

# **ANEXO D**

**REPORT OF THE**

**TASK FORCE**

**VOLUME MEASURES FOR NON-MARKET SERVICES**

**NACE L**

September 1998

## CONTENTS

<b>EXECUTIVE SUMMARY</b>	<b>3</b>
<b>1. Introduction</b>	<b>5</b>
<b>2. Different approaches to estimate the volume of output of non-market services of NACE L</b>	<b>6</b>
<b>3. Input methods</b>	<b>7</b>
<i>3.1. Wages and salaries in cash</i>	<i>7</i>
3.1.1. Wage rate method	8
3.1.2. Average wage method	9
<i>3.2. Other inputs</i>	<i>10</i>
<i>3.3. Other special issues</i>	<i>12</i>
<b>4. Output indicator methods</b>	<b>14</b>
<b>5. Conclusions</b>	<b>16</b>
<b>Annex 1: Participants, Meetings and Documents</b>	<b>17</b>
<b>Annex 2: Comparing different input approaches to calculating constant price values for wages and salaries</b>	<b>20</b>
<b>Annex 3: Output indicators: the case of Finland</b>	<b>35</b>

## EXECUTIVE SUMMARY

The aim of the Task Force was to investigate the existing and potential methods for estimation of output of NACE L (Public Administration, Defence and Compulsory Social Security) in constant prices, to classify methods in A, B and C methods, where particular attention should go to those methods than can be applied from 1999 onwards, and improve the comparability of GDP in constant prices.

For NACE L, two broad types of methods can be distinguished to estimate the volume of output: *input methods*, in which the volume of output is approximated by summing the volume of the inputs, and *output indicator methods* which are based on information on the volume of services rendered.

Given the short-term aim described above, the Task Force mainly focused on harmonisation of input methods. It considered output indicator methods theoretically superior, but (for NACE L) far from being applicable in a harmonised way. Output indicator methods are theoretically superior, because input methods do not measure output and cannot take changes in productivity into account.

The most appropriate application of an input method is to deflate all inputs separately, taking quality changes of the inputs into account. For each category of inputs (intermediate consumption, wages and salaries in cash and in kind, employers' actual and imputed social contributions, and consumption of fixed capital) recommendations on deflation methods are given.

The most important category of inputs for NACE L is of course wages and salaries in cash. For this category two types of methods are distinguished: *wage rate methods* (based on a sample of representative wages for different categories of employees) and *average wage methods* (based on average actual wages). Using different methods can be a serious source of incomparability unless corrections are made. For this, it is foremost important to decide which elements of wage increases should be in the volume component and which in the price component. A reasonable approach seems to be that any increase in wages which is related to performance of employees is related to productivity, and so should be included in the volume component. So, for example, the effect of promotions should be included in volume. However, on the effect of an aging work force (seniority), the Task Force could not reach agreement.

Both the wage rate and average wage methods can result in the desired price/volume decomposition, if applied in the correct way. The Task Force is not in favour of applying arbitrary productivity adjustments, or adjustments based on disputable assumptions, to the sum of deflated inputs. Such productivity adjustments contribute neither to the quality of the results nor to the comparability.

For the development of output indicator methods, it is essential first to define the elementary (final) outputs of each non-market producer. The next step is to find the most suitable indicators to measure these elementary outputs. It is also vital that defining the outputs should be done by (or at least in agreement with) the producers themselves. The output distinguished should be as homogeneous as possible.

The Task Force concluded that only output indicator methods could theoretically become A methods, provided they satisfy a range of criteria. An output indicator method is a C method if not all of the services produced are covered by the indicators, or if the indicators do not in fact relate to outputs but to inputs or outcomes.

Input methods cannot become A methods. If input methods have to be used they should follow the recommendations of this report. The use of numbers of employees as an extrapolator of base year output would be a C method.

## 1. INTRODUCTION

1. The aim of the Task Force was to investigate the existing and potential methods for estimation of output of NACE L in constant prices, to classify methods in A, B and C methods, where particular attention should go to those methods than can be applied from 1999 onwards, and improve the comparability of GDP in constant prices (see document CN 349 presented at the Working Party on National Accounts in December 1997).
2. NACE L concerns Public Administration, Defence and Compulsory Social Security. The Task Force considers however the recommendations contained in this report more broadly applicable. In particular, the recommendations in section 3 on input methods seem to be applicable to all activities where input methods are used for volume measurement, even market activities.
3. Given the short-term aim described above, the Task Force mainly focused on harmonisation of input methods. It considered output indicator methods theoretically superior, but (for NACE L) far from being applicable in a harmonised way.
4. Although the Commission Decision on prices and volumes discusses deflation methods from the product perspective (using the CPA classification), this Task Force considered it more appropriate for the specific type of output studied to take the perspective of the producer unit. Both the input method and output indicator method discussed, depart from data on producer units.
5. Section 2 compares input and output indicator methods from a theoretical point of view. In section 3 we will formulate how input methods should be applied, if necessary. In section 4 we will look at output indicators, and develop the criteria that they should fulfil. Section 5 gives the classification in A, B and C methods and some general conclusions and suggestions for further research.

## 2. DIFFERENT APPROACHES TO ESTIMATE THE VOLUME OF OUTPUT OF NON-MARKET SERVICES OF NACE L

6. For NACE L, two broad types of methods can be distinguished to estimate the volume of output:

- Input methods

In input methods, the volume of output is approximated by summing the volume of the inputs. Hence, a similar approach is taken as for the estimation of the value of output.

- Output indicator methods

Output indicator methods are based on information on the volume of services rendered. These methods therefore yield estimates of (changes in) output that are independent of the estimates of the (changes in) inputs.

7. In the next sections we will look in detail at both methods and make recommendations concerning an appropriate application. In this section, we will compare the types of methods from a theoretical point of view.

8. First, however, it is necessary to understand what the output of non-market units is. In general, the output of non-market units is the amount of services provided to external users (who can be individuals, other organisations or the society as a whole). It is necessary to define output more precisely for each individual non-market service or unit. This should in principle always be possible, even for collective services. If it would not be possible to define what a particular government unit produces, the usefulness of this unit can be questioned....

9. The possibility of defining output does not mean of course that it is easy to measure it in practice. In particular for NACE L, the problems of defining units that can be measured in practice are not easy to solve, because of the high share of collective services in this branch. See section 4 for a more extensive discussion of defining and measuring output of non-market units.

10. Suppose the output of a non-market producer can be described as the number of units of services delivered (e.g. number of tax assessments completed, number of court cases settled, etc.). Then we get the following relations:

In current prices:

number of units \* unit costs = total costs = output in current prices

In constant prices:

number of units \* unit costs in the base year = output in constant unit costs

(For non-market products, the unit costs plays the role of the price, see ESA95, p. 10.25.)

7. To measure output in current prices it suffices to calculate the total costs (i.e. no calculation of the number of units produced has to be made). Output in constant prices (or constant unit costs) can however not directly be observed. In an output indicator method, the number of units produced and the unit costs in the base year are estimated. In an input method, the output in constant unit costs is approximated by deflating the current costs.

8. Of course, in general, output cannot simply be described as quantity of units. There is the quality dimension as well. So, we should speak about volume of output, and change the above formulas using changes in volume and changes in costs. It will not alter the essence, which is that the unit of output is the same in both current and constant prices, although the approach to measure output is different.

9. To judge the methods for estimating the volume of output of NACE L we need some general criteria. The following criteria are derived from the criteria for the deflation of market output as given by the Commission Decision on prices and volumes:

- Completeness: does the method cover all output?
- Productivity: does the method take into account productivity changes (where productivity relates to both the quantity and the quality effects)?
- Consistency: are the underlying principles of the method consistent with the principles of national accounting?

10. Input methods do not measure output, but approximate output by the inputs. Hence, the completeness criterion can strictly speaking not be applied to input methods. It is also clear that changes in productivity (both in terms of quantity and quality) cannot be taken into account by input methods, for the same reason: there is no independent measurement of output. What can be done in input methods is ensure that all quality changes in the inputs are taken into account, and/or make a separate productivity adjustment to correct the sum of deflated inputs. The problem is that such adjustments are inevitably based on assumptions, which cannot be verified without genuine output data. For, if these adjustments were based on a measurement of output, then this information could also directly be used.

11. Hence, the only way to fulfil the first two criteria is to develop output indicator methods, where the indicators should be constructed such that they cover all of the output and allow for quality changes. For a proper measurement of output and to take account of quality it is necessary to define the elementary units of services of each producer.

12. The conclusion is that, from the theoretical point of view, output indicator methods are preferred above input methods. However, as long as genuine output data are not available or of insufficient quality, input methods will have to be used. On the long run the use of input methods is bound to remain unsatisfactory for the users of the data, as productivity analysis remains impossible. The development of systems of output indicators covering all non-market activities whenever possible would therefore be an important step forward.

### **3. INPUT METHODS**

13. In input methods, the volume of output is approximated by summing the volume of the inputs, as said above. The most important category of inputs for activities in NACE L is normally wages and salaries in cash. We will therefore pay particular attention to the estimation of the volume of wages and salaries in cash in section 3.1. In section 3.2 we will investigate the deflation of the other inputs. In section 3.3 we will look at some special other problems involved.

#### **3.1. Wages and salaries in cash**

14. The breakdown of the value of wages and salaries in cash (for which we will say shortly “wages” from now on) in price and volume components is essentially arbitrary. It depends on whether a particular raise in wages is to be interpreted as a price effect (making the same type of labour more expensive) or a productivity effect (more productive or higher quality of labour for the same price). As long as output is not measured directly, the distinction between price and volume remains based on assumptions.

15. In all member countries of the Task Force wages are deflated. There are no countries using extrapolation of base year wages with employment indicators, nor are there countries that apply an additional productivity adjustment to the deflated inputs. But even without such adjustments the resulting volume figures will show changes in productivity, originating from the method used to deflate the wages. The Task Force noted that the methods used to deflate are widely different, leading to incomparabilities in volume measures.

16. We will distinguish two types of methods of deflation of wages: the “wage rate” method and the “average wage” method. The Task Force believes that all deflation methods are covered by these two categories. Annex 2 describes both methods in more detail and shows some examples of calculations.

### 3.1.1. Wage rate method

17. In the wage rate (WR) method, the work force is classified in a number of different categories of employees. For each category for a representative employee the wage rate change is estimated, by using for example official wage rates. The price indices thus derived for each category are weighted by the wages for that category in the current year, to give Paasche type aggregate indices. Alternatively, the wages of a recent base year could be used to give a Laspeyres index.

18. In general, the stratification of the work force should be such that categories for which significantly different changes in wage rates occur are separately distinguished. In the extreme case that all employees receive always the same wage increase no stratification at all is necessary.

19. The WR method defines the price/volume decomposition *a priori*. If a price change is not specified in the wage rate change for a certain category, it will end up in the volume component. The *default* is therefore the volume component.

20. Of the member countries of the Task Force, Austria, Belgium, France, Greece and Portugal use (a variant of) this method.

21. In *Austria* a general wage rate index is used based on the collectively agreed wage rate changes, and applied to all government employees.

22. In *Belgium* the work force of the government is classified in three different categories (civil servants, police force and defence). For each type of employee, the elementary wage rate index includes the following elements: the adaptation of the annual salaries and holiday and New Year bonuses to the Consumer Price Index and also the periodic increases attributed by government to its employees in the framework of collective agreements. The wage rate index of each category is a weighted average of the elementary wage rate indices: the weights are the compensation of the fixed base year or, if they are not available, the number of employees. This method ensures that the three following elements are always considered as variation in volume: variation in employment, normal increases due to seniority and increases due to promotions.

23. In *France*, the way civil servants are paid makes it easy for the national accountants to distinguish between price and volume components. Each grade of civil servant is placed on a scale, and each level of the scale corresponds with a number of points. The value of one point in francs is fixed by the “finance law”. There are several reasons why wages can increase. First, after discussions between trade unions and the government, it could be decided to increase the value of a point. Or, because of aging or promotions, people “climb” the scale and receive more points. Or, it could be decided to give a certain number of points for some specific jobs or grades. The only variation taken into account in the price component is the first one. All other factors are seen as volume. It could be argued however that the last category, the revaluation of specific jobs or grades, should also be in price.

24. In *Greece*, a sample of the work force is made according to grade. For each grade, for two typical family situations (singles and married with two children), a representative wage increase is estimated. As no data are available on number of employees or wages per category to use as weights, an arithmetic average is calculated. The data used are in fact the data used by Eurostat for the so-called “Article 65” calculations for the remuneration of EC officials (see TFL/8).



25. In *Portugal*, if data about the structure of the civil servants collective pay scale are available, these are used to deflate wages. If they are not available, a general wage rate index is used. The resulting price component includes the change in the wage rates (the reference value is the collective agreement in general government), seniority effects and food allowances. In Portugal, seniority is a fixed amount received by the employee every five years, which is unrelated with the performance of the employee and therefore included in the price component. The volume component includes the changes of the number of employees, promotions and other structural changes in the labour force.

### 3.1.2. Average wage method

30. The average wage (AW) method uses average wages for categories of employees. The starting point is an exhaustive set of data on numbers of government employees and their salaries, in a detailed breakdown of grades and activities. For each of the detailed categories of employees distinguished the average wage increase is determined, so that the wages of each category can be deflated separately. This approach implies that the decomposition into price and volume is implicit in the classification made in grades, activities, etc., and the detail of those classifications: shifts *between* the categories end up in the volume component, while wage increases *within* categories are included in the price component. Hence, the coverage of the volume component is increased when a more detailed stratification is used. The *default* is the price component. The method is much more data-intensive than the WR method.

31. Of the member countries of the Task Force, this approach is followed by Denmark and Finland.

32. In *Denmark*, for each of 446 detailed categories of employees the average salary for one person (or in some cases the average wage per working hour) is calculated in the year *t* and the base year. These indices are weighed together using numbers of employees of the current year as weights to generate Paasche-indices for approximately 90 COFOG-categories.

33. For categories of employees who are paid on a monthly basis the base year index is corrected in case of a general change in working hours. For categories of employees paid on an hourly basis the average wage indices are calculated for one hours work. In this way changes in the hours worked will be measured in the volume component.

34. The categories of employment comprises different kinds of occupation i.e. clerks, policemen etc., different positions i.e. head of sections, head of divisions etc. and for the local government different classes of seniority.

35. Stratifying this way has the consequence that structural changes in the work force on account of promotion, changes in the composition of occupation and partly those on account of the change in average seniority will enter the volume component. All other effects i.e. general wage increases, increased use of personal allowances for night-time work, local allowances, dirt money etc. will enter the price component.

36. In *Finland*, the official index of wage and salary earnings (ATI) is used. This index is based on individual wage data with nearly full coverage considering the central and local government (some 0,6 million employees). The ATI measures developments in average earnings for normal working time.

37. Payments for overtime, holiday bonuses etc. remain outside the earnings concept of the index. This index is, in contrast with the Danish one, a fixed-weighted Laspeyres type index, where annual total earnings of the base year 1990 are used as weights.

38. For the time being the deflators used for NACE L are based on four fixed weighted index series classified by industry (public administration and national defence) and earner group (hourly

and monthly paid employers). From the basic data of the current ATI also deflators based on 54 fixed-weighted subseries can be calculated by adding the occupational group into the classification variables.

### 3.1.3. Conclusions and recommendations

39. The main conclusion from the above is that the AW method has a default allocation of wage changes in the price component, while the WR method allocates wage changes by default in the volume component. Therefore, using different methods can be a serious source of incomparability unless corrections are made, as is shown by calculations for Denmark (see annex 2).

40. It is foremost important to decide which elements of wage increases should be in the volume component and which in the price component. As said above, the price/volume decomposition of wages is essentially based on assumptions about productivity, which cannot be verified without genuine output data. A reasonable assumption however seems to be that *any increase in wages which is related to performance of employees, is related to productivity and should hence be in the volume component*. So, for example, the effect of promotions should be included in volume. However, on the effect of an aging work force (seniority), the Task Force could not reach agreement. It was argued by some that an increase in wage, which is given automatically at given time intervals (be it every year or every five years), without any relation to performance, should be included in the price component. Others thought that the seniority effect should always be in volume.

41. Both the WR and AW methods can result in the desired price/volume decomposition, if applied in the correct way.

42. That is, for the WR method:

- the work force should be stratified such that those categories of employees with different wage rate increases are separately identified;
- the wage rate used for each category should be representative for that category;
- the wage rate should include the general wage rate change, specific wage rate changes for categories of employees, changes in allowances or bonuses that are not related to performance, and possibly changes due to seniority.
- the weights should be total wages of each category of the current year or, alternatively, a recent base year.

43. And for the AW method:

- the work force should be stratified such that those categories of employees of which shifts between these categories should be in the volume component, are separately identified;
- that implies stratifying the work force in at least different grades, different functions/activities, and possibly different seniority classes.

44. The Task Force is not in favour of applying arbitrary productivity adjustments, or adjustments based on disputable assumptions, to the sum of deflated inputs. Such productivity adjustments contribute neither to the quality of the results nor to the comparability.

## 3.2. Other inputs

45. All categories of inputs should preferably be deflated separately following the methods described below for each category.

### *Intermediate consumption*

46. As is stated in ESA95 (§10.45) the calculation of intermediate consumption at constant prices reveals no theoretical problems since intermediate consumption relates only to goods and market services. The calculation is made by deflating current values or on the basis of quantities revalued at base year prices.

47. The calculation of intermediate consumption at constant prices is made taking into account, for most of the participants in this Task Force, the information of the supply and use balances. Separate indices are used for each product. The draft Commission Decision on prices and volumes contains many additional guidelines to deflation of market products.

### *Wages and salaries in kind*

48. It is preferable to deflate separately wages and salaries in cash and in kind. For the deflation of wages and salaries in kind the price indices of the respective goods and services should be used. The Task Force observed that there are some difficulties for member countries in isolating systematically this information.

### *Employers' actual social contributions*

49. The amount of social contributions is usually determined as a percentage of the wage. It is in principle the change in this percentage times the change in the wage rate that should be regarded as the "price change" of the social contributions. That is equivalent to using the same volume index as for wages and salaries. The case is similar as the case described in p. 10.49 of ESA95 for taxes on products.

