurement challenges that will make it difficult to make comparisons of FP across countries and over time. For example, we cannot currently say with any degree of certainty whether a measured change in the level of FP in a population is the result of changes in the numerator, which are more likely to be attributed to reforms that occur within the health system, or to changes in the denominator, which are more likely to be due to changes outside the health sector. Therefore, it is unlikely that tracking changes using the official SDG indicator will provide a meaningful picture of how countries are progressing in terms of ensuring FP. Thus, our ability to monitor progress toward UHC across countries and over time is limited.

Third, while there are many outstanding conceptual challenges underpinning the meaning of FP, which influences how it should be measured, it is clear that the official FP indicator does a relatively poor job at capturing some of these aspects of FP. For example, the CHE fails to account for how households finance OOPs and does not

adequately control for changes in unmet need, nor does it account for the risk households face from OOPs. Therefore, relying upon a single FP indicator for global monitoring is unlikely to provide a comprehensive view of FP. Alternative indicators and measures that do account for such features have been proposed and should be further explored for incorporation into global monitoring frameworks.[28]

While our findings suggest that much more work is needed to understand how alternative indicators could be used to improve the measurement of FP, they also point to the need to urgently focus on improving the quality of the underlying data used to calculate FP. Additionally, countries should consider adopting more internationally standardized survey instruments, including a more uniform number of items, consistent wording, and standard recall periods in order to improve consistency of OOP estimates and comparability within and across countries and over time. It is also necessary to increase the frequency with which these surveys are administered, as well as to expand the number of countries from which it collects data so that trends and changes can be adequately tracked over time. Finally, it also raises the question as to whether a global monitoring framework that puts so much weight on a single indicator to capture all the dimensions of FP is warranted. Regional monitoring frameworks, such as those developed in Europe, or frameworks with multiple indicators of FP have been developed and recommended by other organizations, which may lead to more useful comparisons and provide a more complete picture of country progress toward achieving UHC.

In interpreting our results, we caution the reader that our paper is subject to a number of limitations. First, while we have utilized the RACER framework, which has previously been utilized to evaluate other global indicators, there are likely other frameworks that could also be used for this purpose and may lead to different conclusions. Second, while the RACER framework is relatively comprehensive, it may leave out criteria that other people believe are important in the evaluation of the usefulness of an indicator for global monitoring purposes. Third, while our analysis is based on a thorough review of the literature, our review was not systematic in nature and thus we may have overlooked important data or evidence from other studies that could influence our findings.

Conclusions

The advent of global development agendas, mostly notably the establishment of the Millennium Development Goals (MDGs) and now the SDGs, and the rise of global goal setting, has put a lot of power on indicators to shape the way issues are conceptualized, prioritized, and addressed by countries. This is especially true in the area of global health, where over the past few decades there has been significant growth in the number of indicators used to compare and assess the performance of countries in improving health outcomes.

The official SDG monitoring framework currently advocates the use of only the CHE indicator (at both the 10% and 25% thresholds) to measure and monitor progress toward the FP component of UHC. However, using the RACER framework, we identified a number of serious limitations of the official SDG 3.8.2 indicator, which will likely limit its usefulness in measuring and monitoring progress toward the FP component of UHC. Based on this, we recommended that further research should be conducted in order to better understand how alternative indicators may be more useful and efforts should be launched to both improve and standardize the quality of underlying data as well as the frequency of data from which FP indicators are calculated.

However, given the challenges we have identified, we also raise the question whether or not the concept of FP can adequately be measured by a single indicator. Indicators have become immensely popular over the past few decades and have been used to describe a wide variety of social phenomena. They are attractive due to their apparent ability to present information in simple, countable terms. However, attempting to categorize a complex issue risks oversimplifying it and divesting it of its context, history, and meaning.[20] The current SDG monitoring framework which only recommends a single approach to measure FP is unlikely to fully capture the complexity of this concept and could lead to potentially distorted or misleading conclusions or worse could provide perverse incentives to policy makers. In the context of increasingly complex global health challenges, the reliance on imperfect indicators of FP may actually limit efforts to reform health systems in ways to help countries achieve UHC.

