6 Outputs

Introduction

6.1 In the main part of this chapter we consider the application to new or different direct measures of government output of the principles described in Chapter 4. The last section of the chapter considers the implications for productivity measurement. This is on the basis that the move to direct output measures is justified both by its intrinsic merits and by the obligations placed on the United Kingdom by the Eurostat procedures.

6.2 The chapter deals with output measures in general, illustrating the issues by reference to specific spending areas. Individual spending areas are covered in depth in Chapters 8-11. In Chapter 4, we considered the issues theoretically, starting from the desired objective and seeking to derive principles to be applied. This approach is, we hope, of value to people grappling with these problems in other countries. In this chapter, we turn explicitly to the UK situation. It recognises that ONS does not start from a zero base, but from an existing set of measures covering more than a decade. We believe that these 'first generation' estimates have been pioneering, but that, as improvements become possible, they will become more refined. So, starting from existing ONS practice, the present chapter examines how we can move towards implementing more fully the principles set out in Chapter 4.

6.3 The fact that we are not starting from a zero base means that we have to tailor our recommendations to the point of departure. This is well illustrated by the issue of quality adjustment. In Chapter 4, we argued that the long-run goal is to incorporate an appropriate measure of quality change (Principles A and B). We discuss further the approach to devising quality measures and advise a threshold of acceptability for their use. We also recognise that it may take time to devise robust measures of quality and that it may be some time before the threshold of acceptability is reached. This applies to spending areas where no quality adjustment is currently made. Where, as in the case of Education, a quality adjustment is already in place, the question becomes one of comparing any proposed quality measure with that already in place. It is conceivable that the new measure would fall below the threshold if starting from scratch but that it represents a sufficient improvement over the existing measure to warrant its introduction.

6.4 Implementation of the principles set out in Chapter 4 means that we have to consider separately:

- what changes, if any, should be made in those spending areas where direct measures have already been introduced; and

- whether direct output measurement should be extended to new spending areas.

We begin with the first question, and then turn (much more briefly) to the extension of direct output measures to new functions.

Introduction of New Indicators for Existing Functions

6.5 **Recommendation 6.1:** we recommend current direct measures of output should be improved, where needed, by:

a) widening the coverage of output volume indicators for each function;

b) increasing the level of detail at which output indicators are measured;

c) adopting a more reliable data source;

d) revisions of the weighting process;

e) replacing activity indicators with output measures that reflect changes in quality or outcome attributable to a unit of output;

f) introducing or revising an overall quality adjustment;

g) improving timeliness and in-year indicators; and

h) improving UK coverage by making full use of measures from Scotland, Wales and Northern Ireland.

6.6 For example, in the case of the new measures of general government Health output introduced in the *Blue Book 2004*, there were improvements in several respects. There was a large increase in the degree of detail (1,700 treatment types in place of 16); the new data capture a wider range of NHS activity (adding NHS Direct and Walk-in clinics); the new source data are fully reconcilable with audited accounts; weightings were updated; and both annual and quarterly estimates are timelier. Improvements were therefore made under headings a, b, c, d and g.

6.7 The conditions under which new output measures should be introduced are a matter for careful judgment. On the one hand, ONS is rightly concerned to improve its methods, and to take advantage of advances in methodology and data availability. On the other hand, the measurement of government output is a matter of considerable public sensitivity, and the public is often concerned about departures from previous practice. In our view, ONS, with its Code of Practice and the other elaborate arrangements surrounding national statistics, has achieved a good balance. We were happy with the way in which the Health changes were made, which followed the criteria set out in the Interim Report, which we re-iterate in Recommendation 6.2.

6.8 **Recommendation 6.2:** We recommend that ONS should be satisfied on the following conditions, before introducing a replacement output measure:

a) there should be evidence of significant improvement in one or more of the directions listed above, giving particular emphasis to completeness of coverage and to measures that reflect quality change;

b) an analysis should have been carried out of the relevant output data from past years, with sensitivity testing for possible future changes;

c) the validity of the proposed measure should be tested by those with expert knowledge of the relevant function; and

d) there is assurance of the likely continuation of the key data sources.

