PART II

Chapter 13

Price and Volume Measures

Introduction

Getting more value for money is a common interest of policy makers and consumers. While progress has been made since the release of the first edition of the SHA Manual, there remains concern among policy makers and statisticians that the difficult task of price and volume measurement in health care needs more attention. This chapter provides a methodological discussion and several illustrative examples of the developmental work related to health care prices and volumes.

Health care poses basic conceptual difficulties for price and volume measurement that are common to other service industries, for instance, communications, and computer and information services. Some of these are partly due to well-known fundamental measurement problems in service industries, such as the definition of output, the appropriate choice of unit of measurement, and the monitoring of quality changes.

The study of impacts of health care services on the growth of service industries, overall economic productivity trends and relative prices also depends on the availability of reliable measures for price and volume in health care goods and services. Weaknesses in measuring the price and volume of health care also have implications for the measurement of total GDP growth in real terms. Growth of real private household income might also be underestimated in countries where health care is an important component in the consumer price index (Boskin *et al.*, 1996).

This chapter outlines some of the difficulties in the calculations of the constant-price output of health care industries – services and goods – both for market and non-market production, and discusses methodological problems of international price comparisons in this field. This chapter skips the more technical parts of index numbers and focuses on proposals to improve the measurement of price and volume changes in health care.

The chapter presents an overview of the methods tested in several countries that have carried out projects to improve their countries' price and output measurement in health care, mainly using directly-measured output indicators. Results from these studies indicate that, where elaborate health information systems are in place, improved indicators for price and output measurement are feasible. It is thus important that some agreements are reached on how to advance the information systems in such a way that relevant and desirable indicators can be progressively built.

Measurement of changes in price and volume in the SHA framework

The SHA Manual focuses on the final consumption of health care goods and services, with valuation at purchasers' prices. The aggregate "current health expenditure" combines in a single figure the monetary value of the final consumption of all health care goods and services by residents of a given country during a given period.

Another key indicator of interest to stakeholders is the growth in the value of health spending over time. This change in the observed monetary aggregate can be separated into

the part of growth that stems from a change in volume and the part that is due to a change in price. Growth in volume controls for changes in the price level between two periods in order to identify whether changes in expenditure reflect only an increase in the prices of health services or instead an increase in production, and consequently, improved services, more coverage and the like.

Volume is associated with quantities. In a time series, a volume index summarises the proportionate changes in the quantities of a specified set of goods or services between two periods of time. The quantities compared over time must be for homogeneous items, and the resulting quantity changes for different goods and services must be weighted by their economic importance, as measured by their relative values in one or other, or both, periods. For this reason, volume is a more correct and appropriate term for comparisons in time than is quantity (SNA 2008, 15.13). Similarly, for inter-country comparisons, differences in price levels between countries are removed from the expenditure values to provide a comparison between the volumes in the countries concerned.

The first part of this chapter will focus on this *temporal* volume/price breakdown. For this time series type of analysis, a framework for compiling a consistent and integrated set of price and volume measures covering all flows of goods and services is suggested. In this chapter it will be argued that the basic estimation strategies for measuring prices of market health products also provide an adequate starting point for measuring output and prices for a significant share of non-market services in health care.

Besides, the principles of trend analysis can also be used to measure output across countries while holding time constant, that is, to carry out "cross-country" comparisons. Measurement of changes in price and volume can be analysed from a perspective of change in purchasing power across countries over some basket of final consumption goods and services (*spatial* volume/price breakdown).

A cross-country comparison should show that two countries producing the same number of identical items of a given quality would be measured as having the equivalent volume of output. The similarity with the method used for time series should be quite obvious: a time series comparison within the same country can easily be viewed as comparing two countries. In other words, the method used to compare prices in Denmark in 2000 with Denmark in 2010 should be consistent with the method used to compare Denmark in 2010 with France in 2010. This similarity will be heavily used in this chapter: the quantity and quality indicators that are recommended for use in a time series analysis will be, at least conceptually, the same as the quantity and quality indicators to be used in cross-country comparisons, and vice-versa.

Volume time-series are generally calculated by deflating current price time series by adequate temporal price indices. Similarly, appropriate spatial price indices ("Purchasing power parities") are used to deflate current price aggregates (typically in different currencies) in order to compare the volume of output (or consumption) between countries.

It is important to clarify that a time series of volumes and volumes derived using PPP estimates as deflators are two different measures. In a time series of volumes, the effects of price changes from one period to another are removed to produce volume measures from which rates of economic growth are calculated. In the case of an inter-country comparison, the effects of differences due to exchange rates and those due to different price levels within each country are removed from the expenditure values to provide a comparison between the volumes in the country concerned (SNA 2008, 15.228).

Inter-temporal indexes of price and volume

To use an analogy from the whole economy, much effort is spent on measuring the value of GDP at current prices, but often a more important objective is to derive a measure of the growth of GDP and its components in volume. Growth in volume controls for changes in the price level between two periods.

To express a flow in volume terms, each of the goods and services that are the counterparts of money spent must, in principle, be identified. This is because quantities are additive only for a single homogeneous product: if a hospital carries out 200 (expensive) hip replacement operations and 400 (less expensive) cataract surgeries, it is not informative in economic terms to add these together and say that 600 treatments were carried out. Inevitably, the question therefore arises of how are the 200 hip replacement operations and the 400 cataract surgeries to be added in a meaningful way. This question will arise for all the thousands of goods and services that populate the global concept of health services. The answer traditionally draws on the knowledge that the relative prices of the different goods and services bought and sold reflect both their relative utilities to purchasers and their relative costs for the producers. However, is this relevant, in particular, for non-market services where, by definition, there is no market price that reflects the interaction of consumers and producers?

