water

MDPI

Article

Moving Up the Ladder: Assessing Sanitation Progress through a Total Service Gap

Stuart Kempster * and Andrés Hueso

WaterAid, London SE11 5JD, UK; ahuesog@gmail.com
* Correspondence: StuartKempster@wateraid.org; Tel.: +44-207-793-5119

Received: 30 September 2018; Accepted: 21 November 2018; Published: 26 November 2018

Abstract: The Sustainable Development Goals create ambitious targets for achieving universal access to safely managed sanitation by 2030. The core indicator for SDG 6.2 creates positive incentives for governments, and development partners to invest in the whole sanitation chain, recognising the public health benefits of managing waste beyond initial containment. However, the target and indicators also create risks. Global accountability could be undermined by the challenge of accounting for progress across different service levels below the target of safely managed. There could also be perverse incentives to upgrade existing services, in order to meet the benchmark of safely managed, at the expense of extending basic services to those currently unserved. This paper examines methodological options for calculating a 'total service gap', a measure that would combine data on each rung of the service ladder to quantify how far away each country is from universal safely managed services. It conducts a sensitivity analysis to assess the validity of using uniform service level weights, and finds that this approach could add value to existing metrics. Through alternative data visualisations and other devices, it is argued that the total service gap could help to address the risks surrounding global accountability and perverse incentives.

Keywords: sanitation; monitoring; sustainable development goals; accountability

1. Introduction

1.1. Incentives and the SDG Monitoring Framework for Sanitation

The 2030 Agenda for Sustainable Development, which sets out 17 Sustainable Development Goals (SDGs) and 169 targets, was adopted by the UN Member States in 2015. Target 6.2 of the SDGs calls for universal access to sanitation by 2030, with the associated indicator 6.2.1 being the "proportion of the population using safely managed sanitation services" [1]. The WHO/UNICEF Joint Monitoring Programme (JMP) is responsible for monitoring global progress towards this target, using a sanitation service ladder to benchmark household services as either 'safely managed', 'basic', 'limited', 'unimproved', or 'open defecation' [1].

The explicit intention of setting global goals is to shape the development agenda, by defining priorities and creating incentives for action. At a national level, research has shown that the right incentives are imperative to drive progress towards universal sanitation services [2]. Specifically, 'values-based incentives' (for instance, around modernity and cultural heritage) and 'instrumental incentives' (such as career progression and political return) are found to influence performance at various layers of governance and service delivery [3]. The SDG targets and indicators can influence both these forms of incentive. First, they shape global norms and define what should be considered a 'modern' or 'desirable' sanitation service. Second, they create a framework of indicators, against which the work of individuals and institutions will be judged. However, as well as creating positive incentives, the potential also exists to trigger perverse incentives. The impact of the SDGs as a tool

drive sanitation progress will largely be determined by the extent to which positive incentives are maximized and perverse incentives are limited.

To understand the ways in which the SDG targets and indicators may shape the incentives and priorities of national governments and international donors, it is useful to consider experiences from the Millennium Development Goals (MDGs). A survey of government officials and civil society organizations from across 126 countries found that the MDGs were "moderately influential" in setting national developmental priorities, with the greatest influence being in Sub-Saharan Africa [4]. The main motivations for national governments to engage with the goals appear to have been increased global visibility and influence, and increased allocations of overseas development assistance (ODA) [5]. Whilst the overall impact of the MDGs on development partner priorities was more mixed, there is evidence that the global targets influenced the sectoral allocation of development spending in the early years of the MDGs; MDG-linked investments increased by 76 per cent between 2000 and 2005, compared to a 46 per cent increase for non-MDG investments [4]. While it could be argued this increase represents a correlation rather than a causality, evidence from within individual sectors reinforces the view that development spending was increasingly highly concentrated in areas that were specified in the MDG targets. For instance, the volume of ODA for primary education more than quadrupled between 2000 and 2008, whilst the volume of ODA for secondary education barely changed over the same period [6].

The political process of setting global goals and targets can therefore be seen to influence development priorities by shaping both values-based and instrumental incentives. However, instrumental incentives are also shaped by the technical process of defining and measuring specific indicators. In view of this, global indicators have been framed as a "technology of governance" [7,8]. According to this understanding, indicators can shape incentives in two ways. First, and most straightforwardly, indicators define the technical performance standards against which progress can be monitored. Those who are 'monitored' are then incentivised to improve their performance as measured by the indicator, rather than through a broader conceptualisation of the relevant goal or target. Second, indicators have a "knowledge effect", whereby they can come to define the concepts that they were originally intended to reflect [7]. In doing so, global indicators can lead to the establishment of "policy paradigms" in which certain policy options are elevated, and others diminished, as a result of how the indicator has been defined [9]. In this way, the SDGs risk the development of perverse incentives, as the process of translating complex realities into globally comparable, quantifiable indicators can obscure contextual differences or local peculiarities [8].

The ambition inherent in the core indicator for SDG 6.2 creates positive incentives, but also potentially perverse incentives. The inclusion of the safe management of excreta in the indicator puts a greater focus on public health, and will incentivise vital investments in faecal sludge management and wastewater treatment. However, having a single 'top of the ladder' indicator could create perverse incentives by obscuring the progress made at lower rungs of the ladder. Understanding how these perverse incentives could be avoided requires an appreciation of how global accountability is conceptualised and operationalised within the SDG framework.

1.2. Global Accountability within the SDG Framework

The targets and indicators of global goals only have an impact on national priorities and incentives if there is some form of accountability for their achievement. The text of the Agenda 2030 declaration states that the SDGs will promote accountability to citizens via a "robust, voluntary, effective, participatory, transparent, and integrated follow-up and review framework" [10]. The global architecture for follow-up and review is centred on an annual high-level political forum (HLPF), which is tasked with assessing progress, achievements, and challenges, and ensuring that the agenda remains relevant and ambitious. The HLPF itself is designed to be the culmination of a network of global follow-up and review processes, which include thematic reviews of progress—these are linked to the annual theme of the HLPF and they build on the work of the Economic and Social Commission—and

the Voluntary National Reviews by member states [11]. However, whilst institutional frameworks exist to track and assess progress at the global level, the conceptualisation of 'global accountability' within the SDG architecture remains rather amorphous.

The difficulty of defining global accountability is by no means unique to the SDGs, posing challenges in numerous areas of international governance [12]. Accountability at the national level is defined in terms of answerability, enforceability, and responsibility [13]. However, these dimensions are mostly absent at the global level, and none are fully operational within the global SDG framework. Answerability and responsibility are undermined to some extent by the 'aspirational' nature of global targets, and also by the framing of the agenda itself. The Agenda 2030 declaration holds national governments accountable for national targets—which only need to be 'guided' by the global ambition—with global accountability explicitly subordinate to national processes [10]. It has been argued by Engebretsen et al. [14] that because of the extremely broad framing of Agenda 2030, responsibility becomes both all-encompassing and non-existent—the SDGs are "everyone's business but no-one's major responsibility". These ambiguities weaken responsibility and answerability in the global accountability processes. Furthermore, enforceability is completely absent at the global level, with the follow-up and review processes designed to be voluntary in nature [11].

Because of these challenges, global accountability within the SDG framework has been based largely on the principles of 'mutual accountability' [15]. As articulated in the Paris Declaration on Aid Effectiveness, and later in the Busan Global Partnership for Effective Development, mutual accountability refers to a set of voluntary commitments that rely on trust and partnership to drive progress around shared agendas, rather than on sanctions for non-compliance [16,17]. Within a mutual accountability framework, global mechanisms seek to harness the "power of reputation" by identifying countries or cities that can demonstrate significant progress toward meeting individual targets, and facilitating the sharing of these success stories to inspire action by others [12]. Furthermore, accountability mechanisms should provide high quality analysis to help decision makers to better understand the possible pathways to success.

