

A Review of the Atkinson Review

Richard Murray

Swedish Agency for Public Management

Summary

This paper takes a look at the recommendations of the *Atkinson Review, Final Report – Measurement of Government Output and productivity for the National Accounts* (Atkinson, 2005). The Review forms part of a European Union wide effort to measure output and productivity in the public sector. There have been earlier efforts to measure productivity in the public sector in the US, in Sweden and other countries. At last it seems that these efforts are gaining recognition. Instrumental in settling on a standard for measurements has been to integrate these output measures into the national accounts. However, the national accounts may serve different purposes and one has to decide which one should guide the measurements. There is now a reasonable consensus on that GDP should serve as a measure of production. Some ever present issues in the measurement of government output such as quality still haunts the finding of a sound basis for measurements. This is most apparent in the Atkinson Review, which does not seem to have comprehended the production perspective on national accounts and GDP. In this paper some of the Atkinson recommendations are discussed and criticized. Critique is levied at the recommendation that as far as possible try to base weights on marginal social valuation of outputs and to regard an increase in that valuation as an increase in output. Over all, though, the Atkinson Review codifies principles that seem sound and are almost identical with those used in Sweden already in 1986 (Ohlsson, Blohm and Murray, 1987) and onwards (Murray, 1996). The paper ends with some short comments on problems overlooked.

Introduction

Productivity in the public sector is finally recognised as something of importance for the assessment of the production capacity of a nation. It has been in the air for some time. In 1993 the UN commission on national accounts published a new set of recommendations, System of National Accounts 1993 (SNA 93). In it standards were proposed for the measurement of government outputs. This was followed by similar recommendations by the Commission of the European Union in 1996, European System of Accounts (ESA 95). And eventually a Handbook on Prices and Volume Measures in National Accounts was published by Eurostat in 2001, followed by a European Union Commission decision in 2002 that all member states should include output measures along the recommended standards in their national accounts, starting with year 2006. Statistics Sweden (Statistiska centralbyrån, SCB) and other national, statistical offices around Europe are now working to implement this directive and their results will soon be published.

In the United Kingdom a special commission was set up to review present practices to handle government outputs in the national accounts and to recommend developments both in the short and the long run. It has been headed by Tony Atkinson and has been working for two years. It presented its findings last year in the report *Atkinson Review: Final report – Measurement of Government Output and Productivity for the National Accounts*, Palgrave, Macmillan, 2005.

This paper is a review of the Atkinson Review, with some comments on SNA 93 and ESA 95. Basically, I find the ultimate recommendations of the Atkinson Review both practical and sound. There are, however, arguments and recommendations, especially for the long run, that I want to highlight, since I think that they point in the wrong direction. The arguments I am referring to are fundamental and illustrate the difficulty in finding a sound basis for the measurement of non-market goods. The arguments reflect a debate that has been going on since the first efforts were made to introduce output measures in the government sector in connection with program budgeting (Planning, Programming, Budgeting Systems, PPBS) in the late 1950s. That debate is about whether to measure output or outcome.

Background

Looking back it is depressing that progress in understanding the production of public services has been so slow. In the market sector there is a long tradition of studying production functions, demand for inputs, average and marginal cost functions, elasticities of supply, productivity and technical progress. The non-market sector has gone largely unnoticed. In part this can be explained by general difficulties in measuring the output of services, whether public or private. But in part it must be explained by a completely different perspective on public services apart from private services. Resource use for the production of public services has not been regarded as inputs into a production process, but as an end in itself, in the form of public consumption. Consequently, the production activity in the government sector has not been recognized.

The PPBS approach to managing government programs in the US involved setting targets for and measuring the output of government agencies. Stemming from the PPBS approach a program for output and productivity measurements was set up and conducted by the Bureau of Labor Statistics from 1973. It covered ultimately the outputs of some 2/3 of federal civilian employees. The output and productivity measurements were, however, never integrated into the national accounts of the US. The program was discarded in 1994 (Fisk and Forte, 1997).

In the United Kingdom, as probably the only country in the world, direct measures of government output were an integrated part of national accounts in the 1950s and early 1960s. After critique for being crude and arbitrary they were discarded. Since 1999 direct measures of government output have again been introduced and integrated in the national accounts and at present cover some two-thirds of government outputs. (Atkinson 2005, p.14 and p.17)

In 1975 T.P. Hill addressed problems of measurements of government non-market services and suggested a series of principles (Hill, 1977). His work had been commissioned by the forerunner of Eurostat but led to no results in the form of data collection at the time.