6.9 In the case of Health, ONS demonstrated to the peer review process that there are significant advantages to the new, more detailed, indicator. Before adoption, the proposed new method was compared with the existing estimates over the period 1996 to 2003. The new estimates reflected the increased availability of detailed unit cost and activity data from the Department of Health, and the new measure was devised in close cooperation with Department of Health officials.

6.10 Changes in output indicators may also be necessary to take account of changes in government services or in the machinery of government. We have already referred to the extension of health output measurement to cover new services such as NHS Direct. ONS evidently needs to monitor government initiatives, in order to assess any implications for output measurement. Changes in the organisation of government may also affect output measurement, and this is an aspect of the switch from the (output=input) convention that seems to have received inadequate attention. As has happened in recent years, an activity costing £X million may be transferred from a function currently measured on an (output=input) basis to a function with a direct output measure. Under the 'old' heading, both input and output are reduced by £X million; under the 'new' heading, input increases by £X million, but the impact on output depends on the construction of the direct output indicator.

6.11 **Recommendation 6.3:** we recommend that ONS should monitor changes in government services, and in the machinery of government, with regard to their impact on direct output measurement and the need to add further output indicators or to transfer activities.

Treatment of Collective Elements

6.12 The classification of government output into individual and collective services follows in general the COFOG classification (Appendix B), but it is recognised (ESA 1995, paragraph 3.85) that part of the spending under the functions identified there as 'collective' should be treated as individual. The collection of household refuse, for instance, is an individual service, unlike other elements of collective environmental protection. As already noted, ONS has embarked on the introduction of direct output measures for Public Order and Safety, which is classified as a collective service.

6.13 The reverse is also true. Functions classified as 'individual' also provide outputs that are better seen as collective. A public health campaign against smoking is an example. As was noted in the chapter on Public Order and Safety in the Interim Report, this can be said to apply to a number of activities under this heading, such as fire prevention advice, crime prevention, and indeed the criminal justice system as a whole. Here we are concerned with the implications for output measurement.

6.14 In order to achieve an A grade in the Eurostat Handbook classification, a direct output measure appears to be necessary for collective services, just as for individual services. The Handbook recognises 'the difficulty in defining the output of collective services' (p 35), but does not provide much guidance as to how they are to be resolved. In our view, the considerable obstacles to the development of output indicators mean that it is acceptable to settle for a B classification for the collective elements of spending on functions such as public health, and fire and crime prevention.

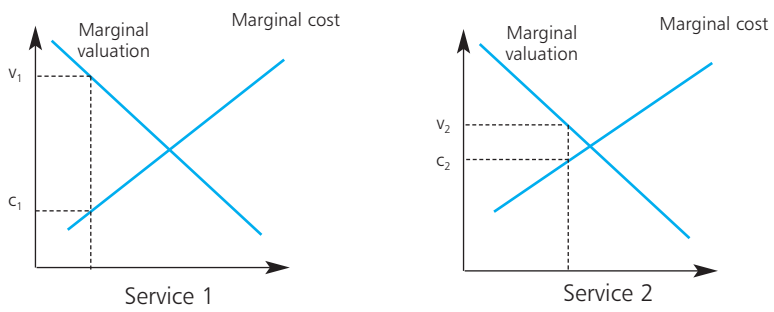
6.15 This, however, leaves open the question as to how this should be achieved. One approach, used by ONS at present, is to assume that areas like public health campaigns, with no direct output measure, grow pro rata to all areas where outputs are measured. This is not an ideal assumption. Another option is to divide expenditure into two, applying direct output measures to part and reverting to an input method for the collective part. But, in order to qualify as a B method, we would need to estimate 'the volume of each indicator separately, taking quality changes of inputs into account'. An alternative offered in the Handbook's B classification for collective services is to use a volume index of activity. In that case, we use a direct method throughout, but are willing to accept activity indicators for areas like fire prevention. We consider that it is acceptable to use an activity indicator for the collective elements included in a function classified overall as individual.