Within the follow-up and review process, there should therefore be space for in-depth qualitative assessments of progress to facilitate peer learning and course correction, as envisaged by the Voluntary National Reviews (VNRs). Analysis of the reviews submitted to date has largely argued that whilst this process has demonstrated governments' political commitment to the SDGs, it has not yet succeeded in providing a platform for peer learning and course correction. The lack of standardisation in content has made it difficult to record and compare the progress across countries, or to identify trends and lessons of success or failure [18,19]. Whilst there is clearly scope to improve the VNRs, a detailed discussion of these processes is beyond the scope of this study. Rather, this paper analyses the quantitative indicators that are used to monitor SDG 6.2 at a global level, and discusses how they can be tailored to better meet the needs of mutual accountability. It is important that these indicators are sensitive to the ambiguities of accountability that are inherent in the SDG agenda itself, and that they are able to provide the "light political tracking" needed to maintain political attention, demonstrate progress over time, and create incentives for all countries to engage, regardless of their circumstances or current levels of sanitation coverage [14,20].

1.3. The Opportunities and Risks of the Global Monitoring Framework for SDG 6.2

The current sanitation indicators support the objectives of mutual accountability, to an extent. The strength of the JMP service ladder is its comparability across countries and its relevance for different levels of sanitation development, reflecting the gradual and often stepwise nature of progress. The ambition of the target reflects the normative criteria of the human right to sanitation, and it has stimulated discourse and action around the management of faecal waste beyond initial containment. Furthermore, the exact framing of target 6.2, which specifies the elimination of open defecation, helps to maintain a focus on both ends of the service ladder. In these ways, the global monitoring framework contributes towards the development of positive incentives for actors in the sanitation sector.

However, there is also a risk that governments and development partners may be faced with perverse incentives to concentrate efforts solely on safely managed services, to demonstrate their contribution to the global target. This creates a danger that those who already have access to basic sanitation services will be prioritised over those currently at the lower rungs of the ladder, for whom reaching safely managed services is more difficult and expensive. For instance, there are already accounts of reduced appetite for promoting and investing in shared sanitation—classified as a limited service—which is widely recognised as an important service option in certain contexts, such as dense informal settlements [21]. These perverse incentives stem from a 'knowledge effect' created by the revisions to the JMP's service ladder following the adoption of the SDG targets. As the custodial agency of SDG 6.2, it was the JMP's responsibility to define what was meant by the term "safely managed service". The original proposition of the JMP's 'Sanitation Task Team' was to include a benchmark for 'basic sanitation' that would include households using facilities shared by no more than five families and by no more than thirty people. However, this recommendation was rejected by the JMP because the household surveys upon which they base global estimates do not often contain data on the number of households sharing a facility. Furthermore, they found it difficult to define an adequate proxy indicator, as the evidence on the relationship between the number of households sharing facilities and their 'safety' was judged to be inadequate. Consequently, the JMP excluded shared facilities from the definition of 'basic' and 'safely managed services', and created the new category of 'limited service' [21]. While this decision can be seen as justified from the perspective of global monitoring, it is important to note that it has produced a policy paradigm in which shared sanitation is seen as less desirable in every context. Yet this paradigm stems from what was technically feasible from a monitoring perspective, not necessarily what was appropriate from a policy perspective. In many ways, this reinforces the criticisms made of the JMP's 'technology-based' monitoring under the MDGs, and the argument that monitoring should instead focus on the intended 'functions' of a sanitation service [22]. With regard to shared sanitation, this is problematic, as there are numerous examples of contexts in which household sanitation is not feasible, and—short of rehousing—high quality shared facilities represent the best option to improve service levels [21,23,24]. Owing to this policy paradigm, the perverse incentives associated with the global indicator for SDG 6.2 are therefore most likely to penalise underserved and vulnerable populations, especially those living in dense informal settlements in the poorest countries.

A second risk of the global monitoring framework relates to accountability. This stems from the difficulty of comparing rates of progress across countries and over time in a way that accounts for changes on each rung of the ladder. As countries have domesticated the sanitation targets, many have included targets for both safely managed and basic services. National governments are primarily accountable for these national targets, and national accountability mechanisms are therefore central to their achievement. These national processes should be supported and reinforced by efforts at the global level. With regard to sanitation, the first major accountability challenge is currently the lack of effective mechanisms at the national level, as demonstrated in two recent global reviews of accountability in the sector [25,26]. However, issues also exist within global mechanisms. As the headline global indicator only tracks safely managed services, it can only paint a partial picture of accountability where the national targets include specific objectives for lower levels of service. While the full JMP service ladder provides a more nuanced understanding than the headline indicator, it can be challenging to compare rates of progress, across countries and over time, in a way that accounts for changes at each rung of the ladder. The service ladder implies a form of ordinal utility, whereby each successive rung of the ladder is preferable to the previous one, but no judgement is made about how preferable. As such, it is difficult to assess progress over time in a consistent and comparable manner. For instance, which has made the greater progress: a country with a five percent increase in safely managed services, or a country with a 10 percent increase in basic services? These types of comparisons between countries and over time are central to operationalising the concept of mutual accountability that is outlined above. For global accountability mechanisms to reflect the progress made at a national level and to drive improvements

in the sector, it is important that they use data that can facilitate holistic comparisons, capturing progress at each rung of the ladder. This is especially important in order to showcase examples where progress at lower rungs of the ladder has been accelerated in order to reach the poorest first.

1.4. Reframing Progress through a 'Total Service Gap'

The two main risks posed by the global monitoring framework both therefore relate to the relative desirability of different service levels. Mechanisms of mutual accountability are undermined by the difficultly of comparing progress over time in a way that consistently accounts for changes made at different rungs of the ladder. Perverse incentives are created by a paradigm that diminishes the value of shared sanitation as a policy option. We argue these two issues could be partially addressed by reframing how 'progress' is defined and communicated.

This paper proposes an approach to combining the data on the SDG service ladder into one composite measure, in order to calculate a 'total service gap'. Rather than assessing progress by the percentage of the population using any individual level of service, the data contained within the service ladder is combined into one single metric to demonstrate how far away a country is from universal safely managed services. A service gap of 0% would signify universal coverage of safely managed services, whereas a service gap of 100% would signify universal open defecation. As a country's sanitation coverage increases, the service gap will reduce, according to relative progress on each rung of the service ladder.

Other options for strengthening monitoring of SDG6.2 have been proposed in the literature, such as those that measure sanitation from a 'multidimensional' perspective [27,28] or from a service-oriented perspective [29]. Both these alternatives produce a revised ladder, with additional dimensions and parameters to be measured. In doing so, they are able to capture more elements of the normative criteria of the human right to sanitation and have scope to be of value if introduced at a national level. Specifically, the multidimensional approach to measuring sanitation poverty, proposed by Gené-Garriga and Pérez-Foguet [28], identifies associations between different 'sanitation deprivations', which have the potential to be of value to national and sub-national decision makers, and to improve the targeting of sanitation policies. In this respect, if adopted at a national level, these alternatives would help to prevent the emergence of the perverse incentives associated with the more simplistic indicators that are used to monitor SDG 6.2. However, the amount of additional data that are required by these approaches means that they would have limited utility for mutual accountability or cross-country comparison at the global level in the short- to medium-term, as these data are not commonly collected across countries. The total service gap, on the other hand, takes the strengthening of mutual accountability as a foundational objective. As such, a key consideration is that it can be calculated immediately, using existing global datasets. In this way, its utility for national decision-making may be lessened in comparison to alternative proposals, but it is better able to address the risks identified within the global SDG agenda.

