HLSP Institute

Scaling up health services: challenges and choices

Catriona Waddington

July 2012

While not a new idea, the term ‘scaling up’ has become increasingly popular in global health. This paper clarifies the way the term is used, presents the objectives of different types of scaling up, and highlights lessons from implementation.

It discusses common challenges and choices that have to be made in the context of scaling up: which interventions should be scaled up; who will champion scaling up and deliver the interventions; how to deal with equity concerns; how to finance scale up; how fast to scale up; sequencing; how to monitor progress.

The paper concludes that scaling up health services as a means to improving health outcomes is not just about increased spending. A number of health system, financial, institutional, legal and social challenges need to be addressed and choices have to be made.

HLSP Institute 10 Fleet Place | London EC4M 7RB | United Kingdom
T +44 (0) 20 7651 0302 | E institute@hlsp.org | w www.hlsp.org/institute

1. Introduction

"Scaling up to meet the need is equivalent to when a large group of people must use a bus to undertake a crucial journey. If the bus is too small, or it goes too slowly, or it takes a wrong turn, or its mechanical problems are not fixed, or it is badly driven, it won't reach its destination in time. Simply pouring in more fuel won't resolve these problems. Government and other players in the countries involved must deal with all the issues if the journey is to succeed." (Rivers, 2008)

'Scaling up' has become a much-used term in global health in the 21st century for a number of reasons. There has been increasing awareness of global inequalities and the number of people without access to essential health services, and that to achieve the health-related Millennium Development Goals (MDGs) requires scaling up in the sense of both 'big' and 'urgent'. Scaling up was a key concept in the 2001 report of the Commission on Macroeconomics and Health. Since then, global health partnerships (such as the Global Fund and GAVI) have helped to significantly increase access to a range of interventions.

Scaling up is clearly a complex topic which raises questions on many levels. Where is the pressure to scale up coming from? Is it best to concentrate on scaling up one intervention or technical programme at a time? What happens when there is pressure to scale up several interventions or programmes simultaneously? Why are there so many stories of well-financed scale ups encountering basic bottlenecks such as the disbursement of money or shortages of staff to perform simple administrative tasks? Is scaling up just about reaching as many people as fast as possible, or are there trade-offs with issues such as equity and sustainability?

This paper does not aim to present a comprehensive discussion of all aspects of scaling up. Instead it focuses on the objectives of different types of scaling up; the importance of identifying constraints to scaling up; identifying existing frameworks and tools to structure discussions about scaling up; and some of the critical choices that have to be made when scaling up. The paper cites examples from different decades – increases in coverage are not a new phenomenon and examples from the past can provide useful insights.

What does scaling up mean? What is it trying to achieve?

Scaling up[1] in the health sector is used to mean 'doing something in a big way to improve some aspect of a population's health'. In its current usage, scaling up is often intended to convey **haste**, **urgency** and the need for a '**special effort**' – this is qualitatively different from 'doing a bit more, but in the style of business as usual'. Within this very broad use of the term, people use it in different ways, including:

- Scaling up **inputs** (mobilising more funds, more staff or pharmaceutical supplies).
- Scaling up **outputs**, by providing more services (to increase **access**, **range** of services) or attracting more clients (**utilisation**). It may refer to any form of service, from hospital to home-based care. The expansion can be either a new or existing service; it can be greater geographical spread or involve a new client group. This version of the term is frequently used in the context of single programmes – but it can equally apply to a multi-programme package of interventions.
- Scaling up in order to produce better service **coverage** or improve **health outcomes** – e.g. to achieve the health MDGs.
- Scaling up a **process** (e.g. performance-based financing) as a means to achieving scale-up in outputs or outcomes.
- Scaling up from pilots to national programmes.
- Scaling up from targeting specific groups to looking at the wider population.

It is important to be clear on what kind of scaling up is being referred to in any given context and how this relates to the ultimate objective of improved health. Moreover, all scale ups, whatever their objective, have to consider **equity** and **sustainability.**

Comprehensive strategies to substantially scale up health outcomes of course involve work in other sectors. For example, reducing child mortality will involve work in the water, education and economic

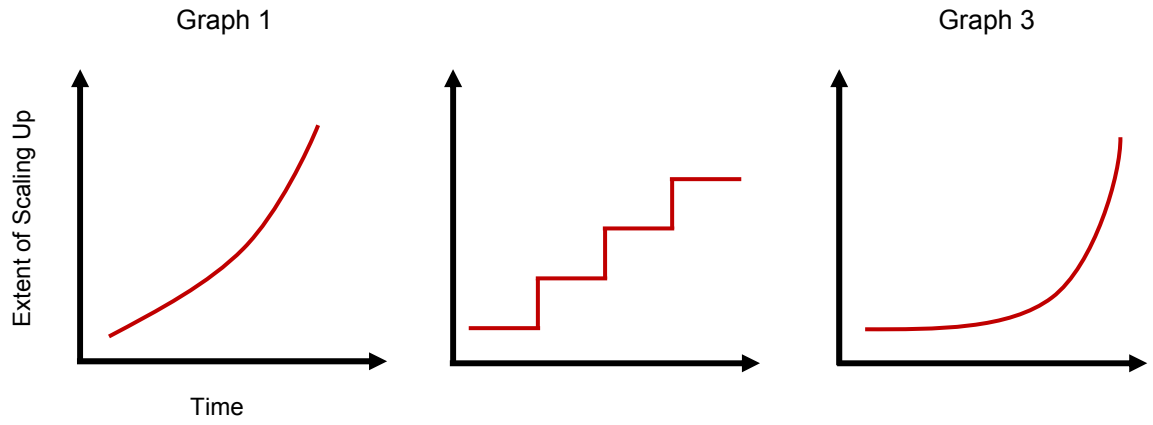
[1] In their 2010 review of scaling up in international health, Mangham and Hanson concurred that there is no one agreed definition for 'scaling up'.

sectors. The paper gives a flavour of possible activities in other sectors, but its primary focus is the health sector. It focuses on scaling up **health services and support systems**.