6.16 **Recommendation 6.4:** we recommend that collective services should be measured by the appropriate international standard, i.e. either a volume index of activity or the volume of inputs, aiming to satisfy Eurostat's requirements for a 'B' method, taking account of quality change of inputs. The same approach should be used for collective elements included in a function classified overall as 'individual', rather than assuming their output changes pro rata to other areas for which there are direct output measures.

Weighting

6.17 An important issue in the construction of a single output indicator for a function is the choice of weights for the different elements that make up the aggregate. Some algebra is necessary. Let us denote by $Y_i(t)$ the output i in a particular function: for example, elective inpatient care or GP consultations. The number of GP consultations in year t, say 2003, is then compared with that $Y_i(1)$ in a base year numbered 1, say 1996. We can then draw conclusions about the growth or fall in GP consultations. But we wish to combine this with the number of inpatients. As required in the SNA, the growth of each item 'must be weighted by their economic importance as measured by their values'. If an inpatient episode is five times as valuable as a GP consultation, then its growth rate is weighted correspondingly. In the case of marketed output, the value may be taken as the purchase price or, equivalently (under certain assumptions, notably competitive supply), the marginal cost. If we denote the marginal cost of output i in year t by $c_i(t)$, then the cost-weighted output in year t, using cost weights in the reference year 1, is $\sum_i c_i(1)\, Y_i(t)$. It should be noted that this requires *marginal* costs: i.e. the additional cost of an extra GP consultation, not the cost found by dividing total cost by the total number of consultations (average cost). In practice, the costs used in output measurement are average costs, and this is the first shortcoming of the method. The Centre for Health Economics / NIESR / NPCRDC project on measuring output and productivity in the UK health care sector (see paragraph 8.49) is investigating the relation between marginal and average costs.

Figure 6.1



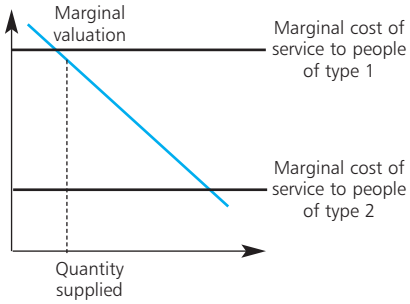
6.18 For non-marketed output, we have the further problem that the output valuation cannot be observed, and there is no reason why it should coincide with the marginal cost. This is illustrated in Figure 6.1. (This graph is based on the earlier Figure 4.1, but allows for a rising marginal cost of supply.) In neither case is the output being supplied to the level where the marginal value of an additional unit is equal to the marginal cost of supply. The quantity supplied is indicated by the dotted vertical line, associated in one case with marginal cost of c_1 and marginal valuation v_1, and in the other case with c_2 and v_2, respectively. Ideally, we would like to use the marginal valuations, denoted by v_i, so that weighted output is $\sum_i v_i(1)\, Y_i(t)$. However, this requires indicators of output values that are comparable across the different i. In some cases, such indicators may be at hand. This may be the case where there are marketed services that provide an alternative. In the case of road use, for example, we may attach value weights to passenger miles and to freight tonne miles, based on the alternative costs of using rail. It should be noted that this valuation would not take into account any difference in the cost imposed on the road system (this would be c_i), so that even if heavier lorries damaged roads more per tonne this would not affect the output (although it would show up in higher maintenance inputs). A second example would be the provision of personal care by social services, where there is a parallel market. If people are willing to pay p_i for daily care, then this can be used for the marginal valuation. It is sometimes objected that the parallel market is artificial, dominated by public purchasing, or that, in other cases, prices are monopolistic. The basis for the pricing is irrelevant; it is sufficient that consumers are willing to pay those prices (providing, of course, that there is a sufficiently wide market to obtain reasonable price observations). The prices may not be equal to marginal costs, but that does not matter if we are seeking to measure v_i, not c_i.

