The economics of scaling up: cost estimation for HIV/AIDS interventions

Lilani Kumaranayake

Background: The scaling up of HIV/AIDS programming has been one of the most extensive undertakings in international public health. Yet decision-makers are encountering significant uncertainties about financing and the need to understand programming costs at different scales of delivery.

Objectives: To review the economic methodologies for examining costs and variation by scale. To summarize and synthesize the current evidence related to the provision of HIV/AIDS interventions and scaling up.

Methods: We used a review of economic methodologies to generate a conceptual framework for classifying existing data, looking at both short-run and long-run perspectives. A review of the literature was performed using PubMed and available grey literature. Factors facilitating comparison and generalizability are highlighted.

Results: There is growing evidence of scale variation among the costs of HIV/AIDS interventions. Scale variation has been found to explain 26–70% of cost variation across locations for similar interventions. Average costs may become larger or smaller as the volume of services expands, depending on the level of coverage and type of intervention. Key constraints to scaling up include infrastructure investments and cost results need to be interpreted in this light.

Conclusions: Evidence to date suggests that cost efficiencies associated with scale may reflect different ways of delivering services at higher volumes, including lower quality outputs. There is still, however, an extremely limited economic evidence base and mechanisms to integrate economic analyses into routine programme monitoring are recommended.

AIDS 2008, **22** (suppl 1):S23–S33

Keywords: costs, economics, HIV/AIDS, scale

Introduction

The scaling up of HIV/AIDS programming in resource-constrained settings has been one of the most extensive undertakings in international public health. Just in the 5-year period 2001–2005, the number of people receiving antiretroviral treatment in low and middle-income countries increased fivefold, from 240 000 to 1.3 million [1]. The current speed and level of global resources focused on the HIV/AIDS problem are unprecedented. International commitments towards the goal of universal access to comprehensive HIV prevention, treatment, care and support by 2010 were endorsed by leaders of the G8 countries in 2005, and subsequently supported at the UN General Assembly high level meeting on AIDS in June 2006. These commitments occur within the broader context of the millennium development goal aiming to halt and begin to reverse the spread of HIV/AIDS by 2015.

The goals of universal access and the millenium development goals place further programming emphasis on the rapid scaling up of both HIV/AIDS prevention and treatment efforts. Decision-makers are, however, encountering significant uncertainties related to costs, financing and sustainability for scaled-up programmes. From a policy perspective on the resources that may be required for expanding services, decision-makers need to understand what constitutes the programming costs at different levels of scale by mode of delivery as well as the extent to which experiences from different countries and circumstances can be generalized.

From the London School of Hygiene and Tropical Medicine, London, UK; and Dalhousie University, Halifax, Canada.
Correspondence to Lilani Kumaranayake, Dalhousie University, Halifax, Canada.
Tel: +1 902 494 2026; fax: +1 902 494 6917; e-mail: phpulkum@lshtm.ac.uk

A number of reviews have already highlighted the lack of data specific to costs and the cost-effectiveness of HIV/AIDS interventions [2–5]. To date, estimates for global resource requirements [6–9] have largely relied on the assumption that average or unit costs are invariant to the scale of delivery [10], reflecting the lack of evidence related to scale.

How important are scale effects to costs? In this era of rapid scale-up, this question is central for determining, first, appropriate levels of financing and expenditures for programmes and, second, the extent to which resources are used efficiently. The aim of this paper is to review the economic methodologies that have been used to examine costs and variation by scale and, in doing so, to summarize and synthesize current evidence related to the costs of providing HIV/AIDS interventions and scaling up.

Methods

Background to economics of scale

Economists define a cost as the value of resources used in producing a service. Costs reflect the process involved in producing a service, and are a function of the mix of inputs used and the way they are combined (technologies). Generally, it is assumed that there is some degree of substitutability between inputs, so that there are different ways or technologies by which to produce a service. The economic or opportunity cost of using a resource is the value of a good or service were it to be used elsewhere [11]. The total cost of an activity represents the cost of all the inputs used to provide a service (including staff, materials and buildings). Scale describes the extent or level of activity and output at which an intervention is operating. The scale of a project can thus be measured in a number of ways: by the coverage of activity (e.g. the proportion or percentage of people reached by an intervention); by the volume of output of these activities (e.g. the total number of condoms provided, total number of persons trained); or more simply by the level at which the activities are undertaken (e.g. community, district or national) [5].

The total cost of an activity falls into two categories: fixed costs refer to inputs whose quantity does not change as the level of output changes. For example, once a building is rented or built for a voluntary counselling and testing (VCT) site, it is fixed even though the number of people receiving VCT increases. Variable costs refer to the costs of inputs that change as the level of output changes, such as the testing kits for counselling, which will depend on the number of clients. Economic theory also distinguishes between the short and long-run. In the short-run, the amount of fixed inputs cannot be changed. In the long-run, the level of all inputs can be changed.

Even interventions characterized by higher proportions of fixed costs will reach limits to the level of coverage possible and will eventually require additional fixed inputs. Ultimately, the VCT centre will become so crowded that further investments in space will be required. This phenomenon is referred to as diminishing returns. When are these limits reached? From an operational perspective, this point can be judged by a number of criteria (e.g. quality, waiting-time). Cost theory also suggests how to judge capacity limits for existing levels of inputs using measures of average and marginal costs.