Assessing progress through a total service gap has the potential to address the challenges that are identified within the global monitoring and accountability framework. The progress of countries could be compared across time while accounting for changes at each rung of the service ladder in a more consistent manner, feeding into stronger mutual accountability mechanisms. A composite measure of progress would limit perverse incentives by helping to create a broader policy paradigm that values lower levels of progress, such as shared sanitation. By reframing progress in terms of the 'distance to universal services', the total service gap provides a useful communications device to help maintain political attention on sanitation in a crowded global accountability space. Importantly, this can also help to draw attention to various forms of inequality, the elimination of which is a key tenant of the SDGs. The total service gap aims to add value to existing metrics, avoiding the creation of an additional measure of 'coverage' that may confuse the existing discourse around levels of 'safely managed', or 'at least basic' services.

Despite the potential benefits of such a measure, the methodological challenge of setting appropriate weights for each service level could undermine its effectiveness or impact. Ideally, the weights would be set using evidence about the relative benefits of each service level. However, there is currently insufficient evidence in the literature to be able to do this objectively. Previous studies have assessed the impact of drinking water and sanitation on diarrhoeal disease [30], reviewed the relative health outcomes of shared and household sanitation [31,32], and compared the direct health gains from household sanitation, with the external benefit of neighbourhood sanitation [33]. However, no existing study provides the quantitative evidence that would be required to set weights for each service level on the basis of the relative health benefits. Furthermore, setting weights based on health impacts alone would be reductionist in itself, as it omits the important non-health benefits of sanitation [34]. As such, in calculating a total service gap, the weights afforded to each service level must be set subjectively. Under the framing of a total service gap, setting the weights for open defecation and safely managed services is non-contentious. Open defecation represents a complete absence of sanitation service, and safely managed services are the international gold standard; the service level weights can therefore be set at 1 and 0, respectively. The most straight-forward way to set the weights for the remaining three service levels would be to assume a uniform increase in benefits as one moves up each rung of the service ladder: a weight of 0.75 for 'unimproved', 0.5 for 'limited' and 0.25 for 'basic'. The strengths of using uniform weights are conceptual simplicity and ease of communication, which would be beneficial, given the political issues that the total service gap is intended to address. However, this model could be criticised for not accurately representing the relative benefits of each level of service, and therefore painting a misleading picture of progress. Weights could instead be set by expert opinion to better reflect the relative benefits of each level of service, or through the use of multivariate techniques to construct a more empirically grounded model. However, the increasing complexity that is required by each of these options could make the communication of the total service gap more challenging, and ultimately reduce its ability to address the risks outlined above.

This paper examines whether using uniform service level weights is an appropriate method for constructing a 'total service gap'. It does so through a sensitivity analysis, comparing this approach to three ideal-type models to assess the extent to which altering the weights impacts the total service gap, and therefore our understanding of countries' relative performance. It goes on to examine the extent to which assessing sanitation progress through the lens of a total service gap can help to address the risks associated with the global monitoring framework for SDG 6.2, with specific reference to strengthening mutual accountability and tackling perverse incentives.

2. Methods

2.1. Calculating the Total Service Gap

To calculate the total service gap, each service level is assigned a weight, which is multiplied by the percentage of the population using that level of service, according to the data from the JMP's 2017 SDG Baseline Report [1]. This gives five component values, which are added together to produce the total service gap:

$$\text{Total service gap} = (SM\% \times SMw) + (B\% \times Bw) + (L\% \times Lw) + (U\% \times Uw) + (OD\% \times ODw)$$

where SM is 'safely managed', B is 'basic', L is 'limited', U is 'unimproved', OD is 'open defecation', and w is the weight assigned to each service. The weights range from 0–1, producing a total service gap that is expressed as a percentage. A total service gap of 100% represents universal open defecation, and a gap of 0% represents universal safely managed services.

A challenge in creating a composite metric for the whole sanitation ladder is the number of countries that currently lack data on safely managed services. Foa and Tanner [35] identify three methods for dealing with missing data in composite indexes. The first is casewise deletion—simply

removing those countries that lack data. However, there is a strong correlation between countries that lack data on safely managed services and those that have low levels of sanitation coverage. As such, casewise deletion would exclude the countries where the risks associated with perverse incentives are most apparent. The second method is to impute missing values. The JMP take this approach in producing regional and global estimates, by calculating the regional population-weighted average for indicators within a set of "master regions" (as defined by the UN Statistical Division's M49 Level 2 classification) [36]. "WatSan Clusters" have also previously been used to impute missing data for analysis of the water and sanitation MDGs [37]. However, as data on safely managed services is currently only available for three countries in Sub-Saharan Africa and no countries in South Asia—the two regions which bear the brunt of the sanitation crisis—neither of these methods would produce imputed estimates that would be sufficiently robust for comparative national analysis. Indeed, the JMP do no publish their imputed country-level figures, in recognition of these limitations. The final approach is to only use existing data for the estimation of the index, but to supplement this with an estimated margin of error based on data gaps in individual countries. The advantage of this method is that it allows scores to be estimated for the maximum number of countries, but it can make direct comparisons more challenging where margins of error overlap.

Given that the underlying aims of the total service gap are to strengthen mutual accountability and to tackle perverse incentives, including the maximum number of countries possible, is an important consideration. As such, the third approach is considered to be the most appropriate method in this context. However, rather than a 'margin of error', we include a 'margin of uncertainty' to account for the gaps in data on safely managed services:

$$\text{Margin of uncertainty} = (B\% \times SMw) - (B\% \times Bw)$$

It is important to note that there is also a large degree of uncertainty inherent within the JMP estimates themselves. This uncertainty stems from sampling issues in the household surveys from which the JMP draws data, and from the linear regression that is then applied to calculate national estimates. These issues have been addressed previously by Bartram et al. [38]. The 'margin of uncertainty' used in this paper does not account for the uncertainty of the estimates produced by the JMP. Rather, it reflects the 'uncertainty' inherent in the JMP's classification of 'at least basic services'. That is, in countries where data on safely managed services are not available, the JMP's estimates only indicate the proportion of the population using a sanitation service that meets the threshold of a basic service; the proportion of the population using a sanitation service which exceeds this threshold remains uncertain. The 'margin of uncertainty' used in this paper reflects this ambiguity, and therefore it only exists for countries with no JMP estimate for safely managed services. It is calculated by assuming that all basic services are safely managed, and therefore indicates the maximum value by which a country's service gap could decrease if all basic services were in fact safely managed.

2.2. Weighting and Sensitivitiy Analysis

To assess the appropriateness of uniform service level weights, a sensitivity analysis was conducted using the four models presented in Table 1. The service level weights for Models B, C, and D were set by the authors to emphasise the differences in the assumptions of each model. These represent simplified versions of models that could be produced through alternative weighting methodologies.

Model A results from assuming uniform incremental benefits as a household moves from one level of service to the next. Model B assumes that the greatest relative benefits are associated with people gaining access to any form of improved facility. Model C assumes that the greatest relative benefits are associated with gaining access to a household sanitation facility, emphasising individual health and safety. Finally, Model D assumes that the greatest relative benefits are associated with achieving the safe management of the whole sanitation chain, with an emphasis on wider public health.

For each model, the total service gap was calculated for every country in the JMP database. The results were then compared to the JMP's measure of 'at least basic services'—in the absence of global data on safely managed sanitation, this measure is increasingly being used to facilitate cross-country comparisons against the SDG sanitation targets [39]. First, the country rankings produced by each model were compared to the rankings that were produced using the 'at least basic services' metric by calculating Spearman's rank correlation coefficients. Second, countries that demonstrated significant sensitivity to changes in service level weight were identified and analysed to understand the drivers of this sensitivity. Third, each model was analysed to determine the extent to which it captures the progress made across the entire service ladder, with the metric of 'at least basic' again used as a common benchmark for comparison. This analysis is used to discuss the utility of uniform service level weights in Section 4.

Table 1. Service level weights for four 'total service gap' models.