In the simple market situation, index numbers are designated to decompose changes in value aggregates into their overall change in price and volume components. A price index can be used to deflate the value of products that have been transacted on the market, and this yields a volume measure of the goods or services transacted. Instead of deflating values with a price index, one could also set up an index of volumes or quantities produced and consumed directly.

A price index can be calculated as a weighted average of the proportionate changes in the prices of a specified set of goods and services between two periods of time. Similarly, a volume index can be calculated as a weighted average of the proportionate changes in the volume of a specified set of goods and services between two periods of time.

The two most commonly used indices are those of Laspeyres and Paasche. The former is used mainly to calculate changes in volumes while the latter is used to calculate changes in prices.[1]

Three things are worth noting about these indices:

● First, price or quantity changes of the products are weighted with expenditure shares and these expenditure shares – measured by market observation – reflect the joint, equilibrium valuation of each product by consumers and producers. Thus, market prices and quantities reveal the interaction of consumer preferences and producer costs.

● Second, in the simple model at hand, changes in quantities Y_i^t/Y_i^{t-1} and in prices P_i^t/P_i^{t-1} are simply measured by comparing them between periods – implying that the units of measurement for Y_i^t are the same as the units of measurement for Y_i^{t-1} and that the set of products is stable – product i has to exist in both periods to be compared.

● Third, the simple presentation here also makes the implicit assumption that there is exactly one measured unit of quantity that constitutes Y_i or one measured price that constitutes P_i. In practice, this is rarely the case. The P_i's are un-weighted averages of individual items which constitute an elementary price index. Similarly, the Y_i's for a volume index are actually product counts. How individual products ("items") are

Box 13.1. **Laspeyres and Paasche indexes**

Volume indices of the Laspeyres type are shown below where P_i^t stands for the price of product i in period t and Y_i^t for its quantity:

$$Y^{L,t} = \sum^N \left(\frac{P_i^{t-1} \times Y_i^{t-1}}{\sum^N P_i^{t-1} \times Y_i^{t-1}} \right) \frac{Y_i^t}{Y_i^{t-1}} \qquad [1]$$

In the formula [1] above, quantity changes Y_i^t/Y_i^t–1 for elementary commodity groups are aggregated by weighting them by their value shares in period t-1. The index can be interpreted as the change in value at constant prices of a basket of products whose composition is kept fixed as it was in the reference period.

The Paasche price index is defined as follows:

$$P^{P,t} = \frac{\sum_{i=1}^N P_i^t Y_i^t}{\sum_{i=1}^N P_i^{t-1} Y_i^t},$$

which can be written in terms of the price changes of the individual commodities, as follows:

$$P^{P,t} = \left[\sum_{i=1}^N \frac{P_i^t \times Y_i^t}{\sum_{j=1}^N P_j^t \times Y_j^t} \left(\frac{P_i^t}{P_i^{t-1}} \right)^{-1} \right]^{-1} \qquad [2]$$

Similarly to [1], the price index written in [2] can be defined as the change in the price of a basket of products whose composition every period is updated and the quantities of the new period t are applied to the prices in t and t – 1.

The volume index [1] can also be obtained indirectly, by dividing the change in value by the Paasche price index (so-called deflation method):

$$Y^{L,t} = \left(\frac{\sum_{i=1}^N P_i^t Y_i^t}{\sum_{i=1}^N P_i^{t-1} Y_i^{t-1}} \right) \Big/ P^{P,t}$$

grouped is a question that has to be answered with respect to the purpose of the price or volume index. A useful criterion for grouping individual items is that they potentially satisfy the same or similar consumer needs or that they are substitutes from a consumer perspective. Conversely, if different items are not interchangeable from a consumer perspective, they should be treated as different products.

The considerations above already indicate that in a non-market context, prices for goods or services are not in general observable. It could be argued that this may also be the case for market services, where a price mechanism is observable but not necessarily a price per unit of output. For example, if health services are defined as completed cycles of treatment, the price for a particular completed treatment may not be directly observable,

because the observable pricing mechanism may be geared to individual activities or providers rather than to a complete treatment. There is also the issue of health insurance: prices that are observed in the health services market may not be indicative of consumers' willingness to pay if these are covered by health insurance. This leads us to conclude that differences in measurement between the "market" and "non-market" provision of health services may be less pronounced than is sometimes assumed.

In the literature, deflation procedures are often exclusively associated with market producers. This reflects the idea that constructing a price index requires the presence of market prices and that the latter are directly associated with market production. While this argument is correct, things are less clear-cut if one allows for a more comprehensive meaning of "deflation". In particular, "deflation" can be understood as applying a true market price index, but it can also be understood as applying a unit cost or "quasi price index" (Schreyer, 2008).

In some countries, hospitals and other providers of medical services are considered market producers because they receive reimbursement that, on average, covers their costs. In such cases, a quasi price index consists of average revenues per treatment. One notes, however, that the fact that there are revenues does not imply that there is a competitive market where prices necessarily carry signals about consumer preferences.