The average or unit cost is calculated by dividing the total cost by the units of output or services produced. The marginal cost is the change in total cost of producing one more unit of output, e.g. seeing one more person at the VCT centre. Figure 1 demonstrates these theoretical cost relationships. Economies of scale or scale efficiency are said to be present if average costs decrease as the level of output increases. Economies of scale may be present as a result of indivisibilities in how the project is operated or specialization and the division of labour that requires a large volume of output to be beneficial. For example, a minimum level of fixed inputs is needed to run the VCT centre (e.g. the building), regardless of whether one or 100 people are seen. The average cost of seeing one client is far more than the average cost of seeing 100 people, as the fixed costs are spread over fewer people (or outputs). The lowest cost per person reached is described as the minimum average cost and is the point of scale-efficiency. If marginal costs are lower than average costs, then increasing the number of people seen will lower the average cost per person. Figure 1a indicates economies of scale if the intervention is producing a volume of services to the left of the 'Min' (minimum) point, indicating that it would be cheaper to increase coverage through the expansion of current services rather than replication. To the right of Min however, diseconomies of scale are present and the cost per person reached increases with increases in volume. Average costs are higher than marginal costs, so increasing output by one unit will mean that average costs will now increase. In this case, it might be more efficient to start another centre of size 'Min' (the scale-efficient volume of services) rather than continue expansion at this centre.

Theoretically, it is assumed that there are U-shaped average and marginal cost curves. In practice, however, that is not necessarily the case. Depending on the intervention or project and its cost structure, one may have only economies or diseconomies of scale, as shown in Fig. 1b.

Framework of economic methodologies

Economic methodologies were first reviewed in order to generate a conceptual framework for classifying empirical data, looking at both short-run and long-run perspectives. These methodologies were then used to classify the studies found in a literature review.

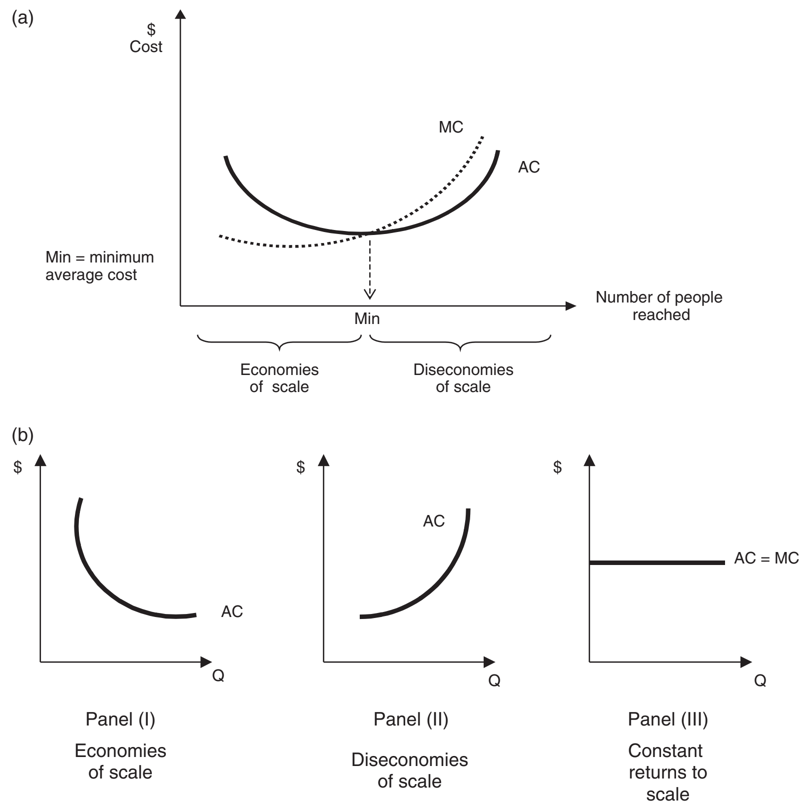

Fig. 1. Relationships between average and marginal costs. (a) Theoretical relationship between average and marginal costs; (b) range of average cost curves. AC, Average cost; MC, marginal cost; Q, output. Data from Varian [12].

Knowing the relationship between average cost, marginal cost and output levels can demonstrate how costs change as coverage/scale changes. In practice, the average or marginal cost curves are not fully observed and rely on estimates from collected data. Our framework classifies studies according to timeframe and methods used in the cost analysis, allowing for short, medium and long-run periods that are distinguished by the extent to which all inputs can vary. Coverage or scale limits are apparent in the short and medium-run, with the short-run constrained to the current capacity of fixed inputs. In the medium and long-run, changes in capacity or systems are possible.

Empirical approaches measure actual costs as programmes change the volume of services. Costs can be measured through time, or using a cross-section of different programmes at various scales to consider the range of costs. The empirical approach helps generate average costs, but is less amenable to measuring marginal costs. In reality, different programmes do not operate with small output incremental differences in scale (such as for each additional person seen). It is generally possible to measure changes in costs associated with larger increments (such as for every 100 people). Empirical approaches can examine how costs vary by scale but cannot systematically measure the extent to which scale, relative to other variables, drives average costs.

Econometric methods estimate the relationship between costs and a variety of factors such as the level of output, allowing for the derivation of average and marginal cost curves as well as a quantification of the degree of scale economies. The method requires empirical data on the costs of the intervention at various levels of output or scale with a minimum of 20–30 data points [12,13]. Then a theoretically specified cost function is fitted to the data, and estimates of the relationship between the costs and scale factors can be estimated. Scale effects of this nature have been described particularly in the hospital cost literature [14], in which econometric methods have been used to consider the optimal size of different types of hospitals and other facilities, as well as measuring if it is cheaper to combine or merge hospital facilities (economies of scope) [15]. In a hospital study of Vietnam [16], different scale effects were found to vary depending on whether the hospital was a district or central type. On the basis of this, policy implications regarding reducing and increasing relative hospital sizes to lower average costs were drawn [16].

Given the paucity of cost data, an alternative approach to determining costs is by using modelling methods to estimate how average or marginal costs change. The earliest cost-modelling study specific to HIV/AIDS [17] used a pure model-base specification of average costs, based on factors such as the target population, and then made assumptions about the relative size of average costs between different types of programmes. Given the lack of reference to existing data and cost structures, however, this pure modelling approach may not be robust to different settings and does not necessarily take into account intervention-specific factors [5,18,19].