50. An alternative, applied in most countries represented in the Task Force, is to deflate the actual social contributions with the same wage rate index as wages and salaries, which implies that it is assumed that the social contribution rates have not changed. Another option is to include employers' social contribution in the wages in the AW method for deflating wages.

### *Employers' imputed social contributions*

51. The most prominent example of imputed social contributions is unfunded civil service pensions, existing in most of the EU Member States. ESA95 makes clear that the value of the imputed social contributions should, in principle, be based on actuarial considerations, similar to those used by e.g. insurance enterprises to calculate premiums. Thus, for civil service pensions, the value of the imputed social contributions should correspond to the provisions that should be made to secure the future pension entitlements of the *current* employees.

52. ESA95 also says that - as an approximation - the total actually paid social benefits (e.g. pensions) in the year can be used. That will only be a reasonable approximation when the ratio between current employees and pensioners remains stable. However, it seems to be quite common in European countries, due to demographic and political developments, that the number of government pensioners is increasing faster than the number of government employees (see also document CPNB/186 for the GNP Committee meeting of June 1996).

53. This problem might not only affect the current price data, but also the constant price data. An example will illustrate this.

		Year t in prices year t	volume index t -> t+1	Year t+1 in prices year t	price index t -> t+1	year t+1 in prices year t+1
Wages and salaries		5000	0.9	4500	1.1	4950

Imputed social contributions	1000	1.05	1050	1.1	1155
Total compensation of employees	6000	0.925	5550	1.1	6105

---

54. Suppose the values of imputed social contributions in years  $t$  and  $t+1$  are based on the actually paid out pensions in these years, and therefore increases as the number of pensioners increases. Suppose the value of the pension is in some way indexed to the current wage rate, so that the same wage rate index (1.1) can be used to deflate wages and salaries and imputed social contributions. This shows that wages and salaries have declined in volume (by 0.9), while the imputed social contributions have increased in volume (by 1.05). In principle, however, the volume of imputed social contributions should follow the volume of wages and salaries, as it is intended to represent the future pension liabilities of the *current* employees. The total volume of compensation of employees is as a result too high, and so will be the productivity growth figures.

55. In such a case would it be more appropriate to take the same volume index for the imputed social contributions as for the wages and salaries (0.9). This would give a totally artificial price index, but would at least give a more correct productivity figure. The best solution however is to correct the current price figures for imputed social contributions, in which case the above problem would not exist.

#### *Consumption of fixed capital*

56. The calculation of consumption of fixed capital at constant prices does not raise any theoretical problems. According to ESA 95 (§ 10.53) the volume and price measures may differ depending on whether a perpetual inventory method is used or not. In the first situation, which implies the calculation of the stock of fixed capital goods at constant prices, the resultant price and volume indices can be adopted. If there isn't a perpetual inventory method, it is possible to use the price indices derived from data on gross fixed capital formation by product. In this matter, it is also stated that the age structure of the capital goods acquired must be taken into account.

### **3.3. Other special issues**

57. In an input method the volume of the total output of the unit, not the individual outputs, is approximated by summing the volumes of the inputs. To derive from total output of non-market producers at final consumption expenditure of government and NPISHs, three categories need to be taken into account: market output, own-account capital formation and purchases that are supplied to households without any transformation (ESA95, p. 3.79).

#### *Market output*

58. A market producer cannot produce non-market output. A non-market producer, however, can produce market output. The distinction between market and non-market output is therefore in the first place determined by the producer. A non-market producer often has a (small) amount of market output, of which the costs cannot be separately identified.

59. The market output has to be separately deflated. The principles for the deflation of market output of non-market producers are the same as for the deflation of market output in general as stated in ESA95 and the draft Commission Decision on prices and volumes. Care should be taken to ensure that the price index used covers all costs of the market production. Using the charge paid by users as a price indicator can create biased results.

*Own-account capital formation*

60. The deflation of own-account production is a more general problem which could not be dealt with by this Task Force.

*Purchases that are supplied to households without any transformation*

61. These are purchases by general government or NPISHs of goods and services produced by market producers that are supplied to households without any transformation. Hence, this category also concerns market output so that for the principles of deflation reference can again be made to the draft Commission Decision.

#### 4. OUTPUT INDICATOR METHODS

62. Output indicator methods are based on a direct measurement of the output of non-market units. None of the Task Force member countries apply output indicator methods currently in the national accounts. Finland, however, has started a research project some years ago to develop a system of direct measurement. Information on that project and some first results can be found in annex 3.

63. In section 2 it was concluded that output indicator methods are theoretically superior to input methods. This Task Force however had the mission to look at harmonization of methods at the short-term. Output indicator methods for NACE L will need still many years of development before they can be used as harmonized method. Therefore, the Task Force concentrated on the input methods, in particular the deflation of wages and salaries. In this section, the conclusions of the Task Force on the use of output indicator methods in general are given.

64. In order to measure real productivity output has to be measured directly and independently (in practice in physical terms) so that changes in output can be separated from changes in inputs. The first thing to do is to define final output, to define what is produced and to define which products are the final products. Next step to proceed is to ask how to measure these final products and which are the most suitable indicators for the measurement.

65. Because of the diversity of the activities in general government (in particular in central government) each (local kind-of-activity) unit has to specify its final output and output indicators by itself. In each case output consists of the services provided to the customers outside the organisation (=measurement unit). Consumers are citizens, organisations (other than producer unit in question) or, for example, the community as a whole.

66. Defining output and developing output indicators by or in collaboration with the producers themselves is also essential for the acceptance of the measurement. It is not desirable that statisticians on their own decide what the output of a government unit is. In many countries government units are obliged to adapt their administration in a more output-oriented way, including definition of their output, in the context of new management programs, new methods of financing, etc. Statisticians could try to connect to the information coming available from this kind of work that is already going on in many units.

67. The services taken into account should be exactly final products in their nature. Goods and services that are produced for use inside the organisation are not included in final output. Output and outcome have to be distinguished. Outcome is the result of different kinds of activities, whereas output is that which is produced. Output indicators should be selected so that they describe the final output provided to the customers and not the outcome, or, for example, activity.

$$\left\{ \begin{array}{l} \text{Function} \\ \text{Objective} \rightarrow \text{Activity} \rightarrow \text{Output} \rightarrow \text{Outcome} \\ \text{Aim} \end{array} \right.$$

68. Sometimes the identification of final products may be difficult and in the absence of real output indicators in describing output substitute (proxy) measures for output may have to be used.

69. The definition and the measurement of collective services is a difficult task. One reason for this is that the customers for collective services are not very clearly definable - nor are services' impacts on individuals. 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 organisations carry out the function of their activity and guarantee the achievement of the goals of activity. This concerns especially the definition and the measurement of output of collective services - but to understand the production aspect and producer's side in the production of all non-market services is the step forward.

70. For the measurement of output to be reliable enough output indicators should cover the whole output or at least the most important and essential final products which make up most of total output. For some particular units, it could perhaps suffice to make a representative sample instead of having exhaustive observations.

71. If the quality of a final product changes slightly, the measured quantity needs to be corrected for the quality change. If the quality of a certain final product changes notably, the said product will become a new one due to the changes in its quality.

72. Classification of the final products is vitally important. Output has to be divided into sub-groups composed of as homogeneous final products as possible. Biases may occur in the results of measurement that does not distinguish between different types of final products and weight them by their own weights. As weights should be used the cost shares or alternatively work time shares of each final product.

## 5. CONCLUSIONS

73. In section 2 we formulated the following general criteria to judge methods for estimating the volume of output of NACE L:

- Completeness: does the method cover all output?
- Productivity: does the method take into account productivity changes (where productivity relates to both the quantity and the quality effects)?
- Consistency: are the underlying principles of the method consistent with the principles of national accounting?

74. Only when all these criteria are fulfilled a method can be called an A-method (for the definition of A, B and C-methods see the draft Commission Decision of prices and volumes). It was concluded that input methods cannot satisfy these criteria, and thus cannot be an A-method. If input methods have to be used, they should follow the recommendations given in section 2. The use of numbers of employees as an extrapolator of base year output is a C method.

75. Only output indicator methods can in theory become A-methods, provided the indicators satisfy the following criteria:

- they should cover all services produced by the producer that are provided to external users, and only those;
- they should be related to the output of each service. They should certainly not relate to the inputs, nor to the outcomes;
- they should take quality changes into accounts, by e.g. recognizing new products;
- they should be weighted by the costs of each service of a recent base year;
- they should be defined as homogeneous and detailed as possible;
- they should be consistent over time, that means they should measure the same output every year.

76. The Task Force considered that in case a method does not satisfy one or both of the first two criteria, this method should be called a C-method. Any other method would then be a B-method.

77. To attain full coverage of the output will in most cases require individual measurement of all outputs of all units. That is in general more difficult than attaining full coverage in price measurement where samples easier suffice. Hence, developing A methods for non-market output can generally be expected to be more difficult than developing A methods for market output.

78. This Task Force was operating with the short-term aim of providing recommendations on methods that can be applied as from 1999 onwards. As output indicator methods for activities in NACE L seem a long way from being generally applicable, this Task Force concentrated on the harmonization of input methods.

79. For output indicator methods the requirements for an appropriate application were specified (in general terms). This Task Force however did not go in detail into the definition and measurement of output of specific non-market producers within NACE L. This should be a task for further research. It will be necessary to define the output for each non-market producer, and to specify one or more indicators to measure this output. This system of indicators should fulfil the criteria set out above, but should also be easy to apply in a comparable way in different countries.



## ANNEX 1: PARTICIPANTS, MEETINGS AND DOCUMENTS

*List of participants to Task Force “Volume measures for non-market services (NACE L)”:*

Austria	Mr R. SCHWARZL
Belgium	Mr C. MODART
Denmark	Mr. T.R. GRAVERSEN
Finland	Mrs M. NIEMI
France	Mrs V. MADELIN
Greece	Mr. T. PAPANIKOLAKOS and A. KRASSADAKIS
Portugal	Mrs P. BORGES
Eurostat	Mr. P. KONIJN (chair)

*List of meetings:*

19-20 January 1998: Luxembourg

26-27 March 1998: Vienna

3-4 June 1998: Lisbon

*List of documents:*

- Eurostat: “First document for Task Force “Volume measures for non-market services (NACE L)” (TFL/1)
- Eurostat: “Draft Commission Decision laying down the principles for measuring price and volume measures ...” (TFL/2)
- Eurostat: “Harmonization of constant price data”, doc. CN 333, rev. 2, discussed at the October 1997 meeting of the Working Party on National Accounts (TFL/3)
- O. Lehtoranta and M. Niemi: “Measuring public sector productivity in Finland; progress report”, paper presented at the OECD-UNECE-EUROSTAT meeting on national accounts, Paris, June 1997 (TFL/4)
- C. Modart: “Compensation of employees in volume for NACE L and M in Belgium” (TFL/5)
- B. Kazemier: “Volume measurement of government output in The Netherlands; some alternatives”, 1991 (TFL/6)
- B. Kazemier: “Volume measurement of government output; Dutch practice since revision 1987”, paper presented at the OECD-UNECE-EUROSTAT meeting on national accounts, Paris, June 1997 (TFL/7)
- Eurostat: “Methodology and definition; Art. 65 and Annex XI of the Staff Regulations”, 1997 (TFL/8)
- M. Niemi: “Measuring government sector productivity; productivity change between years 1995 and 1996 in central government” (TFL/9)
- T.R. Graversen: “The use of average wage indices to calculate constant price values for non-market producers in the Danish National Accounts” (TFL/10)
- M. Niemi: “Volume measurement - constant price data; practice and methods by sectors - NACE L” (TFL/11)
- Eurostat: Agenda for the second meeting, 26-27 March 1998, Vienna, Austria (TFL/12)
- Eurostat: Annotated agenda for the second meeting (TFL/13)
- Eurostat: Report of the first meeting 19-20 January 1998, Luxembourg (TFL/14)
- Eurostat: Report of the first meeting of the Task Force “Volume measures for Health and Social Work (NACE N)”, 12-13 February 1998, Luxembourg (TFL/15)
- Eurostat: Draft note of the first meeting of the Task Force “Prices and volumes for education”, 19-20 February 1998, Luxembourg (TFL/16)
- P. Borges: “Volume and price measures for non-market services: general government” (TFL/17)
- Eurostat: “Deflation of imputed social contributions” (TFL/18)
- Eurostat: Proposal for the structure of the final report (TFL/19)
- Eurostat: “The Council Regulation on Structural Business Statistics and Non-market services” (TFL/20)
- V. Madelin: “Volume measures in the collective non-market services: the French method” (TFL/21)
- Krassidakis: “The measurement of constant prices in NACE L in Greek National Accounts” (TFL/22)

- D. Caplan: “New measures of public sector output” (confidential) (TFL/23)
- Eurostat: “The 1996 Comparison Exercise: Salaries”, paper for the Working Group “Purchasing Power Parities”, March 1998 (TFL/24)
- ÖSTAT, ICP unit: “ECP’96/II: Comparison of non-market services (productivity adjustment: finding the truth)” (TFL/25)
- Statistics Finland: “The index of wage and salary earnings; handbook for users” (TFL/26)
- Eurostat: Report of the second meeting 26-27 March 1998, Vienna, Austria (TFL/27)
- Eurostat: Agenda for the third meeting 3-4 June 1998, Lisbon, Portugal (TFL/28)
- Eurostat: Draft report of the second meeting of the Task Force “Volume measures for Health and Social Work (NACE N)”, 2-3 April 1998, Paris, France (TFL/29)
- Eurostat: Draft note of the second meeting of the Task Force “Prices and volumes for education”, 23-24 April 1998, The Hague, The Netherlands (TFL/30)
- T. Graversen: “Comparing different input approaches to calculating constant price values for wages and salaries” (TFL/31)
- K. Hayes: “The output of non-market services (TFEDUC/27)” (TFL/32)
- First draft of the final report (TFL/33)
- M. Niemi: “The comparison of input methods vs. output indicator method; the case in Finland” (TFL/34)
- P. McCarthy: “Input methods and the impact of changes in productivity” (TFL/35)
- ÖSTAT, ICP unit: “Comparison of non-market services at cross roads (experience, considerations, proposals)” (TFL/36)
- R. Schwarzl: Note on market output of non-market producers (TFL/37)

## ANNEX 2: COMPARING DIFFERENT INPUT APPROACHES TO CALCULATING CONSTANT PRICE VALUES FOR WAGES AND SALARIES

By *Timmi Rølle Graversen*, Statistics Denmark

When calculating the constant price value of production for the general government by the input method it is in theory not possible to measure any change in productivity. Productivity here is defined as the constant price value of gross value added (GVA) divided by the volume of employment (number of employees or working hours).

However, there will usually be an observable change in productivity in the results. This originates from the method used to calculate the constant price value of compensation of employees. This kind of productivity measure is based on assumptions implicit in the calculation method.

Different approaches to calculating constant price values of compensation of employees have been discussed in the Task Force - two of these are: the wage rate method (WR) and the average wage method (AW).

The two methods are illustrated below in a simplified form to show the differences and similarities. Examples are given to show how the results differ between the two methods and that if the level of detail in the calculation increases the results of the two approaches converge.

### *Structural changes in the work force*

Changes in the current and constant price values of compensation of employees can originate from three sources. Changes in the wages, changes in the volume of employment and changes in the structure of the work force. Changes in the wages are price effects and changes in the volume of employment should be volume effects. If calculated correctly, for non-market output, the changes in wages and changes in volume of employment will not affect productivity because output is calculated as the sum of costs. The structural changes however will result in a measured change in productivity and the size of this will depend on the method used to calculate the constant price values.

Even in the extreme case when there is a constant total work force and the wage rates do not change from one period to another there can be a change in the total sum of wages and salaries. This happens when employment shifts between different categories with different absolute wage rate levels.

**Table 1: Example of structural changes in compensation of employees**

	Wage rate t	No. employees t	Total wage cost t	Wage rate t+1	No. employees t+1	Total wage cost t+1
Judges	100	20	2000	100	50	5000
Policemen	50	80	4000	50	50	2500
Total	-	100	6000		100	7500

As GVA in general government is approximately the sum of wages a structural change such as in table 1 will result in an increase in GVA in current prices. If a wage index of 100 is used (i.e. there is no change in wage rates) to deflate the total wage costs the increase in constant prices of GVA will be proportional to the current price values.

In principle, the structure of the work force is determined by the number of categories with different absolute wage levels. Every movement of employees between such categories will result in movements in the current price value of total wages and GVA.

The question when calculating constant price values is whether structural changes of this kind should enter the volume component or the price component. The productivity effect that is

measured by the input method is based on assumptions. These assumptions should be valid in the sense that there is a plausible argument in favour of this productivity effect.

If all differences in wage levels are due to proportional differences in productivity then every structural change in the work force should influence the volume component. In the example above this means that judges are twice as productive as policemen are and are paid double wages. If on the other hand there are differences in wage levels that are not connected to productivity, structural changes between such categories should affect the price component.

#### *The wage rate (WR) method*

The WR method uses general changes in the wage rates ( $\Delta wr$ ) for each category of employees. These changes in wage rates are weighed together with the share of total wages for each category in a base year, to generate Laspeyres indices for different functions of government.

The important question with the WR method is whether the change in the wage rate used is equal to the change in the actual average wage rate. Any change in actual average wage rates different from  $\Delta wr$  will not be included in the wage index and ends up in the volume component. In the WR method the coverage of the elements in the price component can be increased by stratifying the work force and using different wage rate changes for each stratum. All changes in total wages and salaries which is not contained in  $\Delta wr$  will be measured as a volume effect.

#### *The average wage (AW) method*

The AW method uses average actually paid wages for categories of employment calculated by dividing total paid out wages per category by the number of employees in each category. Elementary indices, formed by dividing the average wage of  $t+1$  by the average wage of  $t$ , are weighed together by the number of employees of each category in  $t+1$  to generate Paasche indices for different functions of government. If there are structural changes in wages and salaries within a category of employees, i.e. the average wage changes due to changes in the structure of the work force, then these will end up in the price component.