ORCID

Karen A. Grépin ⓘ http://orcid.org/0000-0003-4368-0045
Bridget R. Irwin ⓘ http://orcid.org/0000-0002-5932-1891
Benjamin Sas Trakinsky ⓘ http://orcid.org/0000-0002-3238-4614

References

1. World Health Organization. Health financing: financial protection. World Health Organization; Published 2019. [accessed 2019 Feb 12]. https://www.who.int/health_financing/topics/financial-protection/en/.
2. Gertler P, Gruber J. Insuring consumption against illness. Am Econ Rev. 2002;92(1):51–70. doi:10.1257/000282802760015603.
3. Scott JW, Raykar NP, Rose JA, Tsai TC, Zogg CK, Haider AH, Salim A, Meara JG, Shrime MG. Cured into destitution: catastrophic health expenditure risk among uninsured trauma patients in the United States. Ann Surg. 2018;267(6):1093–99. doi:10.1097/SLA.0000000000002254.
4. Alam K, Mahal A. Economic impacts of health shocks on households in low and middle income countries: A review of the literature. Global Health. 2014;10(1):21. doi:10.1186/1744-8603-10-21.
5. Wagstaff A, Flores G, Hsu J, Smitz, MF, Chepynoga, K, Buisman, LR, van Wilgenburg, K, Eozenou, P. Progress on catastrophic health spending in 133 countries: A retrospective observational study. Lancet Glob Heal. 2017;6:169–79. doi:10.1016/S2214-109X(17)30429-1.
6. Wagstaff A, Flores G, Smitz MF, Hsu J, Chepynoga K, Eozenou P. Progress on impoverishing health spending in 122 countries: A retrospective observational study. Lancet Glob Heal. 2017:5–10. doi:10.1016/S2214-109X(17)30486-2.
7. Hsu J, Flores G, Evans D, Mills A, Hanson K. Measuring financial protection against catastrophic health expenditures: methodological challenges for global monitoring. Int J Equity Health. 2018;17(1):1–13. doi:10.1186/s12939-018-0749-5.
8. Saksena P, Hsu J, Evans DB. Financial risk protection and universal health coverage: evidence and measurement challenges. PLoS Med. 2014;11(9):e1001701. doi:10.1371/journal.pmed.1001701.
9. Moreno-Serra R, Millett C, Smith P. Towards improved measurement of financial protection in health. PLoS Med. 2011;8(9):e1001087. doi:10.1371/journal.pmed.1001087.
10. World Health Organization, World Bank Group. Monitoring progress towards universal health coverage at country and global levels: framework, measures and targets; 2014. https://apps.who.int/iris/bitstream/handle/10665/112824/WHO_HIS_HIA_14.1_eng.pdf?sequence=1.
11. World Health Organization Regional Office for Europe. Can people afford to pay for health care? New evidence on financial protection in Europe. Copenhagen, Denmark: WHO Regional Office for Europe; 2019. https://apps.who.int/iris/bitstream/handle/10665/311654/9789289054058-eng.pdf?sequence=1&isAllowed=y.
12. Thomson S, Evetovits T, Cylus J, Jakab M. Monitoring financial protection to assess progress towards universal health coverage in Europe. Public Heal Panor. 2016;2:249–400.
13. Reddock J. Seeking consensus on universal health coverage indicators in the sustainable development goals. J Health Serv Res Policy. 2017;22(3):1–5. doi:10.1177/1355819617704676.
14. McIntyre D, McKee M, Balabanova D, Atim C, Reddy KS, Patcharanarumol W. Open letter on the SDGs: A robust measure for universal health coverage is essential. Lancet. 2016;388(10062):2871–72. doi:10.1016/S0140-6736(16)32189-4.
15. O'Donnell O. Financial protection against medical expense. Amsterdam; 2019. doi:10.1093/acrefore/9780190625979.013.27.
16. McPake BI. Crunching health expenditure numbers: important but treacherous terrain. Lancet Glob Heal. 2018;6(2):124–25. doi:10.1016/S2214-109X(18)30007-X.
17. Van Damme W, Meessen B, Por I, Kober K. Catastrophic health expenditure. Lancet. 2003;362(9388):996–97. doi:10.1016/S0140-6736(03)14375-9.
18. Prah Ruger J. Catastrophic health expenditure. Lancet. 2003;362(9388):996–97. doi:10.1016/S0140-6736(03)14376-0.
19. Zeller RA, Carmines EG. Measurement in the social sciences: the link between theory and data. Cambridge: Cambridge University Press; 1980.
20. Engle Merry S. The seductions of quantification: measuring human rights, gender, violence, and sex trafficking. Chicago (Illinois): The University of Chicago Press; 2016.
21. European Comission. Impact assessment guidelines; 2005. https://ec.europa.eu/agriculture/sites/agriculture/files/sfs/documents/documents/sec2005-791_en.pdf.
22. European Commission. Indicators to measure social protection: implications for EC programming. Brussels/Belgium: European Commission; 2017.
23. International Cooperation and Development (DEVCO). Integrating the environment and climate change into EU international cooperation and development: monitoring and indicators. Luxembourg: European Union; 2016.
24. Gerdes H, Bassi S, Policy E, Mazza L, Srebotnjak T, Porsch L. D2.2 final report: evaluation of indicators for EU policy objectives. Ecologic; 2011.
25. McIntyre D, Thiede M, Dahlgren G, Whitehead M. What are the economic consequences for households of illness and of paying for health care in low- and middle-income country contexts? Soc Sci Med. 2006;62(4):858–65. doi:10.1016/j.socscimed.2005.07.001.
26. Flores G, Krishnakumar J, O'Donnell O, Van Doorslaer E. Coping with health-care costs: implications for the measurement of catastrophic expenditures and poverty. Health Econ. 2008;17(12):1393–412. doi:10.1002/hec.1338.
27. Flores G, O'Donnell O. Catastrophic medical expenditure risk. J Health Econ. 2016;46:1–15. doi:10.1016/j.jhealeco.2016.01.004.