Scaling up – smooth, stepped or great leap?

A useful analytic device is to visualise scaling up. If scaling up is about radical change, it may not be enough to just do 'more of the same'. There may be institutional, legal or policy issues that need to be addressed before scaling up can proceed beyond a certain point. The three graphs in Figure 1 represent a useful mental device for thinking through how scaling up might develop in a particular context. Box 1 gives practical examples of each.

Figure 1. Scaling up – smooth, stepped and great leap



Graph 1 shows smooth scaling up. Here the underlying health system and immediate environment are able to cope incrementally with more resources and more activities. Graph 2 shows a series of steps which have to be climbed if scaling up is to progress. Graph 3 (the great leap) shows a situation where a significant block needs to be surmounted.

These graphs are obviously an oversimplification, and the three examples deal with very different lengths of time. They do not illustrate the problem of 'diminishing returns': as a scale up approaches its targets, it becomes more difficult to make and maintain progress. Nevertheless, this mental device is a useful start to thinking through the issues – what blocks (or bottlenecks, or constraints) are there? Which 'shape' of scale-up is mostly likely to apply to a particular situation?

Box 1. How scaling up can develop: three examples

Reducing maternal mortality in Sri Lanka – 'smooth' scaling up
The maternal mortality ratio (MMR) in Sri Lanka gradually dropped from 2,136 per 100,000 live births in 1930, to 486 in 1950, 121 in 1973, and 27 in 1996. Graph 1 can be used to represent this almost continuous increase in maternal survival. An analysis highlighted many factors which facilitated the drop. These included a long standing system for the civil registration of births and deaths; relatively high levels of female literacy; and a declining fertility rate (from the 1950s). This was complemented by a long history of training midwives in well-defined competences. Broad service delivery strategies changed over time as the MMR dropped. To begin with, there was a focus on expanding access, especially in under-served areas. Later, the emphasis was on utilisation and on removing financial and other barriers. More recently, quality of care has received close attention. Key events and explicit scaling up phases do not feature prominently in the story. Rather it is a story of good health sector decision-making and implementation over a long period of time in a conducive environment (Pathmanathan, 2003).

The management of childhood pneumonia in Nepal – scaling up in steps
In Nepal, access by the under-five population to community-based management of childhood pneumonia increased in three phases after 1986. Research into effective interventions took place in

one district in 1986-89, resulting in major publications in 1991. A programme in four districts tested the effectiveness of treatment by existing female community health volunteers from 1995 to 1997. In 1999, this became part of community-based IMCI (Integrated Management of Childhood Illness), meaning that community-based pneumonia case management could be part of routine annual programming for the first time. Moving from each phase to the next required formal Ministry of Health approval, which took time – this is scaling up in steps (Dawson et al, 2008).

Abortion in South Africa – rapid increase in services following new legislation
South Africa greatly relaxed legal restrictions on abortion during the first trimester by enacting legislation in 1997. As a consequence women's access to safe abortion services increased dramatically in a very short time. The legislation played a significant role in the 91% decline in deaths related to unsafe abortion between 1994 and 2001 (Grimes et al, 2006).

2. Setting objectives and identifying constraints

Setting objectives

The way in which a particular scaling up exercise will be implemented depends on where the pressure to scale up is coming from, its objectives and perceived constraints. Pressure for change may come from different quarters – from politicians, donors or health service managers, from potential beneficiaries or society at large. Sometimes ordinary people influence scaling up decisions; this is well illustrated by the role played by civil society groups in promoting the right to HIV treatment.

Objectives influence the nature of scaling up activities. For example, a scaling up exercise with coverage as its prime objective would focus on different activities than one focusing on quality. A scale up aimed at quickly reducing incidence may have more of an incentive to reach high-risk clients than an exercise with coverage as its principal objective. And if sustainability is an explicit objective, there is an added incentive to institutionalise the scale up through measures such as supportive policies and budgets and including the intervention(s) in pre-service training curricula.

Objectives are the ambitions of a particular scale up exercise. Because scaling up is about significant change, it is also important to look at the effect on *other* parts of the health system – effects tend to 'ripple out' beyond the target services or audience. If one intervention or programme is significantly scaled up, how does this affect other interventions or programmes and what is the impact on service delivery and overall health outcomes? Scaling up one part of the system when there is no spare capacity can only be at the expense of other activities. The key is to understand these trade-offs and ensure that the right overall decisions are made.

The problems caused by neglecting sustainability, and the challenges of balancing different objectives and any trade-offs between them, are illustrated later ('How fast can scaling up happen?').

Identifying constraints

Every scale up has its own stories of challenges – i.e. areas which have slowed down progress. For example:

- **Amount of money available.** Macroeconomic conditions and policies may either enable or constrain planned health services scale up. Rapidly growing economies may generate resources for more health facilities, staff or a larger range of services. On the other hand, public budget ceilings set by a Ministry of Finance may limit Ministry of Health staff recruitment.
- **Money getting where it is supposed to.** Even when money is available, it can be difficult to set up a system of smooth-flowing disbursement to the districts, NGOs or other partners which will eventually spend the money.
- **Managerial capacity.** Management capacity may be weak at facility, district, provincial and/or national level. Good, proactive management that is able to deal effectively with practical problems is essential.
- **Coordination and communication.** Scaling up requires that a lot of people in a lot of places are well-informed about (and support) the relevant interventions. They may want to adapt the intervention(s) to suit their own local values or circumstances and need to know enough about the

technical and financial aspects to be able to do this properly. This communication is often a challenge, especially in decentralised countries where local governments run health services and make significant resource allocation decisions.