On the other hand changes in values on account of shifts of employment between categories will not affect the index and consequently end up as a volume component. With the AW method the coverage of the volume component can be increased by stratifying the work force and calculating the average wage for each stratum.

In short, the difference between the WR and AW method is: with the WR method everything that is not specified in the changes in wage rates ends up in the volume component - the default for unspecified/unknown changes in wages and salaries is the volume component. Using the AW method everything that is not specified in the stratification ends up in the price component - the default is price effect. The AW method is more dependent of a detailed stratification than the WR method and it is necessary to know the number of hours worked/persons employed with the AW method.

#### *Examples of the AW and WR method*

At the end of this annex some examples of calculations with the AW and WR method are shown. The set-up is a very simple general government sector who only employ judges and policemen. Judges are paid more than policemen are and the difference can be regarded as a difference in productivity. There is also a difference in wages between employees in the capital and in the province. This could be a wage level difference that is not related to productivity but to cost of living of the employees. So in this example there are two possible sources for structural change in wages and salaries: profession and region.

In each of the five examples the calculations are performed for three levels of detail to show three levels of 'statistical knowledge'. In every example the most detailed calculation is the 'true' story but not necessarily the correct one for constant price calculation. That depends on the assumptions made on the two kinds of structural effects.

In the examples total productivity is calculated for each level of detail in the lines a.6, b.8 and c.12. Different results in productivity shows the deviation between the methods under the different levels of detail. For simplicity the total size of the work force is constant in all examples. The first example (I) shows that when there is an identical increase in wage rate and no structural changes the two methods yield the same results. The level of detail has no influence on the calculation.

The second example (II) shows us that even with different changes in wage rates the results are the same. The level of detail has no influence on the results when wages for t are used as weights for the indices on level a and b. If the number of employees are used the results will be different for the WR method. In this case stratification will improve the results.

In the third example (III) we have structural changes and now there is different results for the two methods. The WR method is not sensitive to the level of detail and the changes in wages and salaries from the structural changes of the work force towards relatively less employment in the province is treated as a volume effect. The AW method shows that when the dimension of the structural change (regional) is not in the stratification (III.a and III.b) the structural effect does not effect the productivity because it ends up in the price component. Only if the dimension of structural change is included in the stratification (III.c) is productivity affected. At the most detailed level the results are equal for the two methods.

In the fourth example (IV) there are structural changes in both dimensions. The wage rate change is the same for all categories. In this example the AW methods sensitivity to the level of detail becomes even more apparent. In IV.b it is seen that when the stratification is only in the 'kind of employment' dimension the structural effect from the changes in regional employment ends up in the price component and does not affect productivity. We can also see that the results of the WR method are not sensitive to the level of detail as long as the changes in wage rates are the same for all categories. Again on the most detailed level the results are equal.

In the fifth (V) example there is structural changes like in ex. IV and there is different changes in wages rates like in example II. Now the WR method is sensitive to the level of detail. The AW method is not influenced by the different changes in wage rates and the results are equal to those in example IV.

We can conclude that with the WR method the level of detail only matters if there are different changes in wage rates for various categories of employment. The AW method is very sensitive to the stratification. Structural changes between categories will end up in the volume component if these categories are distinguished in the stratification.

#### *Comparing empirical results of the WR and AW method*

It is possible to compare the results of the two methods with Danish data when we narrow the example to the information for the central government.

Apart from the National Accounts figures there are two statistical sources that measure the WR changes for the central government.

The first is the official general *wage regulation percentage*. All wages and salaries of the central government are adjusted twice a year with the same general percentage. The wage regulation percentage comprises the results of collective bargaining and a regulation of 80% of the part of the wage change in the private sector that is beyond the changes from collective bargaining.

The second source is a detailed wage statistics of central government wage rates for different categories of occupation. This statistics is calculated with a constant structure of employees in different categories of occupation i.e. changes in seniority or the number of employees between categories will not affect the results.

Comparison of the two alternative statistical sources shows that the results are almost the same. There is a small variation which means that for the period 1991-95 some categories of employment had wage changes that differ from the *general wage regulation percentage* by approximately 0,01 to 0,05 %. When differences are that small the analyses will not improve in quality by using the detailed statistics on categories. So below the general *wage regulation percentage* is used to compare with the National Accounts figures.

The comparison is made for the period 1991-95. This period is chosen for simplicity because in this period there is no change in the length of the working week.

In table 2 below the indices for the calculated annual wage regulation percentage are shown. As it is mentioned above the wages and salaries are adjusted twice a year. These adjustments were used to calculate monthly indices and these monthly indices are averaged to make indices for each year.

**Table 2. General change in wage rates for the Danish central government sector (WR-method)**

	<u>1991</u>	<u>1992</u>	<u>1993</u>	<u>1994</u>	<u>1995</u>
	Index 1991=100				
Wage regulation percentage	100,00	102,09	103,40	104,74	105,94

Table 3 below shows the results from the National Accounts calculations with the AW method. The results are shown for 8 categories of government functions. It is clear that there are significant differences in the results. Over the four-year period the difference in the measured wage increase is approximately 2%. The comparison confirms that the WR method will underestimate the price component compared with the AW method.

This is partly due to an increase in average seniority. Seniority is a price component in the National Accounts figures for the central government because the stratification does not include levels of seniority. Seniority does not affect the wage regulation percentage. One known effect is that the seniority scale has been shortened for office clerks so that the young staff has had a wage increase through a general lift in seniority. Another explanation of the differences is an increasing use of different kinds of personal allowances. When there is an increasing use of allowances the general wage regulation percentage will not be affected, but the average wages within groups of employment will raise and consequently the wage index calculated with the AW method.

**Table 3. Wage indices for central government in the Danish National accounts (AW method)**

	<u>1991</u>	<u>1992</u>	<u>1993</u>	<u>1994</u>	<u>1995</u>
	Index 1991=100				
Public: General public services	100,00	102,38	103,67	105,25	106,95
Public: Defence	100,00	104,06	104,89	106,18	107,37
Public: Education and research	100,00	102,35	103,90	106,01	107,25
Public: Health	100,00	104,96	106,32	108,54	109,98
Public: Social security and welfare services	100,00	103,20	105,55	108,22	106,48
Public: Housing and community amenities	100,00	102,91	104,88	107,78	109,53
Public: Recreational and cultural services	100,00	102,46	103,92	106,16	107,66
Public: Economic services	100,00	103,17	104,67	107,27	108,44

Using the AW method catches a greater price variation between the different sectors of government activity. The WR method does not measure this because the official wage rates are all adjusted with the same percentage, and any additional effect on wages and salaries is measured as a volume effect by the WR method.

#### *Summary*

It is shown that the WR method will have a tendency to underestimate the price component while the opposite is the case for the AW method. This is confirmed by comparing the Danish figures calculated with both methods.

With the WR method everything that is not specified in the changes in wage rates ends up in the volume component - the default is volume component. Using the AW method everything that is not specified in the stratification ends up in the price component - the default is price effect. The AW method is more dependent on a detailed stratification than the WR method and it is necessary to know the number of hours worked/persons employed with the AW method.

If changes in wage rate are the same for all groups, then it makes no difference to stratify with the WR method. The challenge with the WR method is to capture effects from structural changes in the work force that are not related to productivity. If higher nominal wages are paid to employees in the capital than in the provinces then a shift in employment towards relatively

more employees in the capitol will increase the compensation of employees. With the WR method this will be captured in the volume component, but should end up in the price component. To capture this effect in the WR method you have to correct the increase in WR with the increase that originates from movement to the capitol. On the other hand the AW method depends heavily on stratification. If there are structural changes within a stratum then it will end up in the price component. To capture this as a volume component the level of detail has to be increased.

The significant difference between the two methods imply that the choice of method is also a choice of how much productivity change is assumed in the general government. The choice of method should therefore reflect plausible assumptions about the structural element in changes in compensation of employees and the effect on output.



## Examples

### I: Identical increase in wage rate, no structural changes

	t				t+1			
	N	w	W	□ w%	N	w	W	
Judges – capital	10	2000	20000	10%	10	2200	22000	
Judges – provinces	30	1800	54000	10%	30	1980	59400	
<i>Judges total</i>	<i>40</i>	<i>1850</i>	<i>74000</i>	<i>10%</i>	<i>40</i>	<i>2035</i>	<i>81400</i>	
Policemen – capital	15	600	9000	10%	15	660	9900	
Policemen – provinces	70	500	35000	10%	70	550	38500	
<i>Policemen – total</i>	<i>85</i>	<i>518</i>	<i>44000</i>	<i>10%</i>	<i>85</i>	<i>569</i>	<i>48400</i>	
<b>All employees</b>	<b>125</b>	<b>944</b>	<b>118000</b>	<b>10%</b>	<b>125</b>	<b>1038,4</b>	<b>129800</b>	

N = number of employees

w = wage rate (annual pay with constant working hours)

W = total wage costs (N\*w)

□ w% = percentage increase in official wage rate. Totals and subtotals are averages weighed by t wages

	AW method		WR method	
	t	t+1	t	t+1
<b><u>Ex. I.a - no classification</u></b>				
(1) Total wage costs	11800	129800	118000	12980
	0		0	
(2) Number of employees	125	125	125	125
(3) Average wage cost	944	1038,4	944	1038,4
(4) Wage index	100	110,0	100	110,0
(5) Constant price wage cost	11800	118000	118000	11800
	0		0	
(6) Productivity - (5)/(2)	944	944	944	944
<b><u>Ex. I.b - classification by category of employee</u></b>				
(1) Total wage costs	11800	129800	118000	12980
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges	100	110,0	100	110,0
(4) Wage index - Policemen	100	110,0	100	110,0
(5) Constant price wage cost - Judges	74000	74000	74000	74000
(6) Constant price wage cost - Policemen	44000	44000	44000	44000
(7) Constant price wage cost - total	11800	118000	118000	11800
	0		0	
(8) Productivity - (7)/(2)	944	944	944	944
<b><u>Ex. I.c - full classification by all categories</u></b>				
(1) Total wage costs	11800	129800	118000	12980
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges, capital	100	110,0	100	110,0
(4) Wage index - Judges, provinces	100	110,0	100	110,0
(5) Wage index - Policemen, capital	100	110,0	100	110,0
(6) Wage index - Policemen, provinces	100	110,0	100	110,0
(7) Constant price wage cost - Judges, capital	20000	20000	20000	20000
(8) Constant price wage cost - Judges, provinces	54000	54000	54000	54000

(9) Constant price wage cost - Policemen, capital	9000	9000	9000	9000
(10) Constant price wage cost - Policemen, provinces	35000	35000	35000	35000
(11) Constant price wage cost - total	11800	118000	118000	11800
	0			0
(12) Productivity - (11)/(2)	944	944	944	944

## II: Different increases in wage rates, no structural changes

	t				t+1			
	N	w	W	□ w%	N	w	W	
Judges - capital	10	2000	20000	10%	10	2200	22000	
Judges - provinces	30	1800	54000	20%	30	2160	64800	
<i>Judges total</i>	<i>40</i>	<i>1850</i>	<i>74000</i>	<i>17%</i>	<i>40</i>	<i>2170</i>	<i>86800</i>	
Policemen - capital	15	600	9000	30%	15	780	11700	
Policemen - provinces	70	500	35000	40%	70	700	49000	
<i>Policemen - total</i>	<i>85</i>	<i>518</i>	<i>44000</i>	<i>38%</i>	<i>85</i>	<i>714</i>	<i>60700</i>	
<b>All employees</b>	<b>125</b>	<b>944</b>	<b>118000</b>	<b>25%</b>	<b>125</b>	<b>1180</b>	<b>147500</b>	

N = number of employees

w = wage rate (annual pay with constant working hours)

W = total wage costs (N\*w)

□ w% = percentage increase in official wage rate. Totals and subtotals are averages weighed by t wages

	AW method		WR method	
	t	t+1	t	t+1
(1) Total wage costs	11800	147500	118000	14750
	0		0	
(2) Number of employees	125	125	125	125
(3) Average wage cost	944	1180	944	1180
(4) Wage index	100	125,0	100	125,0
(5) Constant price wage cost	11800	118000	118000	11800
	0		0	
(6) Productivity - (5)/(2)	944	944	944	944

### Ex. II.b - classification by category of employee

(1) Total wage costs	11800	147500	118000	14750
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges	100	117,3	100	117,3
(4) Wage index - Policemen	100	138,0	100	138,0
(5) Constant price wage cost - Judges	74000	74000	74000	74000
(6) Constant price wage cost - Policemen	44000	44000	44000	44000
(7) Constant price wage cost - total	11800	118000	118000	11800
	0		0	
(8) Productivity - (7)/(2)	944	944	944	944

### Ex. II.c - full classification by all categories

(1) Total wage costs	11800	147500	118000	14750
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges, capital	100	110,0	100	110,0
(4) Wage index - Judges, provinces	100	120,0	100	120,0
(5) Wage index - Policemen, capital	100	130,0	100	130,0
(6) Wage index - Policemen, provinces	100	140,0	100	140,0
(7) Constant price wage cost - Judges, capital	20000	20000	20000	20000
(8) Constant price wage cost - Judges, provinces	54000	54000	54000	54000
(9) Constant price wage cost - Policemen, capital	9000	9000	9000	9000

(10) Constant price wage cost - Policemen, provinces	35000	35000	35000	35000
(11) Constant price wage cost - total	11800	118000	118000	11800
	0			0
(12) Productivity - (11)/(2)	944	944	944	944

### III: Identical increase in wage rate, structural changes within regions

	t				t+1			
	N	w	W	□ w%	N	w	W	
Judges - capital	10	2000	20000	10%	30	2200	66000	
Judges - provinces	30	1800	54000	10%	10	1980	19800	
<i>Judges total</i>	<i>40</i>	<i>1850</i>	<i>74000</i>	<i>10%</i>	<i>40</i>	<i>2145</i>	<i>85800</i>	
Policemen - capital	15	600	9000	10%	70	660	46200	
Policemen - provinces	70	500	35000	10%	15	550	8250	
<i>Policemen - total</i>	<i>85</i>	<i>518</i>	<i>44000</i>	<i>10%</i>	<i>85</i>	<i>641</i>	<i>54450</i>	
<b>All employees</b>	<b>125</b>	<b>944</b>	<b>118000</b>	<b>10%</b>	<b>125</b>	<b>1122</b>	<b>140250</b>	

N = number of employees

w = wage rate (annual pay with constant working hours)

W = total wage costs (N\*w)

□ w% = percentage increase in official wage rate. Totals and subtotals are averages weighed by t wages

	AW method		WR method	
	t	t+1	t	t+1
<b><u>Ex. III.a - no classification</u></b>				
(1) Total wage costs	11800	140250	118000	14025
	0		0	
(2) Number of employees	125	125	125	125
(3) Average wage cost	944	1122	944	1122
(4) Wage index	100	118,9	100	110,0
(5) Constant price wage cost	11800	118000	118000	12750
	0		0	
(6) Productivity - (5)/(2)	944	944	944	1020

#### **Ex. III.b - classification by category of employee**

(1) Total wage costs	11800	140250	118000	14025
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges	100	115,9	100	110,0
(4) Wage index - Policemen	100	123,8	100	110,0
(5) Constant price wage cost - Judges	74000	74000	74000	78000
(6) Constant price wage cost - Policemen	44000	44000	44000	49500
(7) Constant price wage cost - total	11800	118000	118000	12750
	0		0	
(8) Productivity - (7)/(2)	944	944	944	1020

#### **Ex. III.c - full classification by all categories**

(1) Total wage costs	11800	140250	118000	14025
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges, capital	100	110,0	100	110,0
(4) Wage index - Judges, provinces	100	110,0	100	110,0
(5) Wage index - Policemen, capital	100	110,0	100	110,0
(6) Wage index - Policemen, provinces	100	110,0	100	110,0
(7) Constant price wage cost - Judges, capital	20000	60000	20000	60000
(8) Constant price wage cost - Judges, provinces	54000	18000	54000	18000
(9) Constant price wage cost - Policemen, capital	9000	42000	9000	42000

(10) Constant price wage cost - Policemen, provinces	35000	7500	35000	7500
(11) Constant price wage cost - total	11800	127500	118000	12750
	0		0	
(12) Productivity - (11)/(2)	944	1020	944	1020

#### IV: Identical increase in wage rate, structural changes within regions and type of employment

	t				t+1		
	N	w	W	□ w%	N	w	W
Judges - capital	10	2000	20000	10%	40	2200	88000
Judges - provinces	30	1800	54000	10%	10	1980	19800
<i>Judges total</i>	<i>40</i>	<i>1850</i>	<i>74000</i>	<i>10%</i>	<i>50</i>	<i>2156</i>	<i>10780</i>
							<i>0</i>
Policemen - capital	15	600	9000	10%	60	660	39600
Policemen - provinces	70	500	35000	10%	15	550	8250
<i>Policemen - total</i>	<i>85</i>	<i>518</i>	<i>44000</i>	<i>10%</i>	<i>75</i>	<i>638</i>	<i>47850</i>
<b>All employees</b>	<b>125</b>	<b>944</b>	<b>118000</b>	<b>10%</b>	<b>125</b>	<b>1245,2</b>	<b>155650</b>

N = number of employees

w = wage rate (annual pay with constant working hours)

W = total wage costs (N\*w)

□ w% = percentage increase in official wage rate. Totals and subtotals are averages weighed by t wages

	AW method		WR method	
	t	t+1	t	t+1
<b><u>Ex. IV.a - no classification</u></b>				
(1) Total wage costs	11800	155650	118000	15565
	0		0	
(2) Number of employees	125	125	125	125
(3) Average wage cost	944	1245,2	944	1245,2
(4) Wage index	100	131,9	100	110,0
(5) Constant price wage cost	11800	118000	118000	14150
	0		0	
(6) Productivity - (5)/(2)	944	944	944	1132

