Second Annual UHC Financing Forum
Greater Efficiency for Better Health and Financial Protection

Background paper (Forum edition)

*Washington, D.C.*
*April 20-21, 2017*
This is a background paper to the “Second Annual UHC Financing Forum: Greater Efficiency for Better Health and Financial Protection”. This paper sets the stage for the presentations and discussions at the Forum and was prepared under the guidance of the Forum Technical Working Group. The information provided in this document does not necessarily represent the views or position of the organizations represented on the Technical Working Group.
Contents
I. Executive Summary ........................................................................................................3
II. Introduction ..................................................................................................................6
III. What is efficiency? .......................................................................................................7
IV. Sources of inefficiency .................................................................................................9
  1) Doing the right things ...............................................................................................11
     i. Mix of health interventions and packages .........................................................11
     ii. Prevention versus treatment ............................................................................12
     iii. The balance between governance, administration, public health functions and personal services .................................................................13
     iv. Inter-sectoral and multi-sectoral action .............................................................13
     v. Financial protection and service coverage .........................................................14
  2) Doing the right things in the right place .................................................................15
  3) Doing things right ......................................................................................................17
     i. Health Services ...................................................................................................18
     ii. Health Systems and the Health Financing Component ....................................22
V. Identifying the most important sources of inefficiency ..............................................24
VI. Measuring and monitoring inefficiency ....................................................................25
  1) Macro-efficiency .......................................................................................................25
  2) Efficiency in components of the health system .....................................................27
VII. Strategies to improve efficiency .................................................................................32
  1) Technical options for improving efficiency ...........................................................33
  2) What we do not know ...............................................................................................43
     i. Paying for results .................................................................................................43
     ii. The private sector and efficiency ......................................................................46
     iii. Human resource strategies for efficiency .........................................................46
     iv. Costs of improving efficiency ...........................................................................47
     v. Political economy issues .....................................................................................47
VIII. Efficiency and equity ................................................................................................48
IX. Conclusions ................................................................................................................49
X. Recommendations .....................................................................................................53
XI. References ................................................................................................................55
I. Executive Summary
The nature and extent of inefficiency in health

The concept of efficiency in health includes doing the right things, in the right settings, and in the right way. Efficient health systems produce the best possible health and financial protection outcomes from the available resources.

Inefficiencies exist in health systems everywhere:

- Doing the wrong things (e.g. funding high cost, low impact interventions but not fully funding low cost, high impact interventions, particularly in tertiary care settings.)
- Doing the right things in the wrong settings (e.g. relying on hospitals rather than primary health care.
- Doing things badly (e.g. leakages and waste in terms of pilfering of medicines through the supply chain and medicines left to expire or stored in poor conditions.)

The magnitude of inefficiencies can be staggering: taking all forms of inefficiencies into account, countries can waste an estimated 20% - 40% of their health resources, missing the opportunities to use resources more efficiently to achieve much more.

Sources of inefficiency

The various sources of inefficiency in health have been well established. Many are linked to the main cost-drivers in health systems: medicines, human resources for health, and health facilities and infrastructure, particularly hospitals.

The nature of inefficiencies, however, differ by setting. Countries might be relatively efficient in one area and less efficient in another. For each country to understand how to reduce their inefficiencies, they must assess its most important causes (using the type of checklist provided in this document), then decide which are most feasible to address technically and politically.

Countries then need to develop strategies, a time-table for change, and a strategy to monitor progress.

An important implication is that countries will need to choose indicators to track their progress that relate to the areas of inefficiency they are about to tackle.

Improving efficiency

Many options for improving efficiency exist, and all countries can take actions to achieve better health and financial protection using their available resources.

Some of the options lie in the health financing system. They include raising revenue more effectively, using the tax system to reduce consumption of products that harm health, reducing fragmentation in pooling, ensuring that pooled funds purchase the interventions that deliver the greatest impact for the money, and modifying provider payment mechanisms to encourage both efficiency and quality.

Other solutions require actions in the wider health system. For example, improving medicines-related efficiency certainly involves the ability to buy at the lowest cost, but may also include the capacity to test and ensure quality throughout the distribution chain, to modify regulations or legislation to encourage the use of generics, and demand-side strategies to overcome concerns in prescribers and patients about the quality of generics, and to encourage rational use of all medicines.
The politics of implementing policies to improve efficiency can be complex because changes may face opposition by powerful stakeholders.

Some of the possible solutions are more likely to produce rapid returns than others. These include, but are not limited to:

- taxation to reduce harmful product consumption;
- improving budget flexibility to improve timely release and expenditure of available funds;
- introduction of generic medicines policies, and transparent, competitive bidding (where possible) for purchases.
- allow health workers at lower levels to take on more responsibility as appropriate (task shifting)

Others may need short-term investments but require longer-term commitment before the results start to be seen. For example, moving towards active purchasing (purchasing based on an explicit assessment of needs, prices and value) requires staff skills and computerized information systems that might require time to develop.

To accompany these actions, international collaboration is needed to continually search for technologies that “shift the frontier”, identifying further opportunities to improve health and financial protection at low cost. This may involve translating existing technologies to low-cost settings, or developing new approaches, such as vaccines for Hepatitis C and HIV/AIDS.

**Controversies and missing information**

Despite knowing a lot about the nature of inefficiency and possible technical solutions, there is still a surprising amount that is not known, or that engenders substantial disagreement.

Remarkably little is known about what works at the systemic level. How can hospital efficiency be improved? What is the appropriate role of the private sector in improving efficiency? What types of incentives keep staff motivated and ensure quality, but are affordable? The evidence is simply not yet conclusive, which means these questions cannot yet be answered definitively.

Part of the problem is that many studies of specific interventions, for example, results-based financing, have focused on service utilization and quality, with little examination of costs. This research trend makes it very difficult to determine whether these interventions are a good use of scarce resources even if they work, and if they can be financed in the long term.

In other cases, the methods for undertaking the necessary analytical work are not particularly useful. For example, health technology assessment based on cost-effectiveness analysis is frequently used to help guide decisions about the right intervention mix. It is appropriate for high-income countries where the question is whether small changes in expenditure, on top of well-established existing packages of services, are warranted. This type of HTA is not particularly useful in many lower income settings because the changes needed are not marginal. Techniques for assessing what mix of interventions should be available at primary level, for example, have the capacity to take into account the fact that costs and impact vary with factors such as scale and scope of interventions, and whether the staff needed to deliver cost-effective interventions are available. HTA is also of limited use in assessing the appropriate mix between personal and population based health services, and between governance and public health functions and population-based and personal health interventions. These areas need further development.
A final limitation to countries wanting to take on the efficiency agenda is that the data needed to formally assess the major causes of inefficiency and to monitor progress is extremely limited. This is partly because few countries currently undertake regular assessments of their efficiency, so the indicators that are available are mostly collected for other purposes.

**Recommendations**

**Countries**

- Undertake an assessment of the major causes of inefficiency and those that are feasible to change in the short, medium and longer-term.
- Develop and implement a strategy for improving efficiency in the short to medium term – this should be part of a health financing strategy although some of the actions will need to extend beyond health financing.
- Start to put in place the background investments to ensure the longer-term options can be undertaken – e.g. legislation, consultation, computerized information systems, staff skills.
- Undertake both political and technical analysis to identify which reforms have the greatest chance of success, then build support and negate opposition.
- Develop a set of efficiency indicators specific to the country’s main causes of health inefficiencies, and develop an agenda for achieving more for the available resources.
- Invest in methods to collect indicator data and to evaluate progress regularly.
- Identify areas of possible inter-sectoral or multi-sectoral actions that would achieve the largest health impacts, and the political feasibility of influencing other sectors to implement them (perhaps in collaboration with the Ministry of Health). This would help the Ministry of Health target the key ministries and make the best use of their own limited time and resources.

**International community (including researchers in all countries)**

- Routinely assess the costs as well as impact of efforts to improve efficiency so that countries can determine the efficiency and financial sustainability of different options for improving efficiency.
- Develop an agenda to identify the cost-effectiveness of efforts to redress health inequalities as part of the efficiency and equity discussion.
- Develop methods which can be used to help countries determine which of the myriad of inter-sectoral or multi-sectoral actions to improve health should be given priority with the limited time and financial resources available to a Ministry of Health.
- Continue to invest in the technologies that might “shift the frontier” of possibilities”, identifying further opportunities to improve health and financial protection at low cost, such as vaccines for Hepatitis C and HIV/AIDS.
II. Introduction

The series of Annual Forums on Financing Universal Health Coverage (UHC) focus on health financing actions to smooth the path towards UHC – defined formally as the ambition that all people will obtain the health services they need, of good quality, and without financial hardship linked to paying for them out-of-pocket (WHO 2010a). The first Forum in 2016 dealt with revenue generation, or how to raise funds to meet the needs and demands of the population for good quality health services and financial protection, which are key components of UHC. This is important to countries at all income levels, although the absolute need for additional revenue is particularly high in poorer countries. In some countries, the shortage of funds is so severe that meaningful progress towards UHC will not happen without substantial increases in revenue generation. This is why revenue generation was the topic of the first Forum.

The second Annual Forum on Financing for UHC turns to the question of how to use the available resources in the most efficient way. The extent of inefficiency and waste in health can be staggering, and even if the estimates are relatively imprecise, they give an idea of the order of magnitude of the problem. A commonly quoted figure is that between 20% and 40% of all health resources might be effectively lost to various forms of inefficiency (WHO 2010; Chisholm & Evans 2010). Along similar lines, the OECD recently reported the results of a number of studies on waste (one component of inefficiency), and showed in selected OECD countries, somewhere between 20% and 50% of health expenditures are likely being wasted due to inefficiencies (OECD 2017).

Improved efficiency (achieving more with the available resources) enables countries to obtain greater coverage, and to deliver quality health services and financial protection for the same expenditures. It can also improve health outcomes. For example, a recent IMF working paper suggests that African countries could raise life expectancy at birth by about five years on average if they used their health resources more efficiently (Grigoli & Kapsoli 2013).

Improved efficiency can sometimes also save money or reduce the rate of increase of health expenditures – this has been called “bending the curve” that depicts the relationship between health expenditures and GDP over time (Coady, Francese and Shang 2014; OECD 2017). This, however, does not always happen. Many efficiency reforms require upfront investments before they start to show the improvements in health and/or financial protection that a reorganization of what is done, or the way things are done aspires to achieve.

Recognizing that the health sector must compete with other sectors in the allocation of public finances, Ministries of Finance have sometimes been reluctant to increase allocations to Ministries of Health which are perceived as either not fully spending the funds they already have, or not using them efficiently (Gillingham 2014; Tandon et al. 2014). Investments in other sectors are sometimes considered to offer greater value for money. Therefore, improving efficiency in health can also help to convince Ministries of Finance to allocate more public funds to health: by achieving more with existing resources, additional resources may become available.

Improving efficiency and health financing strategies is just a part of progress towards UHC. Within the health sector, further requirements include: sufficient motivated health workers of the right type located close to people; good quality infrastructure, appropriately located; a focus on health service quality; sufficient essential medicines and other health products; high quality leadership and governance; and timely, accurate information. Progress can also be facilitated by promoting key inter-sectoral actions: both actions in other sectors that improve health, and actions in the health sector that improve income, education and development more generally – which in
turn feedback into health improvements. While recognizing this, the focus on this Forum is on health financing and achieving more with the available resources.

The main objective of the paper is to highlight what is known and what is not known about the main causes of inefficiency in health and what can be done to reduce them. The agenda of the Forum was developed in response to this analysis, focusing on areas where either knowledge is still inadequate for informed decision making, or where there is controversy or disagreement among experts.

The next section of this paper (Section 2) begins by defining the economic concept of efficiency and how it has been used in health. Section 3 then identifies the common sources of inefficiency in health, an important starting point for countries seeking to achieve more with the available resources. Section 4 considers indicators that can be used to identify which forms of inefficiency are the most important in each setting before developing policies to redress them, and then to track progress in reducing inefficiency.

Section 5 turns to identifying recognised technical solutions plus areas where there are still important gaps in knowledge. It also considers what is known about “how” to ensure the desired strategies are implemented taking into account the political economy of efficiency reforms.

Making a health system more efficient does not necessarily mean that it will become more equitable. For example, expanding coverage by first targeting people in isolated areas rather than those in more densely populated settings might not be the most efficient option in terms of improving population health levels or providing financial protection, yet a country may consider this approach to be desirable on equity grounds. The bulk of this paper focuses purely on efficiency but turns to the question of equity and possible trade-offs with efficiency in Section 6.

The final section (Section 7) summarizes the main findings of the paper, highlighting areas where there is insufficient evidence to guide policy, where there is controversy, and where there are possible quick wins in terms of improving efficiency.

III. What is efficiency?

Formal definitions of efficiency from economics and health economics are summarized in Box 1. In essence, a health system that is efficient produces the mix of health services that maximizes the outcomes society expects from its health system, usually population health improvements, uses the mix of inputs that costs the least, and combines these inputs to produce the maximum possible outputs. It would not be possible to get more health for the same level of expenditure by either changing the mix of inputs, getting more out of the chosen mix of inputs, or producing a different set of health goods and services.

Box 1: Formal Definitions of Efficiency

General economics defines three types of efficiency. Technical efficiency is achieved when a particular set of inputs achieves the maximum possible output(s). Technical efficiency could be achieved with a very expensive set of inputs, so productive or production efficiency is when the inputs used to produce this output have the least cost, while allocative efficiency requires the production of the set of outputs that people value the most for the given resources (e.g. see Hollingsworth 2008).
Health economics has typically defined only two sorts of efficiency - allocative and technical. Allocative efficiency has been defined as requiring that the health goods and services produced are those that maximize society’s objectives for the health sector, usually translated into some measure of population health status or improvement.\(^1\) Technical efficiency is taken to mean achieving the most with the available inputs. The question of what mix of inputs in the production process is the least cost mix (production efficiency in general economics) is frequently subsumed, sometimes implicitly, into technical efficiency and sometimes into allocative efficiency (Yip & Hafez 2015; see Cylus, Papanicolas & Smith 2013 for a useful discussion of the various types of inefficiency and how they have been used in health).

Rather than use the formal definitions of efficiency as an organizer for the discussion, we follow a framework suggested by Yip & Hafez (2015) which focuses on the key policy questions facing countries seeking to improve the efficiency of their health systems:

1. Doing the right things (allocative efficiency: what mix of interventions maximize health outcomes for the available resources?);
2. Doing them right (a combination of technical and productive efficiency: are the mix of inputs the lowest cost mix, and do they achieve the maximum possible outputs?).

We add an additional category – doing the right things “in the right places”. Decisions on which care setting services should be delivered have a significant impact on the ability of health systems to improve or maintain health and financial protection. Common trade-offs are between: the different levels of care (e.g. community, primary, secondary, tertiary); day versus inpatient care; long-term institutional care versus home care; and social care versus medical care. Adequate coordination and continuity of care, both within and across levels, are critical elements to ensuring that services are delivered in the right care settings.

Although high-income country governments often urge line ministries to improve efficiency in the face of budget cuts, or to restrain expenditure growth, the main objective of improving efficiency in low and lower-middle income countries is \textbf{not} to reduce overall spending in health or cut budgets. It is to make better use of available resources to achieve faster progress towards UHC, better health and greater financial protection. This is why the paper frequently uses the term “achieving more with the available resources” to describe the term “efficiency”.

This has two implications for the way efficiency is considered in the context of UHC. First, the traditional way of considering efficiency focuses on health outcomes for patients or a population. In this context, increasing coverage with health-sector interventions is only one pathway to improving health outcomes. Multi-sectoral or inter-sectoral approaches are also important and need to be considered in any assessment of the most efficient way of improving population health. They are discussed in Section 3.

Second, the concept of UHC acknowledges that people value not only the health improvements that result from appropriate use of health services but also the assurance that using health services will not result in severe financial losses. This “value of insurance” - where insurance is interpreted in the broad sense of a reduction in the risk of people needing to find the funds for high, unexpected health costs in the future through some form of prepayment and pooling - has been well accepted in economics as having two components. The first is the welfare benefit associated with people knowing they will not suffer unexpected financial losses in the event of illness, and the

\(^1\) Sometimes the assumed maxim is population welfare (see Palmer & Torgerson 1999) or the value of the health improvement (an interpretation of QALYs).
second is the welfare benefit of knowing that they will be able to afford to use health services that would have been unaffordable in the absence of insurance (see Nyman 1999).²

To date, little attention has been paid to understanding this broader concept of efficiency in the context of a society seeking to improve both aspects of UHC at the same time – coverage with quality health services of all types, and coverage with financial protection. A recent exception is the development of an extended cost-effectiveness analysis, which seeks to understand the implications of various choices of intervention mixes, not only on health outcomes but also on impoverishment, one consequence of a lack of financial protection (see Verguet et al. 2013; Verguet et al. 2014; Verguet et al. 2015; Shrime et al. 2015). This is also discussed further in Section 3.