- **Laws and rules.** A policy, law or simply an administrative procedure may effectively block progress. For example, a change in prescribing policy may be needed to allow nurse-based delivery of antiretroviral therapy (ART).
- **Sufficient demand.** There may be limited demand to match the scaled-up supply. So, immunisation coverage cannot be scaled up if there is a widespread belief that immunisations cause unwanted side effects, and the use of interventions related to sexually transmitted infections will be limited if there is a strong stigma attached to using such services.
- **Opposition.** Virtually every scale up has its critics. For example, health workers may resist changes in their working practices, interest groups may object to resources being channelled to other groups, and the private sector may see its profits being challenged.

An early stage in scaling up is to identify bottlenecks such as these. (Travis et al, 2004). Several existing frameworks and tools may help to specify where potential bottlenecks lie – for example, insufficient inputs, lack of managerial or technical capacity, little political push or local ownership. Box 2 summarises some checklists and frameworks which can be adapted to suit particular circumstances.

Box 2. Frameworks and tools for thinking systematically about scaling up

Various frameworks and tools can help stakeholders to think through scaling up issues systematically. No single tool or framework is perfect. Different tools address different aspects of scaling up, and frameworks can, of course, be adapted to specific situations.

Within the health sector, there are frameworks that provide people with different ways of looking at constraints. One way is to look at the performance of health systems' core functions or 'building blocks' and their links to service outputs and outcomes – for example using WHO's health system framework (WHO, 2007a). Another way is to identify constraints by level of the system, and the extent to which increased health sector funding can reduce different constraints (e.g. the framework developed by Hanson et al, 2003). Jacobs et al (2011) apply a framework of supply- and demand-side access barriers to two examples from Cambodia and analyse why certain interventions fail to tackle specific barriers.

There are also tools from the more specific scaling up literature. ExpandNet has developed a number of tools, including *Nine Steps for Developing a Scaling Up Strategy*. Gericke's intervention complexity model (2005) addresses scaling up issues by looking at various dimensions of complexity. For instance, how complex are the human resource and management support requirements for a particular intervention?

There are also non sector-specific tools to help people navigate the political and institutional environment in which health systems operate, and manage change. These include the Open Systems Model, Force Field Analysis and a number of Change Management tools. A useful overview of these tools can be found in Wilson et al (2003).

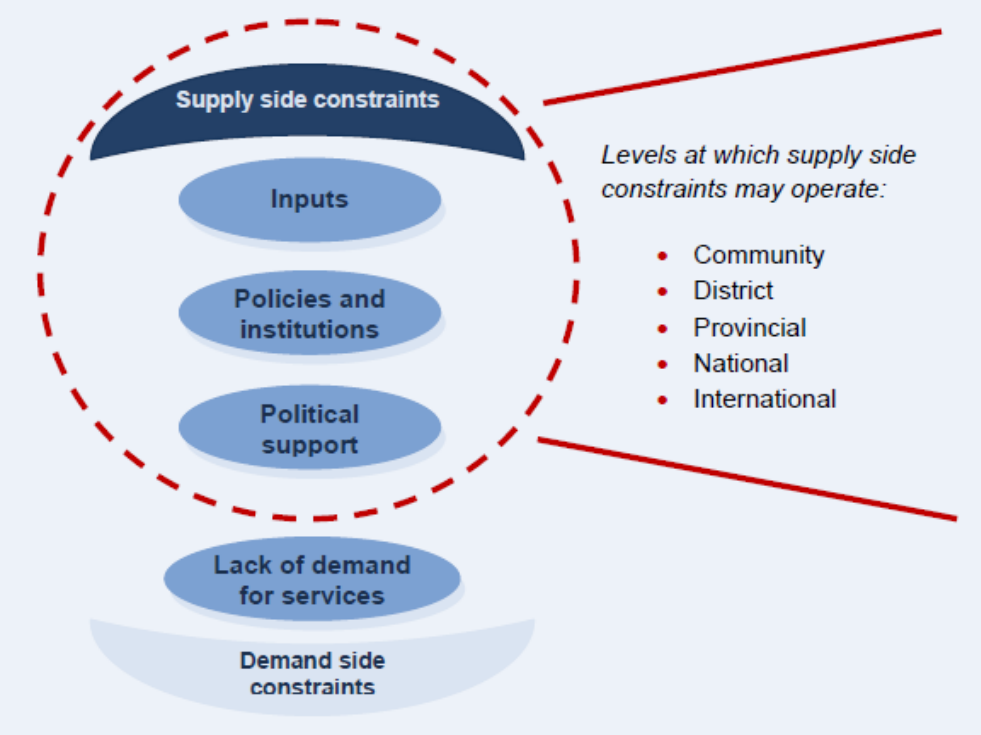
Figure 2 illustrates one possible broad framework for the identification of both supply and demand side constraints. Questions which could be asked within the framework include:

- **Inputs**: What are the main inputs and in what quantities will they be required (e.g. money, workforce, drugs)? Are there any recent changes that may mean more resources are available?
- **Policies and organisations**: What public and/or private sector organisations will be involved in the scale up? What laws, policies and rules do they operate under? Are these enforced? Are any changes needed, and how difficult will it be to mobilise the required support to do so? Do the organisations involved have the requisite capacity and incentives to plan, manage and implement? To whom are they accountable?
- **Political support**: Is there sufficient political support for the scale up? In addition to politicians themselves, groups such as trade unions, women's associations and civil society organisations may

be important to help mobilise support for change. Is political support linked to new resources? If so, when might these be available? Is the support sufficient to offset any opposition to the scale up?
- And underpinning all this, is there a **demand for the intervention(s)?**

These questions can be asked at the community, district, provincial, national and international levels.