A hybrid data-modelling approach uses a combination of data and knowledge about the project to consider the costs of scaling up a project [20]. It might involve establishing a point of minimum average cost and using knowledge from the project and economic theory to hypothesize about marginal costs. This approach requires country or project-specific knowledge. The simplest form of modelling is to use a linear projection, multiplying an empirical average cost associated with a programme with a factor reflecting activity at a larger scale (e.g. if the unit cost per person is US$5 for 100 people, the increase in expanding the programme for another 25 people is US$125).

Methods used to conduct literature review

A systematic review of all published cost and scale studies of HIV/AIDS interventions was completed in the first quarter of 2007. The review included all prevention and treatment interventions (e.g. the treatment of sexually transmitted infections; STI) but excluded tuberculosis treatment costs. The following procedures were used in the systematic review. First, a detailed search was conducted in PubMed using the following search terms: ‘cost AND HIV’; ‘economics AND HIV’. No date/time limitation was specified. Using this strategy, 7045 articles were identified. In addition, two reports and two articles in the press were identified from contacts and the internet and grey literature. As shown in Fig. 2, a system of exclusion criteria was applied resulting in 34 publications from low and middle-income countries. Descriptive information about the type and setting of intervention, method of estimating costs and the influence of scale were examined. Only studies that contained data related to both cost and scale were included. We classified the studies with the conceptual framework based on our review of economic methodologies, according to the timeframe and method of cost analysis.

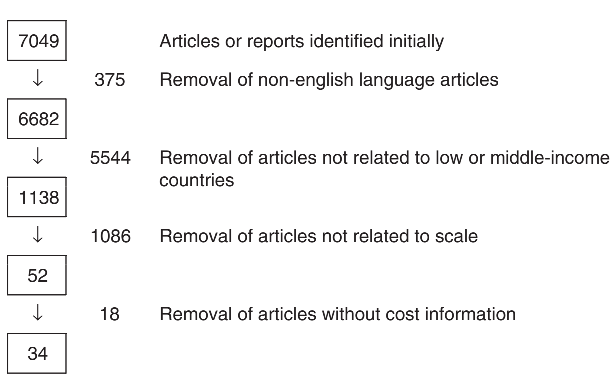

Fig. 2. Results of application of exclusion criteria.

Results

Classification by framework of economic methodologies

Thirty-four studies contained cost information related to HIV/AIDS programming activities and scale and were classified according to the conceptual framework (Table 1). Almost 60% of the studies used modelling methods ($n = 19$), with another 11 studies containing empirical data on scale and costs and four studies undertaking econometric analysis. More than 80% adopted a short-run perspective ($n = 28$), with others allowing for changes in fixed investments [17,24,33–36]. Only one study considered long-run changes [24]. Seven of the modelled studies were cost-effectiveness analyses, in which costs were modelled to correspond to different levels of population coverage of interventions [17,26–28,33–36]. Six studies undertook multi-country estimation of resource requirements for scaling up interventions [6–9,21,23–25], which were recently reviewed in two survey papers [10,19], and four studies estimated the costs of increasing coverage for particular strategies within an individual country [22,29–31]. Many of the studies have used constant average cost assumptions to model resource requirements associated with scaling up [6–9,21–25]. Three studies allow for changes in average costs in the short-run [29–38], and five allow for changes in the medium and long-run [17,24,33–36].

The cost analysis for the Commission for Macroeconomics and Health (CMH) used three scenarios, including the costs of delivering services as well as required investments in infrastructure and broader systems strengthening [24]. Scenario 2007A looked at the possible levels of delivery with an emphasis on expanding the lower levels of the district health system, e.g. health post and outreach services within current capacity. Scenario 2007B estimated the costs for rapidly scaling up priority programmes, assuming substantial investments in existing health systems at all levels of service delivery, but limited by the extent to which these investments can take place over 5 years. Scenario 2015 estimated the costs of achieving the implementation of priority programmes at increased levels of coverage, with investment in existing health systems, over a 13-year period [24]. In 2015, infrastructure costs were estimated to be 17% of HIV prevention, 63% of HIV/AIDS care and 25% of antiretroviral treatment costs, assuming the expansion of care-related infrastructure. Without any fixed investments, the CMH study found that maximum achievable increases in coverage would be constrained to 40% for HIV prevention interventions outside the health sector, 20% for health system-related HIV prevention interventions, 25%

Table 1. Overview of existing studies related to scale and HIV/AIDS programming.