Service Level	Model A Uniform Increase in Benefits	Model B Greatest Benefits from Access to Improved Facilities	Model C Greatest Benefits from access to Household Facilities	Model D Greatest Benefits from Moving to Safely Managed Services
Safely Managed	0.00	0.00	0.00	0.00
Basic	0.25	0.20	0.20	0.50
Limited	0.50	0.40	0.70	0.70
Unimproved	0.75	0.90	0.90	0.90
Open Defecation	1.00	1.00	1.00	1.00

3. Results

3.1. Impact on Comparative Rankings

Spearman's rank correlation coefficient (rho) determines the strength of correlation between two sets of ranked data. As such, it can be used to assess, on a macro-level, the extent to which different weighting models impact on the comparative ranking of countries' sanitation progress. The Spearman's rank correlation coefficients presented in Table 2 demonstrate that the country rankings of each model are all very strongly correlated with the country rankings of 'at least basic services'. This shows that the choice of weighting model has very limited impact on countries' relative performance, as judged by comparative rankings.

Table 2. Correlation between the 'total service gap' and 'at least basic services'.

Model	Rho (n = 223 Countries)
Model A	0.985
Model B	0.987
Model C	0.989
Model D	0.982

3.2. Impact on Individual Country Performance

Whilst having a limited impact on the overall rankings of countries, the four models do produce significantly different outcomes for some individual countries. Table 3 shows five countries that are particularly sensitive to changes in service level weights, to demonstrate the impact that different models could have on how performance is perceived at a national level.

Table 3. Total service gap (TSG) and margin of uncertainty (MU) for selected countries using Models A–D.

Country	Model A Uniform Increase in Benefits		Model B Greatest Benefits from Access to Improved Facilities		Model C Greatest Benefits from access to Household Facilities		Model D Greatest Benefits from Moving to Safely Managed Services	
	TSG	MU	TSG	MU	TSG	MU	TSG	MU
Bangladesh	46	−12	46	−9	53	−9	67	−23
Ethiopia	77	−2	84	−1	86	−1	89	−4
Ghana	58	−4	54	−3	71	−3	75	−7
Kazakhstan	26	−24	21	−20	21	−20	50	−49
Papua New Guinea	68	−5	77	−4	78	−4	83	−9

The total service gap for Ethiopia and Papua New Guinea is lowest in Model A. Because these countries have high levels of unimproved services (Ethiopia 59%, PNG, 65%), their total service gap increases in Model B. Conversely, Ghana's total service gap decreases between Model A and Model B, owing to high levels of limited services (57%). In Bangladesh, there is also a high level of unimproved services (33%), but in Model B, this is offset by a high level of limited services (22%) and the statistical absence of open defection. As a result, the total service gap for Bangladesh does not change between Models A and B (although the change in weights does reduce the margin of uncertainty). The total service gap for Kazakhstan decreases slightly in Model B, owing to the greater weight afforded to basic services.

The total service gap for Bangladesh and Ghana increases significantly in Model C, owing to a less favourable weighting for their high levels of limited services. The total service gap for Ethiopia and PNG also increases in Model C, but as limited services are less common, the increase is smaller in magnitude. Kazakhstan has even lower levels of limited services (2%); therefore, its total service gap remains unchanged between Models B and C.

Model D sees an increased service gap for all countries, except those with 100% safely managed services. However, the greatest increases are witnessed in countries where the coverage of basic services is highest—for instance, Kazakhstan (98% basic) and Bangladesh (47% basic).

This highlights that not only can changes to service level weights have significant impact on results for individual countries, but also that this impact is not uniform across countries. How the total service gap metric responds to different weights is highly dependent upon the mix of service levels used in a country.

3.3. Ability to Demonstrate Progress across the Whole Service Ladder

A central hypothesis of this paper is that the metric of 'at least basic services' obscures important progress made on limited services, and that a total service gap would be able to address this problem. However, the four models differ in their ability to showcase examples of progress across the whole sanitation ladder. In Figure 1, each model is plotted against at least basic services. The dotted lines represent a perfect negative correlation between the two metrics. Where a country falls on or near this line, the total service gap portrays a picture of sanitation progress that is very similar to the measure of at least basic. Yet where a country falls below this line, it signifies that the total service gap is able to present a more holistic snapshot of sanitation progress. For instance, Ghana is highlighted in Figure 1. In each of the models, Ghana lies below the dotted line, owing to high levels of limited services. However, the distance to the line decreases significantly in Models C and D. Furthermore, the total number of countries lying below this line also decreases significantly in Models C and D, when compared to Models A and B. This indicates that Models A and B are best able to showcase progress across the whole service ladder.

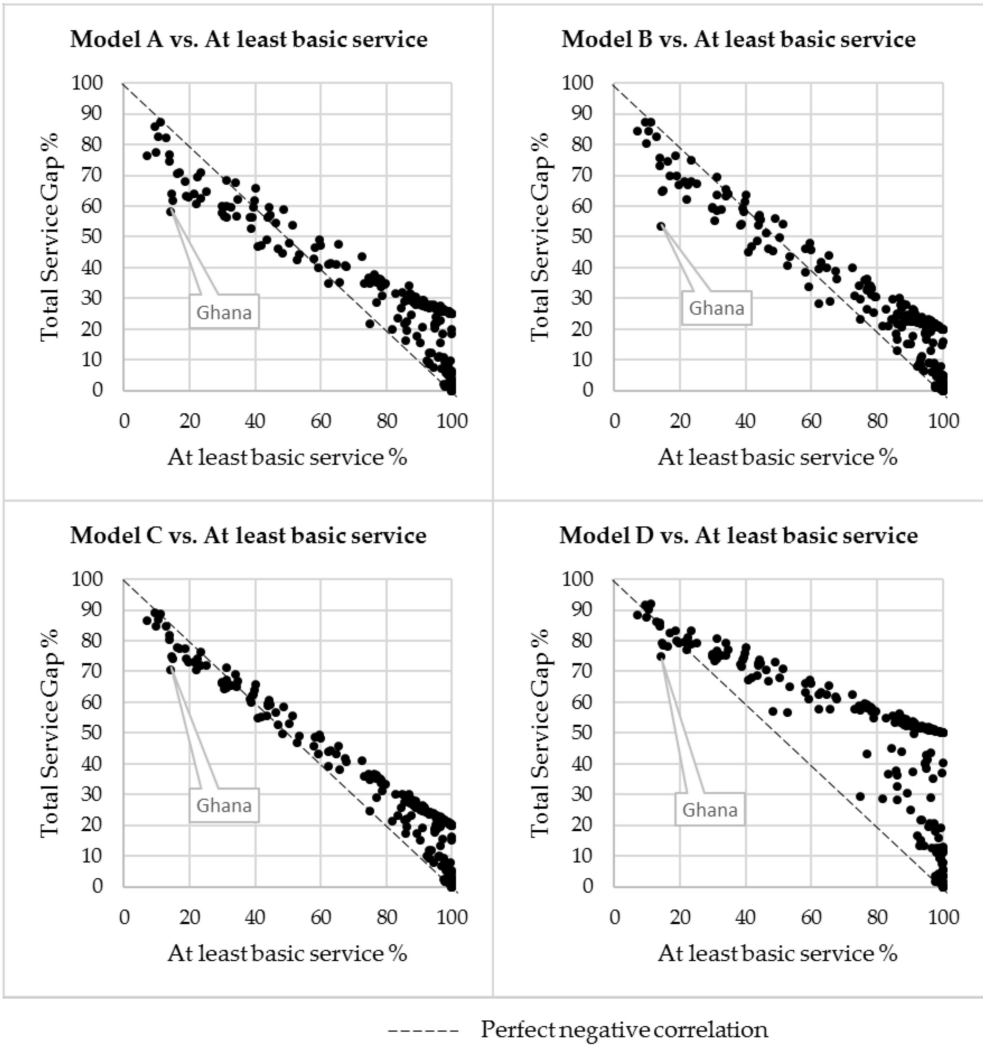

Figure 1. The correlation of the 'total service gap' with 'at least basic services' for Models A–D.