Figure 2. Broad framework for identifying potential constraints to scaling up



3. Design and implementation: critical choices

Scaling up generally involves changes in many parts of the health system. We saw in Box 2 that several frameworks and tools can help address the issues systematically. This section discusses some of the most frequent and topical challenges and choices that have to be made:

- The interventions to be scaled up
- Who will champion scaling up and deliver the interventions
- How to deal with equity concerns
- How to finance scale up
- How fast to scale up
- Sequencing of scale up
- How to monitor progress in order to make corrections where needed.

The interventions to be scaled up

Single or multiple interventions?
Scaling up can involve anything from a single intervention to a broad package of activities. Single intervention scale up may be driven by an urgent or new health need, the mandate of a funder, a local political issue, or a new technology which can be added to existing services. Whole-package scale up often occurs because of a change in a country's political and/or economic circumstances – for example a new government with a real commitment to improve service provision for the poor.

There is no right or wrong answer to the question of single or multiple interventions – they entail different advantages and risks. A broad package may potentially be more cost-effective, but scaling it up across a population can be complex and slow. Scaling up a single intervention (or several related

interventions within one programme) may be simpler to implement, but may have consequences for other parts of the health system. This is because 'narrow' scale ups often use resources which are also involved in delivering other interventions, such as health workers, health facilities or drug distribution systems. These external consequences (the effects that extend beyond the actual scale up) may be negative (for example, the financial incentives attached to dealing with the scaled up intervention may not reflect its local epidemiological importance) or positive (e.g. a general improvement in laboratory services). Box 3 shows examples of external consequences related to human resources.

Box 3. Scaling up and human resources – positive and negative effects

- Work on mainstreaming HIV and AIDS may involve a large percentage of the public sector health workforce in training activities. This is often appreciated, both for its own sake and for the perks (such as per diems) which it brings. In this example scaling up HIV and AIDS activities has a positive effect on the wider workforce.
- AIDS workers may have access to more financial incentives than workers in other technical areas. This can cause resentment and problems for trade unions.
- Workers in other technical areas may feel dispirited because issues they care about are neglected. For example a hospital doctor in Ethiopia said: "With the scale up of ART services in hospitals, patients with chronic illnesses like diabetes, hypertension and other internal illnesses are not given attention".
- In Malawi, a number of health surveillance assistants took the opportunity to train in voluntary counselling and testing. For the individuals, this was effectively a promotion, but the service as a whole lost experienced health surveillance personnel.

(Global HIV/AIDS Initiatives Network, 2008)

The issue is complicated when several narrow scale ups happen simultaneously. If the scale up is being done in a relatively short timeframe (as is often the case), the tendency is for each separate scale up group to focus inwards on its own needs and goals. Without strong government leadership, this can lead to the creation of parallel support systems or to an unregulated competition for scarce resources, such as the time of doctors and nurses.

A variation of the single/multiple intervention debate is when scaling up takes place by grafting additional interventions on to existing service provision. For example, in recent years many outreach immunisation services have added other interventions, such as Vitamin A.

Different types of intervention offer different opportunities and challenges to scale up.
Some general rules of thumb help to design a realistic scale up strategy. Scaling up is easier if:

- **There is a strong evidence base**. Good evidence about the benefits of the intervention (the more localised the better) can help with both the initial decision to scale up and with implementation, especially if there is a significant cadre of doubters or agnostics about the value of scaling up. Routes to scale up differ: some scale ups start with a pilot, others go straight to national scale. Evidence is used differently in these different paths (Janovsky and Peters, 2006).
- **There is demand for scale up**. Does demand exist for the intervention at the political, provider and client levels? When properly managed, demand can help create and maintain momentum for significant change. Some interventions require more demand creation efforts than others. Box 4 gives an example of simultaneously scaling up supply and demand.
- **The intervention is compatible with the norms and values of the population and its health services**. If a radical change is required in the basic way of doing things (rather than just a quantitative increase), scaling up may be slower and require more intensive efforts. Scaling up innovations which challenge existing norms (e.g. about sexual behaviour) face particular challenges.
- **Quick results are possible**. Observing quick results helps to spread demand for further scaling up.
- **The intervention is simple**. Simplifying and standardising interventions (in terms of technical and managerial requirements) facilitates scaling up. Some interventions are by their nature relatively simpler than others and more amenable to standardisation – for example immunisation.

- **Costs are low**: The larger the overall cost of the scale up relative to existing expenditures, the more work may be necessary to mobilise funds or to persuade local decision-makers that the scale up should be a national priority. Costs to consumers are also relevant, since it is more difficult to scale up an intervention which consumers find expensive.

The following are examples of how action in one or more of these areas helped create favourable conditions for scale up.

Stronger evidence. The growth of community-based management of pneumonia in Nepal was informed by evidence (see Box 1). A Technical Working Group was convinced of its effectiveness by 1993. The second, larger, research phase aimed to provide evidence to government officials who doubted that female community health volunteers could effectively manage pneumonia. There is often demand for a second round of research with a larger population in more 'normal' circumstances – people want to be convinced about local practicalities.

Anticipating and managing potential opposition. The Malawian NGO Banja La Mtsogolo (BLM) has worked on scaling up health care in prisons. Recognising the unpopularity of the cause, the scaling up plans included political-level advocacy from the start. A Steering Committee was set up both to monitor the initial project and to be involved in advocacy for prisoners' rights. BLM worked with parliamentarians from the Health and Legal Affairs Committees (BLM, 2007).

Simplifying programmes. HIV programmes have emphasised the importance of standardising and simplifying protocols and procedures for testing and counselling, prevention, management of opportunistic infections and ART (WHO, 2006). Scaling up antiretroviral therapy became easier as the intervention became simpler and more standardized. Simplified record-keeping was an important aspect of national scale up of the Matlab approach to family planning in Bangladesh in the 1980s (Yamey, 2011).