#### **Ex. IV.b - classification by category of employee**

(1) Total wage costs	11800	155650	118000	15565
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges	100	116,5	100	110,0
(4) Wage index - Policemen	100	123,3	100	110,0
(5) Constant price wage cost - Judges	74000	92500	74000	98000
(6) Constant price wage cost - Policemen	44000	38823,5	44000	43500
(7) Constant price wage cost - total	11800	131324	118000	14150
	0		0	
(8) Productivity - (7)/(2)	944	1051	944	1132

#### **Ex. IV.c - full classification by all categories**

(1) Total wage costs	11800	155650	118000	15565
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges, capital	100	110,0	100	110,0
(4) Wage index - Judges, provinces	100	110,0	100	110,0
(5) Wage index - Policemen, capital	100	110,0	100	110,0
(6) Wage index - Policemen, provinces	100	110,0	100	110,0
(7) Constant price wage cost - Judges, capital	20000	80000	20000	80000
(8) Constant price wage cost - Judges, provinces	54000	18000	54000	18000
(9) Constant price wage cost - Policemen, capital	9000	36000	9000	36000

(10) Constant price wage cost - Policemen, provinces	35000	7500	35000	7500
(11) Constant price wage cost - total	11800	141500	118000	14150
	0		0	
(12) Productivity - (11)/(2)	944	1132	944	1132



## V: Different increases in wage rates, structural changes within regions and type of employment

	t				t+1		
	N	w	W	□ w%	N	w	W
Judges - capital	10	2000	20000	10%	40	2200	88000
Judges - provinces	30	1800	54000	20%	10	2160	21600
<i>Judges total</i>	<i>40</i>	<i>1850</i>	<i>74000</i>	<i>18%</i>	<i>50</i>	<i>2192</i>	<i>10960</i>
							<i>0</i>
Policemen - capital	15	600	9000	30%	60	780	46800
Policemen - provinces	70	500	35000	40%	15	700	10500
<i>Policemen - total</i>	<i>85</i>	<i>518</i>	<i>44000</i>	<i>38%</i>	<i>75</i>	<i>764</i>	<i>57300</i>
<b>All employees</b>	<b>125</b>	<b>944</b>	<b>118000</b>	<b>25%</b>	<b>125</b>	<b>1335,2</b>	<b>166900</b>

N = number of employees

w = wage rate (annual pay with constant working hours)

W = total wage costs (N\*w)

□ w% = percentage increase in official wage rate. Totals and subtotals are averages weighed by t wages

	AW method		WR method	
	t	t+1	t	t+1
<b><u>Ex. V.a - no classification</u></b>				
(1) Total wage costs	11800	166900	118000	16690
	0		0	
(2) Number of employees	125	125	125	125
(3) Average wage cost	944	1335,2	944	1335,2
(4) Wage index	100	141,4	100	125,0
(5) Constant price wage cost	11800	118000	118000	13352
	0		0	
(6) Productivity - (5)/(2)	944	944	944	1068

### **Ex. V.b - classification by category of employee**

(1) Total wage costs	11800	166900	118000	16690
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges	100	118,5	100	117,3
(4) Wage index - Policemen	100	147,6	100	138,0
(5) Constant price wage cost - Judges	74000	92500	74000	93438
(6) Constant price wage cost - Policemen	44000	38823,5	44000	41435
(7) Constant price wage cost - total	11800	131324	118000	13497
	0		3	
(8) Productivity - (7)/(2)	944	1051	944	1080

### **Ex. V.c - full classification by all categories**

(1) Total wage costs	11800	166900	118000	16690
	0		0	
(2) Number of employees	125	125	125	125
(3) Wage index - Judges, capital	100	110,0	100	110,0
(4) Wage index - Judges, provinces	100	120,0	100	120,0
(5) Wage index - Policemen, capital	100	130,0	100	130,0
(6) Wage index - Policemen, provinces	100	140,0	100	140,0
(7) Constant price wage cost - Judges, capital	20000	80000	20000	80000
(8) Constant price wage cost - Judges, provinces	54000	18000	54000	18000
(9) Constant price wage cost - Policemen, capital	9000	36000	9000	36000

(10) Constant price wage cost - Policemen, provinces	35000	7500	35000	7500
(11) Constant price wage cost - total	11800	141500	118000	14150
	0		0	
(12) Productivity - (11)/(2)	944	1132	944	1132

### **ANNEX 3: OUTPUT INDICATORS: THE CASE OF FINLAND**

In Finland there has been the aim for improving productivity in the government sector for a long time. It is therefore evident that there is a need for measures by which the real productivity for government services can be estimated.

#### *The project on measuring public sector productivity in Finland*

In 1995 the project for measuring public sector productivity was set up at Statistics Finland. The aim of the project is to develop a measurement and monitoring system for government sector production and productivity by using index theory and an output indicator method.

In Finland the definition of final products and output indicators and the output measurement have advanced quite well in recent years. One reason for this is the fact that all agencies in, for example, the central government had to adopt a result-oriented administration in 1995 at the latest. The measurement of performance including the measurement of final products provided is an essential part of the result-oriented administration.

The first productivity measurement was carried out in 1996 and at that time only the productivity for central government was measured. Productivity measurement in local government was started in 1997.

#### *The central and local government*

The central and local government services are treated separately in the project. The central government services, most of which are collective services, are very diversified. That is why the final output and output indicators have to be specified by each agency itself. Statistics Finland collects input and output data from agencies annually. In local government, municipalities for example, the same kind of services for their inhabitants is produced. These services are mostly individual services in their nature (such as health and social care, education). The output indicators used are applicable more widely than the indicators applied to the output measurement in the central government. Statistics Finland annually collects the data on municipalities' activities.

#### *Measurement issues and the methods applied in the measurement*

The productivity measurement for central government services proceeds from the micro level to the macro level. At first, input, output and productivity indices are compiled for each unit. A unit is an agency, bureau or other organisation. The units covered by the measurement represent various kinds of economic activities in central government, such as public administration, education, defence and research and development.

Although the output indicators vary from one unit to another the method used in the output aggregation at the micro level is the same for each unit: the Tornqvist index is used in the aggregation of the growth rates of final products in each unit. The Tornqvist index is also used in the aggregation of the measurement results of all units when the aggregate output, input and productivity changes are measured. So far, quality changes have not been explicitly taken into account in the measurement of changes in output. However, to a certain extent, output and productivity measures compiled for government units may include changes in the quality of output due to the application of an index formula with a flexible weighting structure.

In addition to the facts mentioned above there are many other questions relating to the measurement of output and to the applicability of output indicators. One question is how to allocate output into separate periods in the situations where the production process of output is very prolonged. Questions arise about which would be the best suited indicators to describe the output.

Although the final products and output indicators have already been defined for a large part of government in some cases the lack of information on the weights of the final products is a problem when measuring output and productivity. The cost or work time shares for the final products cannot at the moment be compiled in adequate accuracy by all units (in the absence of market prices the cost and work time shares can be used as weights for proportional changes of the quantities of final products).

*The first results in the central government*

The productivity measurement for central government has been carried out three times so far. The two measurements carried out in 1996 and in 1997 were made for about 50 units and in 1998 for 56 units. These units covered about 40 per cent of the compensation of employees in central government. The coverage of the measurement has been improved slightly every year, as an increasing amount of adequate data has become available.

The measures compiled so far cover individual services as well as collective services provided. The units included in the reportings cover a variety of activities and the characteristics of the units vary quite a lot. Some units are small having only one type of final product and some units are large providing many types of different products. See the end of this annex for a list of examples of units and indicators.

The results indicate that the growth rates of output and productivity vary extensively between different government agencies. The quantities of final products may also fluctuate quite widely from year to year.

On the one hand this may be caused by the general factors affecting the variation in production of services, such as changes in the demand. On the other hand, in some cases, some changes observed in the quantity of output may be caused by the application of inaccurately and deficiently defined output indicators.

Some aggregate output and productivity growth rates in central government according to the first measurements are presented in the table below.

<b>Some aggregate output and productivity growth rates in central government in 1995 and 1996</b>		
	1995 (1994=100)	1996 (1995=100)
The weighted average of the growth in output		+ 4,9 %
labour input		- 1,7 %
total input		+ 4,1 %
labour productivity	+ 3,0 %	+ 6,8 %
total productivity*	- 0,5 %	+ 0,8 %

\* total productivity is approximated by cost efficiency in real terms and it is measured as the ratio of output changes to the changes in cost in real prices.

**Some aggregate output and productivity growth rates  
in central government in 1995, 1996 and 1997**

	1995 (1994=100)	1996 (1995=100)	1997 (1996=100)
The weighted average of the growth in			
Output		+4,9%	+4,2%

Labour input		-1,7%	+4,2%
Total input		+4,1%	+2,3%
Labour productivity	+3,0%	+6,8%	0,0%
Total productivity*	-0,5%	+0,8%	+1,9%

\*total productivity is approximated by cost efficiency in real terms and it is measured as the ratio of output changes to the changes in costs in real prices

Labour productivity is in this case measured as output divided by labour input in full-time equivalents. "Total productivity" is the difference between the growth in the volume of output and the growth in the volume of costs. The latter measure includes in costs also labour costs (but excludes investments so is more or less in line with national accounts definitions). Hence, it depends on how labour costs are deflated. The deflated total costs give an indication of what the result of the input method would have been for these units. If labour is deflated such that structural changes in the work force are in the volume component, then the figure for total productivity gives the "extra" productivity included in the volume of output figures but not yet included in the deflated input data.

It is very important to specify further the definitions of final products and output indicators so that they would be more and more applicable to output and productivity measurement. To improve the usefulness of the results it is essential to widen the coverage of the measurement.

The data collections and the measurements carried out so far indicate that final products and output indicators are definable for the most part of government sector. Also the output and productivity of collective services seem to be measurable.

## SOME EXAMPLES OF OUTPUT INDICATORS IN CENTRAL GOVERNMENT

### *Consumer Ombudsman's Office*

- number of petitions for market court
- marketing instructions
- (number of) contractual terms negotiated
- statements on legislative initiatives
- cases solved individually
- replying to written inquiries

### *Courts such as*

*The Supreme Court, Courts of Appeal, District Courts, Provincial Courts and Supreme Administration Court*

- number of cases settled

### *Helsinki City Police Department*

- the output indicators of public order and security, such as tasks directed toward the protection of property and the individual
- crime prevention measured by the number of crimes solved
- the final products of traffic safety
- number of permissions  
(the number of passports, identity cards, driving licenses and firearms licenses)

### *Housing Fund of Finland*

- decisions for loan and interest subsidies measured as the weighted number of decisions

### *National Board of Patents and Registration*

- number of patents
- number of utility models
- number of trademarks
- number of pattern rights
- company register cases
- association register cases
- enterprise mortgage cases

### *National Food Administration*

- number of letters guiding supervision
- number of administrative decisions and memos
- number of publications
- number of statements
- number of training events
- materials for instructions, new ones

*Prison system*

- prisoner-days

*Prosecutors' Offices and*

*Distrain Offices*

- number of cases dealt with

*State Audit Office*

- number of annual audits
- supplementary audits
- international audits
- expertise activities, statements

*Tax Administration*

- numbers of private persons, agricultural entrepreneurs and entrepreneurs and corporations subject to income and property tax
- number of supervised registered employers
- numbers of primary producers and entrepreneurs subject to value added tax
- the output indicator of real estate tax

*Universities*

- number of degrees completed (generally separated into graduate and post graduate degrees)
- adult education and continuing education measured for example in days or number of courses (depending on the university)
- number of publications (research)

**FINAL REPORT OF THE TASK FORCE  
“PRICES AND VOLUMES FOR EDUCATION”**

September 1998



## CONTENTS

1	INTRODUCTION
2	BACKGROUND TO THE RECOMMENDATIONS
2.1	Classifications
2.2	Definition of output
2.3	The criteria behind the recommendations
3	THE ALTERNATIVES
3.1	Output indicator methods
3.2	Input methods and comparisons with output indicator methods
3.3	Price methods
3.4	Quality adjustments
4	IMPLEMENTING OUTPUT INDICATOR METHODS
5	RECOMMENDATIONS
5.1	Output indicator methods
5.2	Input methods
5.3	Price methods
5.4	Quality adjustments
6	FURTHER RESEARCH
6.1	Quality adjustments
6.2	Other research and reviewing the recommendations
Annex A	List of task force documents
Annex B	Market education: prices and volume indicators
Annex C	Trial calculations for the task force countries
Annex D	Example on input methods and computerisation
Annex E	Letter from Statistisches Bundesamt and response by the task force

## 1 INTRODUCTION

### *Mandate*

1. The mandate for the task force was given by the National Accounts Working Party at its meeting of December 1997 (document Eurostat B1/CN 349e). In summary, the purpose of the task force was to investigate the existing and potential methods for estimating the output of education at constant prices, both market and non-market, paying particular attention to methods that can be applied from 1999, with the aim of improving the comparability of GDP in constant prices.

### *The task force*

2. The task force comprised Keith Hayes (Eurostat), Frits Bos (Statistics Netherlands), Holger Breiholz (Statistisches Bundesamt), Christina Liwendahl (Statistics Sweden), David Caplan (ONS UK), Deborah Guerrucci/ Federico Nusperli (ISTAT), Anne Gouezel-Jobert (INSEE), Maire O'Mahony (CSO Ireland) and Paul McCarthy (OECD). It met three times: in Luxembourg (19-20 February 1998), The Hague (23-24 April 1998) and London (15-16 June 1998). The list of task force documents is at Annex A.

### *Co-ordination*

3. The task force worked in parallel with two others, on prices and volumes for health and for other non-market services (NACE L). The work of the three task forces was co-ordinated, the aim being to come to consistent conclusions where this was possible.

4. In the course of their work members of the task force consulted specialists on harmonised consumer price indices, purchasing power parities, functional statistics for education and structural business statistics, with the aim of taking into account developments in their fields.

### *Focus of the report*

5. The task force showed a clear preference for output indicator methods over input methods. This was despite the fact that most Member States currently use input methods. However, as explained in this report, in particular in Annex E, the representative of Statistisches Bundesamt did not share this preference.

6. The preference for output indicator methods over input methods is in the end a judgement. For the reasons explained in this report, the task force considered that despite the difficulties, output indicator methods are better from a conceptual viewpoint, feasible from a practical viewpoint and offer the chance to make progress.

## 2 BACKGROUND TO THE RECOMMENDATIONS

### 2.1 Classifications

7. Non-market education services are identified on the expenditure side of the national accounts through the expenditures incurred by government in providing such services. However, non-market education services are classified on the output side of the accounts on the basis of the

major activities carried out by producing units. These units are classified into one of two branches: "education" (which also includes units providing marketed education services) and "public administration" (which includes units formulating policy on education, enforcing standards and regulating schools, etc).

8. A distinction needs to be drawn between activities, expenditures incurred, and how units are classified to different industries on the basis of the major activities in which they are engaged. The UN publication "International Standard Industrial Classification of All Economic Activities" (Series M, no. 4, Rev 3) provides a detailed presentation of the issues involved in allocating units to industries on the basis of their major activities, so no further details will be provided here.

9. The major point of relevance to this Task Force report is to note that different classifications are used in different circumstances. As a result, special steps have to be taken to ensure that all relevant education activities are included consistently between the expenditure and output sides of the accounts.

10. Broadly, government outlays on education are the major element of government final consumption expenditure on education. They include expenditures on services provided to individual students as well as those on services provided on a collective basis. The classification involved is the "Classification of Functions of Government" (COFOG). It is important to note that this classification is an "end-use" or "functional" classification rather than one based on the types of activities being undertaken within particular units. The main components of COFOG Division 04 (Education)<sup>1</sup> are:

- 04.1 Pre-primary and primary education
- 04.2 Secondary education
- 04.3 Tertiary education
- 04.4 Education not definable by level
- 04.5 Subsidiary services to education
- 04.6 Research and development
- 04.7 Education affairs and services n.e.c.

11. The education branch (or industry) on the output side of the accounts is more narrowly defined in terms of the range of education services included but it is broader in the sense that, unlike government final consumption expenditure, the education branch includes market as well as non-market education services. The appropriate classification is the "International Standard Industrial Classification of All Economic Activities" (ISIC) or the NACE Rev. 1. The main components of NACE Division 80 (Education) are:

- 80.1 Primary education
- 80.2 Secondary education
- 80.3 Tertiary education
- 80.4 Adult and other education.

Activities associated with the administration of education are included in ISIC and NACE Division 75: "Public Administration and Defence; Compulsory Social Security".

## **2.2 Definition of output**

12. There are two distinct issues relating to the measurement of non-market output in the national accounts: valuing the output in current prices and measuring the output at constant prices.

---

<sup>1</sup> It should be noted that COFOG is currently in the course of being revised.

13. For the current price estimates, education is no different from other non-market services, with the estimates being obtained simply as the sum of the costs incurred (i.e., wages and salaries + intermediate inputs + consumption of fixed capital + other taxes less subsidies on production).

14. While the method of obtaining current price values is straightforward, obtaining constant price estimates of output is more difficult, both conceptually and in practice. The SNA93 identifies a number of issues relating to the measurement of non-market output (see par. 16.133 to 16.141). Based on the concepts contained in those paragraphs and in ESA95 (par. 10.41 and 10.42), the task force settled on the following working definition of the output of education:

The educational output is the quantity of teaching received by the students, adjusted to allow for the qualities of the service provided, for each type of education. The quantities should be weighted together using data on the costs of (or prices for) the education provided.

15. The task force concluded that the preferred measure of educational output, for both market and non-market education services, is the number of hours spent by students in being taught, which is consistent with paragraph 10.42 of ESA95. A simple example was produced (see Annex B) to show that student-hours will provide the correct measure of output in both the market and non-market sectors, provided they are adjusted for quality changes.

16. The task force recognised that data on student-hours may be difficult to obtain, since the data generally available relates to numbers of students doing particular courses. The example at Annex B also shows that different results will be obtained if the current price values are deflated on the basis of information on the numbers of students rather than on student hours. The implication is that the data on student numbers should be adjusted to take account of any changes in educational arrangements which result in a significant change to the number of student-hours involved in a course.