### IV. Sources of inefficiency

Countries cannot make informed choices about which inefficiencies they can tackle without identifying the most important causes in a given setting; the pragmatic way of achieving this is to start with one of the checklists, which are based on an analysis of the literature on country experiences. Table 1 provides a checklist based on an approach taken by WHO in 2010 (WHO 2010a). This checklist has been modified to incorporate more recent experience and some managerial and administrative inefficiencies that were not considered in the WHO report, drawing on recent work by the OECD (OECD 2017).

The table is organized according to the three key policy questions discussed earlier: doing the right things, doing them in the right places and doing them right. The section on doing things right builds from health system inputs (medicines and other medical products, health workforce, infrastructure and equipment) to the outputs and outcomes they produce (health services). It also includes inefficiencies within the overall health system management, organization and governance and within the health financing system.

**Table 1: Common Causes of Inefficiency**

| **Doing the right thing** | Inefficiency could result from an imbalance between: population-based promotion and prevention versus personal and curative services; high cost, low impact health services versus low cost, high impact services; governance and public health functions versus other health services. Not doing the right thing in the context of UHC can manifest also as inadequate attention to financial protection compared to the availability and quality of health services, or vice versa. |
| **Doing things in the right place** | Inefficiencies would commonly include services being provided at higher level institutions that could be done with the same quality but with lower costs at lower levels of the system (e.g. avoidable outpatient specialist visits and emergency department visits), avoidable inpatient admissions or longer than necessary length of stay, or acute inpatient surgeries which could be performed in day-care. |

² For more information on the extent of health shocks on consumption and the ability to smooth consumption with insurance, see for example, Limwattananon et al. 2015, Wagstaff & Lindelow 2010, Chetty & Looney 2006.
Doing things right – inefficiencies can be linked to higher-cost inputs chosen, or inputs not achieving their maximum potential

A. Inputs

1. Medicines:
   a) Under-utilization of generics or paying too much for any specific medicine;
   b) Use of ineffective medicines, the wrong medicines, or using them at the wrong time;
   c) Overuse or unnecessary use.

2. Infrastructure (e.g. health facilities) and equipment:
   a) Inappropriate health facility size, particularly hospitals, for optimal efficiency;
   b) Under or over-capacity in health facilities;
   c) Equipment that is purchased and cannot be repaired or is not used optimally.

3. Personnel: inappropriate mix between different cadres; located in the wrong places; demotivated workers with low productivity (e.g. low visits per health workers per day, high rates of absenteeism); poor quality of care provided.

4. Inappropriate mix of inputs: e.g. health workers but no medicines or other medical products, a lower cost mix of inputs is possible to achieve the same outputs.³

B. Outputs and outcomes

5. Health services:
   a) Unnecessary, tests, procedures, treatments/surgery compared to need;
   b) Medical errors and low quality care, including doing the interventions at the wrong time (e.g. late) meaning the inputs and outputs do not achieve the desired outcomes;
   c) Underuse of needed health services (prevention, treatment, rehabilitation, palliation; includes medicines) leaving patients vulnerable to unnecessary disease, suffering, and possible increased medical costs subsequently.

C. Health System Structure, Organization and Governance including the Health Financing System

6. Waste (including expired medicines), corruption, fraud.

7. Sub-optimal public financial management practices including late disbursements from the Ministry of Finance, a large number of line items or inflexibility across lines, low budget execution rates.

8. Inefficiency in raising revenues (for health), particularly when revenue raising for health is independent from general government revenue collection.

9. Fragmentation in the system: in pooling, but in the broader health system as well - e.g. procurement, supply chains, laboratories, service delivery. This can be associated with domestic decisions such as establishing separate insurance schemes for different population groups, or to decisions made by external partners to bypass existing national systems and establish parallel systems and structures e.g. financial flows, audit, M&E, service delivery, laboratories.

10. Administrative inefficiency: higher-than-necessary costs for the services offered, including in health insurance agencies.

³ There are some overlaps inevitably between the various forms of inefficiency. For example, health workers without medicines or diagnostic tests might be linked to administrative efficiency, while using higher cost inputs to achieve results that lower cost inputs could achieve is one aspect of health workforce management as well.
1) **Doing the right things**

With resource scarcity, doing the right things requires explicit decisions about which interventions should be available, and within that, the balance between prevention and treatment, governance and public health functions and personal health services, and what types of inter-sectoral actions to improve health should be pursued. These issues are discussed briefly in turn following the order of Table 1.

i. **Mix of health interventions and packages**

Most of the literature on doing the right things has focused on the appropriate mix of health interventions using cost-effectiveness analysis. The literature on the cost-effectiveness of specific interventions or a group of interventions for a particular health problem is vast, too vast to reference fully here (e.g. Cambiano et al. 2015; Edejer et al. 2005; Mock et al. 2015; Kim et al. 2015; Ortegon et al. 2012; Roze et al. 2015; White et al. 2011). There is less analysis comparing the efficiency of funding a set of different types of interventions across priority health problems, the sort of analysis required if countries are going to develop an essential package of health services from scratch or modify their package based on the best possible evidence. From the few analyses that are available, the evidence shows that many countries are not fully funding health interventions that are low cost, high impact in terms of population health benefits (such as childhood immunizations) while high cost, low impact interventions (such as forms of tertiary care for chronic diseases) do receive funding (Chisholm et al. 2012; Evans et al. 2005; Laxminarayan et al. 2006; Jamison et al. 2006; WHO 2017a).

In these cases, reallocation of resources towards the low cost, high impact interventions would improve population health for the same expenditure. There are few explicit analyses of the likely gains of doing this. One exception is the WHO (2010) estimate that switching between interventions in this manner could produce the same health benefits at between 16% and 99% of the current costs (depending on the health problem being studied and the country).

There are a number of technical and practical problems with much of the cost-effectiveness literature that limit its value to countries wishing to “do the right thing” by changing their intervention mixes or developing a package of interventions to which all people will be guaranteed access. First, much of the analysis asks if new interventions should be funded with no guidance about what interventions should be reduced if resources are scarce. The fundamental question of which mix of interventions would most improve population health with the available resources is rarely asked. Second, the cost-effectiveness of any given intervention depends on many location-specific factors, including cost structures, disease patterns, the population age pyramid, what other types of interventions are being undertaken that might interact in terms of costs or effects, and current levels of coverage. It is not necessarily appropriate to extrapolate the results of a study undertaken in one country to other countries, whose variables may differ substantially. Third, costs and effects can change over time with changes in disease patterns and costs, and technological innovations.

It would be difficult for any country to undertake studies of the cost-effectiveness of all possible health interventions – promotion, prevention, treatment, rehabilitation and palliation, as well as personal and population-based – in their own settings, and update them frequently, even

---

4 By modify, we mean add and subtract. Traditional incremental cost-effectiveness analysis looks at what services should be added, but rarely is guidance given about what should be reduced to make way for the new intervention when additional resources cannot be found.

5 These estimates did not take into account the transaction costs involved in making the changes.
with good technical capacities. International initiatives such as the various iterations of the disease control priorities project (DCP) and the WHO-CHOICE project have sought to collate the information on cost-effectiveness from a wide variety of interventions in ways that could help countries, particularly those with limited technical capacity, understand which interventions could be given high priority in their settings (Jamison et al. 2006; Mock et al. 2015; Ortegon et al. 2012; WHO 2017a). WHO-CHOICE also reports results at different levels of coverage, taking into account possible interactions between interventions in terms of costs and effects.6

There are several well-accepted lessons from this type of work. In countries with a high burden of communicable disease and high maternal and child mortality, packages involving prevention and treatment for maternal and neonatal care, child health, HIV/AIDS, tuberculosis and malaria are highly cost effective (Laxminarayan et al. 2006). Most countries currently seek to make these packages universally available, even at low levels of income per capita. Other interventions that seem to be relatively cost-effective (e.g. prevention of traffic accidents, reduction of tobacco use, surgical wards in district hospitals, treatment of acute myocardial infarction and some secondary prevention for cardiovascular events) are not yet, however, universally available in low and middle-income countries. The reasons why they are not widely available are considered further in Section 6.

This highlights another problem with this literature – it rarely considers the resource envelope. Financial constraints prevent most low and lower-middle income countries assuring universal coverage with all the interventions that the literature suggests are cost-effective: this is a failure of the literature more than a failure by countries.

Money is not, however, the only constraint, at least in the short-to-medium term. Others include the number, skill mix and location of health workers, and the type of health infrastructure available, including location and type of hospitals, primary care facilities and community facilities. Efforts to develop a set of guaranteed health services must take into account all of these constraints. This requires a relatively complicated decision-making process. Formally, it would require some form of programming model rather than simply comparing cost-effectiveness ratios, although a second-best alternative is to compare the non-financial resource needs of any proposed set of cost-effective interventions with the available non-financial resources to assess the feasibility of implementation.

**ii. Prevention versus treatment**

Part of the cost-effectiveness literature has focused on prevention either by itself or in comparison to treatment of various types (e.g. Diabetes Prevention Program Research Group, 2012; Chisholm et al. 2012; Granich et al. 2012; Jamison et al. 2006; Kuyken et al. 2015; WHO 2017a). The results are not particularly surprising: some types of prevention are very cost-effective and some are not. Some types of treatment are cost-effective and some are not. Prevention is not always “better” than cure in the sense of producing more health for the money, although often it is. Sometimes, though not always, expenditure on prevention results in a net saving of future treatment costs.7 The decision about funding prevention or treatment needs to be taken on a case-by-case basis, and each country will end up with its own mix depending on factors such as disease patterns, cost structures, and health worker capacities. It is not possible, therefore, to offer generic guidance as to what share

---

6 This allows for economies and diseconomies of scale, economies of scope, and interactions between interventions in terms of effectiveness.
7 That is the present value of the costs of the intervention are lower than the present value of the savings in the future costs of treatment, rehabilitation and palliation.
of health expenditures should be allocated to prevention and promotion rather than subsequent
treatment, rehabilitation and palliation.

iii. The balance between governance, administration, public health functions and personal
services

Improving and maintaining health requires a number of actions beyond ensuring prevention,
treatment, rehabilitation and palliation at the individual level. Population-based promotion and
prevention (such as public health education campaigns), for example, is something that is unlikely to
be achieved without the involvement of government. The set of additional activities that require
government involvement (generally including funding), are often called essential public health
functions or services. These can be defined with differing degrees of aggregation (e.g. WHO 2017b
and 2017c; CDC 2017) and generally include: disease and outbreak surveillance and control;
population-based health promotion; linking people to personal services; developing the health
workforce; setting and enforcing standards in service delivery; and health research.

Sometimes the governance and administrative activities required to keep a health system
functioning are also included, including developing and implementing plans for the health sector,
developing and enforcing legislation and regulations as appropriate, interactions with the
community and promoting inter-sectoral action to improve health. Countries struggle to find the
appropriate balance between finance for routine administration, governance and the other public
health functions in the face of continual demand for personal services for individual people.

Scientifically valid guidance on the appropriate balance to ensure efficiency is difficult to find.
The cost-effectiveness literature examines some of these functions, mostly relating to population-
based health promotion or legislation and regulation to control things that are harmful to health
such as tobacco or traffic (e.g. Ortegon et al. 2012; Goetzel et al. 2014; Mason et al. 2014; Gordon &
Rowell 2015). Beyond arguing that many interventions are cost-effective or that more health
promotion of various types should be implemented (e.g. Lobstein et al. 2015), there is little guidance
what personal health services should be cut to make room for the increased expenditure in these
areas.

iv. Inter-sectoral and multi-sectoral action

It has long been understood that many factors outside the health sector influence health
(such as income, education, inequalities, environmental degradation and gender and social norms)
and that health also influences many external factors (such as the ability to earn and to go to school)
(Grossman 1976; Cumpor 1984; Wilkinson 1997; Commission on Social Determinants of Health 2008;
Marmot et al. 2008). As a result, there is increasing literature arguing that a “whole of government”
or “multi-sectoral” approach is needed to complement health service delivery to improve health and
reduce health inequalities (e.g. Marmot et al. 2008; Carey, Crampend and Keast 2014; WHO 2014).
This builds on the already large literature arguing for “inter-sectoral” actions (the health ministry
working bilaterally with other sectors) to improve health (e.g. Dahlgren 1994; WHO 1997; Adams et
al. 2014; Davies et al. 2014).

There is also a growing literature documenting where inter-sectoral or multi-sectoral actions
have been taken, and how more could be facilitated (e.g. Anaf et al. 2014; Larsen et al. 2014;
Dawson, Huikuri & Armada 2015; de Andrade et al. 2015; De Leeuw & Peters 2015). As yet, however,
there is very little information that can be used to guide governments in deciding whether it is more
efficient to allocate more resources to other sectors at the expense of financing the health sector. Part of the problem is that actions outside the health sector can improve health alongside other aspects of human welfare, such as educational attainment, the environment and income earning capacities. Cost-effectiveness analysis captures only the impacts on health so is not an appropriate tool to guide such broad decisions.

Cost-benefit analysis has sometimes been used in these cases, converting all benefits including reduced mortality and morbidity into money terms. Lives saved are translated into a monetary value, frequently using a method known as the value of a statistical life which is based on an assessment of the value that individuals place on a marginal change in their risk of death, most commonly revealed by the salary premium they would accept to work in a more dangerous occupation (e.g. Kniesner, Viscusi and Ziliak 2014; Laxminarayan et al. 2014; Viscusi 2015). The approach has a strong following for its use in valuing health benefits associated with interventions such as transport improvements, and in the assessment of health interventions. However, the appropriateness of statistical life value methods are widely debated by health economists and disliked by health professionals who consider that human life is priceless (e.g. Alberini & Ščasný 2013; Angevine & Berven 2014).

A narrower perspective for multisectoral or inter-sectoral analysis is to determine which of the possible options (which are not currently being undertaken) the Ministry of Health should focus on in its effort to convince other sectors to take action to improve health. Most of the existing literature on the social determinants of health implies that the health sector should seek to have all of them implemented, but a minister’s time is scare, as is the time of the ministry staff and the funds they would have available to support these actions. From the health perspective, it is more efficient to target the activities in other sectors that offer the greatest health improvements for their investments in money and time. The literature offers little guidance on this, with the recent exception of a paper that sought to identify the health and non-health interventions that had the biggest impact on life expectancy in 54 lower income countries since 1990 based on a form of regression analysis (Hauck, Martin and Smith 2016). Interestingly, of the non-health inputs, gender equality had the biggest impact on life expectancy while primary school enrolment also had an important effect. This is a promising start, but more work needs to be done to help individual countries understand, looking forward, what type of specific interventions outside the health sector are likely to improve health the most. For example, most countries reached close to universal primary enrolment during the MDG era, so increasing enrolment more would not have large impacts in the future.

v. Financial protection and service coverage

As argued earlier, discussions about doing the right things in the face of financial constraints in health have generally assumed that the outcome of interest is an improvement in health. When it is recognized that people also value financial protection, there is another efficiency trade-off: between using scarce funds to increase coverage (and/or quality) with existing health interventions, thereby improving health, or improving financial protection by reducing out-of-pocket payments. Little attention has been given to how to address this trade-off to date.

A recent exception has been called “extended cost-effectiveness analysis”. It examines the impact of different types of interventions on health outcomes as in standard practice, but also considers their impact on a financial protection indicator such as the incidence of impoverishment linked to out-of-pocket health payments (Verguet et al. 2013; Shrim et al. 2015; Verguet et al. 2015a; Verguet et al. 2015b). The information on both types of outcomes are presented separately rather than seeking to put relative weights to the two components.
To use this type of analysis, decision-makers would need to consider pairs of outcomes (health improvement and financial protection), and implicitly attach weights to the relative value of each when deciding what types of health services to support. There are some problems with this approach. For example, it is even more contextual to a specific country than normal cost-effectiveness analysis because out-of-pocket payment levels and their distribution across types of health services, and incomes vary greatly. The analysis has also been applied to health interventions and not yet to interventions aimed specifically at reducing out of pocket payments or strengthening pooling arrangement. However, it is an interesting development that can contribute to understanding at least some of the trade-offs involved in decisions about efficiency and doing the right things.  

2) **Doing the right things in the right place**

Doing the right things also requires an assessment of which care settings are the most efficient for services to be delivered. If there are no effects on quality, services should be delivered in the least costly care setting. Common examples of services being provided in inappropriate care settings include (i) acute inpatient admissions which could have been avoided or shortened through the availability of adequate ambulatory care (including day care), home-based nursing care or social care and (ii) outpatient specialist or emergency care visits for services which could have been provided in primary care. Although some of these service delivery and organization models are not yet widely available in lower-income countries where patients (particularly the poor) sometimes do not obtain any type of formal care at all, experiences from more developed countries can provide important lessons to prevent common setbacks and design challenges in achieving greater efficiency as these lower-income countries develop their service delivery systems.