Reducing costs. Scaling up TB control in India, where many poor people use private providers, has involved reducing costs to patients. A public-private partnership (private provider, public subsidy and technical support) has substantially reduced costs incurred by patients and their attendants in the private sector, and TB outcomes have improved (Floyd et al, 2005).

Box 4. Scaling up demand *and* supply: the HIV response in Brazil

Brazil's scaled-up HIV response included work on demand and supply, on prevention and treatment. In the 1996-2002 period AIDS-related mortality fell by 50% and hospitalisations by 80%. Brazil is perhaps best known for reducing ART drug costs through domestic production and international negotiation, but there have been a range of activities aimed at increasing demand for HIV-related products, as lyrically described by Okie:

> "In Brazil this past February, during the week before Carnaval … citizens who ventured out to catch a bus, buy a beer, or mail a letter were likely to be reminded by their government to use condoms. Postal consumers received condom brochures along with their stamps. Public health officials contracted with Coca-Cola distributors to deliver condom posters to bars along with the soft drink. In a television commercial on the country's most popular soap-opera network, a famous comedian riffed on strategies for remembering not to leave home without a condom. In the north-eastern city of Recife, banners on buses proclaimed, 'On or off the float, *camisinha*' (literally, 'little shirt, the street term for condom). During one lunch hour in Recife's business district, a prostitutes' organisation working with the local health department on prevention of HIV infection staged a pro-*camisinha* demonstration, passing out free condom samples to spectators."

Other demand-creation activities included free HIV testing, backed up by a national media campaign featuring popular athletes, entertainers and models, and the 1996 law providing free antiretrovirals to all eligible citizens. Activist groups were heavily engaged, reminding the government of its constitutional promise of health care as a right for all Brazilians and encouraging high risk groups to demand their entitlement. (Okie, 2005)

Actors and coalitions: who will champion scaling up and deliver the interventions?

While there are few certainties in scaling up, two generalisations about actors (that is, who does what) are fairly robust:

- Scaling up generally requires a highly committed group (or groups) of champions to generate support and push it along.
- Scaling up generally involves multiple organisations working on service delivery, financing and/or stewardship functions (such as co-ordination and regulation). Responsibility for these functions can be shared out in many different ways, and it is easy to see how scaling up often involves quite complicated coalitions of government, private providers, civil society groups, national and international funders and agencies providing technical support. The ultimate responsibility for coordination and oversight obviously rests with national governments.

This section concentrates on the role of champions and on the range of providers which can be involved in delivering the scaled up services.

The role of champions
A group of institutions and individuals committed to a scaling up exercise may be more or less formally constituted and more or less formally working to a scaling up plan. Whatever the exact situation, these are the champions, leaders and drivers of initiating and implementing scaling up. On paper, the list of organisations involved in championing the scale up and actual delivery may look very similar. The key point is that championing and delivery are very different *functions*, which may or may not both occur in the same organisation.

Championing requires a ‘can-do’ positive mentality and a willingness to manage proactively. Any scale up will face political, technical, logistical and administrative obstacles which need to be dealt with in a constructive and timely way. The groups driving scale up thus require a variety of skills, from the ability to win over local support and forge effective coalitions, competence in technical areas, management and training to a talent for resource mobilisation and advocacy.

To promote effective collaboration among scale up partners, it may be beneficial for all the major partners to be represented. As an example, the team associated with the increase in pneumonia management in Nepal (described in Box 1) came in the form of a technical working group which included government employees, local specialists and international agencies (UNICEF, USAID and WHO). In Thailand, a core coalition involved technical experts and government bureaucrats working with civil society groups and political parties to influence adoption and implementation of reforms linked to universal coverage (Mills, 2007). For major scale up, such as many ART roll outs, driving teams are likely to be needed at different levels, and civil society organisations may play a major role in programme design and implementation, as well as in advocacy.

Collaboration can be difficult, with institutional frictions and cumbersome processes to be navigated. Individuals are often instrumental in smoothing over difficulties. Bhattacharya (2004) describes the roles of WHO and the health ministries of central and state governments in smallpox eradication in India in the 1960s and 1970s, and the complex interplay and changing coalitions between them. WHO expended huge energy on diplomatic contacts with the Prime Minister’s office, the states’ chief ministers and the federal and state health ministries. The Prime Minister was sometimes approached directly because her support was deemed vital.

Which delivery organisation(s)?
'Delivery organisations', which actually deliver the scaled up activities, may be any combination of central and local government, private providers, social marketing organisations etc. Rapid expansion may require looking at radical alternatives – all delivery options, and their pros and cons, should be considered. The choice varies with context. The private sector may have good logistics and be responsive to clients; NGOs may be appropriate when an intervention requires considerable local participation and adaptation; starting with some local governments which have shown a particular interest has its advantages; effective central government has a wide reach and responsibility for national policies. Different delivery organisations are suited to different interventions – for example the for-profit private sector is not well suited to delivering some types of intervention. Box 5 illustrates how this notion of **comparative advantage** can be applied.

Box 5. ITNs: a role for the commercial and public sectors

The WHO Position Statement on insecticide-treated mosquito nets (ITNs) sees a vital role for the public financing of ITNs, but recognises that the type of organisation with the best distribution system varies from country to country. It states: "Where strong commercial markets exist or are developing, they should be encouraged: they can provide important benefits, ensuring longer-term access and enhancing management of logistics and education efforts" (WHO 2007).

This comparative advantage analysis has been put into practice in Tanzania, where there has been a targeted subsidy, plus support to the commercial sector from a social marketing programme (Mulligan, 2008).