4. Discussion

4.1. Examining the Validity of Uniform Service Level Weights

With any composite measure such as the total service gap, there is a trade-off to be made between simplicity of communication and the empirical rigour of the methodology. The fact that the overall country rankings change very little between different models strengthens the argument for using uniform service level weights, as opposed to developing models based on more complex weighting methods. That is, the difficulties that could be faced in communicating more complex methods would likely outweigh any benefits that would be accrued through greater empirical precision of the model, given the lack of sensitivity to changes in weights at a macro-level.

However, the deeper analysis highlights that while the impact on overall rankings is minimal, the results for certain countries can be effected greatly by changes to the weights. It suggests that the perception of overall sanitation performance is shaped greatly by how the relative benefits of each service level are understood. On the one hand, this weakens the argument for the use of uniform service level weights, and strengthens the case for a more empirically grounded approach to determine the relative benefits of each service level. However, the analysis also highlights that countries' responses to different models are heterogeneous, and highly dependent upon the mix of service levels that are used in a given context. When viewed through the lens of mutual accountability, there is clearly a political dimension to how service level weights are determined. With regard to shared sanitation,

certain models could reward or penalise countries based on structural factors, such as the density of urban settlements. From this perspective, the use of uniform service level weights could offer a 'politically neutral' option to minimise potential contention and to maximise the possibilities for the total service gap to be taken up as a tool for mutual accountability.

A final consideration is the ability of the total service gap to demonstrate progress across the whole sanitation ladder, which was identified as a key weakness of existing global indicators. Figure 1 demonstrates that the total service gap can serve this function, but the extent to which it adds value to the existing metric of 'at least basic services' is determined to a large extent by the weights of the service levels. Notably, the Model A performs well in this respect.

In the absence of an objective and evidence-based method for setting the weights, the use of uniform service level weights represents a valid method to create a total service gap. Its conceptual simplicity will assist with communication, and its 'neutrality' will reduce possible contention around weighting methodologies. Its ability to capture progress across the whole service ladder means that it can add value to the existing measure of 'at least basic services'. The remainder of the paper examines possible applications of this metric, and discusses how it could address the risks identified in the global monitoring framework, with reference to strengthening mutual accountability and challenging perverse incentives.

4.2. Applications of the Total Service Gap: Alternative Visualisations

An initial application of the total service gap could be alternative data visualisations. Figure 2 shows one possibility: the 'sanitation wheel'. The JMP service ladder provides a good snapshot of sanitation services that are used in a country. However, with no judgement on the relative value of each service level, it can be difficult to visualise progress over time. In the 'sanitation wheel', the total service gap is represented by the white space. Progress at each rung of the ladder reduces the service gap, reducing the amount of white space in proportion to the service level weights.

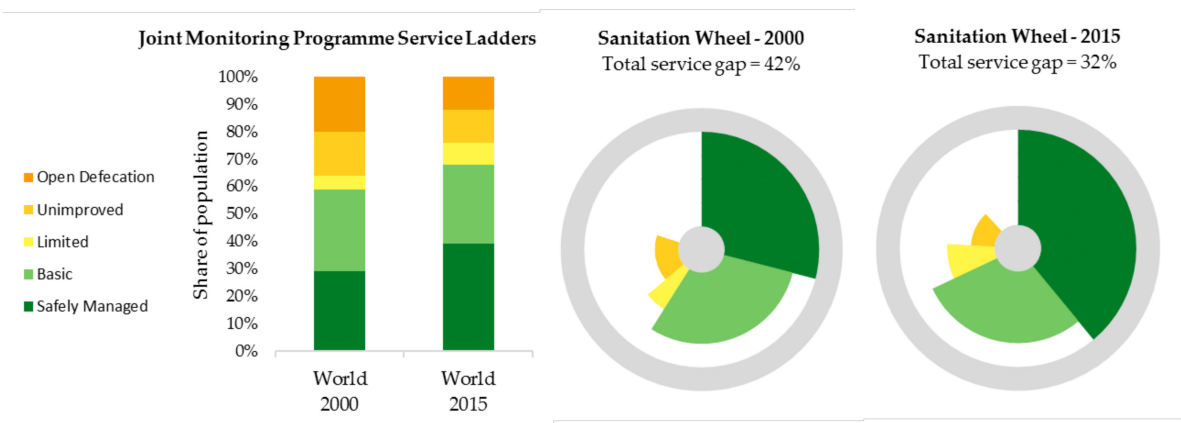

Figure 2. Comparing the JMP service ladder with the 'Sanitation Wheel', using estimates of global sanitation services in 2000 and 2015.

This type of data visualisation could help support stronger mechanisms for mutual accountability by helping to demonstrate progress made over time, while drawing attention to the comparative gap to the goal of universal safely managed services. It can also help to tackle perverse incentives by demonstrating the value of progress at each rung of the ladder. In this way, the 'sanitation wheel' could be a useful device to communicate the concept of 'progressive realisation' in ways that promote pro-poor policy choices.

Comparing progress across countries is a central facet of mutual accountability mechanisms. However, the majority of countries are not yet able to produce data on safely managed services. As a result, the threshold of 'at least basic services' has been used by the JMP [1] and others [39] to compare progress. This allows for the comparability of data between countries, but risks diluting the ambition of the safely managed target, and does not capture progress made on shared sanitation, which is important in many contexts. Using a total service gap to compare progress helps to overcome the issues with 'at least basic'. However, as with any composite measure, it can obscure a number of differences when comparing across countries. For example, Yemen and Malawi have an identical total service gap of 49%. Yet the nature of sanitation progress has differed significantly in each country: Yemen has made greater progress on basic services (59%, compared to 43% in Malawi), while Malawi has made greater progress in eliminating open defection (6%, compared to 20% in Yemen). A further challenge is the lack of data on safely managed services, which limits the potential for some cross-country comparisons in the short-term where margins of uncertainty overlap. However, the inclusion of a 'margin of uncertainty' could itself be a useful device for advocacy and communications. For instance, the margins of uncertainty shown in Figure 3 demonstrate that for countries such as Chad, with a high service gap (86%) and a low margin of uncertainty (−2%), upgrading existing basic services to safely managed will have a limited impact on the overall performance. In these context, the greater focus should be on extending basic, and where necessary, limited services. However, in countries such as Myanmar where the total service gap is smaller (41%) and the margin of uncertainty is larger (−16%), the benefits of investing in safely managed services—or in collecting additional data to demonstrate that existing services are safely managed—becomes more apparent. In this way, the total service gap can help to visualise advocacy messages around a pro-poor progressive realisation of the human right to sanitation.