Market and non-market measures for health care

Market output is output that is sold at prices that are economically significant. Thus, for market health services, the value of output in current prices can be measured by the value of sales of these services. However, health is one of the most common examples of services provided by government free of charge or at prices that are not economically significant and thus constitute non-market output. A price which is not economically significant is deliberately fixed well below the equilibrium price that would clear the market. It is defined as a price which has little or no influence over how much the producer is willing to supply and which has only a marginal influence on the quantities demanded.

The measure of the volume of health care goods and services consumed by patients should not be affected at all by the status, whether market or non-market, of the provider of the service. Thus one objective is to encourage the compilation of consistent measures of health output, whether these services are provided by market or non-market producers. Hill (1975) formulated this idea as follows:

"*It is proposed as a matter of principle that the basic methodology used to measure changes in the volume of real output should always be the same irrespective of whether a service is provided on a market or on a non-market basis. This is not to say that the actual numerical measures would not be affected by whether the service is market or non-market, because different weighting systems would be involved, but at least the methods of measurement should be conceptually similar*" (page 19).

To achieve such convergence, an obvious first approach would be to try to find out whether components of the consumer price index (CPI)[2] could be used to adjust the weights (value shares) in a volume index for the non-market sector. Indeed, the existence of expenditure by households implies that there should be a CPI calculated for these flows, and thus it seems there already exists some experience in calculating a volume/price split. The question then arises as to whether this experience cannot be directly applied to the non-market sector.

The extension from the market sector to the non-market sector, while possible in some cases, is generally more complicated. A problem occurs when the services provided by the market sector are different from those provided by the non-market sector. For example, in some countries, the hospital market sector may be mainly dealing with standard interventions, for which it is easier to calculate a volume measure, while public hospital services also cover more complicated interventions, for which volume calculations are also more difficult. Thus, the market sector may not be representative of the entire sector. Even if many outputs of non-market producers are equivalent to those of market producers, the efficiency of these types of providers may differ.

Another reason for this is that the expenditures on social transfers in kind to individual households by government or non-profit institutions are excluded from the scope of the index, as they are not incurred by households. In the health sector, many of the goods and services consumed by households are financed and paid for by government units or non-profit institutions. This is why the CPI is of limited use to deflate health expenditures whenever social transfers in kind represent an important proportion of health spending.

The United States is a case in point. A large part of expenditures on health in the United States is directly attributed to household expenditure and provided by "market producers". These expenditures are therefore covered in the CPI. The volume of health services in the national accounts is then obtained by applying a mixture of the relevant CPI and Producer Price Index (PPI)[3] deflators to health expenditure (US Bureau of Economic Analysis, 2009). The choice is based on which deflator provides the best coverage. However, the extensive literature on the difficulty of quality adjustment for health services in CPIs and PPIs shows that the basic problems with price indices are the same as with volume indices in other countries.

The basic measurement methods used for both the market and the non-market sector should be consistent. Consistency concerns in particular the aggregation method, where the same type of weights should be used to combine quantities or prices of services into volume or price indices. The main issue is to ensure that the quantities, prices, or unit costs refer to a unit of output, and not to units of inputs. Otherwise, the objective of measuring volumes of outputs cannot be achieved.

Products classification for price and volume measures

The objective of this section is to provide guidance on the measurement of the volume of health services provided by the health care industry. Ideally, the unit of output would capture complete treatments, and would take into account quality change in the provision of treatments. This measurement of health care output would then be able to differentiate among price, quantity and quality changes. In practice, rather than quantifying "complete treatments", it is common practice to define output in the health sector by counting the amount of each type of activity[4] that is undertaken in each health care setting.

A complete treatment refers to the pathway that an individual takes through different health providers in order to receive full and final treatment for a disease or condition. This definition of the target measure, otherwise known as a disease-based estimate of health care output, is similar to that used in the *Eurostat Handbook on Price and Volume Measures* (2001), Berndt *et al.* (2001), Aizcorbe and Nestoriak (2008) and Triplett (2003). We note, however, that the notion of a complete treatment is not always applicable.

A first limitation arises with regards to measuring complete treatments. Conceptually, "complete" is understood as a complete treatment pathway across the health care system. As an example of a complete treatment pathway, consider a hip replacement operation. The pathway approach would imply aggregating all services or procedures associated with the intervention on the condition, whether these are received from primary care services such as a general practitioner, specialists, at hospitals, or at rehabilitation centres. Thus, using the pathway approach would entail collecting data on outputs from a number of health care providers and aggregating them in a meaningful way. This is very challenging.

There are additional reasons why the principle of complete treatment is difficult to implement:

● The total output of an industry is based on summing up outputs of various service providers (establishments), and therefore a complete treatment is hard to capture if service provision cuts across several establishments. Even if it were possible to observe complete treatments, if there are several service providers involved (*e.g.* hospitals and outpatient services), there would be no simple way to allocate the overall service to the different participating units, and yet this is a requirement for national accounts purposes.

● Most data retrieval systems do not have the capacity to link the treatment of an individual across institutions to enable measurement of the complete treatment. Data would be required on both expenditures (value of inputs) and the services received. Thus a health care pathway approach imposes demanding data requirements, as patient records have to be linked across activities and institutions. Even within institutional settings, data may not be appropriately linked.

● The beginning and end point of a treatment pathway is observable in the case of acute health conditions, but unclear for chronic health problems or for medical conditions that give rise to long-term care and services provided in nursing homes. Many of the diseases associated with ageing and most psychiatric conditions are chronic, long-term conditions, and the patient may be treated for more than one illness or problem within a period. Thus, the boundaries of the complete treatment would be unclear.