Evidence of avoidable hospital admissions include admissions for conditions that are not severe enough to warrant an admission (and thus could be treated at less costly care settings) as well as for ambulatory care-sensitive conditions which could have been averted through the provision of adequate preventative care in lower and less costly care settings. For example, the large variation in admissions for low-mortality conditions (e.g. asthma exacerbation) from the emergency department indicates that some of these admissions may be unnecessary (Sabbatini, Nallamothu and Kocher 2014;). In addition, studies show that costly hospitalizations for ambulatory care sensitive conditions (e.g. diabetes) could be avoided through access to adequate prevention and early treatment in primary/community care (James, Berchet and Muir 2017; Rosano et al. 2013). There is wide variation in age and sex-standardized hospital admission rates for Asthma and COPD (both ambulatory care sensitive conditions) across OECD countries (Figure 1). Asthma admission rates are highest for Korea, the United States and Slovak republic while COPD admission rates are the highest for Hungary and Ireland. In Estonia, avoidable chronic obstructive pulmonary disease (COPD) and asthma admissions constituted 76.9% of admissions for lower chronic respiratory disease, while avoidable congestive heart failure (CHF) and hypertension admissions comprised 84.3% of admissions for hypertension and other forms of heart disease (World Bank 2015).

---

8 Another more technical issue is that the extent of impoverishment due to out-of-pocket health payments does not really reflect the ex-ante value of the financial protection, but more the result of not having sufficient financial protection.
A prolonged length of stay in hospitals is another indication of care that is not being delivered in appropriate care settings. Often patients can remain in the hospital due to the need for nursing or palliative care, which they are unable to access in other care settings. For example, in Estonia, about 32.6% of hip fracture patients remain hospitalized beyond the international standard of 28 days, while the percentage of stroke patients remaining in acute inpatient care beyond the international standard of 56-days is 6.91% (World Bank 2015). Home-based nursing care has been shown to be more efficient (achieving at least the same quality of care at lower cost) than institutional care for the frail elderly (de Jonge et al. 2014). While the frail elderly often prefer to remain at home, they cannot easily travel to use office-based primary care services. This sometimes results in hospitalization that would not have occurred in the presence of home-based primary care services (e.g. Stall, Nowaczenski and Sinha 2013). Similarly, the availability of social care services (institutional or home-based) which provide services such as assistance with activities of daily living has been shown to significantly affect the extent of delayed discharges from hospitals (National Audit Office 2016). Inadequate coordination with or a lack of access to social care leads to a “default to doctor” phenomenon, which in turn creates additional pressures on medical staff time, which they are ill-equipped to handle (Dorell 2015).

Surgery for many conditions may be performed in less costly day care settings instead of acute inpatient care settings, while achieving the same results. For example, almost all cataracts could be removed with ambulatory surgery. Despite these opportunities, day care surgeries are not fully used, even in high-income countries. Although well over half of the OECD countries now conduct over 90% of cataract surgeries in day care settings, some still have rates substantially lower – 31% reported for Poland, 50% for Hungary and 72% in Austria (James, Berchet and Muir 2017).

There is also considerable evidence of unnecessary use of outpatient specialist and emergency departments in both low and high income settings, where unnecessary is defined as visits which could have been treated in less costly levels of the system with no reduction in outcome. For example, in Estonia a recent study showed that approximately 20% of visits with outpatient specialists for diabetes patients and nearly 70% of visits for hypertension patients could have been
treated in primary care\(^9\) (World Bank 2015). Emergency department visits per 100 population varied from 10 to 70 in OECD countries in recent years suggesting inefficiency in at least some countries. Record reviews from a number of these countries suggested that between 12% and 56% of these visits could have been treated in less costly settings without any decline in quality of care (James, Berchet and Muir 2017). Bypassing less costly primary care providers for more expensive hospital-based care is also a common phenomenon in low and middle-income countries including China, Kenya, Namibia and Tanzania (Wu et al. 2016; Nshimirimana et al. 2016; Low et al. 2001; Kahabuka et al. 2011).

The delivery of care in inappropriate care settings is driven by a number of factors. The literature suggests that these factors may include:

- System design factors (e.g. organizational separation of health and social care, weak gatekeeping functions of primary care providers, lack of access to primary care providers after hours, etc.);
- Contracting and financing (e.g. weak financial incentives for solving medical problems in primary care, strong financial incentives for increasing the volumes of outpatient specialist and inpatient care, etc.);
- Clinical and professional mechanisms and process (e.g. lack of adherence to clinical guidelines and pathways); and
- Health system inputs (e.g. shortage of home nursing providers, lack of fully functional e-platform for electronic medical records and patient referral coordination) (World Bank 2015).

In Estonia, weak primary care including weak management of patients with chronic diseases, low adherence to evidence-based practice, limited service scope and knowledge gaps were shown to be particularly important contributors to avoidable specialist visits and hospital admissions. As a result of these factors, low trust of primary care providers among patients may contribute to self-referrals and bypassing. In Kenya, for example, patients preferred to bypass the less costly primary care gatekeepers and go directly to more expensive hospitals because of perceived poor communication, long waiting times and being treated without dignity and respect in primary care (Nshimirimana et al. 2016).

Adequate coordination and continuity of care within and between care settings (e.g. receipt of follow-up care after a hospital admission) is also critical to preventing further deterioration in patient’s health, which in turn may require repeat outpatient specialist or emergency department visits, or readmissions to acute inpatient care. This includes coordination with end of life palliative care where managed care by palliative teams has been shown to reduce hospitalization rates (e.g. Reyniers et al. 2014; Seow et al. 2014).

3) **Doing things right**

Once decisions have been made about which interventions should be available, where they should be available, and the balance between expanding services and expanding financial protection, the next question is how to get the most out of the different types of interventions for the lowest cost. Here we discuss briefly the elements included in Table 1 under “doing things right”.

---

\(^9\) This indicator looks at specialist visits by patients whose conditions are considered uncomplicated based on the primary diagnoses made. Of these, visits were considered avoidable if patients presented to a specialist not specified in national Estonian guidelines. If several visits were billed under the same claim (e.g., pertaining to one care cycle), the decision on whether these visits were avoidable was made based on the primary diagnosis code assigned to the claim.
### Health Services

Probably the best documented source of inefficiency from Table 1 relates to medicines which account for somewhere between 20% and 60% of all health expenditures in low and middle income countries, and around 18% in high income settings (Niessen & Khan 2016). Widespread over-prescription and overuse of medication, and incorrect prescription and use, particularly of antibiotics, is well documented (e.g. Kalyango et al. 2015; Li et al. 2012; Holloway et al. 2013; Mao et al. 2015; OECD 2017). Holloway (2011) and Holloway and Dijk (2011) suggest that globally, less than 50% of patients receive appropriate medication (compared to treatment guidelines) for their conditions: even fewer in low and middle income countries - 40% of patients in the public and 30% in the private sectors respectively.

Low adherence to recommended therapy wastes resources because of an increased need for subsequent medical care (Pereira et al. 2014; Ryan et al. 2014; Ali, Abou-Taleb and Mohamed 2016; Choudhry et al. 2016). There are many determinants, but low adherence is more likely with long-term therapy than for an acute episode, and also where out of pocket payments for medicines are relatively high. An increasing literature on the affordability of medicines suggests that, in general, affordability declines with levels of national income per head (e.g. Cameron et al. 2009; Cameron et al. 2012; Jiang et al. 2015; Iyengar et al. 2016; Khatib et al. 2016). In low and middle income countries, medicines frequently have to be paid for out-of-pocket, so the lack of affordability translates into reduced adherence to a full course of treatment and, for some people, an inability to purchase and benefit at all from needed therapy (Lu et al. 2011; Niessen & Khan 2016).

The lack of access to medicines that people need is also related to many countries paying too much for medicines. Prices for the same medicines vary substantially, even across European and OECD countries, for both generics and brand name medicines (Cameron et al. 2012; Simoens 2012; Vogler & Kilpatrick 2015; OECD 2017). While some variations in prices can be expected given the differences across countries in the size of the market (population, disease prevalence) and transport costs, Iyengar et al. (2016) illustrate the remarkable variability in prices across OECD countries. As an example, the price for a course of Sofosbuvir (for Hepatitis C) ranged from US$37,729 to US$64,680, with a median of US$42,017. In a study of 46 largely low and middle income countries, public sector prices for selected generics were been 5 and 17 times above the international reference price, with originator brands almost 30% higher (Cameron et al. 2012). The gap between the prices paid and the international reference price in the private sector was even greater. Similar results were found more recently in China (Jiang et al. 2015). Other studies reveal countries where brand name prices are no higher than the equivalent generic, suggesting those countries are paying too much for their generic medicines (Cameron & Laing 2010; Vogler & Kilpatrick 2015).

Reducing overuse and inappropriate use, and reducing the negotiated price for medicines, frees up resources that can be reinvested in ways that improve health or financial protection, although sometimes the pharmaceutical industry lowers the price of some medicines, but increases the prices of others. Policies to replace originator brand purchases with generic or biosimilar medicines also result in savings. While a number of countries

---

10 Even if these countries can negotiate lower prices (not all can), the difference in price between richer and poorer countries is less than the difference in average household incomes meaning that affordability falls with declines in national income per capita.

11 OECD (2017) points out that price comparisons can be difficult because of differences in packaging across countries as well as the secret nature of some of the discounts that countries negotiate.

12 “A biosimilar (also known as follow-on biologic or subsequent entry biologic) is a biological medical product which is most an identical copy of an original product that is manufactured by a different company. Biosimilars
have generic policies of various types, market penetration is much lower than optimal from the perspective purely of freeing scarce health resources by switching from higher cost to lower cost, equivalent effect, options (Dylst, Vulto and Simoens 2014; Hassali et al. 2014; OECD 2017). Estimates of the proportional cost saving from switching from specific brand name to generic medicines include 51% in Pakistan and 53% in China (Cameron & Laing 2010); between 4% and 23% in Austria (Heinze et al. 2015) and between 11% and 73% in 17 low and middle income countries (Cameron et al. 2012). In dollar terms, Haas et al. (2005) suggested that the US could save around $9 billion annually by introducing a generics policy, while Mulcahy et al. (2014) suggest the savings from switching to biosimilars would be lower - around 4% of current spending on biologicals - but still substantial at around $44.2 billion from 2014 to 2024.

Inefficiencies related to infrastructure have largely focused on hospitals, and less frequently on lower level health facilities. The considerable variation in efficiency (measured essentially as the ratio of health facility outputs to their inputs) across health facilities has been extensively documented in countries at all income levels through the use of frontier production function analysis (e.g. Kirigia et al. 2011; Besstremyanaya 2013; Kiadaliri, Jafari and Gerdtham 2013; Du et al. 2014; Jehu-Appiah et al. 2014; Kittelsen et al. 2015). Inefficiency in this context generally means that the same throughput of patients could be handled with fewer inputs (compared to the most efficient hospitals). A number of the studies in lower income countries have identified “over-staffing” in inefficient hospitals (Kirigia and Asbu 2013; Kirigia, Sambo and Lambo 2015), which can be linked to the inappropriate mix of inputs described in Table 1 as one of the causes of inefficiency. Overstaffing could also, however, be interpreted as showing that the same staff and infrastructure could cope with more patients, suggesting under-utilization of the infrastructure and staff.

Some of these studies also examined economies of scale or optimal hospital size. Frequently, inefficient hospitals are smaller than their efficient counterparts, although a study from South Africa showed that some hospitals were “too large” and some “too small” compared to the efficient set of hospitals (Preyra & Pink 2006; Kristensen et al. 2012; Leleu, Moises and Valdmanis 2012; Kirigia, Sambo and Lambo 2015). A study from two provinces in Canada showed that the optimal hospital size varied by province so it is reasonable to assume variation across countries (Asmild, Hollingsworth & Birch 2013). In some of the seminal work on hospital cost functions and economies of scale in the USA, Lave and Lave (1984) suggest that it is difficult to identify optimal size of hospitals from cross-sectional data because the nature of small and larger hospitals can be quite different. They argue that the optimal size depends on the scope and complexity of the services offered, so there is little generic guidance that can be given to countries about the optimal size. Detailed studies controlling for scope and complexity of services would need to be done in each setting. Questions of possible “over-staffing” could, however, be examined in other ways discussed more in Section 4.

The available data from low and middle-income countries on hospital occupancy rates also suggests substantial inefficiencies can exist in district hospitals. A study of 18 countries in 2007 reported an average bed occupancy rate in district hospitals of 55%, ranging from 20% to 98% (Chisholm et al. 2010). In Botswana, district hospital bed occupancy rates in 2009 were between 40% and 61%, but at the same time the two referral hospitals had occupancy

---

are officially approved versions of original “innovator” products, and can be manufactured when the original product’s patent expires”, Wikipedia, accessed 8 February 2017.

These methods are discussed more thoroughly in Section 4 in relation to tracking progress in improving efficiency.
rates of 143% and 222% respectively, suggesting efficiency and probably quality gains from encouraging admissions into district hospitals rather than referral hospitals (Seitio-Kgokgw et al. 2014).

Health care deemed to be inappropriate, ineffective or harmful, sometimes called “unnecessary” care, has been widely documented in high income countries (OECD 2017; Saini et al. 2017; Brownlee et al. 2017). Unnecessary treatment or investigations are those which have little likelihood of improving the patient’s quality or duration of life or which have more chance of doing harm than good. The list of common forms of unnecessary treatment includes imaging for lower back pain and headaches, antibiotics for upper respiratory tract infections, preoperative testing for low risk patients, cardiac imaging in low risk patients, induction of labour, caesarean section and some types of cancer screening (Hurley 2014; OECD 2017; Brownlee et al. 2017).

Ineffective care includes: the use of vitamin and mineral supplements aimed at preventing cardiovascular disease; antipsychotic prescription in older patients which increase their risk of falls; or interventions undertaken at the wrong time, or not at all. An extreme version of ineffective care is care that is harmful to patients, including medical errors. Ineffective and harmful care can lead to subsequent outpatient visits, hospital admissions and other forms of treatment that could have been avoided, and at the extreme, to avoidable deaths. Reducing all forms of unnecessary, inappropriate and harmful patient management saves money (sometimes also improving health outcomes) that can be re-invested into interventions that improve health and/or increase financial protection.

The evidence of over-use is frequently complicated by the fact that patients differ in terms of need, but for many of these examples, the evidence of wide variations in their use across countries, adjusting for population differences, is taken to imply over-servicing in at least some of the countries. For example, the number of CT examinations per 1000 population varied from 31.9 in Finland, to 254.7 in the USA, with 15 of the 28 countries above the OECD average of 131.6 (2014 data; OECD 2017).

For Caesarean section and the use of antipsychotics in elderly patients, there are standards of need that enable an assessment of overuse – 15-20% of deliveries for Caesareans and 0% for anti-psychotics (reported in OECD 2017). In 2014, rates of delivery using Caesarean section in OECD countries varied from 15.3 per 100 live births in Iceland to 51.1 in Turkey, with an OECD average of 27.5. The average rate increased between 2007 and 2014, although it fell in 12 of the 32 countries for which information was available (OECD 2017, Figure 2.2).

In non-OECD countries, rates of Caesarean section are also increasing and in many are already above the recommended range (Ye at al. 2015; Betran et al. 2016). The average across the countries classified as "less developed” using the Human Development Index in 2014 was already almost 21%, ranging from 1.7 to 56.4 per 100 live births (Betran et al. 2016). Of the estimated 6.2 million unnecessary Caesarean sections annually across the world, 50% occur in Brazil and China alone (Berwick 2017).

Scattered examples of other types of unnecessary, inappropriate and harmful use of health services in low and middle-income countries exist. Presumptive treatment of fevers with anti-malarial medication in some countries wastes resources and is no longer necessary, given the availability of inexpensive rapid diagnostic tests (Ochodo, Garner and Sinclair 2016). Other examples include: the overuse of medicines when prescribers sell them (officially or unofficially) and the longer length of hospital stay linked to payment per day (Chen et al. 2014; Gao, Xu and Liu, 2014; Rahman et al. 2014; Zhang et al. 2015b); substantial overuse of antibiotics, for example, in children with acute diarrhoea, in India and Thailand (Brownlee et al. 2017); and unnecessary cardiac procedures in Brazil and India (Brownlee et al. 2017). There is no reason to believe that overtreatment and inappropriate
treatment is any less in low-income settings than in high-income countries for people who can pay or whose costs are covered by government or insurance.

Adverse events that could have been prevented are also documented in a number of OECD countries; these can lead to the need for additional treatment and sometimes result in unnecessary deaths (OECD 2017). Rates of preventable adverse events in hospitals across the 10 studies reported by the OECD ranged from 1.0% to 8.5%. Similar evidence from low and middle income countries is less readily available, although there is evidence that a number of countries, including China, are developing approaches to improve patient safety and considering staff perceptions of the problem as the basis for developing possible solutions (e.g. Wang et al. 2014; Zhou et al. 2015).