6.19 In other cases, we may not be able at present to apply estimates of v_i. In the case of health care and adult social services, this is the subject of on-going research. So, for the present, the only feasible approach appears to be to continue to use cost weights. Nonetheless, it is clear from Figure 6.1 that this gives a rather different pattern of weights. In the case shown, they are negatively correlated. An expansion of service 1 will be given less weight under cost weights and more weight if a marginal valuation is applied.

6.20 In considering output measures, we have argued for a fine differentiation. Should the same argument be applied to services whose costs differ? Suppose, for example, in the field of Personal Social Services, people require more hours of care in order to achieve a specified level of functioning. A service is provided by a local authority to ensure that people achieve that level of functioning. Should we differentiate people according to cost? This situation is illustrated in Figure 6.2 for the case where the demand for the service is the same for the two types of person. Following the parallel with the private sector discussed in paragraphs 4.6-4.10, where the service is specified in that way, the answer is that we should not differentiate. Rural and urban postal deliveries cost different amounts, but the output is the total number of letters delivered. To differentiate between people of types 1 and 2 in Figure 6.2 would mean that changes in the mix of clients would cause cost-weighted output to rise or fall, whereas the marginal valuation is the same in both cases.

6.21 The case discussed in the previous paragraph referred to the situation where the demands were identical. It could of course be that people of type 2 live next door to their grown-up children and have less need for care, but this is a different issue. We are then saying that there is variation around the 'averaged' marginal valuation per unit shown in Figure 6.2. A service may raise output by reallocating its services. This may require increased, but unrecorded (in the National Accounts) input from the family, an example of the interface between the paid and unpaid economy to which we referred in Chapter 1 (paragraph 1.3).

Figure 6.2



Joint Products and Joint Producers

6.22 Government services may produce more than one output. Schools are educating children and providing childcare. These are not joint products like wool and mutton, since parents cannot opt for only one part of the package. The products cannot be separated on the cost side, so that there is no way of combining them in a cost-weighted output index, but if we use marginal value weights (for example, applying the market price of childcare), then in principle the values can be added. Lest this appear an over-radical proposal, we should note that its impact is limited to the relative weights applied to output indicators. Placing a value on childcare would increase the relative weight on the number of younger children.

6.23 The inverse problem is that government services may be a joint input into the final outcome. Many factors contribute to outcomes in health, education, law and order, and social protection. We have referred on a number of occasions to the problem of identifying the contribution of public spending. Here we are concerned with the attribution of changes in outcome to different spending functions. How much of reduced crime is due to the police force? How much is due to the courts? And should part be attributed to education or children's social services?

6.24 In general, this is a major problem. It may be very difficult to calculate the relative contributions. For this reason, it may be necessary to remain with cost weights. On this basis, the attribution is of the change in the contribution to outcome, and it may be reasonable to assume as a first approximation that the relative contributions of different services are constant over time.

6.25 **Recommendation 6.5:** we recommend that the ideal approach to developing a single aggregate output measure for a function is to weight together different elements by weights based on their marginal valuation. This requires indicators of output values that are comparable for different components. If that is not possible, it may be necessary to use marginal costs. In practice, average costs may be the only information available. Cost weights may be most appropriate where an outcome is affected by several government services and it is not possible to calculate the value of relative contributions.

Quality Adjustment

6.26 In Chapter 4 we discussed three ways in which the measurement of quality in the National Accounts could be approached.

- First, we can differentiate services.
- Secondly, we can define the volume measure in terms of the degree of success.
- Thirdly, the volume measure may be based on the level of activity but the contribution to outcomes introduced in the form of a quality adjustment.

6.27 **Recommendation 6.6:** we recommend that ONS choose on a case by case basis whether to measure quality by differentiation of service, success of activity or attributable contribution to outcome, having regard to:

- the nature of the service;
- the extent to which the service is, or should be, differentiated; and
- the degree to which the change in outcome can be directly and confidently attributed to the service concerned.