Given the difficulty with compiling complete treatments, estimates of health care output usually occur at the institutional level. Thus a narrower view of a treatment is that defined by the type of health service. This measure captures the full treatment only within an institution and generally by function or type of service. Norway and the United Kingdom have, for instance, adopted this practice. Dawson *et al.* (2005) compiled an output index with 1700 categories of NHS activity, including primary care. This aligns with standard practice in national accounting.

Thus, rather than reasoning in terms of complete pathways of treatment across the health system, the output measures proposed in what follows are best thought of as episodes of treatment of particular diseases provided by a given institutional unit. Furthermore, this measurement objective will mainly be applicable for curative care, whereas other measures will have to be targeted for long-term care and other specialised services where it is difficult to establish when an episode of treatment is complete. For example, inpatients in nursing homes do not generally receive treatment for a specific illness or illnesses that have an obvious start and end point. For such institutions, a strong case can be made that the output is defined by the processes of the institution of care, not by a treatment. The same applies to chronic and progressive health conditions where the patient faces a slow, variable and unpredictable progression of a disease.

Direct measures of output in these units will have to rely on measures such as the number of particular processes (procedures, consultations, etc.) or the number of patients treated in various institutional settings. There are advantages in continuing within this framework until the linkage of patient treatment across institutions becomes available.

The present considerations suggest that it is best to treat the measurement of output of medical services by type of health care provider. A limitation of this approach is that it is not able to capture substitution effects between providers. For instance, if treatment of a disease moves from inpatient treatment to outpatient treatment at the same facility, this shift and the ensuing consequences for the unit costs of output will not be captured.[5]

Measurement by provider industry

For practical reasons, the following discussion is organised by health care provider. Firstly, as already emphasised in the previous section, data may not be available to measure health volume by disease across health care providers.[6] Secondly, some health care institutions produce quite different products from others (for example, curative treatments of an acute problem versus ongoing treatment for a chronic condition), which makes it natural to discuss output measurement for each type of provider separately.

The discussion will distinguish between hospitals, residential long-term care facilities and providers of ambulatory health care, following the HP categorisation. For each type of provider, the measurement of health volume output requires identifying a set of homogeneous products and a set of weights (value shares) that can be used to aggregate changes in these products (direct volume measurement) or to aggregate unit cost changes in these products (deflation approach – see above).

Hospitals

Typically, hospitals provide inpatient as well as outpatient treatments. Within a hospital, different outputs can be captured by the identification of treatments. Although this may not always be possible in practice, it is desirable from a conceptual viewpoint to separate inpatient and ambulant treatments of the same disease so as to be able to capture effects of substitution between inpatient and outpatient treatments provided by the same hospital.

To measure treatments, there are two common patient classification systems, both of which attempt to deal with the heterogeneity of hospital output while making comparisons between hospitals possible. The first system is the International Classification of Diseases (ICD) which was originally developed as a basis for mortality statistics, and which thus refers to diagnoses. It is used to classify diseases and other health problems recorded on many types of health and vital records, including death certificates and hospital records. The ICD underlies the development of diagnosis-related groups (DRG) categories. Responsibility for updating the ICD lies with the WHO; ICD-10 was endorsed in 2004.

The most widely available categorisation of hospital inpatient services is provided by DRGs (Fetter, 1991). They were developed with the explicit objective of creating relatively cost-homogeneous groups in order to compare hospital performance. Instead of providing a cost for each component of a hospitalisation, DRGs give a composite bundle of hospital services a single predetermined cost or reimbursement rate. This amount includes all activities from which the patient benefits in the process of the treatment, including nursing care, drugs, imaging and the hotel amenities of care.

DRG systems are attractive for the measurement of volume output because they provide information on (unit) costs per type of treatment and on the number of treatments carried out: in other words, the basic ingredients for a (quasi) price or a volume index.

A typical DRG system comprises a large number (500-1000) of categories. By construction, each category stands for a relatively homogeneous service and thus, in principle, construction of a unit cost or of a volume index from the most detailed level of categories is desirable. This is, however, not always possible. One main reason is that DRG systems are updated on an ongoing basis, with some categories being aggregated and others disaggregated, making comparisons between periods difficult.[7]

Given cost weights and the number of treatments, either a unit cost (quasi-price) index or a direct volume index can be constructed. The choice between these alternatives is often governed by data availability. When indices are constructed such that the weight reference period proceeds the latest period for which (quasi) price changes are to be measured, it is easier to construct a unit cost index than a direct volume index.

The matching of services has its limits when comparable products do not exist in comparison periods or when new services only gradually diffuse in practice. This is the case even when using relatively sophisticated output measures such as DRGs. One possible way to deal with the problem of categorisation changes in the output indices involves imputing values – by deflation or inflation – where cost data are missing for any particular period of time (Castelli *et al.*, 2008).

Residential long-term care facilities

For residential long-term care providers, the notion of a complete treatment is not very meaningful. Elderly and other long-term stay patients tend to have complex clinical presentations characterised by disability, dependency and multiple pathologies. The start and end point of the condition is not clear, and the fact that the condition is chronic means that incentives for reducing the length of stay as under a DRG system would be inappropriate. A classification for nursing home patients, Resource Utilisation Groups (RUGs),[8] has been developed (Carpenter *et al.,* 1995), as DRGs are of little value for chronic care patients and those without straightforward clinical conditions.