**Under-use of necessary services** exists alongside overuse globally, in the same country, and even in the same patient (Saini et al. 2017). Underuse is probably more widespread in low and middle-income countries than in high-income settings. Indirect evidence of underuse is found in the figures that 1.5 million children die each year from vaccine preventable deaths, and an estimated 84% of pre-term deaths are preventable with appropriate care (Glassziou et al. 2017).

More direct evidence comes from an assessment of particular types of interventions. Most of the annual estimated unmet need for surgical interventions of 320 million procedures was in low and middle-income countries (Glassziou et al. 2017). For example, in the “least developed” group of countries based on the Human Development Index, the average rate of Caesarean section in 2014 was only 6% of all deliveries compared to the recommended 10-15% (range across countries from 1.4 to 41.1%). In the African countries for which data were available, the average rate was only 7% suggesting that many women who need a Caesarean section still do not obtain this important life-saving intervention.

Underuse is inefficient in the sense that by failing to access appropriate, low-cost, effective interventions, many of the affected people will need to use more health services in the future, at a later stage in the natural history of the disease, leading to unnecessary deaths and morbidity.

Problems with the health workforce have been widely documented. In high income countries, insufficient domestic production has led to the need to import health workers from other countries, which can have unintended effects on efficiency if the arrivals do not have a good command of the local language or culture (Aluttis, Bishaw and Frank, 2014; Crisp and Chen 2014). In low and middle-income countries, there is a severe shortage of health workers. Only 5 of 49 high-need countries are achieving the minimum threshold of 23 nurses, doctors and midwives per 10,000 population needed to deliver an essential set of maternal and child health services (Crisp and Chen 2014; WHO 2017e). The shortage of health workers is much more acute in remote and rural areas, a trend not limited to lower income countries (Morrell et al. 2014; Abimbola et al. 2015; WHO 2017e).

The mix of skills also varies considerably across countries, suggesting possible inefficiency at least in some settings. For example, India has a ratio of approximately one allopathic doctor to each nurse and midwife (Rao 2014). Although there is no gold standard, most countries have considerably fewer doctors per nurse and midwife – e.g. Indonesia had 0.16; Thailand 0.23 at the time of Rao’s India study (Rao 2014; WHO 2017f).

Community health workers became popular in the 1970s in an effort to move services closer to the people who needed them. Their number declined in the 1980s, but a recent surge in interest has occurred, linked to the need to get services close to people and help mitigate shortages in other types of health workers (Perry, Zulliger and Rogers 2014). A number of
apparent successes have been reported, including lady health workers in Pakistan (Yousafzai et al. 2014); Ethiopia is in the process now of expanding the number and distribution of their community extension workers (Nejmuadin et al. 2011). A large number of studies of these workers’ efficacy have produced inconsistent results. WHO is developing a set of guidelines for community health worker programs and last year called for expressions of interest to undertake systematic reviews of the various studies to help it in this process (WHO 2017g).

Other problems affecting the efficiency of health workers in low and middle income countries are linked to motivation, absenteeism, retention and dual practice (Abimbola et al. 2015; Hotchkiss, Banteyerga and Thanareny 2015, Witter et al. 2015). Absenteeism reduces the number of services the health worker can provide, while high turnover of staff requires retraining and relearning that also reduces efficiency (Daouk-Öyry et al. 2014). Limited information is available on the extent of absenteeism – two studies from Tanzania suggested substantial lost productivity from absenteeism. Kurowski et al. (2004) estimated that there was a 26% reduction in health worker productivity because of unexplained absences and breaks, while Manzi et al. (2012) reported that 44% of the staff were not available at the time of the study visit, and nurses worked only for 57% of their allotted time because of breaks taken while on duty. Comparable data on these factors are not publicly available across countries, so it is difficult to make generalizations on their importance – although health ministries will want to track these variables in their search for greater efficiency.

In 2010, salaried health workers accounted for just over 42% of government health expenditures globally, lower in Africa and South-East Asia compared to Europe and the Americas (Hernandez-Peña et al. 2013). In terms of total health expenditures, remuneration of health workers in the government sector accounted for 34% while independent health worker remuneration accounted for another 10% globally. When funds are short, ministries of health frequently prioritise paying health workers, leading to reports of health workers being in post, but deprived of the inputs such as medicines that they need to do their work (Moszynski 2016). However, there is little systematic data on this type of inefficiency in the mix of inputs across countries.

ii. Health Systems and the Health Financing Component

Waste in the form of fraud and corruption occurs in health systems at a number of levels (see OECD 2017). In service delivery, patients can make wrongful insurance claims and providers can bill for patients or services they did not provide. The opportunities are particularly high in procurement, including the bidding process, and in the subsequent distribution of the inputs that were procured. In human resource management, this extends to taking bribes in return for an offer of employment, or employing friends and relatives who might not be the best person for the job.

A Transparency International Report quoted by OECD suggests that a third of respondents across the 28 participating OECD countries believed their health systems were corrupt or extremely corrupt, with NGOs, the military and the education system believed to be less corrupt than the health system (Transparency International 2013; OECD 2017). Forty-five percent of global respondents (103 countries) considered their health systems to be corrupt or extremely corrupt, and in this case health fared worse than religious organizations and the media, in addition to NGOs, the military and the educational system.

A recent study of fraud in the health sector suggests that somewhere between 3% and 10% of health expenditures are lost annually, with a mean of 5.6%. Extrapolating to the world, corruption and fraud is estimated to cost the world $426 billion annually, resources that
could be used to improve coverage with needed health services and with financial protection.¹⁴

**Poor public financial management** practices also limit the efficiency of government health systems. There are many problems, including unpredictable budget allocations, fragmentation in revenue streams and funding flows, unpredictable or late disbursement from Ministries of Finance to line ministries and from central ministries to sub-national units, low budget execution rates, and inadequate financial accountability and transparency (Brixii et al. 2012; Fox et al. 2013; Cashin et al. 2017). These practices result in the available money not being spent, in poor spending because money arrives late or cannot be transferred between line items, or in leakages because of limited transparency. To illustrate only one part of this, recent public expenditure reviews from 6 African countries suggest underspends of the approved budget range from $10 to $120 million a year (WHO 2016). In per capita terms, this translates into losses of potential spending of between $1 and 3.50 annually. Data from two countries allowed a comparison of execution on salaries versus other types of expenditure. Almost the entire budget for salaries we spent, while expenditure on the inputs needed to keep the health system running ran at around 40%.

Additional sources of waste in a system are linked to administrative forms of **administrative inefficiency** such as higher than necessary administrative costs, bureaucratic red tape and delays, complex systems that take time to negotiate, and increasingly health care providers spending their time on administrative issues rather than interacting with patients or the population (OECD 2017). Information on the extent of administrative costs in health systems is difficult to find for low and middle-income countries in particular. It exists largely for health insurance administration but even then. For example, in the early 2000s, social health insurance administration costs averaged 3.8% of total insurance expenditures in 15 high-income OECD countries, data coming from country health account studies. The lowest proportion was somewhere under 2% in Estonia (ranging from 1.1 to 1.9% depending on the year) and the highest around 7% in Luxembourg (yearly range from 6.6 to 7.0%). Two of the three middle-income countries for which data were available had administration costs of less than 3% of total insurance expenditures (Georgia and Turkey) but in Mexico they were almost 17%.

Administrative costs in private insurance, again using country health accounts data, were in general substantially higher, consistently over 10% in 14 of the 23 high-income countries for which data were collated, reaching as much as 30%. On the other hand, in New Zealand they were only 5%. The variability in the range of administrative costs across countries and types of insurance suggests some are substantially more efficient than others and there are possible gains to be made from reducing the share of insurance expenditures going to administration.

The final source of inefficiency discussed here is fragmentation within systems. In financing systems, this can manifest itself in terms of multiple payers and purchasers for different types of health services or different parts of the population. This is frequently caused by fragmentation in fund pools – health insurance co-existing with government financing and provision, or multiple insurance pools each of which purchases services for their clients. Such fragmentation leads to higher administrative costs and higher prices than could be negotiated by a single large payer. Fragmentation can also be an obstacle to equitable

---

¹⁴ Global health spending in 2014 was US$7.6 trillion (WHO Global Health Expenditure database).
coverage - poorer population groups are covered by schemes that are less well funded than those for the formal sector or government employees, for example, so offer a smaller range of health services (Tucmeanu 2014; Meng et al. 2015; Kutzin, Yip and Cashin 2016).

In broader health systems, examples of fragmentation include laboratories for tuberculosis separate to those for HIV/AIDS; multiple procurement, purchasing, distribution, accounting, monitoring and service delivery systems; or patients requiring long term care being pushed between nursing homes and hospitals because different parts of government pay for the different types of care, and each seeks to minimize their own costs (Sidibé & Campbell 2015, Rao et al. 2014). Some of these issues are found in countries at all income levels (e.g. Hall 2015; Lewis 2015), but an additional feature of many low and lower middle-income countries is the fragmentation associated with inflows of development assistance for health where external partners have chosen to establish systems parallel to those that already existed rather than strengthening and supporting existing systems (e.g. Kieny et al. 2014; Panter-Brick, Eggerman and Tomlinson 2014; Gostin & Friedman 2015). The problems associated with how to best preserve the health gains associated with these programs are now being faced in countries transitioning from funding from Gavi, Global Fund and some bilateral aid agencies.

The question of whether decentralization is a form of fragmentation associated with inefficiency is still controversial despite years of experience. Decentralization policies have been widely implemented, not just in health, with the goal of improving one or more of the following: efficiency, service quality, management, responsiveness to local needs and equity (Saltman, Bankauskaite and Vrangbaek 2004). There is mixed evidence. Some studies find that it has been associated with increased accountability of government to citizens, or with improvements in management that then resulted in higher coverage with health services and improved health outcomes (Alves, Peralta and Perelman 2013; Loubiere et al. 2009; Samadi et al. 2013) Other studies show little impact of higher costs and increased inequalities between local government units due to differences in fiscal capacity (Azfar, Kähkönen and Meagher 2001; Atkinson & Haran 2004; Saltman et al. 2007; Langenbrunner, Xu and Chu 2016). There have also been suggestions that renewed central-level financing would improve equity and efficiency in decentralized settings (Langenbrunner et al. 2016).

One of the reasons for contradictory results is that decentralization takes many forms, and can have many components each requiring local government capacity – e.g. raising revenue, planning, purchasing or providing services, monitoring and evaluation, audit. However, the topic remains contentious and will be discussed at this Forum.

V. **Identifying the most important sources of inefficiency**

Not all countries will have each of the sources of inefficiency described in Table 1, and even where they exist, their relative importance will vary across settings. However, some inefficiency exists in every country – every country could achieve more with the available resources (WHO 2010a). There is little global guidance available on how to identify the most important sources in a specific setting, which is why this section has spent some time explaining the common causes of inefficiency and their sources. This can be a starting point for countries thinking through which ones are the most important in their settings, and which they would tackle first. Indicators of inefficiency can help provide evidence in support of discussions about the most important sources of inefficiency.
and subsequent policy development, so they are discussed in the next section.

VI. Measuring and monitoring inefficiency

There are two broad approaches to measuring inefficiency in health systems. The first seeks to obtain a single score summarizing the efficiency of a country’s entire health system in a similar way to the measurement of hospital or health facility efficiency described earlier. The second approach focuses on measuring efficiency related to the different components of a health system and the possible sources of efficiency described in Section 3. The former we call macro-efficiency and the latter, micro-efficiency analysis. They are discussed briefly in turn.

1) Macro-efficiency

Figure 1 depicts the relationship between life expectancy at birth and total health expenditure per capita in 2014. Life expectancy rises with health expenditure per capita, though at a decreasing rate. There is also considerable variability around the average regression line with countries with similar expenditures achieving very different levels of life expectancy. This type of analysis, sometimes with levels of attainment on other health indicators (e.g. maternal, child or infant mortality), or with levels of coverage on key interventions (e.g. with childhood immunizations) on the vertical axis, has been used as evidence that some countries are more efficient than others in translating expenditure into health outcomes or coverage (WHO 2010a; WHO 2016).

Figure 1: Life expectancy by health expenditure per capita, 2014

Many factors other than health expenditures obviously contribute to the observed variations in life expectancy, including differences in initial disease burden, population age structures and distribution, cost structures, availability of infrastructure and historical patterns of spending on
health and social services. Without controlling for these factors, it is not possible to state categorically that the observed differences are linked to differences in inefficiency.

More formal methods (most commonly data envelopment analysis (DEA) and stochastic frontier production function analysis (SFA)) have been developed to measure the relative efficiency of country or subnational health systems (or health facilities) and to control for, or examine the impact of, possible determinants of outcome in addition to expenditure (e.g. Hollingsworth 2003, 2008, 2016; Joumard et al. 2010; Pang 2005). Essentially, they shift the regression line up to the top of the scatter plot, obtaining a frontier of the most efficient performers, given their levels of expenditure (and other determinants). Those on the frontier are the most efficient (generally given a score of 1), and those below it are deemed inefficient compared to the best performers, with a score between 0 and 1 depending on how close they are to the frontier.\footnote{Different algorithms are possible in this process – e.g. in DEA, efficiency can be measured in the output space by moving vertically up to the frontier, in the input space by moving horizontally to the frontier, or as a mix of the two.}

In 2016, Hollingsworth identified over 400 studies applying these techniques to health issues in the last 30 years (DEA more frequently than SFA). The methods were initially applied to health facilities, particularly hospitals, but since the appearance of the 2000 World Health Report, which ranked health system efficiency across countries using SFA, they have been widely applied to studies of the comparative efficiency of health systems (WHO 2000, Sun et al. 2017).

Over time, the techniques have become increasingly sophisticated in their treatment of random noise in terms of measurement errors, sample noise, specification errors, cross-country heterogeneity and returns to scale (see Olesen & Petersen 2016; Greene 2004 & 2008; Hamidi & Akinci 2016). DEA can now deal with both multiple inputs and outputs, and recent studies have even merged DEA with SFA for different stages of the analysis of the determinants of inefficiency (e.g. Berenguer et al. 2016). In essence, however, they all treat deviations from the frontier that are not explained by the differences in inputs as inefficiency, then seek to understand the factors other than health system inputs linked to the variations in efficiency.

There are a number of problems with this form of macro-analysis from a practical policy perspective. First, the efficiency scores and rankings are sensitive, sometimes very sensitive, to the model specification and the data used for inputs, outputs and determinants (Frogner, Frech & Parente 2015; Gearhart 2016a & 2016b). When countries are grouped into different categories of performers, there is more consistency – e.g. some countries mostly fall in the group with the highest efficiency while others generally fall in the least efficient group regardless of the model specification (although there is still some movement between groups depending on the methods used) (De Cos & Moral-Benito 2014; Medeiros & Schwierz 2015). In this case, the policy focus can be on the countries that are consistently found to be in the lowest efficiency group although little can be said about countries that move between groups.

Secondly, the models assume that countries on the frontier are efficient, when more micro-level work suggests forms of inefficiency exist in all countries.

Thirdly, none of the models as yet capture the lags that must be important in translating inputs into outputs and outcomes. This is not simply a question of using panel rather than cross-sectional data. It relates to the assumption in both cases that expenditures in each given year produce health benefits in that year rather than over a period of years in the future. To capture this, models where current outcomes are a function of health expenditures in previous years as well as this year would be required.

Fourthly, the methods and approaches are very difficult for the typical policy-maker to understand, so the results are sometimes distrusted (Hollingsworth 2016).
Finally, and perhaps most importantly, the results rarely help policy-makers identify how they can improve efficiency or policies for doing so. This is because the variables that can be entered into the models to explain differences in efficiency need to be available and comparable across countries. Typical examples are the Gini coefficient of income inequality and average levels of educational attainment, variables that are not under the direct control of the health sector.

These models can, however, be a useful starting point to think about which country health systems (or within a country, which subnational systems, hospitals, or primary care facilities) seem to suffer from the most inefficiency. However, to understand the causes of inefficiency and then develop solutions, more micro-variables need to be explored.

2) **Efficiency in components of the health system**

Table 2 reports some of the indicators proposed in the literature that describe aspects of the causes of inefficiency described in Table 1, with the sources also indicated.