### **2.3 The criteria behind the recommendations**

17. In order to assess the alternative methods available, the task force identified several criteria which should be satisfied. A two-stage process was involved. The first stage involved establishing the ideal method from a conceptual viewpoint. In the second stage, the task force examined the data needed to implement them and assessed their practicality.

18. The conceptual ideal would be information on the numbers of hours for which students are taught, classified by all the different types of courses and adjusted for changes in the quality of teaching. Ideally, details on the type of courses would be available separately for every course which has a different cost structure from the others. It was emphasised that estimates based on the numbers of teachers are input rather than output measures.

19. The major criteria underlying the task force's recommendations were based on how well any proposed methods matched this theoretical ideal, taking into account the view that output can be defined, at least conceptually, for all such services. The criteria against which the alternatives were judged were:

- the data should provide complete coverage of all courses of study;
- changes in the quality of teaching should be taken into account;
- changes in productivity should be taken into account;
- changes in the structure of education over time should be taken into account; and
- the estimates for non-market services should be consistent with those for market education services.

### **3 THE ALTERNATIVES**

#### **3.1 Output indicator methods**

20. Output indicator methods involve measuring changes in the volume of production on the basis of the quantities produced, adjusted for any changes in the quality of the goods and services concerned.

21. Output indicator methods are already used or being developed for non-market education in a few task force countries (France, Italy and the United Kingdom). Both the ESA95 (par. 10.41-10.46) and the SNA93 (par. 16.133-16.141) «encourage the development of output volume indicators for non-market output».

#### **THEORETICAL CONSIDERATIONS**

22. The task force on education agrees that output volume indicators should be developed. Output indicator methods are best from a theoretical point of view for describing education services in the integrated set of price and volume measures of a system of National Accounts: the definition of the output of educational services leads to a preference for this method, as explained in Section 2.2.

23. Output indicator methods enable a value to be put on the volume of the output of education services on a similar conceptual basis to the volumes of other national accounts aggregates. In particular, they are conceptually consistent with the volume of other marketed goods and services, including the volume of household expenditure on education services. The same principles apply to the measurement of consumption and production. The use of output indicator methods provides a clear picture of the pattern of productivity change and facilitates comparisons of productivity. Finally, measuring output volumes directly is fully consistent with valuing the output at current prices as the sum of costs.

24. The considerations of the task force were based on the definition in Section 2.2. This contains a description of the single homogeneous product to be considered for education services (both production and consumption) and the principle of weighting together the individual outputs: that is the number of hours spent by students in being taught, by type of education, weighted together using unit costs or prices. These are the essential elements involved in performing the calculations.

#### **APPLYING OUTPUT INDICATOR METHODS**

25. There are three stages in applying output indicator methods to non-market education services:

- choosing the output volume or quantity indicators
- taking account of changes in quality
- calculating the weighting pattern to combine the detailed indicators

The calculations rest on the choice of the appropriate indicators and the stratification of these basic data. The issues involved in taking explicit account of quality change are discussed in Section 3.4.

## STRATIFICATION

26. Section 2.2 states that the ideal indicator of the quantity of teaching services is the number of hours of teaching received per student valued at the base year prices, that is, the costs of output for the non-marketed service in the base year.

27. ESA95, par. 10.25 says: “In the absence of a unit market price, the change in the ‘unit cost’ of a non-market service can be considered as an approximation of the change in price”.

If sufficiently detailed data are used, this indicator will capture changes in the mix of education services. In effect, in obtaining the volume it is necessary to distinguish the volume of output of the various levels of education which are of different quality and which have different costs. If the elementary quantities are homogeneous then one part of the quality effect is taken into account. This is a weighty point in favour of output indicator methods.

28. ESA95 (par. 10.17-10.18) says the following about stratification:

“For the purposes of calculating price and volume measures, it is necessary to use as a detailed product classification as possible so that each product identified has maximum homogeneity, regardless of the level of detail used in the presentation of results ... Also the effects of aggregation have to be considered. Variations in the composition of a flow which imply, for example, a shift in favour of higher average quality have to be recorded as a volume increase and not as a price increase. It follows that for outputs, the effect of shifts between markets with differing prices ... will be treated as changes in volume and not as changes in price. It also follows that a price change for a given flow can occur only as result of changes in prices at the level of individual transactions.”

29. So for the purpose of deflating the output of the industry Education, different relatively homogeneous products should be distinguished. This implies that:

- educational outputs should be distinguished from any non-educational (secondary) outputs where these are significant;
- various types of educational outputs should be distinguished.

30. Examples of secondary non-educational outputs which might be produced by the industry are: the provision of meals; the provision of lodging; transport services to and from school; day nursery and child care; health care; the provision of sports and recreational facilities (outside official school hours); the provision of educational and vocational information; research (for universities). Concerning the latter, official R&D statistics covering research by universities exist in all Member States. However, this task force has not investigated methods for deflating research. It makes no recommendations on this. This will form part of the remit for the task force on prices and volumes for business services.

31. The simultaneous analysis by function, industry and product allows the achievement of the homogeneous product ‘education’. Different products should be distinguished if they are quantitatively important and if sufficient data are available. For instance, research is an important secondary output of universities and so should be estimated separately.

32. The proper stratification of educational output should take account of the different types of education and their differences in quality. This implies that in the stratification:

- different types of output should be distinguished when the (unit-)costs are substantially different, e.g. the study of medicine is in general much more expensive than that of economics. This type of stratification is particularly important when substantial shifts occur.

- different types of output should be also distinguished if different quality indicators are to be used.

33. The precise level of stratification to be used depends to a great extent on the Member State's specific circumstances, such as the way the education system is organized, the quantitative importance of the various types of education, regional differences and the specifications of the data sources. Stratification could take account of e.g.:

- substantial regional differences in unit costs;
- various types of secondary education, e.g. vocational, non-vocational, special schools;
- extra-curriculum lessons.

34. The weights to be used to combine the indicators would be the average cost per student hour in the base year in each category. However, even if data are available for Member States on the amount of teaching provided, obtaining them at a reasonable cost may be a difficult task. The task force noted that the availability of unit costs data would in practice be crucial for what stratification could be used.

### **3.2 Input methods and comparisons with output indicator methods**

35. Input methods use data on the inputs to the production process to estimate the volume of the output of the production process. These can be methods that only use data on employment numbers or the compensation of employees, but they can also take account of the other inputs (intermediate consumption, consumption of fixed capital and other taxes less subsidies on production). Input methods can also contain adjustments for quality changes in the inputs. For non-market education, input methods are currently used in most Member States. When compared with output indicator methods, input methods have attractions and disadvantages, as discussed below.

Data availability

36. Input methods use data that are usually readily available in NSIs, and national accountants usually, but not always, consider the data to be relatively reliable. Moreover the data may be available to the NSIs earlier than output indicator data.

Controversial or uncontroversial?

37. Input methods are fairly easy to understand and to explain to users, though, perhaps, output indicator methods are not difficult either. The long history and wide-spread use of input methods make them uncontroversial to some users, though of course they do attract strong criticism from some (more expert) users of the national accounts.

The same method for all non-market outputs?

38. Using input methods for education services would mean that the same method would be used as for other non-market services, such as for example public administration and defence. On the other hand it does lead to a discontinuity with the methods used for market services, and in particular for market education services.

Quality of inputs

39. Some types of input methods make allowances for changes in the quality of the inputs. For example the increasing experience of a teacher might be reflected in his or her higher salary, which in some input methods translates into a higher measured volume of output. This subject is discussed more fully in the report of the task force on non-market services (NACE L). Agreeing with that task force, the task force on education notes that these allowances for presumed changes in the quality of the inputs are essentially arbitrary and subjective, because the impact of changes on the quality of input on the quantity and quality of output is not measured but assumed.

#### Productivity

40. Input methods have one fundamental disadvantage: they measure inputs rather than outputs. This means that only if no productivity change occurs will they give a proper estimate of the volume of output. No productivity change occurs when the volume of inputs and the volume of output move in the same direction and by the same magnitude. In all other situations, the input method will not give a proper estimate of the volume of output. This applies for example to the two following cases:

- The number of pupils by type of school changes substantially over a period of ten years for demographic reasons (e.g. less young people) and because of changes in the subjects/ type of schools preferred by pupils. However, at the same time, the inputs, e.g. the number of teachers, remains rather constant, perhaps for employment reasons, for reasons of regional policy (small pupil/teacher ratios are acceptable in order to ensure that pupils should not travel too far to their schools) and/ or because it is not efficient to hire and fire teachers immediately when the demand for teaching decreases. The input method does not show any of these productivity changes. The numerical example at Annex D shows how input methods fail to reflect productivity changes in the case where labour inputs become more productive.
- Innovations in methods of education, improvements in the curricula taught and better management by school staff can reduce the costs of education substantially while improving the results in terms of the numbers of pupils taught and their achievements. The input method rules out any such change in productivity being captured in the national accounts.

#### Comparability

41. It might also be argued that, even if input methods are not perfect, because they do not take account of productivity changes, they could at least be the basis of comparable estimates for the Member States: if the same assumption about productivity is made in all countries. For example if all countries assume no change in productivity or all assume +0.5%.

42. The task force did not consider this argument to be valid. A harmonised assumption about productivity does nothing to make the resulting estimates of output more comparable. Its view was that in fact this missing element for productivity change is not the same amount in all countries.

43. In some countries the missing element will be positive and in others it will be negative. In case of a zero productivity change assumption, the missing element is equal to the productivity decrease or increase. The more different the developments in productivity among Member States, the less comparable are the results from using a zero productivity change assumption.

44. A +0.5% productivity change assumption is no better. For a country where productivity is increasing, this assumption yields more accurate results than the zero productivity change



assumption. However, for Member States where productivity is decreasing, this assumption only makes the results worse and less comparable.

#### Practical considerations

45. The task force was conscious that, for assessing the relative merits of input methods and output methods, the theoretical considerations are insufficient in themselves. Equally important are the practical considerations: Is productivity change important in education in practice, and can it be measured?

*Is productivity change important? What is the impact on GDP?*

46. The task force looked at some trial calculations for its member countries. These calculations should be regarded as no more than indicative: they have no official status. They were carried out quickly, and mostly use simple and approximate methods. In most cases the calculations compare output quantities (without quality adjustments) with input calculations.

47. Nevertheless the results are interesting and seem conclusive, and they are summarised at Annex C. They show that the productivity change seems to be important. Measured productivity change is positive in some countries, negative in some others and near to zero in some other countries. The impact of changing from input methods to output methods for education alone could easily change the annual GDP growth estimate in a Member State by 0.1% in either direction. The task force considered this to be a significant amount.

#### Arguments of Statistisches Bundesamt

48. Some further arguments were put forward by Statistisches Bundesamt, as detailed in the letter at Annex E. Those arguments did not gain the support of the rest of the task force for the reasons noted in that annex.

### **3.3 Price methods**

*Where can they be used?*

49. The task force focused on measuring the core part of educational output, that is the output of the school and higher education systems. These are mostly provided on a non-market basis in all of the countries represented in the task force. The consumers are not charged at all or are only charged nominal fees.

50. Nominal fees are charged when these fees are relatively small compared to the cost involved in providing the services. These costs are mainly covered by government transfers, property income or voluntary contributions of households in their capacity as consumers. The fees charged may change or differ substantially purely for political reasons. When this happens, changes in these fees can not be used to measure changes in the volume or quality of the education services provided for the national accounts.

51. However in all countries there exist at least some private schools for children and there are also some other educational institutions providing educational services to adults and children on a market basis. In addition in many countries there are moves towards the charging of fees to students to cover a sizeable part of the costs of providing higher education. Thus although price

methods are not applicable to the main part of education output they are applicable for a minor part.

*How do price methods compare with output volume indicator and input methods?*

52. Section 2.2 showed that using price methods, when the customer pays the full fee, and using volume of output methods should give fully equivalent results. Input methods will give different results because of the productivity issue.

#### Quality adjustment of prices data

53. When using price methods, three types of quality adjustment can be distinguished (see Eurostat, “Report to the Council on Harmonisation of Consumer Price Indices in the European Union”, October 1997, p. 46.):

1. linking, which is equivalent to the assumption that the difference in price between the old item and its replacement is wholly attributable to a difference in quality.
2. same quality adjustment, which treats the replacement as being of same quality as the old item which it replaces. Therefore, the whole of the price difference between the old item and its replacement is recorded in the price index.
3. other quality adjustment, where the value of the quality difference between the old item and its replacement is estimated as somewhere between zero and the whole price difference, or occasionally outside that range. Those estimates could be made by experts, or by using methods such as hedonic regression, overlap pricing, option pricing, the production cost approach or the imputation approach.

54. For output volume indicator and input methods, only quality adjustment by linking is not possible, as this is fully based on treating a price difference as a quality difference. However, several of the other types of quality adjustment, like hedonic regression and the production cost approach are possible.

*Developments in price statistics for education*

55. Work is in progress to develop price statistics for those education services usually paid for by the consumer in the context of the harmonised CPI. This work will set some minimum standards and require the collection of price data (for this small part of education output) in all Member States. However its usefulness for the national accounts will be limited if, as seems likely, the data collected concerns only the parts of the total fees paid by households, and other important sources of funds are not taken into account.

### 3.4 Quality adjustments

*Measuring quality*

56. A number of approaches might be used to estimate the quality of the educational services received by pupils: direct measurements of quality, measuring inputs and measuring outcomes. Each approach has its strengths and weaknesses. Quality adjustment is a task where statisticians must use judgement. The brief discussion of quality adjustment within price statistics in Section 3.3 above indicates this. The SNA93 (par. 16.105 to 16.129) gives a general overview of the possibilities and the choices.

57. Particularly for services, quality is a somewhat nebulous concept. There is no simple definition to be applied. Quality is what the customers perceives it to be, based on what he expects to achieve from receiving the service.

#### *Direct measurements of quality*

58. Can the quality of education be directly measured? In the UK, for example, the Office of Standards of Education is responsible for inspecting schools. It estimates the proportion of lessons which are “very good”, “satisfactory” and “unsatisfactory”. The advantage of this approach is that it is a direct measurement of quality. However, there are considerable difficulties. The information which is obtained is subjective and may not be consistent over time or between schools. It is available only sporadically - inspection is not an annual process. Further, it is difficult to turn the information into a quality function which can be used for adjusting data on the quantity of education delivered.

59. It should also be noted that educational quality must be the favourite subject for (academic) researchers. In some countries it may therefore be possible to find academic researches which give comparisons across time and which could be used as the basis for quality adjustments.

#### *Measuring inputs*

60. The amount of inputs delivered can be used to estimate the likely quality of the output. For example, it can be argued that lower class sizes lead to higher quality education. More computers in the classroom may also improve education. It is therefore possible to look at quality change by looking at changes in inputs. This approach has the advantage that some data is likely to be readily available. Work in Italy has suggested how it can be used to produce a quality function based on data on congestion (students per class) and the availability of teaching and support material.

61. However, there is far from total agreement on how changes in inputs lead to changes in quality. For example, in one country, class sizes may have changed with no discernible impact on outcomes, while in another country there may have been an impact. Further, other changes in the quantity and quality of inputs, such as using better teaching methods, will be ignored because they are not measured.

#### *Measuring outcomes*

62. The point of education is to generate particular outcomes - educated children ready to take their place in the adult world. If the outcomes improve then this may reflect improvement in the quality of educational services. A possible approach is to look at changes in outcomes and use those to come up with an estimate of quality change, which then must be attached to the output quantity indicator.

63. For example, changes in attainment in school leaving exams could be used if the attainment levels are constant over time. This approach has the advantage of measuring something that should directly reflect changes in quality. However, the impact of changing quality may be in part confounded by changes in other factors which affect educational outcomes such as general living standards. Further, attainment measures are unlikely to measure all of the desirable outcomes. For example, increased academic success may be associated with lower social skills of pupils. The level of education skills and knowledge acquired depends also on factors such as time spent studying at home and the students' commitment and motivation: these factors are (in large part) independent of the quality of the teaching services received.

64. As ESA95 (par. 10.26) notes, it is not possible to measure the volume of teaching services received by the rise in the level of education. Nevertheless, outcome data can in certain circumstances tell you something about the quality of the service received.

## CONCLUSIONS

65. The discussion above and in Section 3.3 shows that there is no simple method for making quality adjustments, whether for output volume indicators or for price indices. The difficulties faced for non-market education are also faced for many other types of market goods and services, that is, for all but the simplest products.

66. The choice of the method to use for quality adjustment involves difficult statistical judgements. These are best made against the background of data availability and the specific circumstances of the (suspected) quality change.

67. The task force agreed therefore that there is no “best method” that should be applied in all countries in all circumstances. It did consider though that despite the difficulties, the Member States should be encouraged to make quality adjustments where they believe them to be necessary, using the method that they consider on balance to be best in their circumstances. Member States should not simply assume that quality change is zero without further reflection.

## 4 IMPLEMENTING OUTPUT INDICATOR METHODS

68. It would not be necessary for all of the Member States to apply output indicator methods in exactly the same manner to achieve comparable results. As explained in Section 3.1, it is important when choosing output volume indicators to distinguish between outputs which are of different qualities and which have different unit costs. The more detailed the level of stratification used for the calculations, the better.

69. Nevertheless, as explained below, a standardised set of statistics on education is collected by Eurostat, the OECD and UNESCO. This provides a minimum data set for all Member States.

### *Pupil-hours or pupil numbers?*

70. The output of education services, explained in Section 2.2, is the number of pupil-hours taught, classified by type of course and adjusted for changes in the quality of teaching received. Data on pupil-hours is not easily available in each country, by subject or by level of education, but it is probable that movements in numbers of pupils will approximate movements in pupils-hours very closely. Therefore numbers of pupils themselves should be good indicators in practice in most circumstances.

### *Data on output volume indicators*

71. There are many different types of education. Firstly, there are differences in the level of education illustrated by the different grades within a school, secondly there are different subject matters. In higher or further education the subject matter taught is quite important.