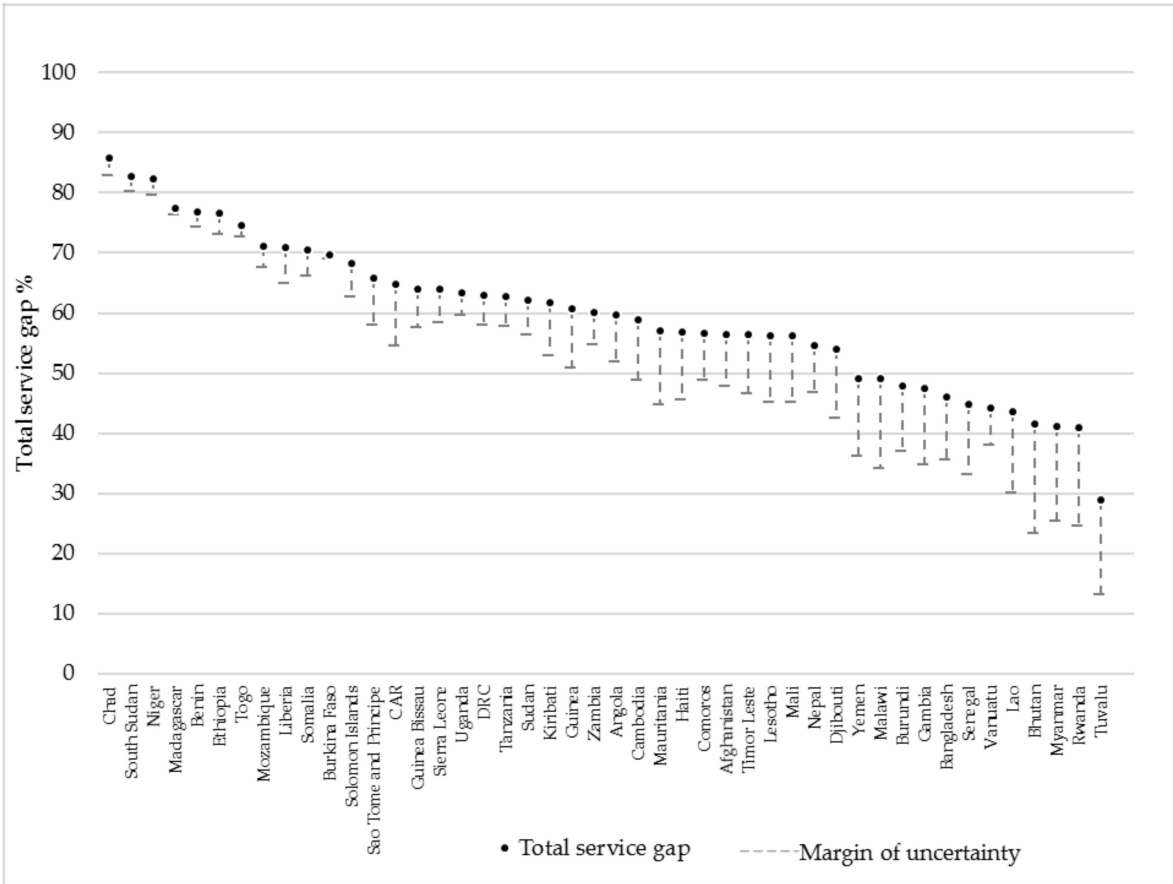

Figure 3. The total service gap for Least Developed Countries in 2015.

4.3. Applications of the Total Service Gap: Sub-National Inequalities

In addition to cross-country comparisons, the total service gap could also be utilised to highlight inequalities at a sub-national level. While the application at a sub-national level is limited by a lack of data on safely managed services, the following examples from Senegal demonstrate how a total service gap could add value. Figure 4 shows the correlation between the total service gap and 'at least basic services' for the sub-national regions of Senegal.

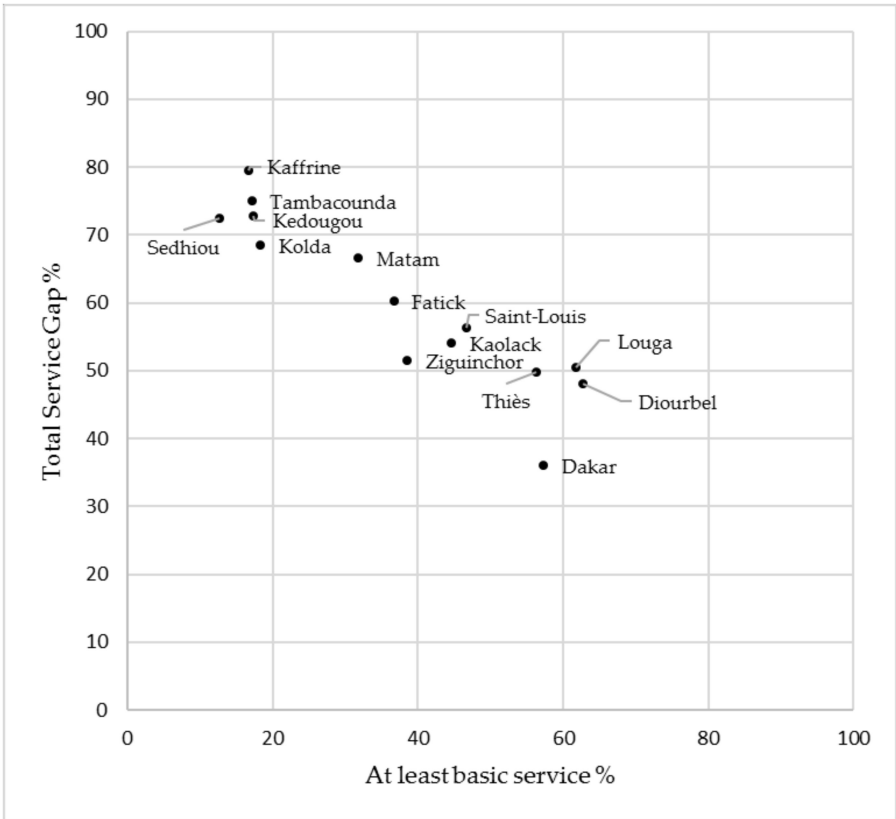

Figure 4. Comparison of the total service gap for sub-national regions of Senegal, 2015.

Figure 4 shows that while Kaffrine, Tambacounda, Kedouhou, and Kolda each have a very similar level of basic services (17%, 17%, 17%, and 18% respectively), their total service gap differs by over 10 percentage points (69% in Kolda and 79% in Kaffrine). Understanding and communicating this difference through a total service gap—as demonstrated in Figure 5—could have benefits for planning and budgeting processes. Similar approaches could also be utilized to better monitor the outcomes of sanitation projects delivered by both governments and development partners.

Figure 6 shows a comparison between the total service gap and 'at least basic services' for each wealth quintile in Senegal (the y-axis for the total service gap has been inverted to allow for comparison). It highlights that the measure of 'at least basic' could potentially downplay the progress made in the lower wealth quintiles. This would be especially true in cases where this progress has been in the form of shared sanitation. The total service gap metric could help to highlight such positive examples of reductions in inequality, which is a key function of global accountability mechanisms.

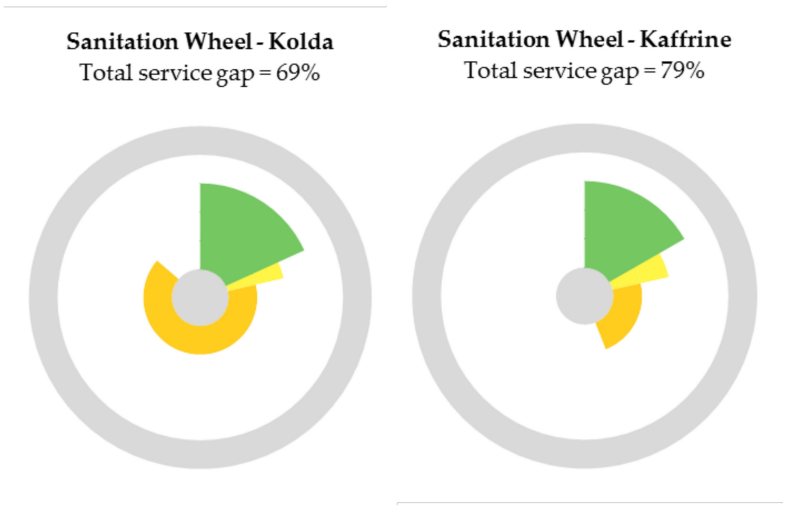

Figure 5. Sanitation wheels for Kolda and Kaffrine, 2015.