6.28 In the case of Health, for example, we can see the effect of the first method. By adopting a detailed classification of outputs, it is possible to represent quality change by structural change within the aggregate. Suppose that there are two treatments. One is of a higher quality and is more expensive. If the output indicator combines the two volume measures with weights according to their cost, then a shift towards increased use of the higher quality treatment will be properly recorded. There will be an increase in expenditure, and a corresponding increase in output. Thus, the move to a more detailed treatment classification (see Chapter 8) has been a step towards taking account of this kind of quality change.

6.29 But a cost-weighted index cannot capture all quality improvements, as may be seen from the example of an improvement in technique such that a lower cost treatment (eg day surgery) can assure the same (or better) outcome as a more expensive treatment. This is recorded as a reduction in output, when no change has occurred (and there has been a saving in inputs). One response to this problem is to seek to weight activities not by cost but by an indicator of the increment in quality of outcome, such as Quality Adjusted Life Years. In Education, the same role may be played by performance on tests. However, as discussed in the section on weighting (see paragraphs 6.17-6.19), this requires that the units be comparable in level terms.

6.30 The second method is based on a simple repackaging hypothesis. Higher-grade petrol gives you 10 per cent more miles to the gallon, so it is equivalent to 1.1 units of the lower grade. In the case of government output, the same can apply. A hip operation may use a superior technique that gives you an extra 10 per cent expected use of the new joint; change in the school methods may give 10 per cent improvement in examination performance. In this case, no comparison of levels across activities is necessary. All that is required is that the *change* in the quality is correctly measured.

6.31 In the third case, the adjustment could be made in two different ways. The first is to choose a particular indicator, or set of indicators, measured regularly, and to apply these on a formulaic basis in making quality adjustments. As discussed in Chapter 9 (see paragraph 9.28), the quality of education could be made a function of the annual results in pupil attainment. An improvement in key stage results, relative to the expected transitions, would be translated into an increase in productivity. The second approach is to employ a range of evidence to make a judgement about the rate of quality change. This approach has the disadvantage that it involves judgements being made ex post (rather than in the ex ante selection of indicators), but the advantage that it is less at risk from changes in examination structures or school inspection.

6.32 Quality adjustment of government output is a particularly challenging area, given the intrinsic difficulty and the relatively limited experience – both in the United Kingdom and in other countries – with such adjustments. It is to be expected that the initial measures will be approximate. It will be necessary to make do with partial information, applying quality adjustments from one part of the service to another for which information is not available (as is the case with volume indicators). At the same time, the greater degree of subjectivity in making quality adjustments, compared with volume measures, and the diversity of dimensions to quality, mean that we should apply a higher standard when judging acceptability. It is essential that the measures employed in the National Accounts should command support from appropriate service experts and from end users. While we believe that the end goal should be to make quality adjustments to all services, we consider that a relatively high threshold should be set before implementation. ONS may well wish to publish quality information in the first instance in some other way – for example, in an experimental series or in an explanatory article or a productivity article, before adjustments are introduced into the National Accounts.

6.33 **Recommendation 6.7:** we recommend that ONS should give priority to work on quality adjustments, but consider that a relatively high threshold should be set for their introduction into the National Accounts; in particular, ONS should not introduce quality adjustments until it is assured that the dimensions covered are sufficiently representative.

Timing and Attribution

6.34 In the previous chapter, we have presented a detailed analysis of the issues of timeliness and periodicity with regard to government inputs. Many of the same issues apply to output indicators. In brief:

a) Much of the information relates to financial years, and has to be converted to a calendar-year basis for National Accounts purposes.

b) Quarterly data are necessary to make this adjustment, and are necessary in their own right to produce quarterly GDP estimates. Quarterly GDP estimates are of lower quality than annual figures and make use of extrapolation. (See *Economic Trends*, November 2001).