In 1982 the former head of Statistics Sweden, Ingvar Ohlsson, formed a research group within The Expert Group on Public Finances (Expertgruppen för studier av offentlig ekonomi, ESO)

with the aim of throwing light on government outputs, productivity and the users of government outputs. Ohlsson had written his PhD thesis on national accounting systems in 1961 and for many years wanted to extend output measurements to the government sector (Ohlsson, 1961). In 1985 (English translation in 1986) the research group published a first collection of results for the period 1960-1980 (Ohlsson, et.al., 1986). A summary of the findings was published by NBER in 1992 (Murray, 1992). Also at that time a new round of productivity studies, commissioned by ESO, was published (Murray, 1994). It contained an enlarged sample of public services, comprising 75 percent of the total of central and local government and covered the period 1980-1992. In it various techniques for measuring productivity were compared, the traditional national accounts technique with the non-parametric technique of data envelopment analysis. Non-parametric techniques had been developed by Färe (Färe, 1988) and others building on the seminal work of Farrell (Farrell, 1957). Among others Bjurek, Blank, Försund, Hjalmarsson and Luoma had performed non-parametric studies of non-market services in the 1990s (See the list of references).

It is only now, that measurements of output and productivity in the public sector is starting to gain recognition and to be encouraged by the broader economist profession. This paper does not try to answer the question why the process of learning has been so exceptionally slow in this field. A range of hypotheses will be left open: Are economists so focused on Pareto efficiency that they cannot discuss degrees of efficiency? And since technology in the minds of economists can only progress, is it because there appears to be negative productivity development in government production that they can't cope with it? Or is it as simple as that economists shun everything that is not emulated on a market? Have national accounts statisticians been reluctant to deal with output and productivity because they are trained within a national accounts tradition of macro economics? Have government auditors been so obsessed with goal achievement that they have ignored to question both the goal and the cost of achieving it?

Instead of trying to find the answers to the question of the slow learning process – which is most likely to be the more important issue – I will review the latest contribution in the field of output and productivity measurements in the public sector – The Atkinson Review – to assess where we stand today. This may also serve as an illustration to the slow learning process, since its findings– with minor differences – are precisely the same as those of the Ingvar Ohlsson group in 1985. Perhaps the clue to the slow learning process lies in the fact that none of the Swedish contributions are referred to in the Atkinson report, despite an otherwise lengthy list of references.

It is because the European Commission demands that the EU member states from the year 2006 start measuring government output and include those measurements in the national accounts leading up to an estimate of Gross Domestic Product that statistics offices around Europe are now doing it. Rather than an awareness by economists and statistical offices of the intrinsic value of doing it.

So, why is the commission so keen on having government output measured? The answer is simple: because member fees, distribution of regional funds and the Maastricht convergence criteria for fiscal management are all based on assessments of GDP per capita. Therefore, these accounts have to be compiled “on the basis of unique principles that are not open to different interpretations.” (EU Commission Decision of 17th of December 2002) Especially the handling of government output differs among member countries, with countries inserting various productivity assumptions – from zero to two percent per year. The Eurostat Handbook

argues that harmonising assumptions about productivity would not make the situation any better. “The more different the developments of productivity among member states, the less comparable are the results from using the same productivity change assumption.” (Eurostat, 2001, paragraph 3.1.2.1) Just to show the impact: adjusting the growth of public services by output measurements for the 1970 decade reduced the growth of GDP from 2 to 1.5 percent per year, due to an overall productivity decline of 1,5 percent per year during that decade (Murray, 1987 a).

It is interesting to note that similar motives guided the development of national accounts after WWII. At that time it was membership fees to the UN and the OECD and the distribution of foreign aid in the Marshall Plan that prompted harmonisation of accounting principles.

What does GDP measure?

The measurement of government output can be done in various fashions, depending on what GDP should measure. Neither in SNA 93 nor in the EU directive is there any explanation only an outright proclamation of a *production perspective*.

National accounts may serve several purposes. Ingvar Ohlsson (1961) distinguishes four:

1. Analysis of the generation of income and the connection between different sectors of the economy regarding economic activity. This analysis serves the purposes of understanding and controlling business cycles.
2. National budget planning is closely related to the first mentioned purpose.
3. Analysis of general welfare.
4. Analysis of production and the connections between different sectors of the economy regarding production (input-output).

From the outset, the theories of JM Keynes focused attention on the first purpose. When ambitions to control the economy rose in the wake of WWII the second purpose came into focus. Neither one of these two purposes calls for measuring government output. What counts in the macroeconomic models is government demand for resources. Hence government is looked upon as a consumer, not a producer, and the use of resources is captured by the concept “public consumption”. When economic growth became a legitimate political concern analysis of welfare with the aid of national accounts soared into the foreground. The Atkinson Review has the welfare perspective, adding that “national income is an indicator of the *contribution* (my emphasis) to welfare of specified economic activities”. (Atkinson, 2005, p.10) The development of national accounts for the purpose of measuring general welfare has always been accompanied by disclaimers. Therefore it is now more accepted to speak of production possibilities. In both a welfare and a production perspective it is of course consequential to measure outputs of all sorts, even in the public domain.