28. Wagstaff A. Measuring catastrophic medical expenditures: reflections on three issues. Health Econ. 2019;28(6):765–81. doi:10.1002/hec.3881.

29. Wagstaff A, Neelsen SA. comprehensive assessment of universal health coverage in 111 countries: A retrospective observational study. Lancet Glob Heal. 2020;8(1):39–49. doi:10.1016/S2214-109X(19)30463-2.

30. Wagstaff A, Eozenou PH-V. CATA meets IMPOV: a unified approach to measuring financial protection in health; 2014. http://documents.worldbank.org/curated/en/114091468149393397/pdf/WPS6861.pdf.

31. Eurostat. Unmet health care needs statistics. Eurostat: Statistics explained; 2019.

32. World Health Organization. Health financing: what is health financing for universal coverage? World Health Organization. https://www.who.int/health_financing/universal_coverage_definition/en/.

33. Wagstaff A, Neelsen S. Medical expenditures: not the only source of financial hardship. Lancet Glob Heal. 2019;8(3):e337. doi:10.1016/S2214-109X(20)30026-7.

34. Zeckhauser R. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. J Econ Theory. 1970;2(1):10–26. doi:10.1016/0022-0531(70)90010-4.

35. Koch KJ, Cid Pedraza C, Schmid A. Out-of-pocket expenditure and financial protection in the Chilean health care system—a systematic review. Health Policy (New York). 2017;121(5):481–94. doi:10.1016/j.healthpol.2017.02.013.

36. Yardim MS, Cilingiroglu N, Yardim N. Catastrophic health expenditure and impoverishment in Turkey. Health Policy (New York). 2010;94(1):26–33. doi:10.1016/j.healthpol.2009.08.006.

37. Joe W, Rajpal S. Unravelling the socioeconomic gradient in the incidence of catastrophic health care expenditure: a comment. Health Policy Plan. 2018;33(5):697–98. doi:10.1093/heapol/czy025.

38. Ghosh S. Catastrophic payments and impoverishment due to out-of-pocket health spending. Econ Polit Wkly. 2011;46(47):63–70. 24 ref

39. Cylus J, Thomson S, Evetovits T. Catastrophic health spending in Europe: equity and policy implications of different calculation methods. Bull World Health Organ. 2018;96:599–609. doi:10.2471/BLT.18.209031.

40. Sas Trakinsky B, Irwin BR, Guene H, Grépin KA. An empirical evaluation of the performance of financial protection indicators for UHC monitoring: evidence from Burkina Faso. Heal Policy Open. 2020. doi:10.1016/j.hpopen.2019.100001.

41. Grépin KA. Private sector an important but not dominant provider of key health services in low- and middle-income countries. Health Aff. 2016;35(7):1214–21. doi:10.1377/hlthaff.2015.0862.

42. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJL. Household catastrophic health expenditures: A multi-country analysis. Lancet. 2003;362:565–72.