How the various types of organisation affect each other is also relevant. For example, a significant increase in the subsidised distribution of an item through the public sector may cause a decline in private sector profits. This may or may not be a good thing, depending on the contribution to improved health of what the private sector was providing (e.g. fake drugs versus good quality ITNs).

Can scaling up be equitable?

Big scale up exercises may magnify existing equity problems in health systems. The hardest-to-reach populations are generally the last to benefit from scaling up exercises, unless deliberate efforts are made to hasten their inclusion. Populations can be hard to reach for geographical, economic and/or social reasons. Extra efforts may be justified on epidemiological grounds (hard-to-reach groups which are also high-risk) or simply on the grounds of fairness. Sometimes a scale up may deliberately exclude equity from its short-term objectives, arguing in favour of 'as many people as possible, as quickly as possible'. For example, in many countries the roll out of the '3 by 5' initiative (to treat three million people living with HIV by 2005) initially focused on increasing absolute numbers on treatment, by beginning with (the largely urban) facilities that were ready to offer treatment. The evaluation subsequently recognised the balance needed between numbers and equity, and emphasised that services should be designed from the start in a way that allows roll-out to low-resource rural settings (Battistella Nemes, 2006).

In immunisation, the Reaching Every District (RED) approach tackles this very issue of scale up-with-equity. RED aims to improve immunisation systems in districts with low coverage. With the RED approach, countries use coverage data to analyse the distribution of unimmunised infants and prioritise districts with poor coverage. Districts are encouraged to make micro-plans to identify and address local problems. In 2005, an evaluation of five countries in Africa that had implemented RED found that the proportion of districts with DTP3 coverage above 80% had more than doubled (Vandelaer, 2008).

How can scaling up be financed?

Not all scale ups require significant financial outlays – at the extreme, fiscal policy changes to increase the price of tobacco or alcohol involve no net government expenditure. But most scaling up in the health sector is expensive. Moreover, because scaling up is by definition not routine, some dedicated resources are necessary until implementation becomes standard practice and costs are financed through routine budgets. This is clearly a huge topic; this paper simply notes the range of possible sources of funds and some of their most important attributes.

Significant scaling up is generally financed through taxation, social health insurance (SHI), public-private infusions of money or international aid (including global health partnerships). Each brings its own challenges. Tax- or SHI-financed scaling up requires a favourable mix of circumstances in the national economic, labour and political scenes, but generally comes with ready-made sustainability and institutionalisation. External financing raises issues of local ownership, sustainability and (if there are multiple funders) aid effectiveness. Earmarked funds in particular may distort local priorities and create incentives for parallel support systems for monitoring, procurement or supervision, which ultimately are hard to sustain.

How fast can scaling up happen?

We have seen how the current usage of 'scaling up' often implies a fast pace of change. Impressive improvements can be achieved in a short time. In Rwanda, for example, a particular form of performance-based funding was introduced in selected districts in 2006. At the end of two years, there had been significant increases in the number of preventive care visits by children aged 23 months or younger (56%) and aged between 24 months and 59 months (132%), as well as a 23% increase in institutional deliveries. Unsurprisingly, it was found that the biggest increases tended to be for services with the highest payment rates (Basinga et al, 2011).

In contrast, in some circumstances scaling up is a long-term undertaking. Box 6 shows how long various countries took over the highly complex scale up of universal coverage. History provides other lessons about pace. A 1996 review of Universal Child Immunisation described how *maintenance* of high immunisation levels could be made difficult by inappropriately fast scaling up: "When targets have been set at unrealistically high levels, they have led to the development of unsustainable immunisation strategies and to the manipulation of data. When their political use has taken precedence over their use in managing programmes, they have deterred effective critical assessment. Immunisation targets should be ambitious but attainable" (UNICEF, 1996).

There is clearly a balance to be struck between the achievements of rapid or explosive scaling up and the advantages of gradual scale up. The current challenge is to use the opportunities afforded by the MDGs, results-based financing and new public-private financing mechanisms, but without responding so fast that quality or sustainability are unacceptably compromised.

Box 6. Scaling up towards universal coverage – a range of timescales

Countries have differed greatly in how long it took from the first health insurance law until legislation was passed to implement universal coverage (which is the ultimate scale up). In Germany – a pioneer in health insurance – it took 127 years, in Japan 36 years. In the Republic of Korea 26 years elapsed, 20 in Costa Rica. However some care has to be taken with direct comparisons, because there can be variation in the interpretation of 'universal coverage', for example in terms of how many services are included and the extent of co-payments. Nevertheless, a number of factors can be associated with a faster speed of change:

- Higher per capita income
- A larger percentage of the workforce in the formal sector
- More urbanised populations and higher overall population densities
- Good national administrative capacity
- Supportive social values
- Good stewardship from government, including open debate about policies relevant to the scaling up which fostered the population's trust in government and other agencies involved.

(Carrin and James, 2004)

What do we know about sequencing scale up?

A frequently asked question is: *Is there a 'correct' sequence of scaling up activities?* Should support systems be strengthened and then interventions scaled up through these systems, or are the interventions themselves the starting point for scale up?

There is, of course, no one answer to this question – as ever with scaling up, it depends on the circumstances. Scaling up health services generally relies on some blend of effective support systems (public or private) and 'special efforts' related to the activity which is being scaled up. A number of questions can help systematic thinking about sequencing:

- What changes (e.g. in staff, pharmaceutical supply systems, policies and regulations) are absolutely essential and what is the quickest way to achieve them?

- Which bottlenecks can be solved in the short term? Criteria such as feasibility, complexity, cost and acceptability of proposed solutions may help make this judgement.
- What changes can wait until later? Are there actions that need to be initiated early on, but may not show results until the medium term?
- What opportunities or entry points are there? Sometimes a timely response to an emerging opportunity can provide a significant boost to a scale up.