Where countries do not have an RUG-type system, the number of occupant days differentiated by the level of care can be used. Many countries have their own care classifications based on an assessment of a patient's care needs, and thus the intensity (and cost) of the care received. If information on costs per level of intensity of care is available, it can be combined with data on the number of occupant days to derive either an approximate direct volume index of output or an approximate unit cost index.

This is, for instance, the case in Denmark. Data on unit costs for different types of residential facilities (nursing homes, sheltered housing, day centres and social centres) exist, if only for the city of Copenhagen. Combined with the evolution of the number of care places, grouped in the same way, this is used to construct a unit cost index for residential and care places.

Providers of ambulatory health care

Medical and dental practices

Medical and dental practice activities are considered market production in many countries. Persons receiving dental services in particular are more likely to be charged an

economically significant price. Medical and dental practice activities each provide both general and specialist services. In the case of both types of services, an appropriate deflator for market output would be a price index such as a CPI component that accounts for the different types of services received and captures – to the extent possible – quality changes.

As a rule, the definition and measurement of hospital-based outpatient treatments remain rudimentary (Castelli *et al.*, 2007). A limited number of countries have developed and used outpatient classification systems. Ideally, all outpatient activity related to one treatment episode would be combined into one measure of output, in the same way as for inpatient activity. The episode would include all consultations, pathology tests, imaging and prescriptions. Developing outpatient DRGs requires a capacity to track patients across outpatient services for the same treatment. To do so, it would be necessary to be able to identify the start and end point of the complete treatment and to have an appropriately supportive legal and information technology framework. There have been limited developments in outpatient DRGs.

Until an international classification system for outpatient care is developed and implemented or national systems become more widespread, basic quantity measures such as number of doctor visits will have to be used in the construction of volume measures, in particular for non-market providers. The *Eurostat Handbook on Prices and Volumes* in the national accounts suggests that outputs should be classified into medically meaningful groups that are as homogeneous as possible. The stratification may take into account the medical content of the output as well as a time dimension. For example, a visit to a general practitioner could be a measure of output. Other quantity indicators would be patient transfers by ambulance, number of pathology tests by broad category of type of test, or the number of prescriptions filled by type. Generally, these data are collected as part of the process of public or private reimbursement.

The *Eurostat Handbook* suggests that general practitioner consultations are measured by the number of visits, but specialist consultations are measured by the first visits only. The reason given for this differentiation is that specialists' visits are more likely to be follow-up visits, *i.e.* ongoing treatment for the same medical condition. This distinction seems arbitrary, as many GP visits are also follow-up visits. In addition, while this recommendation may be applicable to some specialties, it may not be applicable to all. Specialties to which the notion of ongoing treatment would not usually apply include many diagnostic specialties such as pathology, radiology, nuclear medicine, etc.

Offices of other health care practitioners

This category refers to a range of diverse activities such as activities of nurses, midwives, physiotherapists or other paramedical practitioners in the fields of optometry, hydrotherapy, medical massage, occupational therapy, speech therapy, chiropody, homeopathy, chiropractics, acupuncture, etc. Many of these services are provided by market producers. In cases where some of those activities do not have significant prices, it will be necessary to aggregate the output by using relatively basic methods such as the number of consultations, visits or tests.

Other health care providers

This category includes ambulatory health care centres and providers of home health care services. Most likely, output will be aggregated by using relatively basic methods such as the number of consultations, visits or tests.

Overview of measures

To complete the discussion above, Table 13.1 provides a summary of suggested output measures by provider category. The proposed methods are in principle output-based, although the degree to which they constitute fully-fledged measures of output varies and sometimes depends on the level of stratification in implementation.

Table 13.1. **Overview of indicators for volume output of health service providers (HP.1-HP.3)**

HP code	Providers	Output-based methods
HP.1	**Hospitals**	(Quasi) Price index based on DRGs (cost or revenue-weighted) Direct volume index based on DRGs (cost or revenue-weighted) Direct volume index based on ICD categories (*e.g.* number of discharges by category with quantity-weights such as shares in hospital days) Number of days of care* Number of cases/discharges*
HP.2	**Residential long-term care facilities**	(Quasi) Price or unit cost index based on Resource Utilisation Groups (RUGs) or equivalent (cost-weighted) Direct volume index based on RUGs or equivalent (cost-weighted) Direct volume index based on number of days of care by level of care (cost-weighted) Direct volume index based on number of cases by level of care (cost-weighted) Number of days of care* Number of cases/discharges*
HP.3	**Providers of ambulatory health care ****	
HP.3.1	Medical practices	(Quasi) Price index based on number and type of service (cost or revenue-weighted) Direct volume index based on number and type of service (cost or revenue-weighted) Relevant component of the consumer price index if applicable** (Quasi) Price index based on average costs/revenues per service (cost or revenue-weighted) Direct volume index based on number of services (cost or revenue-weighted) Number of services*
HP.3.2	Dental practices	Relevant component of the consumer price index (if applicable) Direct volume index based on number of services (cost or revenue-weighted) Number of services*
HP.3.3	Offices of other health care practitioners	Direct volume index based on number of consultations by type of consultation (cost or revenue-weighted) (Quasi) Price index based on average cost or revenue per consultation (cost or revenue-weighted) Relevant component of the consumer price index (if applicable) Number of consultations* Number of cases treated*
HP.3.4	Ambulatory health care centres	
HP.3.5	Providers of home health care services	

* proxy index.
** services are defined as consultation/visit/treatment depending on the typology of the country.
Source: IHAT for SHA 2011.