43. King G, Gakidou E, Imai K, Lakin J, Moore RT, Nall C, Ravishankar N, Vargas M, Téllez-Rojo MM, Ávila JEH, et al. Public policy for the poor? A randomised assessment of the Mexican universal health insurance programme. Lancet. 2009;373(9673):1447–54. doi:10.1016/S0140-6736(09)60239-7.

44. Powell-Jackson T, Hanson K, Whitty CJM, Ansah EK. Who benefits from free healthcare? Evidence from a randomized experiment in Ghana. J Dev Econ. 2014;107:305–19. doi:10.1016/j.jdeveco.2013.11.010.

45. Limwattananon S, Neelsen S, O'Donnell O, Prakongsai P, Tangcharoensathien V, van Doorslaer E, Vongmongkol V. Universal coverage with supply-side reform: the impact on medical expenditure risk and utilization in Thailand. J Public Econ. 2015;121:79–94. doi:10.1016/j.jpubeco.2014.11.012.

46. Wagstaff A, Lindelow M. Can insurance increase financial risk? The curious case of health insurance in China. J Health Econ. 2008;27(4):990–1005. doi:10.1016/j.jhealeco.2008.02.002.

47. Wagstaff A, Lindelow M, Jun G, Ling X, Juncheng Q. Extending health insurance to the rural population: an impact evaluation of China's new cooperative medical scheme. J Health Econ. 2009;28(1):1–19. doi:10.1016/j.jhealeco.2008.10.007.

48. Bernal N, Carpio MA, Klein TJ. The effects of access to health insurance: evidence from a regression discontinuity design in Peru. J Public Econ. 2017;154:122–36. doi:10.1016/j.jpubeco.2017.08.008.

49. Storeng KT, Baggaley RF, Ganaba R, Ouattara F, Akoum MS, Filippi V. Paying the price: the cost and consequences of emergency obstetric care in Burkina Faso. Soc Sci Med. 2007;66(2008):545–57. doi:10.1016/j.socscimed.2007.10.001.

50. United Nations. IAEG-SDGs: tier classification for global SDG indicators. Sustainable Development Goals. https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification.

51. Gibson J, Beegle K, De Weerdt J, Friedman J. What does variation in survey design reveal about the nature of measurement errors in household consumption? Oxf Bull Econ Stat. 2015;77(3):466–74. doi:10.1111/obes.12066.

52. Beegle K, De Weerdt J, Friedman J, Gibson J. Methods of household consumption measurement through surveys: experimental results from Tanzania. J Dev Econ. 2012;98(1):3–18. doi:10.1016/j.jdeveco.2011.11.001.

53. Lu C, Liu K, Li L, Yang Y. Sensitivity of measuring the progress in financial risk protection to survey design and its socioeconomic and demographic determinants: A case study in Rwanda. Soc Sci Med. 2017;178:11–18. doi:10.1016/j.socscimed.2017.02.001.

54. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev. 2006;63(2):217–35. doi:10.1177/1077558705285298.

55. Heijink R, Xu K, Saksana P, Evans D. Validity and comparability of out-of- pocket health expenditure from household surveys: A review of the literature and current survey instruments. World Heal Organ Discuss Pap. 2011;1–30.

56. Wagstaff, A, Eozenou, P, Smitz, M. Out-of-pocket expenditures on health: a global stocktake the world bank research observer. 2020. https://dx.doi.org/10.1093/wbro/lkz009.

57. Eozenou P, Wagstaff A. FPRO: stata module to compute financial protection indicators for health expenditures. Statistical software components S458457, Boston College Department of Economics; 2018.

58. Wagstaff A, Bilger M, Sajaia Z, Lokshin M. Health equity and financial protection (english). Washington (District of Columbia): World Bank Group; 2011.

59. Houthakker H. The permanent income hypothesis. Am Econ Rev. 1958;48:396–404.

60. Sauerborn R, Nougtara A, Hien M, Diesfeld H. Seasonal variations of housheold costs of illness in Burkina Faso. Soc Sci Med. 1996;43(3):281–90. doi:10.1016/0277-9536(95)00374-6.

61. Raban MZ, Dandona R, Dandona L. Variations in catastrophic health expenditure estimates from household surveys in India. Bull World Health Organ. 2013;91 (10):726–35. doi:10.2471/BLT.12.113100.

62. Friedman J, Beegle K, De Weerdt J, Gibson J. Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania. Food Policy. 2017;72:94–111. doi:10.1016/j.foodpol.2017.08.016.