Time-frame	Modelled	Empirical	Econometric
Short-run	Constant or linear average costs (no adjustment for scale)		
Cannot alter some inputs that are fixed	Broomberg *et al.* (1996) [21] Various countries	Forsythe *et al.* (1998) [37] Thailand	Kumaranayake and Watts (2000) [48] Sub-Saharan Africa
Limits to increasing coverage, dependent on existing infrastructure	Wilkinson *et al.* (2000) [22] South Africa	Thaineau *et al.* (1998) [38] Thailand	Kumaranayake (2004) [49] Malawi
Costs of scaling up associated with increasing those inputs that can be changed	Schwartlander *et al.* (2001) [6]/Bertozzi *et al.* (2004) [23] Various countries	Dandona *et al.* (2005) [39] India	Terris-Prestholt *et al.* (2006) [50] Various countries
	CMH 2007a Scenario (2001) [24] Various countries	McConnel *et al.* (2004) [40] South Africa	Guinness *et al.* (2007) [51] India
	Opuni *et al.* (2002) [25] Latin America and the Caribbean	Carrara *et al.* (2005) [41] Cambodia	
	3 × 5 Initiative Gutierrez *et al.* (2004) [9] Various countries	Terris-Prestholt *et al.* (2006) [42] Uganda	
	UNAIDS [7]/Stover *et al.* (1006) [8] Various countries	Dandona *et al.* (2005) [43] India	
	Marseille *et al.* (1999) [26] Sub-Saharan Africa	Guinness *et al.* (2005) [44] India	
	Rely *et al.* (2003) [27] Mexico	Chandrashekar *et al.* (2006) [45] India	
	Pitter *et al.* (2007) [28] Uganda	Kumaranayake *et al.* (2006) [46] India	
		Marseille *et al.* (2007) [47] Various countries	
	Average costs change by scale		
	Forsythe *et al.* (2002) [29] Kenya		
	Uys and Hensher (2002) [30] South Africa		
	Deghaye *et al.* (2006) [31] South Africa		
	Unclear assumptions regarding scale		
	Boulle *et al.* (2002) [32] South Africa		
Medium-run			
Alter some of the fixed inputs	Rowley and Anderson (1994) [17] Not country-specific		
Increased coverage dependent on extent of change of fixed inputs	CMH 2007b Scenario (2001) [24] Various countries		
Costs also include investment in infrastructure needed to expand delivery and average costs change with scale	Kumaranayake *et al.* (2004) [33] Belarus		
	Terris-Prestholt *et al.* (2003) [34] Sub-Saharan Africa		
	Vickerman *et al.* (2006) [35] South Africa		
	Terris-Prestholt *et al.* (2006) [36] Tanzania		
Long-run			
Can alter all inputs	CMH 2015 Scenario (2001) [24] Various countries		
Potentially large increases in coverage			
Costs include investment in all aspects of infrastructure and average costs change with scale			

for care interventions and less than 10% for antiretroviral treatment [52].

Empirical evidence of scale effects for HIV/AIDS interventions

Table 2 reviews the relatively recent studies that provide actual empirical or econometric evidence related to cost variation and scale. Four empirical studies were time series in nature, tracking a particular programme through two or more points in time [38,40–42]. The remaining seven empirical studies were cross-sectional, looking at different sites or locations for the same intervention. Nine of the empirical studies explicitly commented on changes in costs caused by scale effects.

Among VCT studies, both the empirical studies [39,40,47] and the econometric analysis [48] suggest that significant economies of scale were possible, finding a 27–73% reduction in average costs as output increased 23-fold. These results are attributable to the nature of VCT delivery that is characterized by a relatively low proportion of fixed costs (9–21%) [39,48], and so is consistent with theoretical predictions (e.g. the cost function is similar to Fig. 1b, panel I).

Studies of targeted prevention activities among sex workers also demonstrated scale effects [43–45,47,51], with 38–88% of cost variation attributed to scale [43,44,47]. Unlike VCT, the extent of the scale effect is not unambiguously clear for these interventions; two studies suggest that significant economies of scale are possible over large project sizes ($n = 6379$) [43,47], but two other analyses suggest that diseconomies of scale start to occur at project sizes of 1500–2000, similar to Fig. 1a [44,51]. Closer examination of these studies suggests that these results may be explained by the underlying cost estimates: the study with significant economies of scale also found a relatively low proportion of fixed costs (5.2–8.3%), in contrast to other studies that found much higher levels of fixed costs (11–40%).

Scale effects were also found among STI interventions [37,41,46,47,50], with reductions in average costs documented as programme sizes increased to 13 000 visits [41] or 1400 patients annually [50]. It is more difficult to hypothesize about the general extent and shape of the cost curves for STI interventions because of the lack of econometric studies. The only econometric study used data from a range of cost results from 53 studies in the literature and found small-scale effects [50]. That analysis was, however, limited by variation in data quality and comparability. It is clear that scale effects for STI interventions may exist over relatively large population sizes [42,46], probably reflecting the higher proportion of fixed costs (27–40%) [41,42], and one study found evidence of a U-shaped average cost curve [47]. There was also some indication that prevention of mother-to-child transmission programmes may also face a U-shaped curve [47].

Econometric analysis, using modelled costs of scaling up coverage by 25% in sub-Saharan Africa, suggested that both prevention and treatment interventions would be operating at points of scale inefficiency or diseconomies of scale, assuming no changes to fixed inputs. As expected, diseconomies of scale are much more acute for treatment-related interventions [48].

Discussion

This review unambiguously indicates that the scale and volume of activities are key drivers in determining costs, consistent with standard economic theory [12]. What is less clear is the actual shape of the average and marginal cost curves, which will provide better information on the optimal size of interventions that will fully exploit economies of scale. It might also be the case that as interventions are scaled up, the way the services are delivered (e.g. the production function) might change, for example, two studies pointed to reductions in time spent with clients [43,49]. On the basis of the evidence from Table 2, Fig. 3 shows the relative ranking of HIV/AIDS interventions and likely economies (or diseconomies) of scale effects. Interventions with a small proportion of fixed costs such as VCT, or the capacity to increase to high levels of coverage (information, education and communication) are likely to demonstrate downward sloping average costs for large scales of activity, so the larger the programme the more optimal the size. On the other side of the spectrum, interventions most closely related to health facilities (STI, prevention of mother-to-child transmission) are likely to face diseconomies of scale and the U-shaped average cost suggesting that there is an optimal size for the programme. Econometric methods can help to identify these optimal sizes.

Whereas this review found 15 relatively recent studies yielding information on cost and scale, it is more surprising that given the emphasis on scaling-up interventions since 2001, there is still little empirical cost data collected alongside programming as it expands.

Given the dearth of information, and the importance of this type of data in budgeting and financing considerations, our key recommendation is to develop routine cost-monitoring systems alongside interventions as programming expands. An important aspect of cost monitoring with respect to scale is documenting fixed versus variable costs [19], using consistent definitions of what is contained in fixed costs. Given possible changes in how services may be delivered, it is also important to document service delivery alongside measuring costs.

Table 2. Empirical and econometric analyses of costs and scale.