Distinguishing output from input is common sense in the private sector. In the public sector it creates problems. By the transformation, inputs may be valued differently than outputs. Who is to value the outputs? Who can be considered to have received the outputs (referred to as “the assignment problem”) and how are the outputs to be valued (“the valuation problem”) ? On whose income account are the outputs to be registered? And if they are valued differently from the inputs and there is a profit or a loss, on whose income account is that to be registered?

These problems have combined to promote the solution to regard the government as the receiver of the output and value it according to the cost of the input. “This has given rise to the concept of government consumption.” (Ohlsson, 1961, p.146). It has had the unfortunate consequence that government has been regarded as consumer rather than producer and that, therefore, there has been no interest in finding out about government production. No statistics have been compiled on the basis of production units. No analysis has been undertaken of the transformation of inputs to outputs. Is there technological progress in the production of government services? No one knows! Input-output tables around the world miss the contribution of the public sector. The distribution of income lacks the contribution from public services.

The production perspective changes all of this. The UN standard for national accounts proclaims that “GDP is a measure of production”. (SNA 93, 1.69) It is then clear that outputs must be distinguished from inputs. In a production perspective, logically, the national accounts should account for rates of transformation. This implies that output should be valued according to factor prices, which is what Ingvar Ohlsson recommended for public consumption (Ohlsson, 1961, p.323). Neither in the SNA 93 nor in the ESA 95 are the full consequences of this perspective outlined. There are input-output tables that include the public sector (SNA 93, Table 15.1) but there is no assignment of government output to other sectors than households, e.g. to firms. There are production accounts for the public sector (SNA 93, p. 24) but surpluses and deficits arising from production are not recorded as income anywhere in the system.¹

A measure of production is something else than a measure of welfare or of a contribution to welfare. Rates of transformation are captured by factor costs whereas a contribution to welfare is captured by market prices. The production perspective makes it clear that it is outputs and not outcomes that should be measured.

Principles of measurements

The Atkinson review states that it will adhere to the standards promulgated by the European Union, but, when there is latitude, find a basis for itself.

The review boils down to nine principles. I will comment shortly on each of the first three that are the most important.

Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in the national accounts for market output.

The Review states one very good reason for this principle: a reallocation of production from the public to the private sector or vice versa should not lead to a change in the estimate of national output. This requires not only that output is measured in the same way (in principle, since it may be done with different methods) but also that output is valued alike in the private and the public sectors. Say that a certain treatment in a public hospital costs \$ 10.000. A private hospital would charge whatever clients would be willing to pay, which might be more or less than the costs of labour, capital and other expenditures. Normally, profits would be

¹ Ingvar Ohlsson had all of this in mind when he launched his research group. This may be verified even through the title of the report from the group in 1986: “Public Services: Searchlight on Productivity and Users”.

included in the price, which would then supersede costs. Therefore, transferring hospital services from the public to the private domain would increase the recorded value of the output, without any equivalent increase in welfare.

The problem is that more often than not there are no equivalent private market services and hence no market prices that could be applied as a valuation of the public services. This causes not just one problem, it causes two fundamental problems of measurement of government output.

1. The first problem is of course, the sheer valuation. How much is an appendicitis operation worth?
2. The second problem is the definition of output. What is it that the consumer, who is not identical with the purchaser, hypothetically would be paying for? This entails both identifying the consumer and the product he would be paying for.

Defining output

Nevertheless, I think it is a good principle, as far as it goes. When defining outputs it is a good rule to try to figure out who would be the purchaser on a market where it would be the ultimate consumer that would pay, for what that consumer would be willing to pay and what the product would be. For example hospital treatment: what would be the service a private patient would pay for? Most likely for a consultation, for a therapy session, for an X-ray, for an operation, for a bed-day etc. I think such services are well qualified to be regarded as outputs of a public hospital, not the least in the light of principle A: what is the market analogue?

Strangely enough, such services do not qualify in the eyes of the Atkinson Review, nor in the Eurostat Handbook. Why? Because improved forms of medical treatments may reduce the number of operations necessary and that therefore measured productivity would decrease when in fact it has increased. “Activity is, for example, the number of operations in hospitals or number of patrols carried out by the police...Using activity indicators often does not lead to reasonable productivity numbers....Output is the preferred approach....For hospital services, the output is the amount of care received by a patient.” (Eurostat Handbook paragraph 3.1.2.1) “While activities may be the only available indicator and hence have to be used, they... are an intermediate variable.” (Atkinson, 2005, p. 41). It is difficult to understand distinction between operations etc. and “care” as well as between “activity” and “output”. A more reasonable use of the word “activity” would define it as a *process*, e.g. the act of production. What is needed is a distinction between *intermediate* output and *final* output, where intermediate outputs should be defined as inputs to final outputs. For example, planning, administering, personnel training, booking, procurements of medical supplies, information etc. are all intermediate outputs for a hospital or a general practitioner.