Measurement of quality changes

Price and volume measures of output should reflect quality changes in the health services provided. In other words, only prices, unit values or quantities of the same quality, that is, with the same characteristics, should be compared over time.

A first and important step towards capturing quality change is the correct stratification, *i.e.* a comparison of products with the same or at least similar characteristics. In this way, stratification keeps quality constant if the products included in a particular stratum are relatively homogeneous.

A second way of incorporating the quality of services into output indices would require an adjustment factor that reflects the compliance rate with established procedures by country and disease group. Changes in the proportion, either positive or negative, would indicate where medical practice and procedures have changed to reflect the introduction of new treatments and improvements in the existing practices. There is a limited literature reporting such proportions. For example, by averaging the findings from US preventive care studies, Schuster *et al.* (2005) found that 50% of people received recommended care, and for acute care studies they found that 70% of patients received recommended care and 30% received contraindicated acute care. It is important to note that the achievement of full compliance with guidelines, that is, a rate of 100%, is not necessarily a public health goal. In industrial production processes, uniformity ensures the highest quality of outcomes. The same does not hold in health care, as individual differences and preferences need to be accommodated.

There is a wealth of information on clinical practice guidelines by country but insufficient summary information at this stage on the rate of compliance with best practice. In the United States, the National Guideline Clearinghouse maintains a catalogue of high-quality guidelines published by various organisations (mostly professional physician organisations). In the United Kingdom, clinical practice guidelines are published primarily by the National Institute for Health and Clinical Excellence (NICE). In the Netherlands, the Dutch Institute for Healthcare Improvement (CBO) and the Dutch College of General Practitioners (NHG) have guideline development programmes that use an evidence-based approach. In Germany, the Agency for Quality in Medicine co-ordinates a national programme for disease management guidelines. All these organisations are members of the Guidelines International Network, an international not-for-profit association of organisations and individuals involved in clinical practice guidelines.

A third aspect of quality relates to the impact/contribution of health services on health outcomes. Obviously there is a strong connection between process and output, as treatment guidelines are based on medical evidence about what is efficacious. Health services researchers recommend using both process and outcome indicators for two reasons. First, there is a difference between evidence in research (efficacy) and outcomes in real life (effectiveness). Second, there is frequently a considerable time lag between a process and its impact on the outcome. For example, studies have examined family doctors' compliance with guidelines for hypertension treatment. The outcome associated with this treatment is a reduction in heart attacks and strokes (as an intermediate outcome) and mortality related to cardiovascular diseases (as an ultimate outcome). Reductions in the incidence of these diseases occur over a very long time period, and hypertension treatment is only one factor involved in the ultimate outcome. Among other things, the hypertensive patients themselves have to comply with a lifestyle that includes healthy nutrition and adequate exercise.

This discussion emphasises the use of quality adjustment using process and outcome indicators. It is also noted that there is an "industry of quality measures", but at this stage none appear appropriate for quality adjustment of health volume output. The main reasons for this lack of applicability is that many of the available process and outcome indicators are specific to a country or even institution.

Since quality is multidimensional, it would be ideal to be able to subsume several characteristics of quality into a single indicator that reflects the contribution of the product

to outcome. Alternative means have been suggested to derive a single indicator. The first is to choose the most important dimension only, *e.g.* 30-day survival rate. The second is to use indicators from more than one quality dimension and weigh them as equally important. Third, indicators from a number of quality dimensions can be used but expert opinion should be sought on the appropriate weights (UK ONS, 2008). Finally, measures such as quality-adjusted life years (QALYs) can be used that reduce modifications in health outcomes due to medical care to one dimension, the quality-adjusted gain in time. However, many of these empirical methods are still in a research stage and are normally not robust enough to be applied in the System of Health Accounts.

There are a number of desirable characteristics of indicators that could be used for quality adjustment for volume output for determining the marginal contribution of the health industry to outcome. These are outlined below.

- The quality measure should be aligned with the processes sought by consumers, which would generally be a complete treatment of the disease;
- The adjustment in output should reflect the marginal contribution of the health industry to an outcome. It should not be affected by any other factors that influence health outcomes such as genetic background, income or lifestyle;
- Consumers are ultimately concerned to achieve an improvement in their health outcome. Waiting times and comfort are secondary to improvements in health status. This points to the conclusion that different dimensions of quality should not be given the same weight;
- In many health treatments or processes, there is a time lag before the improvements in health status. Quality adjustment needs to address in a realistic manner the impact of lifetime effects of health expenditures;
- The quality measure should reflect as closely as possible the normal, average or expected effect of the activity on the state of health. Individual capacities to benefit from treatment, or what is known as co-production, should not be counted in the measure of quality-adjusted health volume output;
- International comparison is important, and the indicators and methods of output adjustment should be standardised across countries to facilitate comparisons (Smith and Street, 2007).

To sum up, methods for quality adjustment of output are still under development. Some headway towards capturing quality can be made by using detailed product specifications and following the associated costs and treatments over time. The importance of explicit quality adjustment is undeniable, but until there is a consensus on techniques for adjustment, it will be difficult to put forward a recommendation for an explicit quality adjustment of health volume output in the System of Health Accounts. A similar conclusion has been reached by the United States Bureau of Economic Analysis (BEA), who note that the "BEA will not attempt to account for potential changes in the quality of treatments, a problem where no clear consensus exists on a solution" (Aizcorbe and Nestoriak, 2008, p. 25).