Programme	Description	Findings related to scale and costs	Reference
Prevention of mother-to-child transmission			
	Empirical cost of implementing short course zidovudine in Phayao province, Thailand	5% of pregnant women are HIV positive ($n = 280$). Over 6-month period, programme reached 68% of HIV-infected pregnant women ($n = 190$) with cost of US$0.13 per capita.	[38]
	Empirical cost of 15 programmes in India	Scale variation explains 42% of average cost variation. Upturn or levelling off of unit costs.	[47]
VCT			
	Econometric analysis of empirical costs of VCT in Malawi using monthly data over 34 months.	Number of clients range from approximately 200 to 1600 per month, with fixed costs ranging from 9–21%. Econometric analysis suggested significant economies of scale with index estimate at 1.5 (1, constant returns to scale). Estimates from production function suggest changes in delivery of services as there was a large reduction in time spent with counsellors as the number of clients.	[49]
	Empirical costs of 17 VCT in India using data over 12 months.	Number of clients were 334–7802 per year by centre. Fixed costs were an average 12.4% of total annual costs. Cost per client varied sixfold with 73% of variation attributable to scale. Significant economies of scale suggested by correlation between costs and output.	[39]
	Empirical cost of rapid test VCT in one clinic in South Africa using monthly data over 12 months.	662 Clients in one year, found that costs decreased 66% in months with higher caseload, suggesting economies of scale.	[40]
	Empirical costs of 82 VCT clinics in India, Mexico, Russia, South Africa and Uganda.	Decreasing costs with scale found, with doubling of scale associated with 7–32% lower costs. Contains data from [39] and [40].	[47]
Targeted prevention interventions			
	Empirical cost of 15 HIV prevention programmes among female sex workers in Andhra Pradesh, India, over 12 months.	Range of 803–6379 sex workers reached per programme. Fixed costs were 5.2–8.3% (rent, capital). Found sixfold variation in average costs, with significant relationship between scale and size of programme, suggesting economies of scale. Also found that the time spent with contacts had reduced, as programmes became larger.	[43]
	Empirical cost analysis of 17 HIV prevention programmes among commercial sex workers in Andhra Pradesh and Tamil Nadu, India over 1 year.	Range of 205–2008 sex workers per programme. Fixed costs were 13–40% of costs (median 15%). Fivefold variation in average costs, with 50% of variation in average costs attributed to scale, controlling for factors such as different production functions. Non-parametric analysis finds a U-shaped average cost curve, with minimum average cost occurring at a programme size of 1500 sex workers per year.	[44]
	Empirical cost analysis 15 districts in the southern Indian state of Karnataka, reaching more than 24 000 sex workers over 18 months.	Fixed costs ranged from 11 to 32% of total costs. There was a sevenfold reduction in the average cost per sex worker reached as the scale of activities exceeded 3000 sex workers reached per district. There was a threefold reduction in average costs for later-starting projects.	[45]

(*continued overleaf*)

Table 2 (*continued*)

Programme	Description	Findings related to scale and costs	Reference
	Econometric analysis of 78 state-funded prevention projects and 16 commercial sex worker projects in Andhra Pradesh and Tamil Nadu, India, over 12 months.	Analysis of the expenditure data found economies of scale were not exhausted, a 0.002% fall in total cost for each extra person reached. Estimation using the commercial sex worker cost data found a point of efficient scale at approximately 1750–2000 people reached.	[51]
	Empirical costs of 40 sex worker programmes in India, Russia and South Africa.	38–88% Of variation in average costs explained by scale. Downward average costs, no upturn in average costs found. Includes data from [43].	[47]
	Empirical costs of 22 projects targeting injecting drug users in Russia.	45% Of cost variation explained by scale. A doubling of scale is associated with 34.5% reduction in average costs.	[47]
Treatment of STI			
	Empirical incremental cost analysis of STI clinics in Thailand.	Expanding evening hours at STI clinics would allow treatment of additional 2000 clients at US$31 each.	[37]
	Empirical cost of STI intervention in north-western Cambodia over a 3-year period.	Number of visits per year increased threefold from 4369 to 13 326 with the cost per visit reducing by almost half. Reduction attributed to large increase in visits.	[41]
	Econometric analysis of average costs of STI taken from 53 studies.	Scale ranged from three to 63 693 people per year. Estimates showed a small, but significant scale effect suggesting unit costs decrease with scale.	[50]
	Empirical cost analysis 12 districts in the southern Indian state of Karnataka, reaching more than 20 000 sex workers over 18 months.	The average district cost per person treated was US$27, and there was a sixfold reduction in average costs as the scale of STI treatment increased to more than 1400 patients per district.	[46]
	Empirical cost analysis of 25 projects in India, Mexico and Russia.	Modest scale effect observed with scale variation explaining 42–70% of average costs. Upturn in average costs demonstrated.	[47]
VCT, STI, condom social marketing and IEC			
	Empirical cost of three-arm randomized trial over 4 years individuals in Uganda.	96 000 Individuals targeted within communities with varying interventions. On average 74% increase in uptake led to 34% reduction in average costs, suggesting scale effects. Scale effects constrained in IEC over time, attributed to saturation effect and reduction in uptake.	[42]
IEC			
	Empirical cost of 22 programmes in Mexico.	Proportion of variation explained by scale is 91% with a doubling of scale resulting a 64% reduction in average costs.	[47]
General prevention and treatment			
	Econometric analysis of costs of scaling up programmes in 34 sub-Saharan African countries, looking at relationship with health systems infrastructure and scale for a 25% increase in coverage.	The marginal costs are higher than the average costs suggesting diseconomies of scale. A 25% increase in the marginal costs are approximately seven times higher for care and treatment relative to prevention.	[48]

IEC, Information, education and communication; STI, sexually transmitted infection; VCT, voluntary counselling and testing.