Scaling up is not a matter of systems first, then interventions – or vice versa. It is about developing a practical timetable of activities related to support systems and interventions, then adjusting the mix of activities if needed as one goes along (this is the bus and the fuel of the quotation at the beginning of the paper).

Monitoring and evaluating progress

Scaling up in practice is a continuous stream of decision-making about how to deal with chronic and new constraints to further scale up. Monitoring implementation is crucial for three reasons: assessing progress relative to overall objectives; identifying aspects of the scale up which are not working well; and identifying inevitable 'ripple' effects across the health system. Box 9 contrasts a scale up which was actively monitored with one that was not.

Exactly what is monitored depends on the objectives of a particular scale up. Some programmes such as '3 by 5' or Stop TB have clearly defined targets against which progress is monitored. The link between objectives and indicators is important. For example, assessing whether or not a scale up is equitable requires indicators disaggregated by gender, age, region and so on, something few countries do.

As seen earlier, scaling up can have external consequences – effects beyond the activities which are themselves being scaled up. Monitoring the effects of scaling up is thus not just an internal responsibility, it is also part of the wider stewardship function of governments and coordinated development partners.

Despite its importance, monitoring is often a neglected aspect of scaling up. A review of innovative service delivery strategies in 12 countries noted how rarely routine monitoring was used to inform implementation. This absence of monitoring was even more extreme for strategies that cut across programmes, but where monitoring was organised along programmatic lines (Travis et al, 2004). By contrast, an example of good practice in monitoring is presented in Box 7.

Box 7. Scaling up rapid diagnostic tests for malaria in Senegal: the benefits of good monitoring

Senegal saw a significant decrease in reported malaria cases on a national scale between 2007 and 2009 after implementation of parasite-based diagnosis for malaria, with a corresponding reduction in the consumption of antimalarials. Because the scale up towards universal diagnosis was carefully monitored, the Ministry of Health enjoyed a high degree of certainty on malaria incidence throughout the country. This certainty had multiple benefits: it enabled the Ministry to accurately predict antimalarial drug requirements; allowed resources to be concentrated in areas of higher malaria burden and need; and facilitated the evaluation of interventions such as insecticide treated bednets and indoor residual spraying (Thiam et al, 2011).

Bringing the critical choices together in an overall strategy

As earlier examples show, scaling up generally involves working on several fronts at once. In particular, it is important to think both about 'doing more' (usually expanding the availability of an intervention geographically or to new client groups) and about 'institutional scale up' (e.g. changes related to laws, policies, budget lines and regulation).[2] Inter-sectoral action may be required. In Tanzania (Box 8), a particular point was made of scaling up the *scope* of activities to include all the

[2] The scaling up literature calls these horizontal (expansion) and vertical (institutionalising) scaling up. This is a different from the meaning of 'vertical' as used in the term vertical programme.

interventions specified in the National Multisectoral Strategic Framework for HIV/AIDS. Box 9 brings these critical choices together by providing two contrasting experiences from scaling up family planning.

Box 8. HIV in Mbeya, Tanzania – a multi-pronged scale up

The Mbeya Regional AIDS Control Programme is described as part of UNAIDS' Best Practice Collection. As well as concentrating on access and coverage, the Programme focused on continually expanding the range of activities related to HIV and AIDS in order to be as comprehensive as possible. It encouraged, for example, ART expansion activities, management of sexually transmitted infections, peer education in schools, workplace programmes in the private and public sectors, home-based care supported by government and NGOs, economic support for affected families, political advocacy and tackling stigma (Vogel, 2007).

Box 9. Two contrasting experiences of scaling up family planning in the 1960s and 1970s

India
In the late 1960s rapid population growth was seen as a major barrier to economic development. Introduction of a new contraceptive, the Lippes Loop intrauterine device (IUD), was considered a feasible and effective family planning intervention that would be easy to scale up rapidly. However, after an initially positive response from rural women, acceptance began to decline and IUDs remained discredited in India for many years. The major reasons for this failure were that:

- the needs and rights of contraceptive users were ignored;
- no time was taken to explore the cultural and gender implications of IUD use in traditional rural societies;
- the capacity of the weak service delivery system was not realistically assessed;
- there was no early monitoring or evaluation because success was considered a foregone conclusion.

Bangladesh
In the 1970s, a family planning and maternal and child health experiment initiated by the International Centre for Diarrhoeal Disease Research (ICDDR) began to see a steady rise in contraceptive prevalence. In 1982 the government asked ICDDR to assess whether the approach could be successfully used by the national programme. Further research identified modifications to policy and staffing and helped shape reform in the national family planning programme. Contraceptive prevalence rose from 16% in 1980 to 58% in 2004. The reasons why this scale up succeeded included:

- the way the package of interventions was designed, taking account of local culture combined with open discussions on contraception and service provision;
- scaling-up was phased, and paid attention to quality of care and capacity and system strengthening including supervision and monitoring;
- the original approach was adapted to fit government context, and there was government ownership;
- evidence based policy development;
- a dedicated team of facilitators;
- a participatory process and donor support.

(Simmons et al, 2007; Simmons, 2008)

4. In conclusion: key questions

Scaling up health services as a means to improving health outcomes is not just about increased spending. A number of health system, financial, institutional, legal and social challenges need to be addressed and choices have to be made. In any scale up it is crucial to explicitly consider how to deal with issues of sustainability, equity and the effects of scaling up an intervention (or package of interventions) on the rest of the health system.

The following questions are useful when embarking on any scale up of health services:

Objectives and constraints

- What are we trying to achieve? Who are we trying to reach?
- What are the main obstacles and opportunities? Who are the main opponents?
- How should we respond? What balance needs to be struck between providing more services, better services, more equitable services and early versus longer term results? Should we go for single intervention scale up or make it part of a broader package of activities? Why – what are the pros and cons? What effect will this have on other services?