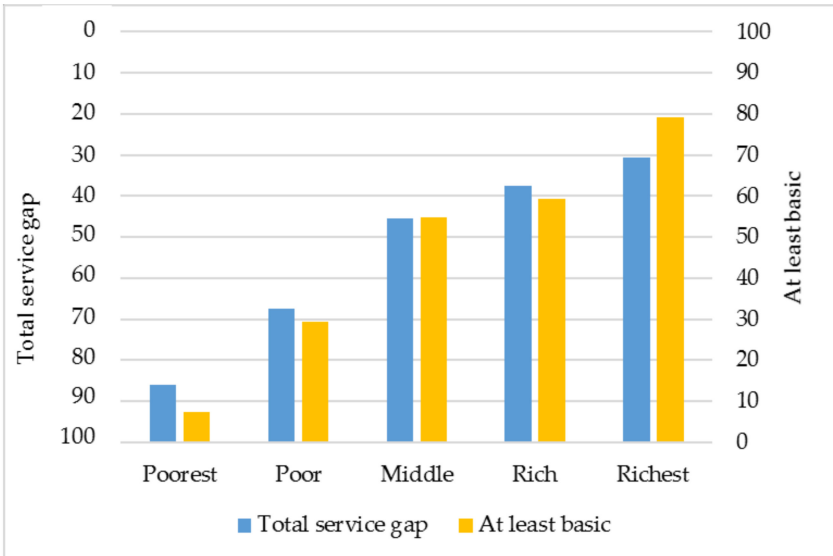

Figure 6. Comparison of the 'total service gap' with 'at least basic services' across wealth quintiles in Senegal, 2015.

4.4. Future Applications: An 'Inequality Adjusted Service Gap'

Building on the wealth quintile analysis, it could also be possible to go a step further and produce an 'inequality-adjusted service gap'. This measure would adjust the value of the total service gap according to the level of equity in services across wealth quintiles. Following a similar methodology as the 'Inequality-adjusted Human Development Index' [40], the total service gaps for each wealth quintile would be combined to quantify the level of inequality using the Atkinson measure of inequality, as defined by the formula:

$$A = 1 - g/\mu$$

where g is the geometric mean and μ is the arithmetic mean of the wealth quintile distribution. Following this, the inequality-adjusted service gap (I-SG) is calculated by combining the level of inequality (A) with the national-level total service gap (TSG):

$$\text{I-SG} = (1 + A) \times \text{TSG}$$

Where there is no difference in levels of access to sanitation across wealth quintiles, the inequality-adjusted service gap would exactly equal the total service gap. However, the inequality-adjusted service gap would rise above the total service gap as the level of inequality increases. In this way, countries are penalised for higher levels of inequality. The countries in Table 4 provide an example of the possible utility of an inequality-adjusted service gap.

Table 4. Composition of the inequality-adjusted service gap.

	Total Service Gap by Wealth Quintile %					Total Service Gap %	Inequality-Adjusted Service Gap %
	Q1	Q2	Q3	Q4	Q5		
Angola	92	82	66	34	26	60	66
DRC	70	66	64	62	52	63	63
Mauritania	94	87	68	47	35	57	61
Uganda	79	71	67	60	47	63	64

The inequality-adjusted service gap is not a measure of inequality in itself, but rather a measure of progress that penalises inequality. For instance, Figure 7 demonstrates that while Angola has a lower total service gap than Uganda and the Democratic Republic of Congo, when adjusted for inequality its service gap becomes comparatively larger. This form of analysis could greatly strengthen mutual accountability for the SDG's 'leave no one behind' agenda, creating incentives not only for increasing the coverage of services, but also ensuring their equitable delivery across wealth quintiles.

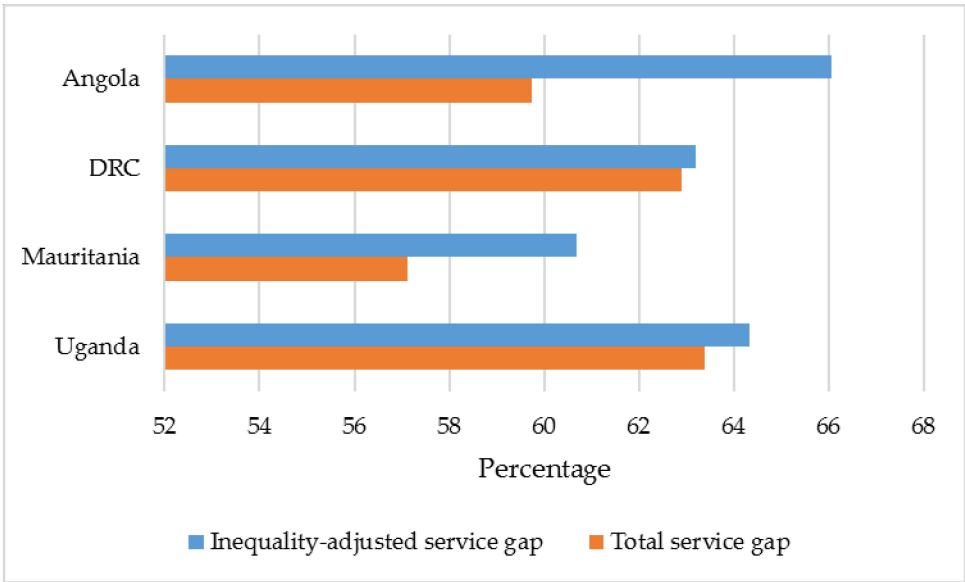

Figure 7. Comparison of the 'total service gap' and the 'inequality-adjusted service gap' for Uganda, Mauritania, Democratic Republic of Congo, and Angola.

5. Conclusions

In the absence of objective criteria for service level weighting, this paper proposes the use of uniform service level weights as the best method to calculate the total service gap. The paper finds that the different weighting models can result in significant changes to the results of specific countries, but they do not lead to significant changes in countries' comparative rankings. The way in which countries respond to different weights is not homogeneous, but is determined by the specific mix of service levels that they have. Viewed from the perspective of strengthening mutual accountability, these results support the choice of uniform service level weights—as a conceptually simple and politically neutral method—in the calculation of the total service gap.

The total service gap makes it possible to develop alternative data visualisations that better communicate concepts such as 'progressive realisation' and 'leave no one behind', providing

information for both practitioners and policy makers. Through such communications devices, the total service gap has the potential to support mutual accountability and tackle perverse incentives through the creation of a broader policy paradigm. Furthermore, the development of the inequality-adjusted service gap can bring greater policy attention to the equity of sanitation services.

Data on individual service levels remains important, and even more granular data are required for planning and budgeting at a national level. At a global level, more and better qualitative analysis is required to facilitate peer learning and course correction, and there remains the need for a concerted effort to fill data gaps on safely managed services, and to define national indicators for 'high-quality shared sanitation' where necessary. However, the total service gap is a metric which could add value to these measures, both at a global and country level.