### Table 2: Possible Efficiency Indicators for the Sources of Inefficiency in Table 1

<table>
<thead>
<tr>
<th>Domain of inefficiency</th>
<th>Indicators suggested</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Doing the wrong things</strong></td>
<td>Share of public spending in remote areas, as % of GGHE</td>
<td>WB FSD; Tandon &amp; Cashin (2010)</td>
</tr>
<tr>
<td></td>
<td>Share of public spending that goes to the poorest 40% of the population</td>
<td>WB FSD</td>
</tr>
<tr>
<td></td>
<td>Share of national spending on pharmaceuticals</td>
<td>Smith &amp; Nguyen (2013)</td>
</tr>
<tr>
<td></td>
<td>% of funding allocated according to a strategic plan for the health sector or according to distribution of burden of disease</td>
<td>Tandon &amp; Cashin (2010)</td>
</tr>
<tr>
<td><strong>Doing things in the wrong place</strong></td>
<td>Share of public spending in primary care, as % of GGHE</td>
<td>WB FSD; Yip &amp; Hafez (2015)</td>
</tr>
<tr>
<td></td>
<td>Number of outpatient visits at tertiary hospitals per 100 population</td>
<td>OECD (2017)</td>
</tr>
<tr>
<td></td>
<td>% of surgeries conducted in ambulatory settings (cataract, tonsillectomy, inguinal hernia repair, cholecystectomy, laparoscopic)</td>
<td>OECD (2017)</td>
</tr>
</tbody>
</table>
### Spending Badly: Inputs

#### Medicines

<table>
<thead>
<tr>
<th>Description</th>
<th>Indicator</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-utilization of generics or paying too much for any specific medicine</td>
<td>Percentage of drugs purchased by the MoH that are generic or generics market shares by volume (%)</td>
<td>WB FSD; OECD (2017); Heredia-Ortiz (2013)</td>
</tr>
<tr>
<td></td>
<td>Average medicine cost per encounter or average number of medicines prescribed per encounter</td>
<td>WHO/INRUD</td>
</tr>
<tr>
<td>Use of ineffective medicines, the wrong medicines, or using them at the wrong time</td>
<td>Cholesterol-lowering drug &amp; antidepressant consumption</td>
<td>OECD (2017)</td>
</tr>
<tr>
<td></td>
<td>Percentage of prescriptions in accordance with clinical guidelines</td>
<td>WHO/INRUD</td>
</tr>
<tr>
<td>Overuse or unnecessary use</td>
<td>Percentage of encounters with an antibiotic or injection prescribed</td>
<td>Desalegn (2013); Hu, Liu &amp; Peng (2003); Wang et al. (2014); Ferreira et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Mean number of drugs/prescription</td>
<td>WHO/INRUD; Bashrahil (2010); Ferreira et al. (2013)</td>
</tr>
<tr>
<td></td>
<td>Percentage of medicines prescribed from an Essential Medicine List or formulary</td>
<td>WHO/INRUD; Desalegn (2013)</td>
</tr>
</tbody>
</table>

#### Personnel

<table>
<thead>
<tr>
<th>Description</th>
<th>Indicator</th>
<th>Source(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inappropriate mix between different cadres; located in the wrong places; demotivated workers with low productivity; poor quality of care provided.</td>
<td>Average ratio of community health worker to population</td>
<td>McIntyre &amp; Meheus (2014)</td>
</tr>
<tr>
<td></td>
<td>Absenteeism rate for health workers</td>
<td>WB FSD, Tandon &amp; Cashin (2010); Heredia-Ortiz (2013); Okwero et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Health worker density in urban vs rural areas</td>
<td>Yip &amp; Hafez (2015); Lannes (2015)</td>
</tr>
<tr>
<td></td>
<td>Ratio of doctors to total health personnel or to nurses and midwives</td>
<td>Heredia-Ortiz (2013)</td>
</tr>
<tr>
<td></td>
<td>Density of physicians/nurses (per 1,000 population)</td>
<td>Heredia-Ortiz (2013); Lannes (2015)</td>
</tr>
<tr>
<td></td>
<td>Staff turnover or retention of health workforce</td>
<td>Dieleman &amp; Harnmeijer (2006); Lannes (2015); Meessen, Soucat &amp; Sekabaraga (2011)</td>
</tr>
</tbody>
</table>

#### Infrastructure (e.g. health facilities) and equipment
<table>
<thead>
<tr>
<th>Inappropriate health facility size, particularly hospitals, for optimal efficiency</th>
<th>Number of hospital beds per 100,000</th>
<th>OECD (2010); Smith &amp; Nguyen (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bed occupancy rate</td>
<td>WB FSD; Heredia-Ortiz (2013); WHO (2010)</td>
<td></td>
</tr>
<tr>
<td>Turnover rate for acute care beds</td>
<td>OECD (2010)</td>
<td></td>
</tr>
<tr>
<td>Wrong scale and scope of hospitals</td>
<td>Heredia-Ortiz (2013); WHO (2010)</td>
<td></td>
</tr>
<tr>
<td>Number of admissions, discharges</td>
<td>WHO (2010)</td>
<td></td>
</tr>
<tr>
<td>Under or over-capacity in health facilities</td>
<td>Number of consultations per doctor or number of outpatient visits or interventions provided per full time equivalent worker or per facility</td>
<td>OECD (2010); Dieleman &amp; Harnmeijer (2006)</td>
</tr>
<tr>
<td>Average PHC service utilization rate</td>
<td>Tandon &amp; Cashin (2010)</td>
<td></td>
</tr>
<tr>
<td>Equipment that is purchased and cannot be repaired or is not used optimally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inappropriate mix of inputs: e.g. health workers but no medicines or other medical products</td>
<td>Availability of essential medicines list or key medicines to practitioners</td>
<td>WHO/INRUD</td>
</tr>
</tbody>
</table>

### Spending Badly: Outputs and Outcomes

#### Health services

<table>
<thead>
<tr>
<th>Inappropriate and unnecessary care</th>
<th>Average length of stay (ALOS) for hospital visits or for specific admissions (following AMI, cancer)</th>
<th>WB FSD; OECD (2010); Australia National health performance reporting; Heredia-Ortiz (2013); Cylus, Papanicolas &amp; Smith (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative stay index (number of days spent in hospital for selected diagnostic-related groups (DRGs) divided by the expected number of days spent in hospital)</td>
<td>Davis et al. (2013); Australia National health performance reporting</td>
<td></td>
</tr>
<tr>
<td>Caesarean section rates</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>MRI/CT scan exams per 1,000 population</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>Average consultation time</td>
<td>WHO/INRUD</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Underuse</td>
<td>WHO/World Bank (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of infants that did not receive 3 doses of DTP3 immunization</td>
<td>WHO/World Bank (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of HIV positive adults and children who do not receive antiretroviral treatment</td>
<td>WHO/World Bank (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of pregnant women who do not receive 5 antenatal care visits</td>
<td>WHO/World Bank (2015); Yip and Hafez (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of people with hypertension treated and controlled</td>
<td>Glasziou et al. (2017)</td>
<td></td>
</tr>
<tr>
<td>Medical errors and low quality care</td>
<td>Unplanned readmissions</td>
<td>Davis et al. (2013); OECD (2017); Australia National health performance reporting</td>
</tr>
<tr>
<td>Proportion of HIV positive adults and children who do not receive antiretroviral treatment</td>
<td>WHO/World Bank (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of pregnant women who do not receive 5 antenatal care visits</td>
<td>WHO/World Bank (2015); Yip and Hafez (2015)</td>
<td></td>
</tr>
<tr>
<td>Proportion of people with hypertension treated and controlled</td>
<td>Glasziou et al. (2017)</td>
<td></td>
</tr>
<tr>
<td>30 day mortality rate (hospital or for specific conditions such as AMI or Ischemic stroke)</td>
<td>Davis et al. (2013); Hussey et al. (2004)</td>
<td></td>
</tr>
<tr>
<td>Rate of healthcare associated infections</td>
<td>Australia National health performance reporting</td>
<td></td>
</tr>
<tr>
<td>Postoperative pulmonary embolism (PE) or deep vein thrombosis (DVT) in hip and knee surgeries</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>Postoperative sepsis in abdominal surgeries</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>Obstetric trauma, vaginal delivery with (or without) instrument</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>Foreign body left in during procedure</td>
<td>OECD (2017)</td>
<td></td>
</tr>
<tr>
<td>Incidence rate for pertussis, measles, and Hepatitis B (vaccine-preventable diseases)</td>
<td>Hussey et al. (2004)</td>
<td></td>
</tr>
<tr>
<td>Maternal mortality, child mortality</td>
<td>Sajedinejad et al. (2014)</td>
<td></td>
</tr>
</tbody>
</table>

**Spending Badly: Health Financing and Health System Organization**

<table>
<thead>
<tr>
<th>Waste, corruption, fraud.</th>
<th>Health budget execution rates</th>
<th>WB FSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of medicines and others (incl. IV fluids) destroyed due to expiration, breakage and/or inappropriate storage conditions</td>
<td>WB FSD; Heredia-Ortiz (2013); Okwero et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>Degree of corruption (international benchmark not specific to health sector)</td>
<td>Tandon &amp; Cashin (2010)</td>
<td></td>
</tr>
<tr>
<td>Are health-specific anti-corruption policies in place?</td>
<td>WB FSD</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Informal payments in health, as % of the OOP expenditure</td>
<td>WB FSD</td>
<td></td>
</tr>
<tr>
<td>% of government health funding that reaches services delivery</td>
<td>Tandon and Cashin (2010)</td>
<td></td>
</tr>
</tbody>
</table>

**Inefficiency in raising revenues for health, particularly when revenue raising for health is independent from general government revenue collection.**

<table>
<thead>
<tr>
<th>Government revenue as % of GDP</th>
<th>Tandon and Cashin (2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of health expenditure with respect to GDP</td>
<td>Tandon and Cashin (2010)</td>
</tr>
<tr>
<td>Health budget as % of total government budget</td>
<td>Tandon and Cashin (2010)</td>
</tr>
<tr>
<td>International health assistance as % of total and government health spending</td>
<td>Tandon and Cashin (2010)</td>
</tr>
</tbody>
</table>

**Fragmentation in the system: in pooling, but in the broader health system**

<table>
<thead>
<tr>
<th>Are there adequate donor coordination mechanisms to align external financing with government priorities, processes, and the health budget?</th>
<th>WB FSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>How are providers paid and does the payment modality create incentives for cost containment, quality of service delivery or provision of services to specific groups of people?</td>
<td>WB FSD, McIntyre and Meheus (2014)</td>
</tr>
</tbody>
</table>

**Administrative inefficiency: higher than necessary costs for the services offered, including in health insurance agencies.**

<table>
<thead>
<tr>
<th>Health sector administrative cost, as % of government health expenditure</th>
<th>WB FSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>What type of budgeting process is used in the country, e.g., input-based or output-based, and how does this affect providers’ / purchasers’ ability to allocate resources to be in line with priorities?</td>
<td>WB FSD; Tandon and Cashin (2010)</td>
</tr>
</tbody>
</table>

The list is long, and certainly not exhaustive. For example, an indicator of the extent of underuse of needed services could be defined for all diseases and types of interventions focusing on that disease, such as screening for certain types of cancers, and subsequent treatment.

Some of the indicators suggested in the literature are also difficult to obtain, particularly in lower income countries, and would require additional expenditures to establish and maintain the systems to collect and analyse them routinely, for example: the extent of informal payments (surveys); share of public expenditures going to the poorest 40% of the population (utilization surveys with information on household expenditures or incomes); percentage of medicines prescribed according to national guidelines (surveys or observation of encounters); wrong scale and scope of hospitals (production function analysis based on intense data collection); and degree of corruption (public expenditure reviews or special forms of audit).

Others are difficult to interpret because a move in one direction does not unequivocally mean an increase or decrease in efficiency. They may still be useful, however, and international yardsticks derived from multi-country comparisons, particularly those of similar countries, could be used to make...
these judgements. Examples are health worker density, distribution of health workers to rural areas, average length of inpatient stay, and the share of national spending on primary care or medicines. Even though it is possible to determine if a country’s scores are higher or lower than that of similar countries, it is difficult to be sure if more or less is good or bad. The analysis is most useful when there are outliers – where the variable in question in one country is substantially higher or lower than in other countries.

Another set is readily easy to interpret and most countries should be able to measure them without great additional expense. From the variables in Table 2, twelve fall into this category. This does not mean that they are the most important to monitor or that the set of variables is the best to understand the overall efficiency of the system, but only that they should be able to be measured routinely in countries at all income levels. Examples include absenteeism rates for health workers, the proportion of pregnant women who do not receive 4 antenatal consultations, infant coverage with 3 dose DTP, post-operative sepsis in abdominal surgery, Caesarean section rates, budget execution rates and bed occupancy rates. Higher-income countries with established measurement systems can obviously monitor a much greater number and other countries will choose to develop the systems to monitor other indicators depending on the areas of inefficiency they choose to tackle.

The advantage of the micro-efficiency over the macro-efficiency approach is that the indicators are easier for policy-makers, health workers, patients and the population to understand. They are also directly relevant to the analysis of the most important causes of inefficiency in a country, which informs the policy debate about what to do about them.

The downside of using a dashboard of indicators is that it can be difficult to understand if a system is getting more or less efficient unless all indicators improve at the same time. Even then, it will not be clear if efforts to improve efficiency in one area result in lower efficiency somewhere else, in areas that are not being measured.

We observe that few countries have developed a specified set of indicators with which they track efficiency and improvements over time. It is important that they do so, perhaps using Table 2 as a starting point and taking into account their capacity to measure and analyse and the costs of obtaining each variable. There may well need to be different sets for different actors – e.g. a hospital manager may require a set of relatively specific indicators while the Minister of Health might require a broader set that summarizes efficiency across the entire health system, but includes less detail on the individual components. Without an appropriate set of indicators, it will be difficult to determine if the strategies to improve efficiency have been successful.

VII. Strategies to improve efficiency

Once the major sources of inefficiency have been identified, the next step is to develop and implement strategies for increasing efficiency. Understanding the reasons why inefficiencies exist can help to identify appropriate strategies for reducing them, so in this section we turn to available technical solutions. The list of possible solutions draws on policies commonly put in place in response to the reasons why inefficiencies persist. Remarkably, there is little known about what

---

16 The World Bank Group is in the process of trying to provide the range of values for many of these indicators so countries can determine where they are in relation to other countries.
works at the system level and the section concludes by identifying some of the key unresolved and controversial issues.

1) **Technical options for improving efficiency**

Table 3 builds on Table 1, suggesting reasons for each possible source of inefficiency, and solutions that have been identified in the literature based on considerable country experience. Solutions directly linked to health financing are highlighted in bold. These solutions have been drawn from an extensive review of the literature reporting on what types of interventions have been implemented with the intention of improving efficiency. The list is, however, unlikely to be comprehensive – the literature is vast and it is possible some options have been missed; some strategies might have been implemented at country level without a formal evaluation; and other evaluations might not have been published or made public. Readers are encouraged to indicate to the authors other options that should be included in any list that countries can use to identify possible solutions to their problems, particularly those for which there is good evidence that they work or do not work.

**Table 3: Policy Options for Improving Efficiency**

<table>
<thead>
<tr>
<th>Source of Inefficiency</th>
<th>Common Reasons for Inefficiency</th>
<th>Possible Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Doing the wrong things: High cost, low impact interventions funded at the expense of low cost, high impact; inappropriate mix between levels of care; prevention, promotion versus treatment; public health and governance functions versus personal services; mix of intersectoral actions or intersectoral versus health services</td>
<td>Difficult to obtain the necessary technical data to guide decisions; no clear yardsticks for deciding what is enough (prevention vs treatment; public health and governance vs personal services); political, health worker or community preferences (types of interventions, levels of care); financial interests (little prevention funded by insurance).</td>
<td>Increased country capacity to generate and use key epidemiologic information on burden of disease to guide decisions; further development of methods for assessment of integrated packages and for incorporation of financial protection; patient empanelment; gate keeping; increased service delivery capacity and quality at lower levels of care; health taxes and financial incentives for personal action on prevention; civil society organizations and professional associations promoting adequate provider behavior.</td>
</tr>
</tbody>
</table>

---

17 This Table builds on an approach taken in WHO (2010).
2. **Doing things in the wrong places:** Inappropriate use of higher level versus lower level settings for inpatient and outpatient care (including day surgery), overuse of health versus social care and institutionalized versus home-based long-term care.

   Insufficient information on the efficiency of delivery of packages at different levels; organizational separation of health and social care, weak coordination of care, including referral systems, poor quality at lower levels of care, financial incentives across care settings that promote care at higher levels.

   Methods to assess the efficiency of packages delivered at different levels of the system; coordination/integration of health and social care; patient empanelment and gatekeeping; clinical pathways; increased service delivery capacity and quality at lower levels of care, including IT innovations e.g., tele-medicine, e-consultations); provider management networks (e.g., primary care networks); appropriate blending of payment methods.