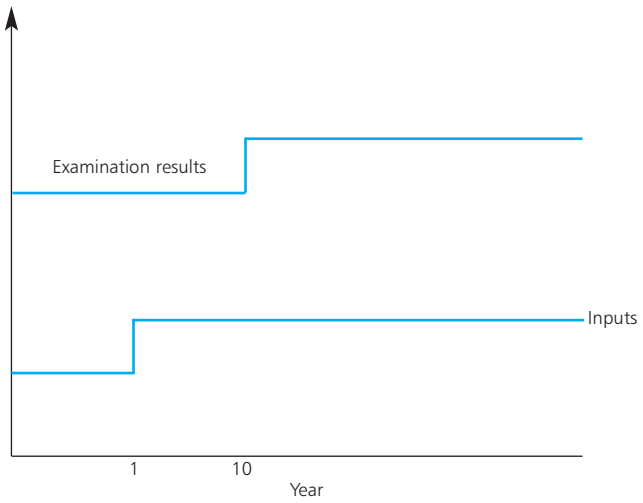
c) There are considerable delays in the supply of financial year data; although the moves towards faster closure of accounts in both central and local government should help reduce these over the next few years.

6.35 The issue of timeliness is a general one for the National Accounts, but we are concerned here specifically with government output. A priori, one would expect that output growth is more stable than private marketed output and that changes can be forecast from legislative and known budget changes. It is not therefore evident that the improvement of quarterly series for government output should be a first call on ONS resources. Best use should be made of timely, audited data, but priority should be given to improving the annual estimates.

6.36 **Recommendation 6.8:** we recommend that ONS should seek to improve the timeliness of annual estimates of outputs of public services (as a greater priority than more accurate estimates for quarterly outputs).

6.37 Data need not only to be timely but also correct, as far as possible, in their attribution of the timing of output. The fact that output is now being measured separately from input means that we have to consider its time path. To which period should output be attributed? The issue may be illustrated by reference to Education. No one would seriously suggest that school output should be recorded as falling during the summer holiday; there should simply be a smoothed annual series. But how should we treat an improvement in education that takes the form of an investment paying off in recorded examination performance at the age of 16? (See Figure 6.3.) Increased inputs take place in year 1 but the improved examination results first occur in year 10. It may be possible to replace the examination measure by use of earlier test scores, and the issue of attribution of timing may be one consideration influencing the choice between different measures. But it is conceivable that the increase in inputs is indeed an investment, the effect not being observable until later. The interpretation of these time paths depends on the particular context and subject matter, and should be discussed in the productivity articles.

Figure 6.3



Output and productivity measurement

6.38 The existence of direct output measures means that they can be divided by the corresponding input measure to derive an indicator of the change in government productivity. However, as we have emphasised in earlier chapters, this is not simply a matter of arithmetic. A number of pre-conditions have to be satisfied. We end therefore the two chapters on inputs and outputs with a resume of the points to be considered before drawing conclusions about government productivity.

6.39 The first requirement is self-evidently that the input indicators satisfy the criteria set out above. The input indicators should be comprehensive, notably including capital services, so as to provide a basis equivalent to that in total factor productivity calculations (see paragraph 2.33). The inputs should be properly deflated, using deflators that satisfy criteria such as those set out in Table 5.1. Appropriate account should be taken of quality change, such as better-trained workers.

6.40 Equally self-evidently, the output indicators should satisfy the criteria set out earlier. They should meet the conditions set out in Principle D and Recommendation 6.2. These include adequate coverage of the range of services, appropriate allowance for quality change, and full geographic coverage.

6.41 It is not however sufficient that the above conditions are satisfied. As we have emphasised in Principle H (Triangulation), we need to look at the two numbers in conjunction. The output and input indicators may each be satisfactory in their own terms but the ratio may not be satisfactory. This may happen on account of measurement errors and their joint distribution. The margin of error (see paragraph 4.72) may be acceptable for each series taken independently but unacceptable for the ratio. There may be issues surrounding the timing and attribution, as just discussed, where the top and bottom of the ratio are affected differently, with consequences that are serious for the implied productivity estimate.

6.42 More generally, we have stressed the need for interpretation. No single ratio, however carefully constructed, can fully capture the performance of complex public services with multiple objectives. As we made clear at the outset, there is a difference between National Accounts estimates of output and performance indicators for the management of the public services.