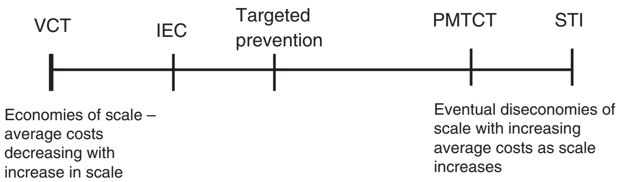

Fig. 3. Scale and HIV/AIDS interventions. IEC, Information, education and communication; PMTCT, prevention of mother-to-child transmission; STI, sexually transmitted infection; VCT, voluntary counselling and testing.

Analyses should focus on monitoring costs through time and across sites, rather than on cost analyses for single years and one location. As data become more available, it is also important to exploit econometric methods to measure the extent of scale economies or inefficiencies, which will assist in the planning and financing processes.

In the absence of hard data, modelling will continue to remain a key method by which to assist financing and budgeting. This review of costs, however, suggests that linear projections that assume constant average and marginal costs will result in both over and underestimates of resource requirements for scaling up. The direction and extent of bias depends on the level of coverage. In terms of the marginal cost, a linear projection approach will only lead to an underestimation of the 'true' cost at relatively high levels of coverage. A linear unit-cost projection will probably underestimate the average cost at either very low or high levels of coverage.

At a national level, the cost relationships may mirror the average cost/marginal cost analysis as shown in Fig. 4. Output is measured in terms of the country population or the proportion of people reached (expressed as a percentage) and the maximum output occurs when there is 100% coverage. Increasing coverage may increase or decrease the average cost per person reached, depending on the level of coverage. Without actual data, it is not clear where exactly the maximum coverage occurs in relation to the cost curves. It could occur at any of the points A, B or C (Fig. 3a), suggesting that there could be very different costs for different countries at 100% coverage. Points A, B and C reflect that scaling up is completed as 100% of the population is covered. Second, if it is assumed for illustrative purposes that all countries have the same cost functions, then the average cost and marginal costs of reaching 100% coverage could vary, simply because of the differences in population size. The average cost per person reached may thus be lower in country B than in country C, even though both have full national coverage. This implies that costs may need to differ between small and large countries in modelling resource requirements.

Another issue in scaling up coverage at the national level is capacity and infrastructure. Within a project or activity, if the limits of the capacity to scale up are reached, then additional fixed inputs are necessary (e.g. physical buildings). Scaling up interventions to a national level raises the question about the infrastructure constraints. It may be impossible to increase coverage beyond a particular level without substantial increases in infrastructure, as shown in Fig. 4b. Therefore, in practice, there may be a large jump or discontinuity in the average cost curve over a particular range of coverage. Tools are currently being developed to model non-linear cost functions better for transportation and supervision costs, as well as fixed costs for health centres [53], and planning is underway to develop new resource requirements modelling for HIV/AIDS interventions to achieve the millennium development goals [54]. Tools and models are, however, limited by the data they contain, and so collecting cost data in relationship to scale must remain a key priority.

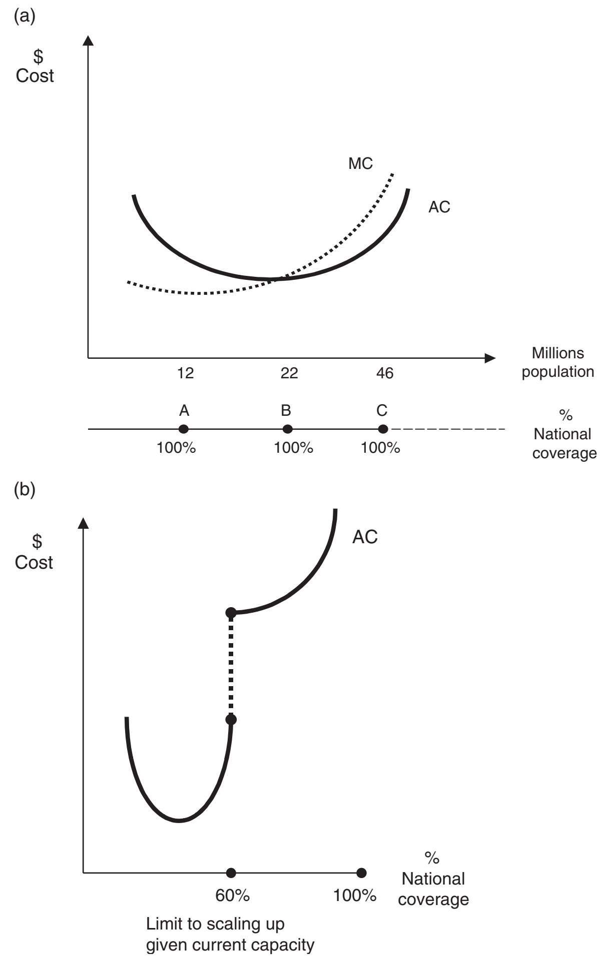

Fig. 4. Cost implications for scaling up interventions. (a) Marginal and average cost curves in the context of national coverage; (b) possible relationship between costs, coverage and capacity constraints. AC, Average cost; MC, marginal cost.

Conflicts of interests: None.