Financing

- What inputs are needed and where will they come from? What will it cost and how will it be financed? Have we got the trade off between short term results and long term sustainability right?

Actors

- What political support is needed, and how can we mobilise it? How can opposition to the scale up be dealt with?
- Which policies and institutions may be affected and what changes and support are required? Who needs to be involved – government, public and private providers, the intended beneficiaries, national and international funders and agencies providing technical support?
- Who should be made responsible for pushing the process along? Are there sufficient champions and good managers?
- What forms of communication will support the scale up?

Time frame and sequencing

- What time frame is envisaged? Is this realistic?
- In what order should things be done – what needs to start sooner and what can wait until later?

Monitoring

- How will progress and effects be monitored – both in terms of intended results, and effects on the rest of the health system?

References

Banja La Mtsogolo (2007). Annual Report 2006.

Basinga, P et al (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. Lancet 377:1421–28.

Battistella Nemes MI et al (2006). Evaluation of WHO's contribution to 3 by 5: main report. WHO.

Bhattacharya S (2004). Uncertain advances: a review of the final phases of the smallpox eradication program in India, 1960-1980. American Journal of Public Health 94 (11):1875-1883.

Carrin, G and James, C (2004). Reaching universal coverage via social health insurance, key design features in the transition period. EIP Discussion Paper 2. WHO.

Dawson et al (2008). From research to national expansion: 20 years' experience of community-based management of childhood pneumonia in Nepal. WHO Bulletin 86:339–343.

ExpandNet (2007). Nine steps for developing a scaling up strategy. ExpandNet/WHO. Draft. http://www.expandnet.net/tools.htm

Floyd K et al (2005). Cost and cost-effectiveness of PPM-DOTS for tuberculosis control: evidence from India. WHO Bulletin 84:437–445.

Gericke, CA et al (2005). Intervention complexity: a conceptual framework to inform priority-setting in health. WHO Bulletin 83 (4).

Global HIV/AIDS Initiatives Network (2008). Briefing Sheet 2: Human resources.

Grimes, DA et al (2006). Unsafe abortion: the preventable pandemic. Lancet 368:1908–19.

Hanson K et al (2003) Expanding access to priority health interventions: a framework for understanding the constraints to scaling-up. Journal of International Development 15: 1–14.

Jacobs, B et al. (2011) Addressing access barriers to health services: an analytical framework for selecting appropriate interventions in low-income Asian countries. Health Policy and Planning 1–13

Janovsky K and Peters D (2006). Improving health services and strengthening health systems: adopting and implementing innovative strategies. An exploratory review in 12 countries. Making Health Systems Work: Working Paper No. 5. WHO/EIP/health systems/2006.2

Mangham, L and Hanson, K (2010). Scaling up in international health: what are the key issues? Health Policy and Planning 25:85–96

Mills, A (2007). Strategies to achieve universal coverage: are there lessons from middle income countries? Health Economics and Financing Programme, London School of Hygiene and Tropical Medicine.

Mulligan JA et al (2008). Costs and effects of the Tanzanian national voucher scheme for insecticide-treated nets. Malaria Journal 7:32.

Okie, S (2006). Fighting HIV: lessons from Brazil. New England journal of Medicine 354:19, 1977–81.

Pathmanathan, I et al (2003). Investing in maternal health: learning from Malaysia and Sri Lanka. World Bank.

Rivers, B (2008). Scaling up to meet the need: overcoming barriers to the development of bold Global Fund-financed programs. Aidspan White Paper.

Commission on Macroeconomics and Health (2001) Macroeconomics and health: investing in health for economic development.

Simmons, R, Fajans P and Ghiron L (2007). Scaling up health service delivery from pilot innovations to policies and programmes. WHO/ExpandNet.

Simmons, R (2008). Presentation to WHO. University of Michigan and ExpandNet, 2008.

Thiam, D et al (2011). Major reduction in anti-malarial drug consumption in Senegal after nationwide introduction of malaria rapid diagnostic tests. PLoS ONE 6(4)

Travis P et al (2004). Overcoming health-systems constraints to achieve the Millennium Development Goals. Lancet 364: 900–6.

UNICEF (1996). Sustainability of achievements: lessons learned from Universal Child Immunization.

Vandelaer, J et al (2008). Reaching Every District (RED) approach: a way to improve immunization performance. WHO Bulletin 86 (3).

Vogel, UF (2007). Towards universal access to prevention, treatment and care: experiences and challenges from the Mbeya region in Tanzania. A case study. UNAIDS (Best Practice Collection).

WHO (2006). Towards universal access by 2010: how WHO is working with countries to scale up HIV prevention, treatment, care and support.

WHO (2007a). Everybody's business: strengthening health systems to improve health outcomes: WHO's framework for action.

WHO (2007b). Insecticide-treated mosquito nets: a WHO Position Statement.

Wilson D and Beaton L (2003). Promoting institutional and organisational development: a source book of tools and techniques. DFID.

Yamey G (2011). Scaling up global health interventions: a proposed framework for success. PLoS Med 8(6).

A note on the scaling up literature

The scaling up literature comes in many guises, including historical accounts, readings on universal coverage, evaluations of global scaling up exercises and the 'innovations' literature, which looks at how to encourage the adoption and dissemination of effective innovations. Other sectors (notably agriculture and rural development) also have a rich literature on scaling up. Of particular relevance to this paper is the work of ExpandNet, a network of public health professionals and scientists focusing on scaling up health service innovations which have been tested in experimental, pilot and demonstration projects. The website (www.expandnet.net) includes guidance tools to assist countries with scaling up.