3. **Spending badly:** higher cost inputs chosen or inputs not achieving their maximum potential

<table>
<thead>
<tr>
<th>Health Service Delivery</th>
<th>Medicines:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.</strong> Medicines:</td>
<td></td>
</tr>
<tr>
<td><strong>a)</strong> under-utilization of generics or paying too much for any specific medicine;</td>
<td>No generics policy; provider/patient perceptions that generics are poor quality; financial incentives to prescribe branded medicines; poor purchasing practices or corruption; lack of knowledge of international prices; high mark-ups or taxes on medicines.</td>
</tr>
<tr>
<td><strong>b)</strong> use of ineffective medicines, the wrong medicines, or using them at the wrong time;</td>
<td>Inadequate regulation/administration to control substandard medicines; poor knowledge by providers; demand or low adherence from patients.</td>
</tr>
</tbody>
</table>

| Generics policy and essential medicines list for health facilities accompanied by quality controls; information on generics to providers/population with quality control system; financial incentives for prescribing generics and not branded medicines; active purchasing with appropriate competitive bidding; centralized procurement; multi-year procurement frameworks; increased transparency in purchases and tenders; zero taxes on essential medicines; monitoring and publication of medicine prices. |

<p>| Increased government capacity to regulate medicines to ensure safety and quality; information exchange for providers and the population. |</p>
<table>
<thead>
<tr>
<th>c) overuse or unnecessary use.</th>
<th>Financial incentives to prescribe and sell more; industry promotion; consumer demand.</th>
<th>Appropriate financial incentives – e.g. separate prescribing from sales; increase information to providers and patients; regulate and enforce standards for industry promotion.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Infrastructure (e.g. health facilities) and equipment:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) <strong>Inappropriate health facility size, particularly hospitals, for optimal efficiency;</strong></td>
<td>Lack of information about appropriate size; patients choose to go to higher levels of care.</td>
<td>Monitoring efficiency of health facilities; gatekeeping; increased service delivery capacity and quality at lower levels of care.</td>
</tr>
<tr>
<td>b) <strong>Under or over-capacity in health facilities;</strong></td>
<td>Too few facilities for the demand or maldistribution; patients choose higher level facilities – over and under capacity coexist; poor management; financial incentives for high admission and long length of stay.</td>
<td>Master plans for streamlining infrastructure; gatekeeping; increased service delivery capacity and quality at lower levels of care; appropriate blending and use of payment methods; improved management capacity with appropriate incentives.</td>
</tr>
<tr>
<td>c) <strong>Equipment that is purchased and cannot be repaired or is not used optimally.</strong></td>
<td>Donations of equipment that cannot be serviced locally or where supplies and maintenance are too expensive; poor procurement practices; corruption.</td>
<td>Refusal of donations of new technology where budgets will not be able to pay supplies and maintenance; improved donor practices; improve procurement practices and controls of corruption.</td>
</tr>
<tr>
<td>3. Health services:</td>
<td>Poor management &amp; control (perhaps linked to insufficient management resources or inadequate information on patterns); financial incentives to over-service; defensive medicine.</td>
<td>Improve management and availability and use of data; clinical guidelines; financial incentives to prevent overuse and to promote quality.</td>
</tr>
<tr>
<td>a) <strong>Unnecessary tests, procedures, or underutilization of these compared to need;</strong></td>
<td>Inadequate provider knowledge; insufficient data collection or use of data by managers; no incentives for quality; poor infrastructure; low quality including hygiene; poor compliance with infection</td>
<td>Continuous training for providers; improved data availability and use (e.g. clinical audits); clinical guidelines; incentives for quality (contracting, provider monitoring, payment systems,</td>
</tr>
<tr>
<td>4. Personnel:</td>
<td>Poor planning; inappropriate training intakes (can be linked to student demand); outmigration or lack of retention of some cadres; resistance by various cadres to less skilled people taking more responsibility.</td>
<td>Health workforce planning based on labour market assessment and links to training intakes; HRH training and recruitment aligned with broader health system's objectives; strategies and incentives to recruit and retain key health workers in remote and underserved areas; skill-task matching, including task shifting;</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>a) Inappropriate mix between different cadres;</td>
<td>Incentives insufficient for some locations;</td>
<td>Revise salary structures and incentives for underserved locations.</td>
</tr>
<tr>
<td>b) Located in the wrong places;</td>
<td>Poor wages and incentive structures; poor management and supervision; poor working conditions; recruitment and promotion not based on merit.</td>
<td>Salary and incentive structures in line with system objectives; regulation of dual practice; improved management, supervision and working conditions; multi-disciplinary teams, eliminate “cronyism” in hiring and promotion (establish clear process for hiring, deploying, and promotion of health personal).</td>
</tr>
<tr>
<td>c) Demotivated workers with low productivity and poor quality of services (e.g. low visits per health workers per day, absenteeism).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Inappropriate mix of inputs: e.g. health workers but no medicines or other medical products.</td>
<td>Poor management or budgeting practices; inflexible contracts with workers.</td>
<td>Improved management of inputs for service delivery; budget practices providing greater flexibility of use of inputs.</td>
</tr>
<tr>
<td>7. Waste (including expired medicines), corruption, fraud. <strong>Low budget execution rates</strong> are a form of waste because the available funds are not used.</td>
<td>Poor procurement, inventory control practices; poor storage and transport; lack of transparency and accountability with appropriate management, audit and legislation; inadequate supervision; poor management or inflexible line item budgets leading to low budget execution (can also related to delayed disbursement from ministries of finance).</td>
<td>Improved procurement, supply chain management, inventory control, storage of supplies of drugs; improved regulation and governance with sanctions for corruption and fraud; codes of conduct; improved public financial management including budget practices and resource tracking and accountability.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8. Inefficiency in raising revenues for health, particularly when revenue raising for health is independent from general government revenue collection.</td>
<td>Efficiency of raising general government revenues is beyond the health sector. Inefficiency in collecting charges levied by ministries of health or health insurance premiums – poor management, poor information systems, lack of motivation of staff to collect revenues, inability to enforce payment of contributions, inability to identify the indigent who do not pay, corruption.</td>
<td>Fee and premium systems enforceable at low cost, improved skills, management and information systems for tracking payments and service use particularly in national health insurance systems; organizational/staff incentives to collect fees/premiums; user-friendly ways for people to pay contributions; methods of identifying the indigent; corruption control.</td>
</tr>
<tr>
<td>9. Fragmentation in the system: in pooling, but in the broader health system as well - e.g. procurement, supply chains, laboratories, service delivery.</td>
<td>Donor practices (developing systems for: channelling and tracking funds; procurement and distribution; employment; service delivery; monitoring and evaluation) outside government structures; power structures in the ministry of health; responsibilities in a federal system (e.g. central government responsible for hospitals, lower level government for other services); pressure from the already insured to maintain their benefits when health insurance for the informal sector is developed.</td>
<td>Organizational or virtual integration as appropriate – e.g. standardized information, budgeting and accounting systems, planning, M&amp;E across fragmented units. Risk adjustment across pools.</td>
</tr>
</tbody>
</table>
10. **Administrative inefficiency:** higher than necessary costs for the services offered, including in health insurance agencies.  

|    | inflexible staff contracts; lack of incentives for efficiency; inadequate training and knowledge of managers and staff; high staff turnover. | contract systems; incentives for efficiency; improved knowledge and capacities of staff; policies and incentives to retain staff where high staff turnover. |

Most of the options considered in Table 3 involve making changes to existing strategies that either obtain more for the existing resources or ensure the same outcomes for fewer resources. They represent movement towards the frontier identified in Figure 1. There are other types of interventions that have the potential to shift the frontier by expanding the technical options for improving health and financial protection. Some examples are provided in Box 2, although the remainder of this section focuses on the options listed in Table 3.

**Box 2: Shifting the Frontier: Innovations in Health**

Alongside efforts to deliver health services, product developers and practitioners are advancing innovations to ensure that people in low-resource settings have access to new approaches for life-saving interventions that vastly expand current options. Many of these innovative technologies, systems, and services seek to provide affordable solutions specifically designed to address the needs of vulnerable populations around the world.

One necessity for the provision of quality healthcare is data management. Digital health information systems break down barriers that prevent technologies and systems from scaling, and enable support for platforms that can be reused, adapted, and built upon. For example, a national electronic immunization registry being developed in Tanzania and Zambia by the Better Immunization Data Initiative automatically sends immunization information to health care workers in advance of vaccination days with information regarding the number of children due for vaccines, which immunizations they need, and the volume of vaccine stock or supplies the clinic should have on hand. This system minimizes missed vaccines. Vaccines are among the most powerful lifesaving tools for children under five. With accessibility to better data, health workers are equipped to make better decisions about vaccine delivery and achieve higher immunization coverage rates.

Similarly, redesigning biomedical technologies for use in low-resource settings will accelerate lives saved and avert health care costs. For example, one innovation under development is better respiratory rate monitors and portable pulse oximeters to improve timely detection and treatment of pneumonia among children under five. Difficulties in diagnosing pneumonia among young children in low-resource settings often lead to unnecessary treatment delays and increased risk of death.

Innovations in diagnostic technologies can have a significant impact when better monitoring leads to expanded coverage. A recent study conducted by PATH, *Harnessing the power of innovation to save mothers and children*, modelled the impact of innovative pneumonia-detection technologies which are more reliable and easier to use than existing tools, estimating that just over one million lives could be saved (PATH 2016).  

---

18 PATH. IC2030: Harnessing the power of innovation to save mothers and children.  
2016. [http://www.path.org/publications/detail.php?i=2647](http://www.path.org/publications/detail.php?i=2647); Modelling was applied toward increased intervention coverage within the context of USAID’s Acting on the Call model.
Innovative approaches are even more important where no tool currently exists to expand coverage for existing products. Extending coverage for the Hepatitis B vaccine would save up to 6 million lives, for the haemophilus influenza type b vaccine would save up to 1.7 million lives, for the pneumococcus vaccine would save up to 1.8 million lives, for the rotavirus vaccine would save 900,000 lives, and for the human papillomavirus vaccine would save 500,000 lives over the next 10 years (WHO 2013). Additional benefits are longer and more productive lives, higher earnings, and averted health care costs, with a potential of up to $44 return across the lifespan of an immunized child for every $1 invested (Ozawa et al. 2016).

Prioritizing innovation within a UHC framework will enable countries to take advantage of a rich set of emerging tools that can enable dramatic change, including reduced mortality and morbidity, inclusive and sustainable growth through cost effectiveness, and impressive gains in health outcomes (Lancet Commission on Investing in Health 2013). Rapid adoption and scale of health innovations is critical to foster these gains, as is a focused strategy and further political commitment (Atun 2012). Through innovation, UHC can be delivered with the greatest health value for money, providing people the opportunity to lead healthier, more productive lives.

An important observation from Table 3 is that only a minority of the solutions are totally within the remit of health financing with its functions of revenue generation, pooling and purchasing/provision. Because this Forum is specifically on health financing, we begin with ways that the health financing function can contribute to improving efficiency. It is not intended to describe all the details of how to implement the various health financing strategies related to efficiency, but to outline the broad areas where there is agreement that the solutions work. Once countries choose the type of strategy they are interested in implementing, technical work would then need to be undertaken to design the specific details, building on a very large literature.

After the health financing options, the paper considers briefly some of the additional strategies, on the grounds that if a country is serious about reducing inefficiencies it will need to use a mix of measures that include, but are not exclusive to health financing.

One part of the efficiency of revenue generation is the cost of enforcement and administration, to government, households and firms, sometimes called administrative efficiency. A second part is the yield generated from a tax, or production efficiency. The third part is the costs of the economy of changes in behaviour and economic output resulting from a particular tax – for example, people working less because of high marginal tax rates on their income, typically called deadweight losses (e.g. Okafor 2012; Thompson, Beatty and Thompson 2012; Scott 2014). Finance departments, sometimes working with external agencies like the IMF, the World Bank and Regional Development Banks, constantly seek to reduce the costs of enforcement and administration and increase yields. Generalizations are that corruption increases administrative costs, and income based

Source: PATH

Taxes are difficult to collect when there is a large informal sector (e.g. Imam & Jacobs 2014). Indirect taxes yield more than income taxes on individuals in these settings. Thresholds at levels that exclude small enterprises avoid high costs of enforcement relative to yields and improve equity. Various types of reforms can increase yield or administrative efficiency including, for example, the establishment of semi-autonomous revenue collection agencies at national and/or sub-national levels (Von Haldenwang, Von Schiller and Garcia 2014).

This is rarely something that is considered by the health sector except when the collection of health insurance premiums or user-charges is part of its mandate. Drawing from the public finance literature, it is difficult to use the income tax system to collect health insurance premiums when there is a large informal sector, so other methods need to be devised. Various innovations from broader taxation could also be useful, including locating agencies that collect premiums closer to the people that will pay them, or licencing facilities that are well distributed in the country to collect them. However, the experience in Thailand suggests that the administrative costs of collection of the original 30 Baht co-payment from the informal sector were high and outweighed the yield, something that can apply also to collection of premiums from the informal sector (Limwattanon et al. 2011). For this as well as equity reasons, there is a broad consensus that health insurance coverage of basic services should be financed from general revenue, while contributions are a condition for eligibility to a wider set of health services.

The general consensus on pooling is that small pools are inefficient in terms of high administrative costs, and they can result in inequities if one part of the population is covered with more and better services and high financial protection than others (e.g. Rao et al. 2014; Meng et al. 2015; Kutzin, Yip and Cashin 2016). It is, however, sometimes difficult to merge existing pools for political reasons, and countries moving towards health insurance from scratch are best advised to avoid creating different pools covering different population groups with different benefits (Hanvoravongchai & Hsiao 2007; Knaul et al. 2012).

The question of how to purchase inputs and services is important to many of the efficiency problems in Table 3. There are four components. The first involves a choice of the appropriate mix of personal health interventions available in the right places, the mix between governance, public health and personal health services, and the appropriate mix of inter-sectoral and multi-sectoral actions in support of health sector strategies. Possible methods to assess the efficiency of the different options, and their weaknesses, were discussed earlier but many countries are in the process of assessing the appropriate mix of services to guarantee to everyone and where they should be provided. Redirecting resources from unnecessary and harmful overused services to low-cost high-impact services that are underused addresses two sources of inefficiency at the same time – over- and under-use – and negative lists, or low value lists of interventions that should not be used or will not be covered by payers are one of the possible options of doing this (Elshaug et al. 2017).

Reducing the proportion of patients who bypass lower level services to use more costly services at secondary or tertiary levels because of their perceived better quality is another part of ensuring that the needed services are provided in the right place. This requires not only having some form of gatekeeping, but also strengthening the quality of primary care services. Box 3 outlines some recent developments on support to countries seeking to develop their primary health care.
Box 3: Supporting Primary Health Care

The Primary Health Care Performance Initiative (PHCPI) is a global partnership to help countries build high-performing primary health care systems in low- and middle-income countries. The PHCPI partners—the Bill & Melinda Gates Foundation, the World Bank, the World Health Organization, Ariadne Labs, and Results for Development—work to support countries to achieve the health SDGs through collecting better primary health care data, unearthing and sharing best practices, and deploying data and evidence to make improvements in the quality, effectiveness and efficiency of primary care services.

High-performing primary health care systems are critical to ensuring that the right things are done in the right settings:

- Primary care is the front line of health – close to people and delivering essential health services like vaccinations, maternal and newborn care, and family planning;
- Good primary care helps patients manage chronic diseases, avoiding unnecessary hospitalizations and care and time away from family and work;
- Primary health workers form the early warning system for detecting and stopping disease outbreaks before they become deadly epidemics – this is the first line of defense against epidemics;
- When the right things are done in the right way, primary health care systems can cover the large majority of health needs in ways that are responsive, safe, of quality and trusted by the people they serve.

High-performing primary health care systems are also the backbone of efficient health systems. In both high and low income settings, it has been shown that strong primary health care prevents many illnesses and catches others early with effective low-cost treatment, thereby keeping people out of hospital and reducing subsequent treatment costs (e.g. Kruk et al. 2010; Kringos et al. 2013).

Effective service delivery means that patients receive the right preventive care or treatment, at the right time, in the right place, and with respect. Doing this requires attention to five PHC systems features (adapted from Starfield 1992):

- First-contact access: Primary health care systems should serve as the entry point into the health care system, where people can access affordable care for most health needs.
- Comprehensiveness: Primary health care systems should deliver a broad spectrum of preventative, promotive, curative and palliative care – for example, through multidisciplinary teams that contain health professionals with varied, complementary skills.
- Continuity: Primary health care systems should support long-term patient-provider relationships – allowing providers to care for patients at every stage of life.
- Coordination: Primary health care systems should coordinate an individual’s journey through complex health systems.
- Person-centeredness: Primary health care systems should be oriented around the needs of people and communities.

The PHCPI repository of promising practices explains how leading low and middle income countries put these five features into practice to improve the quality and efficiency of their primary health care systems: [http://phcperformanceinitiative.org/tools/promising-practices](http://phcperformanceinitiative.org/tools/promising-practices)
The second component involves developing ways of paying for the chosen services (Honda 2014; Maeda et al. 2014; Tangcharoensathien et al. 2015; Xu, Cheng and Colón-Ramos 2015; Bastani et al. 2016; Kutzin, Yip and Cashin 2016). Provider payment has been shown to be a powerful tool to influence the behaviour of providers, and in turn the provision of care. In general, the evidence shows that no single provider payment method delivers value for money in all settings. For instance, capitation encourages efficiency but may lead to under-provision of services and increased referrals. On the other hand, fee-for-service can lead to over-provision of services and can lead to cost escalation. There has been also been growing interest in performance-based financing in health in recent years; although the more recent evidence on quality of care and efficiency is scattered – see the next section of this paper on what we do not know.