The principle has its limits, though, since there are a fair number of outputs that are consumed collectively as pure public goods in the Samuelson sense and where there can be no market analogue. The Atkinson review does not give much guidance on how to handle collective public services but refers to the Eurostat Handbook, which suggests either measures of the volume of input or of “activities”. (Ibid., p. 42-43) The question remains: on what grounds should “activities” be regarded as outputs? Can any “activity” or “intermediate variable” be

regarded as output? Since there is no market analogue, what should be the principle to base identifying and measuring collective service outputs upon?

Paradoxically, the market analogy principle is helpful even when it concerns public goods. I have already made a distinction between the welfare and the production perspective of national accounts. Also when defining non-market services in a production perspective this must not be done arbitrarily but must be done with a view to welfare or to the benefits that the outputs are supposed to bring. As with individual public services one should look for what it is that brings about the benefits or the desired effects of the public good. If it is patrolling the streets that causes the crime rate to drop, such patrolling should qualify as an output as well as investigating crime. On the other hand, physical exercise of policemen, planning, personnel training etc. will not in themselves affect crime rates.

ESA 95 follows a similar reasoning:

“In order to be capable to define the output of collective services, the starting point of the definition should be based on the production aspect. The definition of final products can be derived from the tasks and the activities by which organizations carry out the function of their activity and guarantee the achievement of the goals of the activity.”(Eurostat, 1998 p. 16)

This calls for a separate principle for public goods:

Principle A2: the measurement of government non-market outputs in relation to public goods should be based on an identification of what it is in terms of final outputs that may be considered to cause the desired policy outcomes.

Valuing output

Valuation according to the Atkinson Review should be based on a market analogy. Why? “The key element is that the welfare justification lies in measuring the added value to consumers.” (Atkinson, 2005, p.39) Following this argument valuations reflecting the value added should be used also for non-market services. If there are market prices for similar outputs they should be used for the non-market outputs. If possible find willingness-to-pay estimates! (Ibid., p.89) This is important, the Atkinson Review argues, since “An expansion of service 1 will be given less weight under cost weights and more weight if a marginal valuation is applied.” The assumption, of course, being that marginal valuations exceed marginal costs.

As noted above a production orientation of the national accounts makes market prices less relevant unless they mirror factor costs. The production perspective makes it easier to find a valuation when there are no market prices: costs (preferably marginal costs). If there are market prices I wouldn't hesitate to use them as weights for similar public services, although it could constitute a breach with factor costs, just for the sake of not causing false reallocation effects. Otherwise I would argue that costs, rather than market prices should be the basic norm for the valuation of government output. This is what the Eurostat Handbook recommends and also where the Atkinson review lands, but unwillingly. “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.... Ideally, we would like to use the marginal

valuations...So, for the present, the only feasible approach appears to be to continue to use cost weights.” (Ibid., p. 89)

If GDP is taken to measure production and not the contribution to welfare, weights that reflect factor costs are the most relevant. However, the inclusion of VAT and other commodity taxes in both the public and the private sector would not distort the comparability to any significant degree. Cost weights will tell us the rate of transformation between various public services. So that anyone, regardless of priorities, whether it is day-care centres or PhDs, can assess the production capacity on his own, familiar terms. Also, production capacity is a very relevant way of looking at government that has to cope with this number of births, that number of high-school kids, this number of by-pass operations and that number of fire-brigade alarms. Do we have the capacity to supply government services in sufficient amounts now, next year, in ten years?

Quality

Principle B: the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental contribution of the service to the outcome.

Anyone wanting to measure government non-market outputs has to handle quality. And in doing that, you risk getting tangled up in different approaches that muddle things up. There is the choice between output and outcome. Since there is no price it is tempting to want to measure outcomes instead of output in order to capture the value of the output. And when choosing weights it is tempting to want weights expressing benefits instead of costs. To some extent this is reasonable, but it may also be carried too far.

The ways in which to capture quality are three in the Atkinson Review (Ibid. p.91):

1. differentiate services
2. define volume measures in terms of success
3. quality adjustment of volume measures on the basis of the contribution to outcome

The recommendation is in line with the techniques that have been used throughout the Swedish reports (Ohlsson, et.al.1986, Murray, 1992 and 1994).

Basically there are two ways to handle quality: the first is to adjust volume measures of output, the second is to adjust weights for the output. The reasoning behind quality adjustment, though, is the same.

An example of differentiating services is the measurement of output from the health care sector, which, in the Swedish study covering 1960-80, was differentiated into 312 outputs, outputs being defined as patients treated at various clinics, each with a unique unit cost weight (Murray 1994 a). The idea is that a higher quality – synonymous with a more valuable output – is expressed in a higher weight. This technique of handling quality is by far the most important in all systems of output measurements. It lends itself also to the aggregation of different branches of government production.

Using unit cost weights is, of course questionable. It is readily defended in a production approach to measuring output. But its relevance in relation to outcomes and welfare must be

checked. If, for example, a medical therapy of ulcer is just as efficient in curing the disease as an operation upon the stomach although it is much less costly it should be weighted on par with the operation. By prescribing medicine instead of operations the productivity of the hospital will increase. Such substitutions are very important in raising productivity and must therefore be captured. The weights express rates of substitution rather than rates of transformation.