References

1. United Nations. *The millennium development goals report 2006*. New York: United Nations Department of Economic and Social Affairs; 2006.
2. Bertozzi S, Padian NS, Wegbreit J, DeMaria LM, Feldman B, Gayle H, *et al*. **HIV/AIDS prevention and treatment.** In: Jamison DT, Breman JG, Measham AR, Alleyne G, Claeson M, Evans DB, *et al*., editors. *Disease control priorities in developing countries*, 2nd ed. New York: Oxford University Press; 2006.
3. Creese A, Floyd K, Alban A, Guinness L. **Cost-effectiveness of HIV/AIDS interventions in Africa: a systematic review of the evidence.** *Lancet* 2002; **359**:1635–1643.
4. Walker D. **Cost and cost-effectiveness of HIV/AIDS prevention strategies in developing countries: is there an evidence base?** *Health Policy Plan* 2003; **18**:4–17.
5. Kumaranayake L, Watts C. **Economic costs of HIV/AIDS prevention activities in sub-Saharan Africa.** *AIDS* 2000; **14 (Suppl. 3)**:S239–S252.
6. Schwartlander B, Stover J, Walker N, Bollinger L, Gutierrez JP, McGreevey W, *et al*. **Resource needs for HIV/AIDS.** *Science* 2001; **292**:2434–2436.
7. UNAIDS. *Resource needs for an expanded response to AIDS in low and middle income countries*. Geneva: UNAIDS; 2005.
8. Stover J, Bertozzi S, Gutierrez JP, Walker N, Stanecki KA, Greener R, *et al*. **The global impact of scaling up HIV/AIDS prevention programs in low- and middle-income countries.** *Science* 2006; **311**:1474–1476.
9. Gutierrez JP, Johns B, Adam T, Bertozzi SM, Edejer TT, Greener R, *et al*. **Achieving the WHO/UNAIDS antiretroviral treatment 3 by 5 goal: what will it cost?** *Lancet* 2004; **364**:63–64.
10. Vassall A, Compernolle P. **Estimating the resource needs of scaling-up HIV/AIDS and tuberculosis interventions in sub-Saharan Africa: a systematic review for national policy makers and planners.** *Health Policy* 2006; **79**:1–15.
11. Creese A, Parker D. *Cost analysis in primary health care: a training manual for programme managers.*. Geneva: WHO; 1994.
12. Varian H. *Microeconomic analysis*. New York: Norton and Company; 1998.
13. Vitaliano DF. **On the estimation of hospital cost functions.** *J Health Econ* 1987; **6**:305–318.
14. Wouters A. **The cost and efficiency of public and private health care facilities in Ogun State, Nigeria.** *Health Econ* 1993; **2**:31–42.
15. Preya C, Pink G. **Scale and scope efficiencies through hospital consolidations.** *J Health Econ* 2006; **25**:1049–1068.
16. Weaver M, Deolalikar A. **Economies of scale and scope in Vietnamese hospitals.** *Soc Sci Med* 2004; **59**:199–208.
17. Rowley JT, Anderson RM. **Modeling the impact and cost-effectiveness of HIV prevention efforts.** *AIDS* 1994; **8**:539–548.
18. Watts C, Kumaranayake L. **Thinking big: scaling-up HIV-1 interventions in sub-Saharan Africa.** *Lancet* 1999; **354**:1492.
19. Johns B, Tan-Torres TT, WHO–CHOICE. **Costs of scaling up health interventions: a systematic review.** *Health Policy Plan* 2005; **20**:1–13.
20. Over M. **The effect of scale on cost projections for a primary health care program in a developing country.** *Soc Sci Med* 1986; **22**:351–360.
21. Broomberg J, Soderlund N, Mills A. **Economic analysis at the global level: a resource requirement model for HIV prevention in developing countries.** *Health Policy* 1996; **38**:45–65.
22. Wilkinson D, Floyd K, Gilks CF. **National and provincial estimated costs and cost effectiveness of a programme to reduce mother-to-child HIV transmission in South Africa.** *S Afr Med J* 2000; **90**:794–798.
23. Bertozzi S, Gutierrez JP, Opuni M, Walker N, Schwartlander B. **Estimating resource needs for HIV/AIDS health care services in low-income and middle-income countries.** *Health Policy* 2004; **69**:189–200.
24. Kumaranayake L, Kurowski C, Conteh L. **Analysis of the costs of scaling-up priority health interventions in low and selected middle-income countries.** In: *Macroeconomics and health: investing in health for economic development*. Geneva: World Health Organization Commission on Macroeconomics and Health; 2001.
25. Opuni M, Bertozzi S, Bollinger L, Gutierrez JP, Massiah E, McGreevey W, *et al*. **Resource requirements to fight HIV/AIDS in Latin America and the Caribbean.** *AIDS* 2002; **16 (Suppl. 3)**: S58–S65.
26. Marseille E, Kahn JG, Mmiro F, Guay L, Musoke P, Fowler MG, *et al*. **Cost-effectiveness of single-dose nevirapine regimen for mothers and babies to decrease vertical HIV-1 transmission in Sub-Saharan Africa.** *Lancet* 1999; **354**:803–809.
27. Rely K, Bertozzi SM, Avila-Figueroa C, Guijarro MT. **Cost-effectiveness of strategies to reduce mother-to-child HIV transmission in Mexico, a low-prevalence setting.** *Health Policy Plan* 2003; **18**:290–298.
28. Pitter C, Kahn JG, Marseille E, Lule JR, McFarland DA, Ekwaru JP, *et al*. **Cost-effectiveness of cotrimoxazole prophylaxis among persons with HIV in Uganda.** *J Acquir Immune Defic Syndr* 2007; **44**:336–343.
29. Forsythe S, Arthur G, Ngatia G, Mutemi R, Odhiambo J, Gilks C. **Assessing the cost and willingness to pay for voluntary HIV counselling and testing in Kenya.** *Health Policy Plan* 2002; **17**:187–195.
30. Uys L, Hensher M. **The cost of home-based terminal care for people with AIDS in South Africa.** *S Afr Med J* 2002; **92**:624–628.
31. Deghaye N, Pawinski RA, Desmond C. **Financial and economic costs of scaling up the provision of HAART to HIV-infected health care workers in KwaZulu-Natal.** *S Afr Med J* 2006; **96**:140–143.
32. Boulle A, Kenyon C, Skordis J, Wood R. **Exploring the costs of a limited public sector antiretroviral treatment programme in South Africa.** *S Afr Med J* 2002; **92**:811–817.
33. Kumaranayake L, Vickerman P, Walker D, Samoshkin S, Romantzov V, Emelyanova Z, *et al*. **The cost-effectiveness of HIV preventive measures among injecting drug users in Svetlogorsk, Belarus.** *Addiction* 2004; **99**:1565–1576.
34. Terris-Prestholt F, Watson-Jones D, Mugeye K, Kumaranayake L, Ndeki L, Weiss H, *et al*. **Is antenatal syphilis screening still cost-effective in Sub-Saharan Africa?** *Sex Transm Infect* 2003; **79**:375–381.
35. Vickerman P, Terris-Prestholt F, Delany S, Kumaranayake L, Rees H, Watts C. **Are targeted HIV prevention activities cost-effective in high prevalence settings? Results from a sexually transmitted infection treatment project for sex workers in Johannesburg, South Africa.** *Sex Transm Dis* 2006; **33 (10 Suppl.)**:S122–S132.
36. Terris-Prestholt F, Kumaranayake L, Obasi A, Cleophas-Mazige B, Makhkha M, Todd J, *et al*. **From trial intervention to scale-up: Costs of an sdolescent sexual health program in Mwanza, Tanzania.** *Sex Transm Dis* 2006; **33 (10 Suppl.)**: S133–S139.
37. Forsythe S, Mangkalopakorn C, Citwarakorn A, Masvichian N. **Cost of providing sexually transmitted disease services in Bangkok.** *AIDS* 1998; **12 (Suppl. 2)**:S73–S80.
38. Thaineua V, Sirinirund P, Tanbanjong A, Lallemant M, Soucat A, Lamboray JL. **From research to practice: use of short course zidovudine to prevent mother-to-child HIV transmission in the context of routine health care in Northern Thailand.** *Southeast Asian J Trop Med Public Health* 1998; **29**:429–442.
39. Dandona L, Sisodia P, Ramesh YK, Kumar SG, Kumar AA, Rao MC, *et al*. **Cost and efficiency of HIV voluntary counselling and testing centres in Andhra Pradesh, India.** *Natl Med J India* 2005; **18**:26–31.
40. McConnel CE, Stanley N, du Plessis JA, Pitter CS, Abdulla F, Coovadia HM, *et al*. **The cost of a rapid-test VCT clinic in South Africa.** *S Afr Med J* 2005; **95**:968–971.
41. Carrara V, Terris-Prestholt F, Kumaranayake L, Mayaud P. **Operational and economic evaluation of a NGO-led sexually transmitted infections intervention in north-western Cambodia.** *Bull WHO* 2005; **83**:434–442.
42. Terris-Prestholt F, Kumaranayake L, Foster S, Kamali A, Kinsman J, Basajja V, *et al*. **The role of community acceptance over time for costs of HIV and STI prevention interventions: analysis of the Masaka Intervention Trial, Uganda, 1996–1999.** *Sex-Transm Dis* 2006; **33 (10 Suppl.)**:S111–S116.
43. Dandona L, Sisodia P, Kumar SG, Ramesh YK, Kumar AA, Rao MC. **HIV prevention programmes for female sex workers in Andhra Pradesh, India: outputs, cost and efficiency.** *BMC Public Health* 2005; **5**:98.