Countries should decide which are the most appropriate blends of payment methods use within each care setting, and how these payment method blends align across care settings to achieve its own policy objectives with regards to provision of care, cost and quality, (Cashin et al., 2009). “Blending” different payment methods within care settings is done to encourage certain desired outcomes as well as to mitigate the negative incentives of individual payment methods (Belli and Hammer 1999; Dranove and Satterthwaite 2000). For example, in primary care, capitation and FFS payments are often combined. The FFS payments are used primarily to encourage the provision of certain priority services (e.g. vaccinations), services and procedures which require costly supplies (e.g. injectable medicines), as well as services which lie on the border between primary and specialist care (e.g. wound care, drainage of abscesses, removal of benign lesions) and thus tend to be referred (Robinson 2001). In combination with capitation payments, primary care providers are still encouraged to limit the volume of services provided to achieve cost savings (Langenbrunner and Wiley 2002; Langenbrunner and Somanathan, 2011). Similarly, the blends of payment methods across care settings should be taken into account to ensure that incentives are aligned with health system objectives (e.g. encouraging better management of patients at the primary care level, reducing incentives to increase volumes in outpatient specialist and acute inpatient care, increasing incentives for coordination with primary care after discharge, etc.).

The third component involves strengthening contracting and provider monitoring capacities of the purchaser in order to enforce the intended behaviour changes of new payment methods (Cashin et al., 2009). Provider contracts can help contribute to health system objectives by taking advantage of provisions ranging from ethical codes and data disclosure requirements to cost and volume caps and risk sharing mechanisms to reduce transaction costs and shape provider behaviors. In addition, purchasers can establish and use transparent criteria from whom to contract, for example, accreditation and/or minimum volume requirements to foster quality of care. Provider monitoring enables continuous quality improvement by highlighting areas of care delivery that need optimization, and by setting common targets for these areas. When employing payment methods that encourage cost reduction (e.g. capitation, bundled payments, etc.), use of quality monitoring, in addition to risk adjustment, is essential to ensure that providers are not cutting costs in ways that jeopardize patient health. Provider monitoring can include activities such as clinical audits, regular collection of data on specific quality indicators (e.g. adherence to clinical guidelines, adverse events, patient satisfaction). Based on the monitoring results, purchasers can hold providers accountable accordingly by rewarding higher-level performers and/or sanctioning lower-level performers.

Finally, the fourth component involves empowering patients to hold purchasers and providers accountable through formal representation of consumers in purchasing organizations.

---

23 This is sometimes called strategic or active purchasing.
development of packages of care with formal coverage guarantees, establishment of patient’s rights legislation, charters or ethical codes and the development of a formal mechanism to receive and respond to patient complaints (Busse et al. 2007).

Improved public financial management (PFM) would contribute to reducing many of the forms of inefficiency identified in Table 3. These include reducing the funds returned to the treasury because they have not been used in the financial year, helping to control corruption, fraud and other types of waste, and linking public expenditures more closely to results (Fritz, Sweet and Verhoeven 2014; Barroy, Sparkes and Dale 2016; WHO 2016; Cashin et al. 2017). Despite the considerable attention given to assessing PFM systems over the last decade, it is not clear that reforms have been universally successful (Hepworth 2015). This can be linked to the political economy of reforms, something that is taken up in the next section.

We use medicines to illustrate some of the other strategies from Table 3 that can complement health financing changes to achieve more with the available resources. Many strategies have been shown to increase efficiency in a variety of settings ranging from establishing and enforcing an essential medicines list which uses generics to the maximum extent, increasing the transparency and information available for medicine procurement so that countries pay the right price and corruption is eliminated, developing the capacity to monitor medicine safety and quality not just when medicines are licenced or registered but once they enter the distribution system, eliminating incentives for over-prescription such as separating prescribing from sales, and various types of strategies to improve provider and patient knowledge about generics, appropriate and timely prescriptions, and the importance of adherence (e.g. Holloway 2011; Holloway et al. 2013; Atav et al. 2014; Chen et al. 2014; Hassali et al. 2014; Hurley 2014; Choudhry et al. 2016). It is, however, clear that it is simpler to implement these policies in the public than in the private sector where regulations sometimes do not exist and where they do, enforcement is difficult (e.g. Van Nguyen et al. 2013; Sheikh & Uplekar 2016).

There is simply too much literature describing attempts to improve the efficiency in the other components of health systems to report it all here but a recent WHO publication suggests that, in addition to considering health financing, it is useful to work through the health system functions of generation of human and physical inputs, governance/stewardship and service delivery (Sparkes, Durán and Kutzin, J. 2017). They then suggest looking for inefficiencies common to each function and inefficiencies that prevent the integration of activities across functions. We turn now to discussing some of the areas on which there is no consensus and where further evidence and discussion at this Forum would be valuable.

2) What we do not know

Some recent developments that may improve efficiency in health are still in their infancy. For example, the question of whether electronic medical records improve efficiency, and if so in what circumstances, is still being explored (e.g. Nguyen, Bellucci and Nguyen 2014; Campanella et al. 2015). This type of question in not discussed here where the focus is on areas of disagreement or where there is simply not enough information on long standing questions to make informed decisions.

i. Paying for results

Many experiments have been undertaken with forms of results-based payments that supplement incomes for staff or institutions in return for a set of agreed outputs (e.g. Rudasingwa, Soeters and Bossuyt 2015; Das, Gopalan and Chandramohan, 2016; Spisak et al. 2016). Similarly,
there have been recent experiments with forms of value-based payment\textsuperscript{24} as a replacement for fee-for-service in the US with the broader goal of improving continuity of care and provider coordination as well as reducing the growth of costs (e.g. Conrad et al. 2014; Damberg et al. 2014; Carey 2015; Press, Rajkumar and Conway 2016). An alternative is called population-based payment. Bundled and population-based payments are relatively new, but the preliminary results are promising, although they require the ability to design and implement a risk adjustment system and to monitor results. A brief summary is provided in Box 4.

**Box 4: New Options for Strengthening Provider Payment Mechanisms for Care Integration**

Two new payment mechanisms are currently being piloted, mostly in high-income countries, to improve integration of care. They are (i) bundled payments for acute episodes of care and chronic conditions and (ii) population-based payments (PBPs) covering specific services for defined population groups. These payment methods span across care settings to increase incentives for provider coordination.

**Bundled payments for acute episodes and chronic conditions.** Bundled payments involve a single, fixed rate paid to two or more providers to cover all services delivered for: (i) treatment of an acute episode of care, or (ii) management of the care for patients with a specific chronic condition or disease (American Medical Association, 2016). Bundled payments for acute episodes of care have been piloted in the US and Europe since the mid-2000s. Examples include Best Practice Tariffs (BPTs) in the UK, covering admissions for hip fracture, stroke, cholecystectomy and cataract surgery, and Ortho Choice bundled payments in Sweden, covering orthopedic procedures including hip, knee and spine surgery (Srivasta et al. 2016). These payments typically cover the costs of all inpatient and outpatient services from the initial visit through treatment, recovery and rehabilitation, including post-discharge care and any complications that may result within a certain time period after discharge (Porter and Kaplan 2015).

Bundled payments for chronic conditions have been implemented to help improve coordination of care for these patients and aim to encourage a holistic, long-term perspective rather than one-off encounters or interventions (Srivasta et al. 2016). Pilots of these bundled payments in European countries have focused on both rare, yet high-cost diseases, and on more common chronic conditions. The bundles typically cover all services related to the management of the disease or condition, in line with clinical guidelines and pathways, and can span more than one care setting. For example, under the Dutch bundled payment scheme, services included regular primary care check-ups, additional consultations, imaging, lab tests, examinations (e.g. foot examinations), counselling, medications, psychosocial care, and coordination of specialist services. The costs of complications from these diseases are typically covered outside of the bundle (Porter and Kaplan 2015). The payment is typically time-based (per month or year) since, unlike acute care episodes, the cycle of care for a chronic condition covered by a bundled payment may continue indefinitely.

Bundled payments for acute care episodes have been shown to achieve significant costs savings, with no or little deterioration of quality. The impacts of bundled payments on chronic conditions are less clear. In Portugal costs for treating HIV/AIDS decreased while quality of care was maintained (e.g., patient adherence to medication, controlled infection levels, compliance of providers with treatment guidelines, etc.). However, while quality improvements were observed

\textsuperscript{24} Value-based purchasing “refers to a broad set of performance-based payment strategies that link financial incentives to health care providers’ performance on a set of defined measures in an effort to achieve better value” (Damberg et al. 2014).
for both diabetes care process and patient outcome indicators in the Netherlands, costs increased. Reasons for this increase are unclear, but may be due to delaying the use of specialist care, which could have resulted in more costly care.

**Population-based-payments.** Population-based payment (PBP) models involve setting a prospective benchmark budget, which forms the basis for payments to groups or networks of providers for the provision of all or the vast majority of services for a defined population. These models have been closely related to the emergence of Accountable Care Organizations (ACO) - networks of health care providers that are collectively accountable for the organization, costs and quality of health care for their members – in the United States and elsewhere. Similar to bundled payments, providers are permitted to keep at least portion of the savings generated below the benchmark budget (contingent on achieving specified quality targets) and may be responsible for any costs exceeding the total PBP amount. Because PBPs cover a range of services across providers and are not linked to specific care episodes or conditions, they are thought to promote greater integration of care, a more holistic view of population well-being as well as incentives for innovations to keep costs down (e.g., risk stratified case management, discharge planning, preventive activities, etc.).

The largest ACO pilots are in the US, part of broader reforms mandated by the Affordable Care Act of 2010. For example, Medicare has contracts with over 400 ACOs. Providers forming the ACO typically include primary care providers and hospitals, but can also include specialists, long-term care facilities and home care (Srivasta et al. 2016). In Europe, examples of PBP models for ACOs have been implemented in Germany and Spain.

The range of health services providers are financially responsible for varies across the available models. In the Medicare model, ACOs are financially responsible costs which include inpatient hospital care, skilled nursing care, hospice and home health services as well as hospital outpatient care and doctors’ services. In Germany, providers are responsible for all health care costs for the insured population with the exception of long-term care (Srivasta et al. 2016). The size of the population that is assigned to an ACO also varies widely, ranging from 5000 to 245,000 patients.

PBP implementation is relatively recent and any impacts observed on quality and cost should be considered preliminary. A subset of the Medicare ACOs, known as Pioneer ACOs, were able to achieve improvements in 28 of the 33 required quality indicators including controlling high blood pressure, screening for future fall risk and screening for tobacco use and cessation, for example. On an aggregate level, Medicare ACOs contributed to slowing the growth in health spending, though not all were able to achieve cost savings. In 2012, total net savings for Medicare in amounted to $383 million while the GK model achieved savings of EUR4.6 million (Srivasta et al. 2016).

Source: World Bank, Forthcoming

In terms of the experience with results based payments or payment for performance in lower-income countries, there have been mixed results. Sometimes they have the desired impact and sometimes they do not. They can work for some of the desired outcomes but not for others (e.g. Fox et al. 2013; Mills 2014; Binyaruka et al. 2015; Das, Gopalan and Chandramohan 2016). There is undoubted evidence that they have stimulated staff motivation, quality and efficiency at times, and the more intensive information systems they require have also benefited the wider health system.
(Cashin et al., 2014). However, systematic quantitative assessment or meta-analysis across the experiments is complicated by the fact that each experiment has been different in the type of health workers they target, in the outputs for which payment is triggered, or in the way people are paid, so data strictly should not be pooled without some way of controlling for this heterogeneity (Markovitz and Ryan 2017). It is also not clear the extent to which these experiments are sustainable, as many of them have been funded by development partners, at times independent of a country’s financing system and only with emerging evidence of the program costs (De Bruin et al., 2011; Eijkenaar, 2013; Cashin et al., 2014).

Further work is needed to before it is possible to be sure what payment mechanism, or mix of payment mechanisms, best encourages high staff motivation, quality, and efficiency. To contribute to this, the World Bank Group is currently in the process of summarizing their experiences with forms of results-based financing.

**ii. The private sector and efficiency**

There has been considerable recent interest in the appropriate role of the private sector in health and many organizations have developed strategies for engaging with the private sector (IFC 2011; USAID 2009; World Bank 2013). Systematic reviews of the extensive literature suggest, however, that there is no evidence that either public or private health services are inherently more efficient or of higher quality than the other (Coarasa et al. forthcoming; Berendes et al. 2011; Basu et al. 2012). The key issues and challenges in service delivery, such as substandard patient safety and quality of care, excessive reliance on hospitals and inadequate infrastructure, are common across the public and private sectors. Moreover, deficient safety and quality has less to do with the ownership of the provider than with the incentives faced by the provider. This means that in seeking to expand service coverage, there is no reason to argue that more public, or more private, is the preferred option.

Opportunities do exist to engage the public sector at both hospital and primary care levels and examples can be found of public-private partnerships improving coverage and quality with hospitals. At the primary care level, engagement with the private sector has included contracting, provider networking, implementation of voucher schemes, and inclusion of informal providers in the formal delivery system (including programmes aiming at retraining and formalizing their status) (Montagu & Goodman 2016). However, while there seems to be a wide consensus that more efforts should be undertaken to establish a dialogue and relationship between government and private providers, there does not yet seem to be agreement on how best to do it.

There is more agreement that some innovations in the private sector have benefited the public sector. These include innovations in: business process functions such as marketing, financing, and operating; promotion of health services to the poor through social marketing and service design; redesigning cost structures that allow products and services to be more affordable for the poor, by lowering operating costs through simplifying medical services, lowering unit costs through higher volumes and cross-subsidization; and new operating strategies that increase the availability of services in remote areas, mainly achieved through optimizing human resources, process and product reengineering, and increasing outreach activities (Bhattacharyya, et al., 2010).

**iii. Human resource strategies for efficiency**

Considerable work has been done on human resource strategies in addition to how to pay providers. This includes developing strategies and incentives for: continuous training or retaining where necessary; retaining staff; ensuring there are sufficient staff with the right skills in isolated and disadvantaged areas; and trying to control harmful effects of dual practice (e.g. Asante et al.
Salaries and financial incentives are not the only issue, but they are important in most settings. In low income countries, the key issue is how to address the wide range of issues that would reduce the health worker shortage, improve skill mixes and ensure motivated workers are at work and located where they are needed for the available resources. Where Ministries of Health (and Education) are have limited resources, where should they start? This requires an assessment of where the biggest impact would be obtained for given levels of expenditure, but this information is not easily available – using dual practice as an example, there have been many attempts to control the problems associated with public sector health providers working also in the private sector, but there are few generalizable lessons of what strategies have had sustainable success (e.g. Sandier and Polton 2004).

iv. Costs of improving efficiency

Many studies have evaluated the impact of attempts to improve efficiency in the health system, some reported here and some in the Annex to this document. Rarely are the costs reported. Where they are, information is relatively sparse – for example, Conrad et al. (2014) reported that there are substantial transaction costs involved in introducing value-based payments in the USA including those linked to changes in the computer payment systems (Conrad et al. 2014). There is scattered information that transaction costs might be high for results-based financing in lower income countries, to the extent that Borgi et al. (2015) questioned if the effects justify the costs. All strategies to improve efficiency involve transaction costs, and this information is key for decision makers seeking to allocate scarce resources to inter-sectoral actions, improve the quality and range of services available and increase financial protection – and improve efficiency at the same time.

v. Political economy issues

The health sector is complex, shaped by powerful interest groups and many interests in both the public and private sectors interact and collide on a daily basis (Daemmrich 2013). Efficiency reforms challenge the status quo in the provision of health services, and in their financing and/or organisation, so they naturally trigger broader political, economic and ethical concerns (Roberts et al. 2004). Efficiency reforms cannot, therefore, be viewed only from a technical side and the political economy around their possible success or failure also needs to be understood. Perhaps this is why Fox and Reich (2015) argue that successful reforms are the exception rather than the rule in health.

Formal models of political economy are now being applied to health reforms more frequently to shed light on these processes: for example, models of competing interest groups and voter models (see Hauck & Smith, 2015); political settlement analysis (Kelsall, Hart & Laws, 2016); and stakeholder analysis (Bump et al. 2014). Reich et al. (2016) argue that mapping out the veto points and veto players is key to any reforms – veto points are ‘junctures in the legislative and policy design where reforms can be blocked’, whereas veto players are the ‘individuals or collective actors whose agreement is required to make policy decisions’.