An example of measuring output in terms of success is measuring output of the police in terms of investigated crimes that were solved and detected traffic violations (among other outputs). The idea is, of course, that it is only success that contributes to the ultimate outcome to reduce crime rates. However, this may be disputed, and great care must be taken when introducing measures of success in affecting outcome. In the first place, even the show of the police may contribute to reducing traffic violations. In the second place, there are a number of other factors influencing the frequency of traffic violations. If e.g. teaching does not provide students with any more knowledge, should the output of teaching hours be counted nil?

The most difficult technique is quality adjustments of a specific output. Say that teaching improves so that students learn 25 percent more. The contribution to a desired outcome is attributable to the output of the schools. Does that mean that output should be adjusted upwards by 25 percent? That is the recommendation by the Atkinson Review. (Ibid. p.130-131) I find this issue more complicated.

The national accounts handles a multitude of such instances in the market domain of the economy, but not in the way the Atkinson Review recommends. The most common method is to adjust quality on the basis of the cost increase to improve quality. Suppose teaching costs per student have increased by 20 percent in order to improve learning by 25 percent. Output would then be adjusted upwards by 20 percent, on the ground that the improved learning is worth at least as much as 20 percent more, otherwise it would not have been undertaken. However, this technique of cost adjusting for quality change does not capture possible productivity changes.

Another technique is to use hedonic prices. How much more costly is teaching in schools with 25 percent better results? Suppose costs are 20 percent higher and that there are municipalities that willingly pay for that. This could motivate a 20 percent upward quality adjustment, regardless of the cost increase, thus allowing for productivity change.

To equate output with outcome as is suggested by the Atkinson Review, may be altogether wrong. Diminishing marginal productivity may call for greater than proportional cost increases to raise results by 25 percent and may or may not be paid for willingly. I would argue for a cost based or a hedonic price quality adjustment. Whereas the recommendation on taking account of quality and quality change by the Atkinson Review is this: “a number of dimensions of quality should be measured, with results weighted together by marginal social valuation...” (Ibid. p.120)

Complementarity between public and private output

Principle C: account should be taken of the complementarity between public and private output, allowing for the increased real value of public services in an economy with rising real GDP.

The application of this principle is illustrated by a number of examples. The starting point is a direct measure of education output, student full time equivalents (FTEs). (Ibid., p. 43) However, FTEs do not increase to the extent that inputs in the education production do. Therefore, a new measure is sought. It appears – though the argument is not explicit – that the Atkinson Review finds it strange to accept that taxpayers at large would accept a falling productivity and increasing price of education and that therefore there must be something wrong with the measure. The solution lies in arguing that over time the value of education has risen and that this increase in value represents an increase in output. “...the output of government services rises with the real value of private assets and incomes.” (Ibid., p.46) Another example is the output of fire services that is taken to increase by the increase in real estate values. (Ibid. p.45)

I have difficulties in following this reasoning. I see nothing strange in falling productivity and rising unit costs of education. Increased bureaucracy could be one reason, but also increased difficulties in teaching ever more independently minded students. The latter reason I find could constitute a more reasonable motive for questioning the plain FTE measure. However, changes in production prerequisites are seldom taken into account in output measurements of marketed outputs. Doing that would imply e.g., that the output of crude oil production should be adjusted upwards by the depth of the wells drilled. This reasoning is rightly rejected by the Atkinson Review in connection with discussion of postal services. And it is also noted, rightly, that this is a matter of statistical convention. (Ibid. p.38)

Also, I see nothing strange in taxpayers paying more and more for a service that is becoming more and more costly. It just indicates that with the growth of incomes their willingness to pay increases, which is tantamount to saying that there is a strong, positive income elasticity, so strong that it dwarfs the negative price elasticity of demand. But this increase in the willingness to pay for education has nothing to do with increased productivity.

One could argue, as does the Atkinson Review, that part of the increase in incomes that in turn affect the valuation of education, stems from education.

In the first place: should the output of education be valued, not just by the value of earnings that the educated student receives, which from a welfare point of view would be consistent, but also by future income increases in income in the whole economy? Let aside the difficulties to correctly assess that part of GDP increase that is attributable to education, but such a reasoning would change the principles of national accounts fundamentally. The price paid for a car should not be the value attached to the output of cars but instead the value of all future changes in GDP due to the purchase of that car, including of course external economics, both positive and negative. A principle of national accounts is precisely that external economics should not be included.

In the second place: accounting in constant prices leaves no room for a continuous increase in output on the grounds of a rising value of output. Only if we could imagine accounting educational output in current prices would it be possible to include such a value in the nominal value of output. With a measure of the nominal value of output it would be possible to calculate a value added by deducting the value of purchased goods and services. The nominal value added could then be deflated and used to calculate value added productivity. Value added productivity catches both allocative efficiency (i.e. producing goods and services that are highly valued and purchasing goods and services for input that are lowly valued) and

productivity. But this, I understand, is not something the Atkinson Review considers. I, too, find it too far fetched and it would miss the opportunity to measure genuine productivity, i.e. efficiency of production.