72. Data on pupil numbers by age and by level of education are available in all Member States. The joint UNESCO, OECD, Eurostat questionnaire on educational statistics known as ENRL1, which is used by all of the 15 Member States, is used to bring together this data centrally. The levels of education are standardised using the International Standard Classification of Education (ISCED, UNESCO, 1976)<sup>2</sup>.

73. The data sent to Eurostat by Member States and published by Eurostat (Education across the European Union - Statistics and Indicators), distinguishes the following levels of education (some categories may not exist in some countries):

- pre-school education
- primary education
- lower secondary education
- upper secondary education: general
- upper secondary education: vocational
- higher education (non-university)
- higher education (university)
- other education

74. Ideally a more detailed split would be used for the calculations, both within the primary and secondary education (for instance special schools for disabled) and within higher education, because different relative costs would be incurred in educating different types of students. For example costs per student-hour are likely to be higher per science student than for literature students. Where such additional data is available, its use would of course improve the estimates.

---

<sup>2</sup> ISCED is currently in the course of being revised. The new COFOG (see Section 2.1) will use a breakdown of education based on the level categories of the new ISCED.

#### *Unit costs linked to output indicators*

75. The weights to be used to combine the output indicators would be the average cost per student-hour in the base year in each category or, the average cost per pupil for each level. This implies that the series relating to the total number of pupils within a particular category of education should be weighted by the total costs of providing their education in the base period.

76. However, even if complete data are available for Member States on the amount of teaching provided, obtaining them at a reasonable cost may be a difficult task. In addition, the task force acknowledges that the availability of unit costs data would in practice determine what stratification could be used.

77. The Eurostat questionnaire on the financing of education (the FINANCE1 questionnaire) should lead to better consistency between the quantities of educational services and the costs statistics needed to weight them together, since it uses the same classifications of levels of education. Therefore it should be the case that a minimum detailed stratification could be achieved for all Member States. Where there is a major shift in unit costs, for example as a result of a policy decision, then the weights will need to be reassessed to reflect the changes in quality.

#### *Index formula and base year*

78. The task force also considered the preferred index formula to be used. In the light of the issues described above, it was decided that a Laspeyres volume formula would be the most appropriate.

79. Ideally, costs data should be collected every year, and the base year for the calculations would be the previous year. The same approach should be used for education as for other parts of the accounts. However, obtaining detailed information on the costs incurred by level of education may be costly. If so, the data should be collected at least every fifth year.

## 5 RECOMMENDATIONS

### 5.1 Output indicator methods

80. The Task Force's recommendations are:

An output indicator method is an "A" method if all four of the following conditions are satisfied:

- All output of education is covered. In practice it may not be important if there are some small omissions.
- The stratification distinguishes the main categories of education services having different unit costs, that is at least the following categories where they exist and are significant in a Member State:
  - pre-school education
  - primary education
  - lower secondary education
  - upper secondary education : general
  - upper secondary education : vocational
  - higher education (university)
  - higher education (other).
  - other education (e.g. driving lessons)
- Ideally the base year is the previous year; however if annual data are not available then the base year should be changed at least every five years; the reference year will be the same as used in other national accounts aggregates and should be changed every five years.
- A justified allowance for quality change in teaching is made. It can either be based on a direct measurement of quality, involve the use of data on inputs, or be grounded on outcome data.

81. An output indicator method is a "B" method if all of the conditions for an A method are met except the allowance for quality change is not made.

82. An output method will be a "C" method if the coverage criteria or the stratification criteria are not achieved.

### 5.2 Input methods

83. The task force came to the following conclusions about input methods:

- The report of the task force on non-market services (NACE L) discusses the ways in which input methods can be improved.
- Any input method which does not involve the direct measurement of productivity change should be considered as a C method.
- Input methods can only be considered to be A or B methods if they use the approaches set out in the NACE L report and if they directly estimate and take into account productivity change - using output volume indicators to make this direct estimate of productivity change. i.e. they are not simply based on an arbitrary assumption about the size of the productivity change.

### 5.3 Price methods

84. The task force conclusions about price methods are the following:

- Price methods have a fairly minor role to play in the measurement of educational output for the national accounts, as most of educational output is non-market output.
- Price methods, when the total fees are paid by households, will give equivalent results to using output volume indicators for educational market output. They should be considered as A methods when the requirements concerning coverage and quality adjustments for the harmonisation of the consumer price indices on education have been met (e.g. no automatic linking and a comparison of the quality adjustment method with the standard reference practice). Otherwise they should be considered as B methods.
- Price methods, when the prices collected concern only the part of the total fees paid by households, and this share varies over time, should be considered as C methods for the national accounts (even though such prices may be appropriate for the harmonised CPI).

#### **5.4 Quality adjustments**

85. The task force's recommendations are as follows:

- There are several possibilities for quality adjustments. The choice of the method to use for a quality adjustment involves difficult statistical judgements. These are best made against the background of data availability and the specific circumstances of the (suspected) quality change.
- The task force did not consider that there is a “best method” that should be applied in all countries in all circumstances. It did consider though that despite the difficulties, the Member States should be encouraged to make quality adjustments where they believe them to be necessary, using the method that they consider on balance to be best in their circumstances. Member States should not simply assume that quality change is zero without further reflection.



## 6 FURTHER RESEARCH

### 6.1 Quality adjustment

86. Further research could improve the range of tools available for making quality adjustments. This could focus on:

- definitions of quality
- availability of data
- possible functions linking quality with outcomes and inputs
- better use of the work of price statisticians to inform national accountants

### 6.2 Other research and reviewing the recommendations

#### *Quarterly Accounts*

87. This task force concentrated on measuring the output of education in accordance with the requirements of the Stability and Growth Pact. It has therefore not dealt explicitly with the measurement of the output of education in quarterly national accounts.

#### *Research*

88. An important output of higher education institutions is research, which may be carried out by students or teaching staff. Methods for estimating output volumes for research should be assessed as part of the work of forthcoming task force on business services.

#### *Child care and pre-school education*

89. In order to increase the comparability between countries, the distinction between child care and pre-school education is currently being studied by education statisticians. Output volumes for pre-school education and for child care should probably be measured in a comparable way. The forthcoming task force on personal and social services should consider this issue.

#### *Harmonised Consumer Price Index*

90. Work is currently progressing in the task force on the HICP to incorporate a harmonised approach to including the price of education in the overall index. It is difficult to see how useful the resulting price indices will be for the national accounts in view of the fact that much is supplied at heavily subsidised rates to the consumers. Once they are available, it would be worthwhile reviewing the recommendations on the HICP of education to check their suitability for the national accounts.

#### *Structural business statistics*

91. Research is currently underway in various task forces dealing with the implementation of the Council Regulation No 58/97 concerning structural business statistics. The scope of these surveys may include market and non-market sectors, including education. To date agreement has not been reached on what types of information can be collected. For the purposes of providing useful data for education at constant prices, consideration might be given to getting details on the qualitative aspects of the service and also on sources and amounts of funding.

**LIST OF DOCUMENTS CONSIDERED BY THE TASK FORCE**

Documents for the first meeting:

Fax from Keith Hayes of 20/12/97, enclosing mandate for task force	TFEDUC/0
Agenda and dates, etc. for 1st meeting	TFEDUC/1
Note for the first meeting of the Task Force on prices and volumes for education, Keith Hayes	TFEDUC/2
Draft Commission Decision ... version of 19/12/97	TFEDUC/3
Harmonisation of constant price data (document for the National Accounts Working Party meeting of 29-30 October 1997)	TFEDUC/4
Eurostat statistics on education, Keith Hayes	TFEDUC/5
Report of the first meeting of the task force “Volume measures for non-market services (NACE L)”	TFEDUC/6
Project proposal: Productivity in public education by Ken Tallis, ABS	TFEDUC/7
Compensation of employees in volume for NACE L and M in Belgium.	TFEDUC/8
Non-market education in Sweden, by Christina Liwendahl	TFEDUC/9
Draft Commission Decision ... (Version of 16/2/98 for discussion at the National Accounts Working Party meeting of 26-27 February 1998)	TFEDUC/10
Volume indices of non-market education services supplied by General government sector, by Deborah Guerrucci	TFEDUC/11
Productivity Trends in the Public Sector in Sweden (extracts from a report by the Swedish Ministry of Finance supplied by Christina Liwendahl).	TFEDUC/12
Task force “Prices and Volumes: Education” - note by Holger Breiholz, Statistisches Bundesamt	TFEDUC/13
Volume measurement of Government output; the Dutch Practice since Revision 1987 by Brugt Kazemier, CBS	TFEDUC/14
Briefing on Ireland’s methodology in respect of the output of education, by Maire O’Mahony	TFEDUC/15
Measuring output for education services, by Paul McCarthy	TFEDUC/16
Measuring education output in the UK national accounts, by David Caplan	TFEDUC/17
Council Regulation No. 58/97 of 20 December 1996 concerning structural business statistics	TFEDUC/18

Documents for the second meeting:

Draft note of the first meeting of the task force	TFEDUC/19
Agenda for the second meeting of the task force	TFEDUC/20
Preparatory note for the second meeting of the task force on prices and volumes for education, Keith Hayes	TFEDUC/21
The Council Regulation on structural business statistics and non-market services, Keith Hayes	TFEDUC/22
Report of the first meeting of the task force volume measures for health and social work (NACE M).	TFEDUC/23
Education output, Paul McCarthy	TFEDUC/24
Comments on the note of the first task force meeting from Bettina Knauth, Eurostat E3 (education and training statistics)	TFEDUC/25
Draft outline of the final report, Keith Hayes	TFEDUC/26
The output of non-market services, Keith Hayes	TFEDUC/27
Purchasing power parities for education, Keith Hayes	TFEDUC/28
Draft report of the second meeting of the task force on NACE L	TFEDUC/29
Draft report of the second meeting of the task force on health and social work	TFEDUC/30
Harmonised CPU's for education and quality adjustments in the CPI for education, Keith Hayes	TFEDUC/31
Data available in Sweden and trial calculations, Christina Liwendahl	TFEDUC/32
Trial calculations for Ireland, Maire O'Mahony	TFEDUC/33
Data availability in Germany and trial calculations, Holger Breiholz	TFEDUC/34
The national accounts and economic growth: the case of Dutch education, Frits Bos	TFEDUC/35
Education statistics in the UK, David Caplan	TFEDUC/36
Current prices in the national accounts in Italy, Deborah Guerrucci/ Data available in Italy concerning the education system, Deborah Guerrucci and Federico Nusperli	TFEDUC/37
Price and volume in education services: the French experience, Anne Gouezel-Jobert	TFEDUC/38
Some remarks and data availability in France, Anne Gouezel-Jobert	TFEDUC/39
Education statistics in Ireland (data availability), Maire O'Mahony	TFEDUC/40

Documents for the third meeting and final documents:

Draft note of the second meeting of the task force	TFEDUC/41
Agenda for the third meeting of the task force	TFEDUC/42
Input methods and changes in productivity, Paul McCarthy	TFEDUC/43
Draft guidelines for the inclusion of education in the harmonised CPUs, and background papers	TFEDUC/44
Preparatory note for the third meeting of the task force, Keith Hayes	TFEDUC/45
Draft final report of the task force (version of 5/6/98)	TFEDUC/46
Report to the Council of Ministers on the Harmonisation of Consumer Price Indices (HICPs) in the European Union	TFEDUC/47
Commission Regulation No 1749/96 of 9 September 1996 on the initial implementing measures for Council Regulation No 2494/95 concerning harmonised indices of consumer prices	TFEDUC/48
Advantages of input methods: a response to the German arguments in Section 3.1 of the draft final report, Frits Bos	TFEDUC/49
Some comments on output indicator methods, Christina Liwendahl	TFEDUC/50
Draft note of the third meeting of the task force on non-market services, Paul Konijn	TFEDUC/51
Trial calculations for Italy, Deborah Guerrucci	TFEDUC/52
The national accounts and economic growth: the case of Dutch education, third draft, Frits Bos	TFEDUC/35 (Rev. 1)
Data available in Sweden concerning the school system and trial calculations using output and input approaches, Christina Liwendahl	TFEDUC/32 (Rev. 1)
Draft note of the third meeting of the task force, Keith Hayes	TFEDUC/53
Second draft of the final report of the task force (version of 18/8/98).	TFEDUC/54
Final draft of the report of the task force (28/9/98)	

## MARKET EDUCATION: PRICES AND VOLUME INDICATORS

1. A key issue in determining the best method of calculating constant price output for the non-market component of education services is to identify what actually constitutes the output of the education industry. It is clear that the current price value of non-market output of the education industry consists of the amounts paid out as wages and salaries to all teaching and ancillary staff and the value of intermediate consumption plus consumption of fixed capital. It is also clear that the ultimate aim of the educational process is to improve the knowledge and skills of a country's citizens. However, it is not possible to equate the output of the education industry with the total knowledge gained by students. Additional important factors in acquiring knowledge include the additional study done by pupils outside school; the extent of informal ("on the job") training for people such as apprentices; and the experience accumulated over time by identifying, and thereby avoiding, situations which have caused problems in the past. Thus, the output of the education industry has to be defined more narrowly in terms of the services provided in practice by educational institutions.
2. It is clear that an indicator based on the numbers of teachers is an input rather than an output measure. However, it is less clear what actually constitutes an appropriate measure of output. One way to determine the best indicator of non-market output is to identify the corresponding indicator of market output for the same industry. The following example shows that the market sector output is actually the total number of "student-hours", but there is a need to adjust this measure for changes in the quality of teaching.
3. For the purposes of simplifying the example, let's assume that the output of the market sector in the education industry consists of nothing but running courses and that there are only three separate courses available. The data available relate to three separate years and show the number of students attending each course, the number of hours for which each course runs and the price charged per hour for each course. From this data, it is possible to calculate the amounts received by the industry for running courses, the numbers of student hours by course and a price index relating to the hourly rate charged for each course. The common method of obtaining constant price estimates for market sector activities is by deflation using price indexes and so constant price estimates can be obtained by deflating the business receipts from running the courses by the price index relevant to each course. They can also be calculated by valuing at base year prices the student hours spent attending each course. The table below shows that identical constant price estimates are obtained by price deflation and by directly valuing student hours at base year prices. The conclusion to be drawn is that the preferred output measure for market sector education is in fact student hours. The implication is that student hours is also the output which we should be aiming to measure for the non-market sector.
4. A critical element not taken into account in the example is the extent to which quality is changing from year to year. The price indexes used to deflate the revenues from each course are simple hourly rates, with no attempt being made to adjust for quality differences. In a similar way, the numbers of student hours on which the revaluation is based have not been adjusted for quality change.
5. The implication of the simple example is that the critical issue involved in measuring the output of non-market sector education services is not so much identifying the output to be measured but in determining a means of assessing changes in the quality of the services delivered by teachers. An index of quality change could then be used to adjust the constant price estimates obtaining by revaluing the number of student hours at base year prices.

## EDUCATION - MARKET SECTOR OUTPUT

Calculations based on student-hours

	Year 1	Year 2	Year 3
No. of students			
Course A	250	280	240
Course B	410	480	450
Course C	620	500	530
Total	1280	1260	1220
No. of hours per course per year			
Course A	100	104	103
Course B	80	78	79
Course C	70	65	62
Price (\$/hour)			
Course A	25	27	28
Course B	29	28	26
Course C	30	40	44
Total revenue (\$,000)			
Course A	625.0	786.2	692.2
Course B	951.2	1048.3	924.3
Course C	1302.0	1300.0	1445.8
Total	2878.2	3134.6	3062.3
Price indices (Year 1 = 100.0)			
Course A	100.0	108.0	112.0
Course B	100.0	96.6	89.7
Course C	100.0	133.3	146.7
Total (IPD)	100.0	112.4	116.2
Constant price estimates (\$,000) (using price deflation - Paasche)			
Course A	625.0	728.0	618.0
Course B	951.2	1085.8	1031.0
Course C	1302.0	975.0	985.8
Total	2878.2	2788.8	2634.8
Constant price estimates (\$,000) (quantity revaluing student hours)			
Course A	625.0	728.0	618.0
Course B	951.2	1085.8	1031.0
Course C	1302.0	975.0	985.8
Total	2878.2	2788.8	2634.8

6. However, data on student-hours may be difficult to obtain, since the data generally available relates to numbers of students doing particular courses. The example below illustrates that different results will be obtained if the current price values are deflated on the basis of information on the numbers of students rather than on student hours. The implication is that the data on student numbers should be adjusted to take account of any changes in educational arrangements which result in a significant change in the number of student-hours involved in courses during a year.

## EDUCATION - MARKET SECTOR OUTPUT

Calculations based on annual fees

	Year 1	Year 2	Year 3
Annual fee per student (\$)			
Course A	2500	2808	2884
Course B	2320	2184	2054
Course C	2100	2600	2728
Price indexes (Year 1 = 100.0) - based on annual fees			
Course A	100.0	112.3	115.4
Course B	100.0	94.1	88.5
Course C	100.0	123.8	129.9
Constant price estimates (\$,000) (using price deflation - Paasche)			
Course A	625.0	700.0	600.0
Course B	951.2	1113.6	1044.0
Course C	1302.0	1050.0	1113.0
Total	2878.2	2863.6	2757.0

## TRIAL CALCULATIONS FOR THE TASK FORCE COUNTRIES

The trial calculations shown in this annex should be considered to be no more than indicative. They have no official status.

### Netherlands

Two output indicator based series are calculated, the first stratified by type of education, the second not stratified.

The input method series is taken from the existing Dutch national accounts.