44. Guinness L, Kumaranayake L, Bhuvaneswari R, Sankaranarayanan G, Vannela G, Raghupathi P, *et al*. **Does scale matter? The costs of HIV prevention interventions for commercial sex workers in India.** *Bull WHO* 2005; **83**:747–755.
45. Chandrashekar S, Kumaranayake L, Ramesh BM, Blanchard J, Alary M. **Learning effects on the costs of phased scale-up implementation of targeted HIV prevention among high risk populations in Karnataka, India**. Presented at the *International AIDS Conference*. Toronto, Canada, 13–18 August 2006 [Abstract CDD1243].
46. Kumaranayake L, Chandrashekar S, Washington R, Alary M. **The economics of STI provision in scaling-up HIV prevention among high risk populations in Karnataka, India**. Presented at the *International AIDS Conference*. Toronto, Canaca, 13–18 August 2006 [Abstract MoPe0676].
47. Marseille E, Dandona L, Marshall N, Gaist P, Bautista-Arredondo S, Rollins B, *et al*. **HIV prevention costs and program scale: data from the PANACEA project in five low and middle-income countries.** *BMC Health Serv Res* 2007; **7**:108–116.
48. Kumaranayake L, Watts C. **HIV/AIDS prevention and care interventions in Sub-Saharan Africa: an econometric analysis of the costs of scaling-up.** *S Afr J Econ* 2000; **68**:1012–1033.
49. Kumaranayake L. **Production and cost functions for voluntary counselling and testing: what is the scale effect? A study from Malawi.** Presented at the *International AIDS Economics Network Meeting*. Bangkok, Thailand, July 2004.
50. Terris-Prestholt F, Vyas S, Kumaranayake L, Mayaud P, Watts C. **The costs of treating curable sexually transmitted infections in low- and middle-income countries: a systematic review.** *SexTransm Dis* 2006; **33 (10 Suppl.)**:S153–S166.
51. Guinness L, Kumaranayake L, Hanson K. **A cost function for HIV prevention services. Is there a 'U' shape?** Presented at the *6th World Congress on Health Economics, International Health Economics Association*. Copenhagen, Denmark, July 2007.
52. Kumaranayake L, Watts C, Kurowski C, Conteh L. **Estimating the infrastructure requirements for an expanded response to HIV/AIDS in low and middle-income countries.** Presented at the *XIVth International AIDS Conference*. Barcelona, July 2002 [Abstract TuPeE5106].
53. Johns B, Baltussen R. **Accounting for the cost of scaling-up health interventions.** *Health Econ* 2004; **13**:1117–1124.
54. Nordstrom A, Tan-Torres Edejer T, Evans D. **What will it cost to attain the health MDGs?** *Bull WHO* 2007; **85**:246.