Output or outcome

Stated clearly: quality is different attributes of the product. However, those attributes cannot be defined without taking into account their value for the consumer.² If a bus is painted green, this is an attribute of the product “bus ride”. But it is most likely not an attribute that deserves to be called quality, since it does not matter for bus passengers what colour the bus has. And since what matters to the consumer is important for the definition of quality (and for the definition of output, for that matter) you may infer that output is synonymous with outcome. Outcome, e.g. whether the patient recovers or not, may serve as an *indicator* of quality, but does not in itself constitute quality. When discarding “activities”, such as consultations and operations, as outputs and wanting to define output as “the amount of care” you are stepping onto a slippery path. It may lead you to define output in the health care sector as “whole courses of treatment”³ rather than a number of separate products such as consultations, operations, therapeutic sessions, check-ups etc. In doing that you are definitely way beyond a market analogy. A hospital couldn’t charge for “a whole course of treatment” until the patient gets well, but charges for each activity, sometimes joined in “packages”, such as X-ray, operation and post-operative care. If the operation is not successful – not because of any error on behalf of the hospital – and some other treatment is suggested, this will constitute a new output to be purchased by the patient. It is true that “a whole course of treatments” will capture important quality elements – though not all by far⁴ – but at the expense of introducing new complications. Such complications are measuring output, the production of which extends over many years, the need to take account of the will of the patient in trying various treatments, the behaviour of the patient and other outside factors influencing recovery etc. The suggestion to base output measurements on “the whole course of treatment” would be sensible if interpreted as “the whole standard course of treatments”, i.e. a standard bundle of services. But it would still entail the problem of production and costs over many years before the bundle is complete. Diagnostic related groups, DRG, is precisely that, a standard package of services to treat patients diagnosed to belong to a specific group. However, DRG is a much smaller bundle, not a lifelong treatment story and it may be handled as output with not too big a problem of allocating outputs and costs to the correct period.

It is understandable, but nonetheless wrong, to want to define outputs and its qualities in such a way that the ultimate objective is reached. This is going beyond accounting for production and productivity and belongs to the sphere of welfare and should be handled by social cost-benefit analysis.⁵ “GDP is a measure of production.”

² The Atkinson Review formulates this thus: “The activities are designed to benefit the individual patient. To the extent that they do, the health care provided constitutes the output associated with these input activities.” (Ibid. p.40) The quote is an example of the muddling up of things. It states that it is “the health care provided”, not the “activities” in themselves that constitute the output.

³ “Ideally, we should look at the whole course of treatment for an illness rather than at its components.” (Atkinson, 2005, p. 113)

⁴ See for example the assessment of quality change in the Swedish health care 1960-90 in Murray (1996). Ten different dimensions of medical quality are assessed. In addition there are quality dimensions pertaining to food, lodging, information, entertainment etc.

⁵ Should the use and consequences of guns purchased be included in the measure of output from the weapons industry in the national accounts? The debate on whether to measure outputs as pure products (“activities”) or as

Lancaster(1966) once formulated a meta theory of consumption. It stated that we do not consume cars, bicycles, bus rides etc. We consume transportation by whatever mode is the most cost-efficient and appealing in terms of comfort, speed, safety etc. But there are still markets based on the demand for cars, petrol, bikes etc. The theory provides a vehicle for better understanding of these markets. Until there is a genuine market for transportation – encompassing all modes – outputs of each category will be defined and measured for the inclusion in the national accounts. Over time there are examples of packaging goods and services into new outputs. Instead of hiring architects, bulldozers, bricklayers, carpenters and plumbers for building the house of your dreams you nowadays can order a complete house. Principle A should rule: parallel market output! Which means that as long as there is no market for “transportation” the separate outputs should be recorded and measured.

When “transportation” becomes the unifying unit of output for all modes of transportation, so called reallocation gains will be registered as part of the national output. Depending upon the weights attached to bike and to car transportation it could turn out that just by changing mode from car to bike transportation GDP would increase. It could, depending on the weights, also be the other way around. Now, that is something we are not prepared to include in the national accounts and it is a shortfall of national accounting that most economists will accept. On a lower level of substitution, however, we are keen to include reallocation gains. Were we not to include any reallocation gains from substitution at all we would not record any productivity increase at all. So the question is really, on what level of substitution we are to include reallocation gains. The market cannot be said to provide the “right” answer, but it provides a standard! In cases of non-market outputs when there is no similar market output we have to proceed by careful analogy. My suggestion, which seems to be in line with present practices in the national accounts, is to restrict reallocation gains to situations of *immediate substitution*. In market situations, immediate substitution can take place between any goods having a market price. Hedonistic prices sometimes have to be calculated on the basis of immediate substitution in order to assess price differences and change. In non-market situations we have to figure out the possibilities for substitution ourselves. When one medical practice is as good as another much more expensive they should be counted on par. When training air force pilots in simulators is as good as flying real air planes the two activities should be counted equal⁶.