#### Subsidised education

	GDP factor costs, current prices	GDP factor costs, t-1 prices	Value added, education, current prices	Value added, education, t-1 prices, input method (1)	Value added, education, t-1 prices, output method unstrat- ified (2)	Value added, education, t-1 prices, output method strat- ified (3)	Differ -ence (1)– (2)	Differ -ence as % of GDP	Differ -ence (1)–(3)	Differ -ence as % of GDP
1985	38911 0		18482							
1986	39835 0	39769 7	18821	18966	18264	18349	702	0.18	617	0.16
1987	40141 0	40649 8	19177	19071	18391	18516	680	0.17	555	0.14
1988	41597 0	41124 4	18931	19158	19182	19223	-24	-0.01	-65	-0.02
1989	44132 0	43450 8	18862	18912	18928	18963	-16	-0.00	-51	-0.01
1990	46856 0	45923 0	19656	19053	19261	19553	-208	-0.05	-500	-0.11
1991	49291 0	48027 4	20330	19735	19594	19757	141	0.03	-22	-0.00
1992	51211 0	50156 6	21579	20634	20420	20507	214	0.04	127	0.03
1993	52390 0	51456 5	22321	21812	22130	22210	-318	-0.06	-398	-0.08
1994	55149 0	54089 7	23087	22646	22858	22790	-212	-0.04	-144	-0.03
1995	56878 0	56423 2	24601	23109	23477	23346	-368	-0.07	-237	-0.04

*Millions Guilders*



## Italy

Two output indicator approaches are calculated for the general government sector. Both use pupil numbers by level of education. The first uses in addition a quality adjustment based on data on congestion and teaching and support material available, while the second does not. The input method uses teacher numbers by type of teacher.

	Value added, output indicator method, no quality adjustment (2)	Value added, output indicator method, quality adjusted (1)	Value added, input method (3)	Difference (3)-(1)	Difference as a % of GDP	Difference (3)-(2)	Difference as a % of GDP
1988	59024	58477	58183	-294	-0.02	-841	-0.07
1989	58160	58465	58102	-363	-0.03	-58	-0.00
1990	57641	57641	57641	0	0	0	0
1991	56719	57163	57791	628	0.05	1072	0.08
1992	56090	56154	58114	1960	0.15	2024	0.15
1993	55479	55716	57318	1602	0.12	1839	0.14

*Billions of Lira, constant prices*

## Sweden

For primary schools two output indicator approaches are calculated. Firstly using seven series of pupil numbers having different unit costs in the base year (one of the strata is student hours in native language courses – which have a high unit cost). The second calculation uses no subdivision of pupil numbers.

The input method uses hours worked for employees, deflated intermediate consumption and capital consumption at constant prices.

### Primary schools

	GDP at constant prices	Value added, output indicator method, stratified (1)	Value added, output indicator method, not stratified (2)	Value added, input method (3)	Difference (3)-(1)	Difference as % of GDP	Difference (3)-(2)	Difference as % of GDP
1987	1281602	32853	33466	30794	-2058	-0.16	-2671	-0.21
1988	1310454	32893	33309	31177	-1716	-0.13	-2132	-0.16
1989	1341596	32741	32974	31500	-1241	-0.09	-1474	-0.11
1990	1359879	31986	31986	31986	0	0	0	0
1991	1344697	31359	31523	32090	730	0.05	567	0.04
1992	1325579	30832	31054	30875	43	0.00	-179	-0.01

*Millions of Krona, 1990 constant prices*

For upper secondary schools again two output indicator approaches are calculated. Firstly using a stratified approach to take account of program changes. Secondly using no subdivision of pupil numbers. The input method again uses hours worked for employees, deflated intermediate consumption and capital consumption at constant prices.

#### Upper secondary schools

	GDP at constant prices	Value added, output indicator method, stratified (1)	Value added, output indicator method, not stratified (2)	Value added, input method (3)	Differ- ence (3)-(1)	Differ- ence as % of GDP	Differ- ence (3)-(2)	Differ- ence as % of GDP
1987	1281602	8479	8334	8418	-61	-0.00	84	0.01
1988	1310454	9096	8989	8676	-419	-0.03	-313	-0.02
1989	1341596	9445	9228	9075	-370	-0.03	-153	-0.01
1990	1359879	9447	9447	9447	0	0	0	0
1991	1344697	9048	9013	9527	480	0.04	515	0.04
1992	1325579	9369	9708	9594	225	0.02	-114	-0.01

*Millions of Krona, 1990 constant prices*

#### Ireland

The output indicator approach uses pupil numbers by level of education (4 strata). The input method uses teacher numbers by type of teacher (3 strata).

	Value added, output indicator approach (1)	Value added, input method (2)	GDP	Difference (2) – (1)	Difference as a % of GDP
1988	1063.2	1098.4	24039.5	35.3	0.15
1989	1068.4	1074.1	25565.5	5.7	0.02
1990	1076.3	1076.3	27190.0	0	0
1991	1093.1	1089.1	27481.3	-4.0	-0.01
1992	1117.4	1121.0	28425.8	3.6	0.01
1993	1143.8	1154.4	29108.0	10.6	0.04
1994	1160.4	1181.3	30882.1	20.9	0.07
1995	1167.1	1181.7	33876.1	14.5	0.04
1996	1170.6	1204.9	36161.0	34.3	0.09

*£, millions (IRE), constant prices*

#### Germany

The output indicator approach uses pupil and student numbers by type of educational institution.

The input approach is the one used in the national accounts at present (sum of deflated compensation of employees, intermediate consumption and capital consumption).

	Value added, output indicator	Value added, input method (2)	Difference (2) – (1)	Difference as a % of GDP
--	-------------------------------------	-------------------------------------	-------------------------	-----------------------------

	approach (1)			
1990	75650	76664	1014	0.04
1991	89200	89200	0	0
1992	94617	94205	-412	-0.01
1993	97264	97805	540	0.02
1994	99612	99294	-318	-0.01
1995	100794	100066	-728	-0.02

*Millions of Marks, constant prices*

## France

The output method is the one used in the French national accounts at present. The output indicator uses pupil and student numbers by level of education (13 strata), without any quality adjustment.

For the trial calculations the input method uses a price index of wages based on the actual value of the “grading unit” for civil servants and capital consumption at constant prices.

	Value added, current prices	Value added, output indicator method, n-1 prices (1)	Value added, output indicator method, n-1 volume index (2)	Value added, input method, n-1 prices (2)	Value added, input method, n-1 volume index	Difference (2) – (1)	GDP n-1 prices	Difference as % of GDP
1985	209472	201771						
1986	219158	210946	100.70	214516	102.41	3570	4813350	0.07
1987	230345	220096	100.43	226965	103.56	6869	5182772	0.13
1988	236396	230519	100.08	233281	101.27	2762	5567375	0.05
1989	250322	242152	102.43	245421	103.82	3269	5961591	0.05
1990	270895	254165	101.54	263577	105.30	9412	6310206	0.15
1991	287796	273251	100.87	282801	104.32	9350	6560175	0.14
1992	305372	288489	100.24	296933	103.17	8444	6845753	0.12
1993	322874	307866	100.82	313556	102.68	5690	6909316	0.08
1994	335376	326725	101.19	330450	102.35	3725	7262237	0.05

## MILLIONS OF FRANCS

## United Kingdom

The input method calculations show the figures used in the UK national accounts up to summer 1998.

The output method calculations use pupil and student numbers by level of education (x strata). The quality adjustment is based on educational attainment data.

	Value added index, input method, local authorities and universities only. For	Value added index, output indicator method, no quality adjustment	Value added index, output indicator method, with quality adjustment
--	---	---	---

the latter, data only up to 1992 *			
1986	112.2	87.6	86.9
1987	114.9	86.8	85.6
1988	116.1	86.1	85.0
1989	115.3	86.4	85.5
1990	114.9	87.3	86.5
1991	116.6	90.2	89.6
1992	115.3	93.3	92.8
1993	* 102.0	95.8	95.5
1994	100.4	98.6	98.4
1995	100.0	100.0	100.0
1996	99.8	100.8	100.9
1997		101.8	102.1

*1995=100, constant prices*

*\* Universities were re-classified, out of the public sector, from 1993*

## EXAMPLE ON INPUT METHODS AND COMPUTERISATION

The following example is designed to show the problems involved in obtaining constant price output by deflating inputs. In the example, an artificial drop in constant price GDP occurs as a result of deflating inputs rather than directly measuring the (unchanged) constant price output of government.

The following assumptions are made:

- the whole of government final consumption expenditure is on a single process (e.g., processing social security payments) and neither the numbers of payments processed nor the quality of the output change from one year to the next;
- the productivity improvement achieved by government from introducing computers is responsible for a reduction in employment which cuts the government real wages bill from 90 in year 1 to 70 in years 2, 3 and 4;
- the only imports are computers (year 2) and they all go into government gross fixed capital investment;
- the government depreciates computing equipment over three years.

The example shows how deflating the various inputs into government final consumption expenditure results in a decline in constant price government final consumption expenditure (and thus constant price GDP) because the productivity improvements associated with computerisation are not taken into account in this process. If output indicators, based on the numbers of forms processed, were used to derive constant price output then there would be no change recorded in constant price government final consumption expenditure and constant price GDP would also remain the same.

Of course there is no requirement to use capital expenditure to make labour more productive. Other improvements in working practices or methods can be used to improve labour productivity (such as better focusing on the aims of the production process, better motivation of staff, better organisation within the production unit, simplifying the production process, more focused training). In such cases the cost of labour inputs can be reduced and output remain unchanged, as in the example, while this time there is no change in capital expenditure or in capital consumption.

	<u>Year 1</u>	<u>Year 2</u>	<u>Year 3</u>	<u>Year 4</u>
<b><u>Current price values</u></b>				
Private FCE	500.0	510.0	520.0	530.0
Government FCE (of which)	120.0	106.9	109.1	111.3
- <i>wages</i>	90.0	72.1	74.2	76.3
- <i>intermediate inputs</i>	30.0	30.8	31.7	32.4
- <i>cons of fixed capital</i>	0.0	4.0	3.2	2.6
Government GFCF	0.0	12.0	0.0	0.0
Exports	30.0	30.8	31.5	32.3
Imports	0.0	12.0	0.0	0.0
<b><i>GDP</i></b>	<b><i>650.0</i></b>	<b><i>647.7</i></b>	<b><i>660.6</i></b>	<b><i>673.6</i></b>
<b><u>Deflators</u></b>				
Private FCE	100.0	102.0	104.0	106.0
Government FCE				
- <i>wages</i>	100.0	103.0	106.0	109.0
- <i>intermediate inputs</i>	100.0	102.8	105.5	108.0
- <i>cons of fixed capital</i>	100.0	80.0	64.0	51.2
Government GFCF	100.0	80.0	64.0	51.2
Exports	100.0	102.5	105.0	107.5
Imports	100.0	80.0	64.0	51.2
<b><u>Values in year 1 prices</u></b>				
Private FCE	500.0	500.0	500.0	500.0
Government FCE (of which)	120.0	105.0	105.0	105.0
- <i>wages</i>	90.0	70.0	70.0	70.0
- <i>intermediate inputs</i>	30.0	30.0	30.0	30.0
- <i>cons of fixed capital</i>	0.0	5.0	5.0	5.0
Government GFCF	0.0	15.0	0.0	0.0
Exports	30.0	30.0	30.0	30.0
Imports	0.0	15.0	0.0	0.0
<b><i>GDP</i></b>	<b><i>650.0</i></b>	<b><i>635.0</i></b>	<b><i>635.0</i></b>	<b><i>635.0</i></b>

## TEXT OF A LETTER FROM STATISTICHES BUNDESAMT

Wiesbaden, 7 August 1998

## III B 4 - Br

**Stance of the German Federal Statistical Office in the Final Report of the Task Force on Prices and Volumes for Education**

In the general government sector, the input method produces the fewest errors. Although it does not provide full international comparability, it does allow better comparability than output approaches, which are still provisional.

Output indicator methods suffer from the difficulty of measuring volume changes in state education accurately. This is because:

- \* the definition of output (pupil hours) can only be arrived at by agreement;
- \* not all the necessary data (pupil hours) are available for calculating a given output, which means that indicators (pupil numbers) have to be used;
- \* statistical data on education do not provide a suitable quality indicator.

The use of output indicators limits the scope for comparison between EU Member States, because:

- \* different data (pupil and student numbers, pupil hours, pupil days) are used to calculate output volumes;
- \* the absence of uniform financial data means that outputs are subject to different levels of stratification;
- \* the quality indicators used are not uniform and are highly subjective and vague.

Output indicators only cover the core of the education sector (schools, universities, adult education). The lack of volume data means that the input method would still have to be used in the other areas, so as to avoid some fairly large inaccuracies. This would involve considerable additional cost and effort.

The shortcomings of the output method need to be compared with those of the input method. Where input approaches are adopted, productivity changes are either assumed or not taken into account. Genuine productivity changes, however, are unlikely to be as significant in state education and other public-sector branches as in the private sector, because:

- \* in the public sector, labour is the most important factor of production and cannot be replaced by capital (or not to any significant extent);
- \* institutional barriers limit innovation in the state sector more than in the private sector (lack of competitive pressure, absence of profit-and-loss accounts, less supervision of public employees, other methods used in decision-making procedures in the state sector);
- \* levels of employment in the public sector do not always reflect actual labour requirements, but can also serve other purposes, such as labour market policies.

The extent to which comparability suffers from unequal changes in productivity at international level is therefore limited. As productivity changes are not a major factor, precise, harmonised input methods are an appropriate instrument for measuring the output value of education at constant prices, in accordance with the requirements of the Stability and Growth Pact.

The use of input methods for non-market producers and output methods for market producers is methodologically consistent insofar as the output of both types of producer can, in principle, be determined by adding together intermediate consumption, the compensation of employees, consumption of fixed capital, taxes on production (less subsidies) and operating surplus. In the case of both market and non-market producers, changes in input prices feed through to changes in output prices. The only difference between market and non-market producers is that the latter are assumed not to make any profits (operating surplus = 0). The deflation of the output price corresponds with the deflation of its cost elements.

The requirements in terms of accuracy placed on the Harmonised Consumer Price Index (HICP) should also be applied to GDP at constant prices. By way of analogy to Article 4 of Council Regulation (EC) No 2494/95, in conjunction with Article 7 of Commission Regulation (EC) No 1749/96 and Article 3(3) of Commission Regulation (EC) No 2454/97, input methods can be used for deflation purposes if, compared with using output indicators (which are, in theory, preferable), their effect on the calculated rate of growth of GDP at constant prices is consistently below 0.1 percentage points. If this is so, input methods should at least be classed as B methods, without the need for reliable data on changes in labour productivity.

Even if the requirements of comparability were not adopted as for the HICP, input methods should be classed as at least B methods, given the provisional nature of output methods. Research into output indicators over the last 20 to 30 years has been inconclusive, and the empirical results of trial calculations have proved unreliable and too unstable. Given the political importance attaching to calculations of changes in productivity, continued use of input methods should be allowed.

The introduction of output indicators would entail considerable effort, and therefore also substantial additional funds. The cost/benefit ratio is unlikely to be acceptable, unless the Commission (Eurostat) bears some of the additional cost.



**ADVANTAGES AND DISADVANTAGES OF INPUT METHODS:  
A RESPONSE TO THE LETTER FROM  
STATISTISCHES BUNDESAMT**

For deflating non-market education, the German Statistical Office prefers input methods to output methods, for the reasons outlined in their letter included in this annex. Statistisches Bundesamt put forward a number of arguments, and for the most part responses can be found in the main body of this report.

The task force does not consider this argumentation convincing. For non-market education, output indicator methods should be preferred to input methods, because:

1. In contrast to output indicator methods, input methods can not properly take into account changes in the composition of outputs. For the proper measurement of the volume of output, the various types of output (like basic education, university education and research at the universities) should be distinguished and each deflated. Input methods can not cope with this, as they can only account for changes in the composition of inputs.
2. In contrast to output indicator methods, input methods can not properly take into account productivity changes.
3. Input methods can not properly take into account changes in the quality of the education services when this is achieved by the same total of inputs. Input methods ignore a-priori any such quality change for all types of education. In order to account for quality change, output indicator methods could be combined with input information, but also with other methods. The exact combination can differ for different types of education or even over time.
4. Input price methods do not imitate the price methods used for market output. The input price method refers to the prices of inputs while the price methods used with market output generally refer to the prices of outputs. Market output could be deflated by using output prices and by using input prices. The deflation of market output by output prices is generally preferred (also by the Statistisches Bundesamt).
5. The volume of market output could be measured by using data on volumes of output or by using data on volumes of input. However, measuring the volume of market output by using volumes of output is generally preferred (also by the Statistisches Bundesamt). So, the treatment of non-market output similar to market output implies that the volumes of output should be measured and not the volumes of input.

*report of the taskforce*

**PRICES AND VOLUMES FOR HEALTH**

***NACE N***

**September 1998**

## EXECUTIVE SUMMARY

The aim of the Taskforce has been to investigate existing and embryonic output methods that make consistent constant prices for both market and non-market health services and account for productivity and quality changes.

In Member States the current way of measuring output of non-market health services at constant prices has been to use, as a proxy, input indicators. A first methodology, the input price approach, derives constant prices by deflating current price inputs, which are then summed and the total used as a proxy for the total volume of output. The second method, the input indicator approach, extrapolates inputs in the base year using quantity information. Most often only the change in the general rate of salaries is included in the input price approach and only employment in the health service is included in the input indicator approach. Besides, output methods are used for market health services.

The corner stone of output methods is the definition of the output of health services, which is the same for both market, and non-market activities.

ESA95 is clear that volume changes in the output of individual health services should be measured in terms of the use made of the services and not in terms of the outcome obtained from their use. The objective is to measure the volume of services actually produced according to the following definition of health output: the quantity of health care received by patients, adjusted to allow for the qualities of service provided, for each type of health care.

The quantity of health care received by patients should be measured in terms of treatments. In theory, a treatment should be measured in its entirety. The whole bundle of complementary services constituting a treatment should be taken into account: the medical services, the paramedical services, laboratory and radiological services and, in the case of hospitalisation, the non-medical services such as the provision of food and accommodation.

In practice, the feasibility of measuring complete treatments depends on the degree of fragmentation of the services making up a treatment. The complete treatment applies mainly for inpatient treatments in hospitals or specific treatments provided by specialists, dentists or paramedics. Even treatments received as an inpatient may be fragmented if, for example, the patient is transferred from a hospital to a nursing home.

The Taskforce stated that a wide set of health statistics is already available in Member States so that simple output indicator methods could be implemented soon, for example concerning treatments supplied by general practitioners, dentists, etc. for which data are already available and easy to use.