Author Contributions: Conceptualization, S.K. and A.H; methodology, S.K and A.H; data curation and analysis, S.K.; writing—original draft preparation, S.K.; writing—reviewing and editing, S.K. and A.H.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. WHO/UNICEF. *Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines*; WHO/UNICEF: Geneva, Switzerland, 2017.
2. Northover, H.; Ryu, S.K.; Brewer, T. *Achieving Total Sanitation and Hygiene Coverage within a Generation—Lessons from East Asia*; WaterAid: London, UK, 2016.
3. Mason, N.; Matoso, M.; Hueso, A. *Beyond Political Commitment to Sanitation: Navigating Incentives for Prioritisation and Course Correction in Ethiopia, India and Indonesia*; WaterAid: London, UK, 2016.
4. Sethi, T.; Custer, S.; Turner, J.; Dilorenzo, M.; Latourell, R. *Realizing Agenda 2030: Will Donor Dollars and Country Priorities Align with Global Goals? A Baseline Report*; AIDDATA: Williamsburg, VA, USA, 2017.
5. Sarwar, M.B. *National MDG Implementation: Lessons for the SDG Era*; ODI: London, UK, 2015.
6. Melamed, C. *After 2015 Global Agreement on Development*; ODI: London, UK, 2015.
7. Fukuda-Parr, S.; Yamin, A.E.; Greenstein, J. The Power of Numbers: A Critical Review of Millennium Development Goal Targets for Human Development and Human Rights. *J. Hum. Dev. Capability* **2014**, *15*, 105–117. [CrossRef]
8. Merry, S.E. Measuring the World. *Curr. Anthropol.* **2011**, *52*, S83–S95. [CrossRef]
9. Broome, A.; Homolar, A.; Kranke, M. Bad science: International organizations and the indirect power of global benchmarking. *Eur. J. Int. Relations* **2018**, *24*, 514–539. [CrossRef] [PubMed]
10. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*; United Nations: New York, NY, USA, 2015.
11. United Nations General Assembly. Critical Milestones towards Coherant, Efficient and Inclusive Follow-Up and Review at the Global Level. Report of the Secretary-General; 2016; pp. 1–27.
12. Elgin-Cossart, M.; Chandran, R. *Designing Better Accountability Mechanisms for the 2030 Agenda for Sustainable Development*; United Nations University: Washington, DC, USA, 2016.
13. Ocampo, A. *A Post-2015 Monitoring and Accountability Framework*; Economic Analysis & Policy Division: New York, NY, USA, 2015.
14. Engebretsen, E.; Heggen, K.; Ottersen, O.P. The Sustainable Development Goals: Ambiguities of accountability. *Lancet* **2017**, *389*, 365. [CrossRef]
15. Espey, J.; Walecik, K.; Kuhner, M. *Follow-Up and Review of the SDGs*; SDSN: New York, NY, USA, 2015.
16. Development Assistance Committee. *The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action*; OECD: Pairs, France, 2008. [CrossRef]
17. Fourth High Level Forum on Aid Effectiveness. *Busan Partnership for Effective Development Co-Operation. Proceedings of the Fourth High Level Forum on Aid Effectiveness*; OECD: Busan, Korea, 2011.
18. Together 2030. Voluntary National Reviews: What Are Countries Prioritizing? Available online: https://www.together2030.org/wp-content/uploads/2017/07/FINAL-Together-2030_VNR-Main-Messages-Review-2017.pdf (accessed on 10th September 2018).

19. Sarwar, M.B.; Nicolai, S. *What Do Analyses of Voluntary National Reviews for Sustainable Development Goals Tell Us about "Leave No One Behind"?* ODI: London, UK, 2018.
20. Cross, P. Small steps towards building national-regional-global coherence in monitoring WASH. In *From Infrastructure to Services*; Practical Action Publishing Ltd.: Warwickshire, UK, 2015; pp. 137–155.
21. Evans, B.; Hueso, A.; Johnston, R.; Norman, G.; Pérez, E.; Slaymaker, T.; Trémolet, S. Limited services? The role of shared sanitation in the 2030 Agenda for Sustainable Development. *J. Water Sanit. Hyg. Dev.* **2017**, *7*, 349–351. [CrossRef]
22. Kvarnström, E.; McConville, J.; Bracken, P.; Johansson, M.; Fogde, M. The sanitation ladder—A need for a revamp? *J. Water Sanit. Hyg. Dev.* **2011**, *1*, 3–12. [CrossRef]
23. Cardone, R.; Gambrill, M. How Many People Can Share a Toilet? Available online: https://blogs.worldbank.org/water/how-many-people-can-share-toilet (accessed on 17 September 2018).
24. Cardone, R.; Schrecongost, A.; Gilsdorf, R. *Shared and Public Toilets: Championing Delivery Models That Work*; The World Bank: Washington, DC, USA, 2018.
25. Jiménez, A.; Livsey, J.; Åhlén, I.; Scharp, C.; Takane, M. Global Assessment of Accountability in Water and Sanitation Services Using GLAAS Data. *Water Altern.* **2018**, *11*, 238–259.
26. Van de Lande, L.; Fonseca, C. *Global Review of National Accountability Mechanism for SDG6*; End Water Poverty: London, UK, 2018.
27. Giné-Garriga, R.; Pérez-Foguet, A.; Jiménez Fdez de Palencia, A. Monitoring sanitation and hygiene in the 2030 Agenda for Sustainable Development: A review through the lens of human rights. *Sci. Total Environ.* **2016**, *580*, 1108–1119. [CrossRef] [PubMed]
28. Giné-Garriga, R.; Pérez-Foguet, A. *Measuring Sanitation Poverty: A Multidimensional Measure to Assess Delivery of Sanitation and Hygiene Services at the Household Level*; OPHI Working Papers; OPHI: Oxford, UK, 2018.
29. Potter, A.; Snehalatha, M.; Batchelor, C.; Uandela, A.; Naafs, A.; Fonseca, C.; Moriarty, P. *Assessing Sanitation Service Levels*; IRC: The Hague, The Netherlands, 2011.
30. Wolf, J.; Prüss-Ustün, A.; Cumming, O.; Bartram, J.; Bonjour, S.; Cairncross, S.; Clasen, T.; Colford, J.M.; Curtis, V.; De France, J.; et al. Systematic review: Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: Systematic review and meta-regression. *Trop. Med. Int. Heal.* **2014**, *19*, 928–942. [CrossRef] [PubMed]
31. Heijnen, M.; Cumming, O.; Peletz, R.; Chan, G.K.-S.; Brown, J.; Baker, K.; Clasen, T. Shared Sanitation versus Individual Household Latrines: A Systematic Review of Health Outcomes. *PLoS ONE* **2014**, *9*, e93300. [CrossRef] [PubMed]
32. Heijnen, M.; Eisenberg, J.N.S.; Clasen, T.; Fuller, J.A. Shared Sanitation and the Prevalence of Diarrhea in Young Children: Evidence from 51 Countries, 2001–2011. *Am. J. Trop. Med. Hyg.* **2014**, *91*, 173–180. [CrossRef]
33. Andrés, L.; Briceño, B.; Chase, C.; Echenique, J.A. Sanitation and externalities: Evidence from early childhood health in rural India. *J. Water Sanit. Hyg. Dev.* **2017**, *7*, 272–289. [CrossRef]
34. Pearson, J.; Mcphedran, K. A literature review of the non-health impacts of sanitation. *Waterlines* **2008**, *27*, 48–61. [CrossRef]
35. Foa, R.; Tanner, J.C. *Methodology of the Indices of Social Development*; ISS Working Paper No. 2012-4; International Institute of Social Studies: The Hague, The Netherlands, 2012.
36. WHO/UNICEF. *JMP Methodology: 2017 Update & SDG Baselines*; WHO/UNICEF: Geneva, Switzerland, 2018.
37. Onda, K.; Crocker, J.; Kayser, G.L.; Bartram, J. Country clustering applied to the water and sanitation sector: A new tool with potential applications in research and policy. *Int. J. Hyg. Environ. Health* **2014**, *217*, 379–385. [CrossRef] [PubMed]
38. Bartram, J.; Brocklehurst, C.; Fisher, M.B.; Luyendijk, R.; Hossain, R.; Wardlaw, T.; Gordon, B. Global monitoring of water supply and sanitation: History, methods and future challenges. *Int. J. Environ. Res. Public Health* **2014**, *11*, 8137–8165. [CrossRef] [PubMed]
39. WaterAid. *Out of Order: The State of World's Toilets 2017*; WaterAid: London, UK, 2017.
40. United Nations Development Programme. *Human Development Indices and Indicators: Technical Notes*; United Nations Development Programme: New York, NY, USA, 2018.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).