Inter-country indexes of price and volume

Purchasing Power Parities (PPP) are used to compare levels of price and volume between different regions or countries in the same period of time. These allow for a *spatial*

price/volume breakdown (*i.e.* a price/volume breakdown among countries for a given point in time).

Conceptually, PPPs are very similar to consumer price indexes. PPPs are measures of price level differences across space or, in their most popular form, across countries. Because the prices of goods and services in different countries are expressed in national currencies, the purchasing power parity between currencies of two countries, say A and B, is the number of units of currency of country B (or A) that has the same purchasing power as one unit of currency of country A (or B). Though the PPPs are similar to price index numbers in spatial comparisons, they assume special significance because the PPPs can be used as a conversion factor, in place of exchange rates, in converting various economic aggregates from different countries into a common currency unit (a statistical construct).

PPPs are regularly measured for all components of GDP in the International Comparison Programme (see Box 13.4 below) and in the Eurostat-OECD PPP comparison.[9] In their simplest form PPPs are nothing more than price relatives that show the ratio of the prices in national currencies of the same good or service in different countries. PPPs are calculated in three stages. The first is at the product level, where price relatives are calculated for individual goods and services. The second is at the product group level (*i.e.* basic heading), where the price relatives, which are calculated for the products in the group, are averaged to obtain unweighted PPPs for the group. And the third is at the aggregation levels, where the PPPs for the product groups covered by the aggregation level are weighted and averaged to obtain weighted PPPs for the aggregation level. The weights used to aggregate the PPPs in the third stage are the expenditures on the product groups.[10]

In the ICP and Eurostat-OECD PPP comparison programmes, expenditure on health appears in three parts in GDP expenditure: household consumption expenditure, expenditure of non-profit institutions serving households (NPISH) and government expenditure on health, and they are classified according to 25 basic headings (20 in the case of the ICP).[11]

PPPs for medical goods and services under household consumption expenditure and social benefits in kind are based on a "normal" price collection. This means that the estimation of PPPs starts by selecting a sample of products in each expenditure category to compare their prices in different countries. Prices to be collected should be average prices for the whole country.

For inpatient services, an input method is applied for public but also for private hospitals. This follows national accounts conventions that recommend that the value of the output of institutional units in the health industry is measured by the sum of the costs of production in the case of non-market products (SNA 2008, 6.130). In the input method, PPPs for the compensation of employees are based on a wage comparison of employees, that is, wages are used in the PPP estimation in the same way as normal product prices. PPPs for other inputs are based on proxy PPPs extracted from other parts of the comparison.[12]

A PPP between two countries for a particular expenditure category is a geometric average of all price relatives (parities) formed from a set of product prices belonging to the category. In the ICP and OECD-Eurostat comparison programmes, countries also indicate whether or not products they have priced are representative (so-called starred products). In the averaging, parities based on products that are representative in both countries, get a

higher weight, and parities based on non-representative products (but included in the product list because of their importance for some third countries) are excluded.

In a multilateral comparison, PPPs are derived for all pairs of countries whenever prices are available in both countries. If all countries price the same products and the products are representative everywhere, geometric averages of the price relatives provide directly transitive PPPs (that is, for countries A, B and C, PPP relatives for A/B and B/C are consistent with A/C) for a given basic heading. On the other hand, if a full set of representative prices is not available for all countries, comparisons between pairs of countries will be based on different sets of products, resulting in intransitivity. To make results transitive, an EKS procedure, due to Elteto and Koves (1964) and Szulc (1964), is applied where the final parity for a pair of countries is based on a geometric average of all direct and indirect (via all other countries) parities.

Box 13.2. **PPP indexes**

If there are M countries in the comparison, then a price index between country j and country k is derived as:

$$I_{jk}(EKS) = \prod_{s=1}^{M} \left[I_{js} \times I_{sk} \right]^{\frac{1}{M}}$$

where:

$$I_{js} = \left\{ \prod_{i \in M(s)}^{M} \left[\frac{p_{is}}{p_{ik}} \right]^{\frac{1}{n(s)}} \times \prod_{i \in M(j)}^{M} \left[\frac{p_{is}}{p_{ik}} \right]^{\frac{1}{n(j)}} \right\}^{\frac{1}{2}}$$

where n(s) and n(j) are the number of starred items in countries s and j; M(s) and M(j) are the sets of commodities that are starred in countries s and j.

Despite a long tradition of work in the area, three main problems have to be addressed in the measurement of PPPs. The first issue is to identify products that are comparable across countries: while simple in some cases (*e.g.* a particular type of washing machine), this can be complicated in many other cases, because products are not identical, because there are differences in quality or because products simply do not exist in all countries. The second issue is to ensure the representativeness of the products: whatever price is compared, it has to be the price of a product that is widely and typically purchased in each country. The third issue arises when there is a product, but no meaningful market price for comparison. Issues one and two arise in the comparison of all prices, while issue three arises in the comparison of non-market products. In many countries, health services count among these products.

When goods or services are supplied by a non-market producer such as the government, the prices charged to consumers are significantly below the price that a market producer would charge. In some cases, the price may even be zero. It would make

no sense to compare such prices charged to patients or consumers across countries, as they reflect administrative decisions and not the value of products. It has therefore not been customary in PPP compilations to compare costs of producing non-market goods and services.