Several studies have shown the importance of ‘events’ (e.g. political or economic crises, wars, or natural disasters) in triggering a reform process. Those events are crucial as they tend to challenge the power of interest groups and collective actors within the health system and in some cases, they lead to a breaking point where the system is no longer deemed appropriate. For instance, Reich et al.
(2016) find that the collapse of communism led to a political discussion on social protection systems and design of major health reforms promoting the development of primary health care and the implementation of essential benefit packages. In France, the United Kingdom and Japan, the development of universal coverage financing system was integrated as part of the post-war reconstruction process (Stuckler et al., 2010; Reich et al., 2016). In Thailand, the 1977 Asian financial crisis was the stimulus to develop the UHC Scheme that extended coverage with health services and financial protection to the entire population (Patcharanarumol et al. 2011), while the 2008 financial crisis was the stimulus for a series of reforms including those that increased efficiency in the health system (Hou et al. 2013).

Studies from Kyrgyzstan and Mexico show how health reform became a political agenda that obtained the support of the population, enabling health financing changes to be pursued despite opposition from some interest groups (Kelsall, Hart & Laws 2016; Frenk et al. 2006; Parry and Humphreys 2009). Bump et al. (2014) trace out the opponents and the supporters of recent health financing reforms in Turkey and active steps were taken to build support and reduce the opponents’ support base while Harris (2015) points to the importance of a group of civil servants in Thailand who built support and combatted opposition to the introduction of their universal coverage scheme.

Beyond this, there is little knowledge of what factors allow efficiency reforms to succeed and fail, and how governments strengthened support and overcame opposition. It is not yet clear if there are generalized lessons from experience that would help countries learn from the experience of others.

That being said, it is clear that there can be political obstacles to implementing even the reforms that technically seem the simplest. Different types of hospital reforms to improve efficiency can be opposed by managers, clinical staff, non-clinical staff, or the community (e.g. Galetto, Marginson and Spieser 2014). Closing hospitals or reducing hospitals beds in the face of over-capacity is likely to be opposed by politicians in whose electorate the hospitals are located, by the affected community, and by employees (e.g. Bloom et al. 2015). An important requirement for the successful implementation of efficiency reforms is to undertake a form of political mapping in the design phase, to understand who is likely to support or oppose the reform. Strategies for dealing with opposition, including engagement with key stakeholders, then need to be developed and implemented before, or at the same time, as the reforms themselves.

VIII. Efficiency and equity

Reducing inefficiency is a means of moving more rapidly towards UHC for the available resources, thereby improving population health and financial wellbeing. Efficiency analysis does not typically account for the distribution of coverage and outcomes across population groups but considers aggregate outcomes at the population level. Improving population health and financial protection are an important goals of health system development, but reducing inequity is also critical and there is a large literature on the nature of inequalities and inequities in health and how to improve them (e.g. WHO 2000; de Andrade et al. 2015; Mackenbach et al. 2015; Marmot 2015).

The question of whether there is an efficiency-equity trade-off in health policy has also been widely discussed in the last two decades: drawing on this literature the World Health Report of 2000 stated that ‘equity and efficiency can be easily in conflict’ (WHO, 2000). For instance, some interventions directed specifically to vulnerable groups might be more costly to implement if those groups are located in a remote location or present demographic, cultural or socio-economic constraints. Ebong & Levy (2011) compared the efficiency of facility-based and outreach programs in a universal immunisation campaign in Cameroon. The former was more cost-effective than the latter.
However, although the outreach program to increase population coverage was more costly, it was more effective at reaching vulnerable groups. In the same vein, two studies suggest showed that the most efficient programs for HIV/AIDS in South Africa were not the most equitable (Cleary, Mooney & McIntyre 2010; Verguet 2013). Targeting the poor and the vulnerable came at a cost in terms of the overall population impact.

On the other hand, some strategies that improve efficiency can also have positive distributional impacts. Strengthening investment in primary care is the most obvious example. Primary care interventions are often the most cost-effective way of reducing the burden of disease in low- and middle-income countries compared to other levels of care, and primary care tends to disproportionally benefit the poor (Jamison et al., 2006; Asante et al., 2016). This does not always mean that the rich do not benefit more than the poor from PHC, but that they poor benefit relatively more from PHC than from higher-level services.

Some evaluations of training community health workers and various forms of task shifting have also been suggested to reduce delivery costs, increase coverage levels and benefit the poor and people living in remote areas. Zachariah et al. (2009), for example, assessed three programs of task-shifting to provide ART and treatment monitoring (in Lesotho, South Africa and Malawi), showing a reduction of costs, increased access and improvements in geographic and socio-economic equity.

Benefit-incidence and financial-incidence analysis have increasingly been used to assess which groups benefit the most from a particular policy or financing flow – e.g. a subsidy or voucher scheme or a particular level of care (e.g. Asante et al. 2014; Chen et al. 2015; Asante et al. 2016). While this information is interesting, it is not particularly informative about identifying the most efficient ways of reducing inequity.

An illustration is the relatively large literature on targeting particular vulnerable groups versus a more universal approach. Most of the studies of targeting have focused on the impact, and whether the poor benefit as intended. Sometimes the poor do not benefit in which case there is no need of further analysis (e.g. Coady, Grosh & Hoddinott, 2004). But where the targeted group benefits, questions of costs and efficiency become paramount.

There are many costs involved in targeting - administration of targeting schemes, continuous updating of tools for the identification of the poor, fraud control and resource transfer costs (Coady, Grosh & Hoddinott 2004; Dutrey 2007). These are very rarely reported even in studies that show impact. The evidence on the merits of alternative ways to reduce inequalities is an important missing ingredient to evidence-based policy development in this area.

Finally, equity considerations have influenced the design of health financing strategies over the last decades. Received wisdom in health insurance has been that it works most efficiently when insurance is offered for low frequency, high cost occurrences. People with chronic illness and people living close to the poverty line can suffer financial catastrophe or be pushed into poverty from high frequency, low cost events. Protecting these people from severe financial hardship linked to paying out of pocket for health services is an equity issue at the foundation of the concept of UHC.

**IX. Conclusions**

Inefficiency can be found in the health systems of all countries. In low and middle income countries, it imposes substantial costs in terms of slowing the rate at which they can move towards UHC and improve the health of their populations. In high-income settings it can mean that services are cut or out-of-pocket payments are increased unnecessarily at a time of financial constraint. Achieving more health and financial protection with the available resources is, therefore, an important
complement to efforts to raise the necessary resources for health. Proof that the health sector is getting more efficient may also persuade the Ministry of Finance to allocate more funds to health.

Some country health systems are more efficient than others, but not all suffer from the same types of inefficiency. It is also possible to be relatively efficient in some areas and less efficient in others but every country could do something to ensure they achieve more with the available health resources. A first step is to identify the main sources of inefficiency, the reasons why they persist, and which ones are amenable to technical solutions that are politically feasible to implement. Political analysis to understand the likely opponents and supporters of any particular reform, and to subsequently build support and counteract opposition, is a critical step towards maximizing the probability of success with any reforms that area developed.

Some of the technical solutions that can reduce inefficiency are well known and accepted. Within the health financing system, for example, the need to reduce fragmentation in pooling by merging pools or by starting with a single large pool is one. So are the ideas of choosing the mix of health services that is the most efficient, delivered in the right places; introducing payment mechanisms and systems that encourage quality, efficiency and results based on reliable up to date information; and optimizing public financial management practices.

Outside the financing system, some of the possible solutions related to medicines are well documented although there are also important steps that have been shown to be effective with health workers and infrastructure. These include shifting tasks from doctors where they are in short supply to other types of health workers so as to increase coverage without additional cost.

The solutions of Table 3 were organized according to three main policy questions they seek to address: doing the right things; doing them in the right places; and doing them right. Governments will want to know which solutions offer the greatest chances of an immediate benefit and which ones will reap longer-term benefits. Table 4 summarizes those that are the most likely to deliver short-term efficiency returns.

**Table 4: Possible Quick Wins**

- **Organization and management of care**
  - Establish management networks for (primary) care providers;
  - Establish and enforce gatekeeping at PHC-level;
  - Improve service delivery capacity/quality at lower levels of care (e.g., PHC), including telemedicine.

- **Tax policies:**
  - Implement health taxes and financial incentives for personal action on prevention;
  - Remove taxes on essential medicines.

- **Public Financial Management**
  - Modify budget practices as necessary – e.g. move from line item budgets to more flexible budgeting; timely release of funds;
  - Improve controls to prevent corruption;
  - Improve regulation and governance with sanctions for corruption and fraud;
  - Promote codes of conduct.

- **Payment systems**
  - Develop more user-friendly ways for people to pay contributions.
Human Resources for Health
- Improve health workforce planning with links it to training intakes;
- Improve management, supervision and working conditions;
- Allow health workers at lower levels to take on more responsibility as appropriate (task shifting).

Medicines
- Develop and enforce generics policy and essential medicines list for health facilities with quality control which might include:
  - Limiting financial incentives for prescribing branded medicines
  - Information on generics to providers/population with quality control systems;
  - Remove inappropriate financial incentives – e.g. separate prescribing from sales;
  - Increase information to providers and patients; regulate and enforce standards for industry promotion.
- Active purchasing of medicines with appropriate competitive bidding;
- Increase transparency in purchases and tenders;
- Monitor and publish medicine prices.

Information systems
- Increase country capacity to generate and use necessary information;
- Improve management and availability and use of data.

Infrastructure
- Improve procurement processes for infrastructure;
- Refuse donations where local service cannot be assured or where budgets will not be able to pay for spare parts;
- Ensure appropriate maintenance and cleanliness.

Political economy
- Information sharing with key stakeholders on reasons for efficiency actions.

The politics of efficiency improvements might mean that some of these strategies could take considerably longer than expected in some countries, while others identified in Table 3 might be feasible to implement more quickly. Focusing on quick wins should not, of course, divert attention from some of the key longer-term options such as changing provider payment mechanisms and introducing more comprehensive forms of strategic purchasing. The longer-term options also sometimes will require immediate investments to ensure that they can have an impact. For example, strategic purchasing generally requires strengthening capacities in terms of personal skills and computerized information systems. Changing provider payment mechanisms will frequently require new or modified legislation and a period of consultation with both patients and providers.

Despite the knowledge about technical solutions in some key areas, some important gaps in knowledge and a number of controversies remain:

A. Mix of services

There is a considerable body of knowledge about the costs and effects of different sorts of health services that can be useful for countries seeking to choose the best mix. Despite this, the choice
of a benefits package still relies as much on an assessment of what is feasible given the needs of the population as on evidence of the efficiency of different intervention mixes. The areas where additional information is critical include:

a) **Economies of scale and scope**: Which interventions should be delivered in the same setting with the same staff (scope) and what is the most efficient level of coverage for the available resources (scale)?

b) **The right services in the right place**: What is the efficiency of different services delivered in different places and how to decide what is the appropriate mix at community, primary, secondary and tertiary levels? How to ensure continuity of care across the different levels and across the life-cycle, including the balance between prevention, promotion, treatment, rehabilitation and palliation?

c) **Governance, public health functions and personal services**: What is the efficient balance between these activities and how much should be spent on them?

d) **Expanding coverage with health services, improving quality, and financial protection**: How can the efficiency of these alternatives be assessed and compared in a way that policy-makers can incorporate into their policy choices?

e) **Inter-sectoral or multi-sectoral actions**: What guidance can be offered to ministries of health about where their limited time and money should be focused if they want to influence other sectors to take actions that improve health?

**B. Incentives for efficiency and quality**

a) Are there forms of value-based payment that better encourage efficiency, quality and results than existing payment systems. How much do they cost, do they have unintended negative consequences and could they be routinely incorporated into health systems without external financing? This is not simply an unresolved question but a controversial issue where there are strong proponents and opponents, for example, of different forms of results-based payment.

b) What capacities do countries need to have to begin to purchase strategically and what systems (for example, budgeting, accounting and legal systems) need to be in place to support it? What has been the experience where it a country has sought to move towards strategic purchasing without the requisite systems in place and is it feasible to move forward in very resource-constrained settings?

**C. Measurement**

There are major gaps in the availability of data to identify the extent and sources of inefficiency in countries, particularly lower-income countries, partly because few countries routinely monitor their own progress in reducing inefficiency and partly because routine health information systems produce only a few of the necessary indicators.

a) How can countries determine which indicators are critical to their own efforts to achieve more with the available resources and what are the investment costs of ensuring they can be monitored regularly?

**D. Political economy**

a) Are there generalizable lessons to be learned about why some efficiency reforms work and others fail, or reforms work in one country and not in another?

b) Are there approaches that would be useful for all countries to take when developing the agenda for improving efficiency to maximize the chances of success, such as involving civil society or parliamentarians?
E. The efficiency and sustainability of options to improve efficiency

Little information is available on the costs of the different options for improving efficiency to include alongside the benefits to allow an assessment of the most efficient way of improving efficiency to be determined. Studies of payment for performance provide one example, and most of the literature has simple reported results.

a) How can countries rapidly assess the costs of the various options for improving efficiency? Is there any way the global community can assist?

F. Equity and efficiency

Some types of efficiency improvements can improve equity and some can exacerbate inequities. Sometimes it depends on how the intervention is delivered rather than on the disease or health condition targeted or the tool being used. What is rarely assessed are the costs and effects of different ways of improving equity – recognizing that improving equity is a legitimate goal of government, and they should do it in the most efficient way.

b) How can countries rapidly assess the costs and impacts of the various options for reducing inequalities in health? Is there any way the global community can assist?

X. Recommendations

Every country has inefficiencies in their health systems and every country has technical options for reducing them. Based on an assessment of what is known and what is missing, a number of recommendations for immediate action can be formulated. They are divided into actions that need to be taken by countries and those where the international community including financial partners with low-income countries and researchers can assist.

Countries

- Undertake an assessment of the major causes of inefficiency and those that are feasible to change in the short, medium and longer-term.
- Develop and implement a strategy for improving efficiency in the short to medium term – this should be part of a health financing strategy although some of the actions will need to be broader than health financing.
- Start to put in place the background investments that need to be made to ensure the longer-term options can be undertaken – e.g. legislation, consultation, computerized information systems, staff skills.
- Undertake a political as well as a technical analysis to guide which reforms have the greatest chance of success, then build support and negate opposition.
- Develop a set of efficiency indicators relating to the main causes of inefficiency in the country and the agenda for achieving more for the available resources.
- Invest in methods to collect them and to evaluate progress regularly.
- Identify the areas of possible inter-sectoral or multi-sectoral actions that would achieve the largest health impacts, and the political feasibility of influencing other sectors to implement them (perhaps in collaboration with the ministry of health). This would help the Ministry of Health target the key ministries and make the best use of their own limited time and resources.
International community (including researchers in all countries)

- Routinely assess the costs as well as impact of efforts to improve efficiency so that countries can determine the efficiency of different options for improving efficiency. An assessment of the financial sustainability of the different options is also helpful.
- Develop an agenda to identify the cost-effectiveness of efforts to redress health inequalities as part of the efficiency and equity discussion.
- Develop methods which can be used to help countries determine which of the myriad of inter-sectoral or multi-sectoral actions to improve health should be given priority for the limited time and financial resources available to a Ministry of Health. This would feed back into item 5 above.
- Continue to invest in the technologies that might “shift the frontier” of possibilities”, identifying further opportunities to improve health and financial protection at low cost, such as vaccines for Hepatitis C and HIV/AIDS.
XI. References


progress toward universal health coverage.
http://apps.who.int/iris/bitstream/10665/254680/1/9789241512039-eng.pdf


Honda, A., 2014. What is strategic purchasing for health? LSHTM Research on Line. [http://resyst.lshtm.ac.uk/sites/resyst.lshtm.ac.uk](http://resyst.lshtm.ac.uk/sites/resyst.lshtm.ac.uk)


Knaul F.M., González-Pier E., Gómez-Dantés O., García-Junco D., Arreola-Ornelas H., Barraza-Lloréns M., Sandoval R., Caballero F., Hernández-Avila M., Juan M., Kershenobich D.,


Kringos, D.S., Boerma, W., van der Zee, J. and Groenewegen, P., 2013. Europe’s strong primary care systems are linked to better population health but also to higher health spending. *Health affairs*, 32(4), pp.686-694.


Markovitz, A.A. and Ryan, A.M., 2017. Pay-for-Performance: Disappointing Results or Masked Heterogeneity?. Medical Care Research and Review, 74(1), pp.3-78.


http://dx.doi.org/10.1787/9789264266414-en


Tiemann, O., Schreyögg, J. and Busse, R., 2012. Hospital ownership and efficiency: a review of studies with particular focus on Germany. *Health Policy*, 104(2), pp.163-171. Confirms that private hospitals are not necessarily more efficient than public in gmy.


World Health Organization, 2017e. Achieving the health-related MDGs. It takes a workforce! http://www.who.int/hrh/workforce_mdgs accessed 9 February 2017


World Health Organization, 2017g. Community-based health workers (CHWs) http://www.who.int/hrh/community accessed 9 February 2017


http://apps.who.int/iris/bitstream/10665/185989/1/WHO_HIS_HGF_SR_15.1_eng.pdf