The Atkinson Review tries to push the argument of “whole courses of treatment” even further, when considering that outputs even from different branches of government often congregate to one common impact, i.e. health. However, the Atkinson Review realises that this would be going too far. Because of the difficulties in calculating relative contributions to health, i.e.

impacts, outcomes, effects, benefits etc. has been going on for a very long time. The pioneer defining government outputs, Hill, even changed his position from 1975 to 1977. In 1975 he wrote: “Thus, the output of advocates is not to be measured in terms of the number of cases they win... The service provided by a lawyer is, therefore, no more and no less than the presentation of his client’s case.” (Hill, 1975, p.13) But in 1977 he wrote: “If the pupil’s qualifications and ability are such that he is incapable of understanding and absorbing the teacher’s instruction, there can be no change in his condition as a result of the teacher’s activity and no service is produced in these circumstances. The activity of the teacher is wasted and cannot count as productive.” (Hill, 1977, p 324) It seems to me important to distinguish the definition of output from the way you measure it. In order to capture a quality attribute of the output one may have to and sometimes can use outcome as an indicator of quality. However, this must be done with great care since it may on occasion be just as misleading as it is sometimes illuminating. If outcome reflects the quality of the service produced it is all right. But if the students that by free will participate in the education or that have been assigned to the teacher are not capable of learning anything it is not the fault of the teacher or the school. The quality may be excellent!

⁶ Both examples appear in the Swedish output and productivity studies (Murray, 1994).

from schooling, refuse collection, social protection etc., “it may be necessary to remain within cost weights.” (Ibid. p.90)

Staying within cost weights is very reasonable, but the reason that it is not feasible is poor. A better reason would be to conclude that policy-makers have had their say and in prioritising various branches of government they have – to the best of their minds – allocated resources to various public services according to ideas about their contribution to policy goals such as health. This is not a perfect competition market but an imperfect political market. But I see no other way in which to justify cost weights from an outcome perspective. Of course, in a production perspective, they make sense. Also, as noted above, the production perspective makes it much more sensible to handle collective public services by measuring as output those “activities” that contribute directly to the common benefit or goal.⁷

Overlooked problems

Finally, I will shortly mention some problems that I don’t see treated by the Atkinson Review but that have to be.

1. There is extensive treatment on the choice of deflators. However, whether deflators should be general or specific in relation to the specific kind of production process, the inputs of which are to be accounted for, is not discussed. Results will be different depending on if you deflate police costs with a price index that reflects the increase in police salaries or the general increase in civil servants’ salaries. Do we want to hold the police accountable for its relative salary increase or not? Do the salaries of the police reflect the benefit forgone in other lines of production by hiring these people as polices?
2. Aggregation poses problems that I have not found addressed.
 - a. In the first place, there are agencies and ministries that serve as overhead or support to produce final services. It would be wrong to account for and add their outputs to the final outputs. Their costs should, however, be included in the inputs for the production of the final output. Such is the case for ministries that control agencies or systems of authorities and boards that produce a final output. It is also the case for support functions like premises, computer, stationary and information handling agencies and the like. Final outputs must therefore be identified, that are final in relation to government, not just in relation to the producing organisation. However, there are difficult problems of aggregation, such as in the judiciary, where this principle cannot be followed.
 - b. In the second place, how are outputs used as inputs by private firms to be handled? The output should be deducted from the value added or added to the costs of the output of the private firms. The following table is from Ohlson et.al., 1986.

⁷ The judiciary is an example where the outputs of several agencies congregate to one final output. Outputs from the police may be seen as inputs to the attorney and so on. Should they be treated as one entity with a single output: correctional services? Following the recommendation of ESA 95 regarding collective services their different activities contributing to combat crime should all be added up.

The destination of Government Output, Percentage of Public Consumption, nominal prices

	1970	1975	1980
<u>Input into production</u>			
Private industry	8.5	8.0	7.6
Government	3.0	2.5	2.2
<u>Consumption</u>			
Individual goods	61.6	63.7	68.0
Collective goods	26.9	25.8	22.2

That part of government output that directly benefits the private sector – part of the consumption of roads, harbours, employment agencies, and so on – has diminished its share of public consumption although it has grown in volume. Recalculating the growth of GDP by subtracting this part of the contribution of the public sector to GDP from that of the private sector increases the growth of GDP by 0.14 percent over ten years (Murray, 1992).

A similar calculation was undertaken by Kuznets (1971). He included a much larger part of the government output in what was to be subtracted from GDP as sheer costs to keep society functioning. General administration, police, military defence, and so on were looked on as prerequisites for other kinds of economic activities. Kuznets found that over 100 years the growth of the US economy was lowered by 7 percent.