There are two possibilities for comparing costs, one based on *inputs* and one based on *outputs*. The input-based method, traditionally applied in PPP comparisons of non-market products, consists of comparing the prices of inputs in the production process of non-market services. In the case of health services, an input-based method would, for example, compare the wage rate of a surgeon in different countries. In other words, the price comparison is approximated through a comparison of wages or values per unit of inputs. Apart from the fact that it is notoriously difficult to compare wages across countries (even for the same type of occupation: qualifications may be different, it is hard to control for experience and seniority payments, etc.), the main drawback of this method is that it ignores any productivity differences between countries. In other words, if health services are provided more efficiently in one country than in another, this would go unnoticed in a PPP comparison that is based on the price of inputs.

Box 13.3. **The International Comparison Programme**

The International Comparison Programme (ICP) is a worldwide statistical operation that produces internationally comparable price and volume measures for gross domestic product (GDP) and its component expenditures. The measures are based on purchasing power parities (PPPs). To calculate the PPPs, the ICP holds surveys every five years to collect price and expenditure data for the whole range of final goods and services that comprise GDP, including consumer goods and services, government services and capital goods.

The ICP was first established in 1968 as a joint venture of the UN and the International Comparisons Unit of the University of Pennsylvania, with financial contributions from the Ford Foundation and the World Bank. Starting with a modest project to undertake comparisons in 10 countries in 1970, further ICP rounds have been conducted in 1975, 1980, 1985, 1990 (only partial), 1993, and 2005. The 2005 ICP was the largest ICP round to date, covering 146 economies from five geographic regions: Africa, Asia and the Pacific, the Commonwealth of Independent States, Latin America, and Western Asia, and the countries of the regular PPP programme managed by the OECD and the Statistical Office of the European Union (Eurostat).

The ICP is being implemented as a true global initiative for the second time, with the reference year 2011. It will build on well-programmed activities of a wide network of national and bi- and multi-lateral institutions that will engage in methodological research and review, survey activities and data processing and analysis in the areas of price statistics and national accounts and related fields, with a view to estimating the PPPs of the world's principal economies.
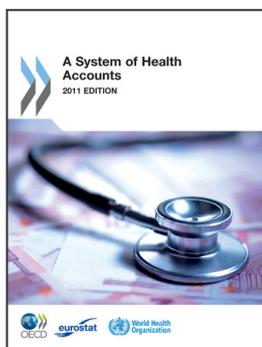

The second option for comparing costs is based on *outputs*. Here PPPs are measured by comparing the costs per unit of output; in the case of medical services, this is typically the *cost per treatment*. The comparison of quasi-prices per unit of output is – in principle – capable of reflecting productivity differences between countries. It is thus conceptually preferable to input-based approaches (SNA 2008, 15.122).

In the health domain, costs per unit of output are not generally readily observable, but there is an alternative source of information that provides valuations of outputs: in many countries, health services are managed through reimbursement schemes, where health providers and health administrations or insurance companies either negotiate reimbursements per treatment or where the government administers reimbursements per treatment. Reimbursement values per treatment or per episode of illness can be used to emulate the role that prices play for other goods and services. We label negotiated or administered rates as "*quasi-prices*" to signal that they are not necessarily the result of market transactions, that they are not prices that apply to transactions between producers and consumers of health services, and that they are not observed.

An OECD Working Paper (Koechlin, Lorenzoni and Schreyer, 2010) reports on the pilot work on PPPs at the OECD that aims to develop an output-based measurement of hospital services PPPs.

Notes

1. For a discussion of index number formulae for consumer and producer prices, see ILO *et al.* (2004a, 2004b).

2. The Consumer Price Index measures changes over time in the prices of consumption goods and services acquired, or used, by households.

3. The Producer Price Index measures average changes in prices received by domestic producers for their output. The output of the services sector and other sectors that do not produce physical products is conceptually within the PPI universe.

4. An activity might be a consultation with a general practitioner, a hip replacement involving a stay in hospital or a visit to an outpatient clinic.

5. See Aizcorbe and Nestoriak (2008) for a discussion about the effect on costs (and real expenditure growth) of treatment substitution among providers.

6. See the discussion on expenditure by disease in this Manual.

7. This is true to a larger extent for cross-country comparisons, given that the DRG classification logics may differ.

8. RUG-III is a hierarchical classification system with a structure that has forty-four groups. Each group is assigned an index score that represents the amount of nursing time and rehabilitation treatment associated with caring for residents who qualify for that group. The forty-four groups fall into seven major categories: rehabilitation (14 final RUG-III groups), extensive services (three groups), special care (three groups), clinically complex (six groups), impaired cognition (four groups), behaviour problems (four groups), and reduced physical function (ten groups). Across most RUG-III categories, patients are divided into a final classification group based on their performance on an index of four activities of Daily Living: eating, toileting, bed mobility and transferring.

9. For a full description of the methods used, the reader is referred to World Bank (*www.worldbank.org/ICP*) and Eurostat-OECD (2006).

10. This methodological description is greatly simplified. Further information can be found in Eurostat-OECD (2006), *Eurostat-OECD Methodological Manual on Purchasing Power Parities*.

11. A detailed description of the basic headings can be found in Annex II of the methodological manual (see note 229). Government expenditure on health is classified according to 17 basic headings, Household consumption according to seven basic headings, and spending by NPISH to a single heading.

12. As an example, PPPs for pharmaceutical products are used as proxy PPPs for intermediate consumption, although the relative difference between prices paid by hospitals and pharmacy prices is not necessarily the same in all countries.