3. It has to be decided from what year weights and constant prices should be. What should be the base year? Should it be the last year in a measurement series of several years? This was the national accounts standard at the time the Swedish studies were undertaken, the motive being that prices and weights should reflect present day priorities and valuations. The thinking at that time was much more oriented towards welfare. Or should it be the first year in the row, which maybe could be motivated by a wish to capture pure production efficiency. However, if that is the purpose of the exercise it would be more in line with up-to-date productivity research to use moving weights, preferably an average of weights from two consecutive years.
4. Other techniques than (as) complete coverage (as possible) should be considered. Statistics provides us with techniques to estimate population parameters by drawing samples. The Atkinson Review advocates “triangulation”, i.e. taking other sources of information into account, in order to assess output and productivity. I wholeheartedly agree, but would also consider using other sources of information for estimating output and productivity to be included in the national accounts. One such source is samples of production units, for which detailed data could be collected. There is a host of data envelopment analyses and other non-parametric and parametric methods to estimate productivity change with. (See for example Hjalmarsson, Bjurek, Försund, Luoma, Blank among the references). Another is price indices for government outputs. In the US health care sector there is a need to have such price indices when insurance companies are to remunerate hospitals for their services. (See for example Berndt E.R. et.al. 1998)

References

- Andreassen, J., Försund, F. och Hernaes, E.*, Produktivitet i statlig sektor, Gjennomgang av litteratur och anbefaling om satsing, Rapport nr 14/89, Senter for anvendt forskning, Norges Handelshøyskole, Universitetet i Oslo
- Atkinson, T.* 2005, Atkinson review, Final report – Measurement of Government Output and Productivity for the National Accounts, Palgrave-Macmillan
- Berndt, E.R., Cutler, D.M., Frank, R., Griliches, Z., Newhous, J.P. and Triplett, J.E* 1998, Price Indexes for Medical Care Goods and Services: An Overview of Measurement Issues, NBER Working Paper No. W6817, December 28, 1998
- Bjurek, H.*, 1994, Essays on Efficiency and Productivity Change with Applications to Public Service Production, ak.avh., Göteborgs universitet
- Blank, J.L.T.*, ed., 2000, Public Provision and Performance, Elsevier.
- Commission of the European Union 1996, European system of National Accounts 1995, (ESA 95)
- Eurostat* 1998, Task Force on volume Measures for Non-Market Services, NACE L (mimeo, September 1998.
- Eurostat* 2001, Handbook on Prices and Volume Measures in National Accounts
- Farrell, M.J.* 1957, The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, Series A, General, 120
- Fisk, D. and Forte, D.* 1997, The Federal Productivity Measurement Program: final results, Monthly Labor Review, May 1997
- Färe, R.* 1988, Fundamental of Production Theory, Springer-Verlag
- Hill, T.P.* 1975, Price and Volume Measures for Non-Market Services (mimeo), Statistical Office of the European Communities, April 1975
- Hill, T.P.* 1977, On Goods and Services, The Review of Income and Wealth, ser. 23, no. 4, December 1977
- Hjerpppe, R. och Luoma K.*, 1997, Finnish Experiences in Measuring and Promoting Productivity in the Public Sector, Discussion Paper 150, Government Institute for Economic Research, Helsingfors
- Kusnetz, S.* 1971, Economic Growth of Nations: Total Output and Production Structure, Harvard University Press
- Lancaster, K.* 1966, A New Approach to Consumer Theory, Journal of Political Economy, April 1966
- Murray, R.* 1987 a, Den offentliga sektorn – produktivitet och effektivitet, Bilaga 21 till LU 87, SOU 1987:3
- Murray, R.* 1987 b, Productivity Measurement in Bureaucratic Organizations, in Bureaucracy and Public Choice, ed. Jan-Erik Lane, Sage Publications
- Murray, R.* 1992, Measuring Public-Sector Output: The Swedish Report, in Output Measurement in the Service Sectors, ed. Zvi Griliches, National Bureau of Economic Research Studies in Income and Wealth, Chicago
- Murray, R.* 1994, Den offentliga sektorns produktivitetsutveckling, Ds 1994:24
- Murray, R.* 1995, Productivity in the Public Sector: The Swedish Experience, in The European Yearbook of Comparative Government and Public Administration, Vol. II/1995, Nomos Verlagsgesellschaft
- Murray, R.* 1996, Productivity Trends in the Public Sector in Sweden, Report to the Expert Group on Public finance
- Murray, R.* 1997, Public Sector Productivity in Sweden, Papers on Public Sector Budgeting and Management in Sweden, vol. 3, Budget Department, Ministry of Finance.

Ohlsson, I. 1961, *On National Accounting*, National Institute of Economic Research, Stockholm

Ohlsson, I., Blohm M. and Murray, R., 1987, *Public Services – a searchlight on productivity and users*, Report to the Expert Group on Public Finance, Ds Fi 1986:13

United Nations, *System of National Accounts* 1993