For more complex treatments such as inpatient treatments a best output method would be to use a classification that assigns inpatients to one of a set of homogeneous groups according to their diagnosis, treatment and cost of treatment. Such diagnosis related group (DRG) type classifications are in the process of being introduced in several Member States. However, more investigation will be necessary on the characteristics of the classifications itself, the handling of changes in the classification and the construction of the volume index with the data provided by the classification and the calculation of cost weights.

The TaskForce also proposed intermediate solutions using statistical sources already available. Of course, data have to meet certain requirements of detail and homogeneity.

An example of an output indicator method is presented if data are available for the number of complete treatments in hospitals but their price or cost, necessary for weights, are unknown. Costs of nursing days, surgical operations etc. are allocated (by cross-tabulation) to complete treatments. The resulting total costs per treatment are used as approximate weights for the number of patients classified by disease and treatment.

An example of a price method is presented using both basic Consumer Price Index (CPI) information and additional information on the average number of inpatient days.

Comparisons with trial calculations between input and output methods show that differences in the growth measure are significant.

Monitoring changes in the quality of individual health services will be difficult, because of the complexity and speed of the changes. In theory, quality should cover both changes in the physical characteristics of treatments and changes in treatment mix. In practice, output methods will, as a first step, only account for quality by taking changes in the product mix into account, while measuring the quality changes in products needs further investigation.

The Taskforce recommends that output methods should be used to estimate the output of human health activities at constant prices. Not only do they have the advantage of providing volume measures of output that are consistent between market and non-market sectors, they also provide a more appropriate decomposition between price and volume than do input methods. This is an important consideration for the interpretation of health figures in the national accounts.

However, it is unlikely that input methods can be abandoned in the short term because statistical resources required for output methods are still being developed. Consequently, the Task Force recognised the inevitability of continuing with input methods for the immediate future, but recommends that there should be greater liaison and co-operation between national accountants and health statisticians with a view to developing appropriate output methods.

## CONTENTS

1- Introduction	p. 5
2- Background to the recommendations	p. 6
2.1- Classifications of output and expenditure	p. 6
2.2- Market and non-market output	p. 6
2.3- Estimating output at constant prices	p. 7
2.4- Definition of output	p. 8
2.5- Criteria underlying the recommendations:	p. 10
3- Recommendations of the Taskforce	p. 13
3.1- Output methods	p. 13
3.1.1- Output indicator methods	p. 13
3.1.2- Price methods	p. 17
3.2- Input methods	p. 21
3.3- Quality changes	p. 23
4- Implementation of the recommendations	p. 24
5- Recommendations for further research	p. 25

### Annexes:

I-	Price index numbers of complementary goods by Henk van Tuinen, Bram de Boo and Jaco van Rijn
II-	Presentation on DRGs classification and its use in Member states
III-	Presentation of the Dutch work on activities crossed with treatments: methodology and statistical sources required
IV-	DRGs used in Norway- An example
V-	List of documents and Taskforce participants

## 1. INTRODUCTION

1. The Taskforce has been mandated by the National Accounts Working Party to investigate existing and embryonic methods for estimating output at constant prices for Section N of NACE and to recommend methods that could be applied by Member States from 1999.(document CN/349). The aim is to improve the comparability of GDP at constant prices across Member States by improving the volume measure of both market and non-market output.

2. Even though the mandate concerns the output of the whole of NACE N, the Taskforce focused on human health activities, where the main difficulties occur, and excluded veterinary activities and social work activities. The Taskforce recommends carrying out further research to ascertain whether its conclusions are also applicable to these activities.

3. The Taskforce recommends using output methods as these give a more realistic measure of the volume of human health activities produced than do input methods. The differences in volume measures between these two methods are significant. This has been demonstrated by many studies and by the trial calculations of the Taskforce. For this reason, input methods are classified as C methods.

4. It is unlikely, however, that input methods can be abandoned in the short term because the statistical sources required for output methods are still being developed. Consequently, the Taskforce recognises the inevitability of continuing with input methods for the immediate future, but recommends that there should be greater liaison and cooperation between national accountants and health statisticians with a view to developing appropriate output methods. The Taskforce also recognised that the development of output methods could not be achieved without an initial commitment of additional resources.

5. Within human health activities, the Taskforce focused on inpatient health care. The output of hospital activities was defined in terms of complete treatments, that is the whole package of complementary medical, paramedical, diagnostic, analytic and non-medical services comprising a treatment. A best output method would use a classification that assigns inpatients to one of a set of homogeneous groups according to their diagnosis, treatment and cost. Such diagnosis related group (DRG) type classifications are in the process of being introduced in several Member States. The Taskforce proposes intermediate solutions using statistical sources already available.

## 2. BACKGROUND TO THE RECOMMENDATIONS

### 2.1 Classifications of output and expenditure

6. NACE N Group 851- *Human health activities* - covers individual health services. Collective health services are covered under NACE L Group 751 - *Administration of the state and the economic and social policy of the community* - and are not in the remit of this Taskforce.

7. Producers of the individual health services identified in NACE N Group 851 are broken down into four classes:

- 8511 - *Hospital activities*
- 8512 - *Medical practice activities*
- 8513 - *Dental practice activities*
- 8514 - *Other human health activities*

8. Final consumption expenditures on individual health services by households, NPISHs and general government are classified respectively in Divisions 06, 13 and 14 of COICOP as follows:

- 06.3.0, 13.2.7, 14.2.7 - *Hospital services*
- 06.2.1, 13.2.4, 14.2.4 - *Outpatient medical services*
- 06.2.2, 13.2.5, 14.2.5 - *Outpatient dental services*
- 06.2.3, 13.2.6, 14.2.6 - *Outpatient paramedical services*
- 14.2.8 - *Public health services*
- 13.2.8 - *Other health services*

9. There is not a one-to-one correspondence between these classes and those of NACE N. Expenditures on the output of producing units assigned to NACE 8514 - *Other human health activities* - are spread over the COICOP classes. In addition, because of the convention that all consumption expenditure of NPISHs benefits households individually, COICOP 13.2.8 - *Other health services* - includes expenditures by NPISHs on collective health services.

10. The two classifications are designed to serve different objectives. NACE classifies producing units by activity, COICOP classifies individual consumption expenditure by purpose. It is necessary to ensure that individual health services are consistently covered and treated on both the production and the expenditure sides of the accounts.

### 2.2 Market and non-market output

11. The production of individual health services covers both market output and non-market output.

- Market output is sold at market prices. Hence, it is valued at current market prices.
- Non-market output is not sold at a market price. By convention, its value at current prices is the sum of its production costs.

12. ESA 95 states that the 50 per cent criterion should be applied to determine whether a producer is a market producer or a non-market producer. If more than 50 per cent of the production costs are covered by sales, the producer is a market producer; if less, the producer is a non-market producer. (ESA 95 para. 3.32)

13. The guidelines vis-à-vis the application of the criterion are ambiguous. They allow producers of individual health services, that receive payments from general government having an explicit link with output and covering more than 50 per cent of production costs, to be classified as market producers when perhaps it would be more appropriate to classify them as non-market producers as there are no identifiable prices linked to the sales. (ESA 95 para. 3.33, 3.34 and 3.36). These points are under discussion within an Eurostat Taskforce (B1-B4 joined) on the delimitation of the General Government sector.

14. In the absence of market prices, output at constant prices of these “market” producers will have to be obtained in the same way as output at constant prices is estimated for non-market producers.

15. Market and non-market output of individual health services are also covered in the expenditure account.

- Expenditure on market output is included in the final consumption expenditure of households, NPISHs and general government.
- Expenditure on non-market output is defined as the output of their producers, that is NPISHs and general government, less receipts from sales.

16. This reinforces the earlier observation about the need to ensure that individual health services are consistently covered and treated on both sides of the accounts.

### **2.3 Estimating output at constant prices**

17. Methods for estimating output at constant prices can be classified as output methods that use prices or quantities relating to output and input methods that use prices or quantities relating to inputs. ESA 95 has an evident preference for output methods since these make it possible to analyse changes in productivity. Generally, input methods assume that the relationship between inputs and outputs is constant and that there are no changes in quality or productivity.

- Market output of individual (health) services at constant prices should be estimated either by deflating current values by price indices or by using base-year prices to revalue the quantities underlying the current price series. When neither deflation with price indices nor revaluation with base year prices is feasible, either because it is difficult to obtain reliable and representative price indices or because the price and quantity data are not sufficiently detailed, volume changes in output should be obtained using quantity indicators of output. If this too is not possible, the volume changes will have to be derived using constant price estimates of inputs or, in the last resort, quantity indicators of inputs. (ESA 95 para. 10.32 and 10.42)
- Non-market output of individual (health) services at constant prices should be estimated by using base-year unit costs to revalue homogeneous quantities of output. This approach requires unit cost and quantity data at a fine level of detail and, for this reason, may not be viable. Instead, volume changes in the output should be based on quantity indicators of output. If reliable and representative quantity indicator of output are not available, the volume changes will have to be derived using input measures. (ESA 95 para. 10.25, 10.42 and 10.44)

18. Output methods need to be based on as detailed a product classification as possible so that each product identified has maximum homogeneity. This is to ensure that changes in quality over time that result from modifications of the physical characteristics of products or from variations in the composition of output are shown as changes in volume and not as changes in price. (ESA 95 para. 10.18)



19. Besides ensuring that quality changes appear as volume changes, disaggregated data also makes it easier to define relevant and representative price indices and output indicators. Output methods at a high level of aggregation may provide estimates that are no better than those obtained using input methods.

20. The constant price estimates of market output and non-market output should be consistent. This means that an output price method or output indicator method should be used for market output and a unit cost method or output indicator method should be used for non-market output. Consistency does not require the method to be the same providing the definition of output is the same. An output price method could be used for market output and an output indicator method for non-market output. Consistency is lost when an output method is used for market output and an input method is used for non-market output.

21. There should also be consistency between the current price estimates and the constant price estimates of output. This is ensured for both market output and non-market output when a price method or output indicator method is used to obtain the constant price estimates. Consistency between current and constant price estimates of non-market output does not require an input method to be used to obtain the constant price estimates even though output at current prices is defined as the sum of inputs.

## **2.4 Definition of output**

22. ESA 95 is clear that volume changes in the output of individual health services should be measured in terms of the use made of the services and not in terms of the final results or outcomes obtained from their use. The objective is to measure the quantities of services actually produced and not to measure the benefit or utility derived from the services. (ESA 95 para. 10.26)

23. In other words, a clear distinction has to be made between the output of individual health services and the health of the community. It is not possible to measure the volume of health services by improvements in the health of the population, though outcome measures may indicate something about the quality of services provided if the impact of other factors influencing community health can be determined.

24. The definition of health output is based on that developed for educational output by the Taskforce on Education.

*The health output is the quantity of health care received by patients, adjusted to allow for the qualities of service provided, for each type of health care. The quantities should be weighted together using data on the costs or prices of the health care provided.*

25. The quantity of health care received by patients should be measured in terms of treatments. In theory, a treatment should be measured in its entirety. The whole bundle of complementary services constituting a treatment should be taken into account: the medical services, the paramedical services, laboratory and radiological services and, in the case of hospitalisation, the non-medical services such as the provision of food and accommodation.

26. In practice, the feasibility of measuring complete treatments depends on the degree of fragmentation of the services making up a treatment. Outpatients consume sets of discrete services that are provided by independent producers. Inpatients consume bundles of services while they are hospitalised, but usually these services can not be linked with the outpatient services they receive before and after hospitalisation. Even treatment received as an inpatient may be fragmented if, for example, the patient is transferred from a hospital to a nursing home.

27. Different measures of treatment are required for the different producers of individual health services. For some, the measure could be relatively simple - for example, the number of consultations for the medical activities of general practitioners. For others, a composite measure of varying complexity would be required - such as the number of treatments by type of treatment for the medical activities of dentists or the number of patients discharged grouped by homogeneity of diagnosis, treatment and cost (or Diagnosis Related Groups) for hospital activities.

28. Monitoring changes in the quality of individual health services will be difficult, because of the complexity and speed of the changes. The quality of treatments depends on the amount of resources provided per patient including the number and mix of medical and paramedical staff, the quantity and sophistication of the equipment at their disposal and the availability of the latest drugs and medicines. The provision of individual health services is now an evolving high technology activity. As a result, both the division of labour between medics and paramedics and health treatment practices are changing rapidly. For example, an increasing number of surgical interventions and drug therapies formerly requiring hospitalisation are now administered as outpatient treatments.

29. Quality changes in individual health services combine positive and negative effects that are difficult to measure explicitly. For example, a reduction in the length of stay in hospital may be indicative of a quality gain from improved medical treatments and procedures resulting from advances in drug therapy and surgical technology; but it may also be indicative of a quality loss because post-operative care previously provided by hospitals is now being provided by family and visiting paramedics.

## **2.5 Criteria underlying the recommendations**

30. The recommendations recognise the need for consistency between the production and the expenditure sides of the accounts, between the current and constant price estimates of output and between the constant price estimates of market output and non-market output. They also recognise that measures of volume changes in output should encompass both the quantity and the quality changes in the output. The proposed unit of measurement for the output of individual health services is a complete treatment.

31. Central to the recommendations is the criterion that output at constant prices should be estimated using either an output price method or an output indicator method for market output and an output indicator method for non-market output.

32. The best output price method would be that which uses the ideal price index. The ideal price index is defined as that which, when used to deflate output at current prices, produces the ideal volume index. The best output indicator method would be the one which uses the ideal volume index. The prices/quantities used to construct the ideal indices have to meet the practical requirements listed below:

- The prices/quantities should relate directly to output. This means that they should refer to complete end products and not to contributory activities or to contributory intermediate or primary inputs.
- The prices/quantities should have sufficient stratification: This means that different prices/quantities should be available for all different product groups comprising output.
- Product groups should have sufficient homogeneity. This requirement will be met if there is only one product in a product group. If there is more than one product within a product group, an additional requirement is that the composition of the product group does not change over time.

- The prices/quantities should be sufficiently representative of the product group. Ideally, this requires an integral observation of the prices/quantities of all items in the product group, but this may not be possible. Prices/quantities may not be available for the product group; or the prices/quantities available do not cover all products of the product group; or the prices/quantities for the products covered are based on a sample. In these circumstances, changes in the prices/quantities that are observed should be representative of changes in the prices/quantities that are not observed.
- The prices/quantities for a product group should account for changes in the quality of products. Changes in values resulting from changes of quality should be excluded from the price index and included in the volume index. The volume index = quantity index x quality index.
- Prices should be combined using the value of output in t as weights. Quantities should be combined using the value of output in t-1 as weights for market output and the total costs in t-1 as weights for non-market production.

33. The requirements are not all independent. For example, the requirements of sufficient stratification and sufficient homogeneity are closely related. Sufficient stratification is also pertinent to measuring changes in quality.

34. The requirements are the same for both prices and quantities with the following exceptions:

- The sufficiently representative requirement is less severe for prices than for quantities. Price changes observed in a sample are expected to be representative for the whole population more often than quantity changes observed in a sample of the same size. This is an important advantage of output price methods over output indicator methods. It is more difficult to get a good coverage of output with quantities than with prices.
- Weights of the current period should be used for output price methods and weights of the base period should be used for output indicator methods. In other words, the price indices should be Paasche indices and the volume indices should be Laspeyres indices.

35. The requirements are a good starting point for the definition of A, B and C methods. The difference between A, B and C methods is that they increasingly differ from the ideal index. Whether a method should be called an A, B or C method depends on how far its volume index is expected to differ from the ideal volume index. Hence:

- A methods (or recommended methods): the volume index obtained by the method is expected to be a good approximation of the ideal volume index. It should take account of quantity and quality changes in output. Quantity should encompass complete treatments. Quality should cover both changes in the physical characteristics of products and changes in the product mix. A methods make it possible to analyse changes in productivity.
- B methods (or methods to be used in case an A method cannot be applied): the volume index obtained by the method is expected to be a reasonable approximation of the ideal volume index. It should take account of quantity changes in output and quantity should encompass complete treatments. It should also take account of quality changes arising from changes in the product mix. B methods make it possible to measure changes in productivity approximately.

- C methods (or methods to be avoided): the volume index obtained by the method is expected to be an unacceptable approximation of the ideal volume index. In other words, it is neither an A method nor a B method.

It should be noted that since the ideal volume index is unknown, the judgment whether a certain method gives a good, reasonable or unacceptable approximation has to be determined by statistical analysts.

### **3. RECOMMENDATIONS OF THE TASKFORCE**

#### **3.1 Output methods**

##### **3.1.1 Output indicator methods**

36. Below an attempt will be made to describe the requirements for A methods and for B methods according to the six aspects discussed in section 2.5 and to give some examples.

###### **3.1.1.1 DIRECT RELATION TO OUTPUT**

###### **• A-methods**

37. A strict requirement for an A method next to quality changes allowance is that the indicators and weighting coefficients used have a direct relation to output. For instance with hospitals output is the complete treatments of diseases. Appropriate indicators are the number of discharged patients by kind of disease and kind of treatment (classified according to a DRG type of classification or similar classification) weighted by their price or cost in t-1. (See Annex II).

**38. OUTPUT INDICATORS BASED ON NUMBER OF CONSULTATIONS (GENERAL PRACTITIONERS, SPECIALIST PRACTITIONERS) BY KIND OF CONSULTATION; NUMBER OF DENTAL TREATMENTS BY KIND OF TREATMENT ETC ARE ALL INDICATORS SATISFYING THE REQUIREMENT FOR A METHOD.**

###### **• B methods**

39. *An example of an approximation is the method presented in the document of De Boer and Zijlmans (see Annex III). This method can be applied if data are available for the number of complete treatments in hospitals but their price or cost, necessary for weights, are unknown. Costs of nursing days, surgical operations etc. are allocated (by cross-tabulation) to complete treatments. The resulting total costs per treatment are used as approximate weights for the number of patients classified by disease and treatment. As long as no direct data are available about the production value or the total costs per complete treatment this method seems to be an acceptable approximation. Of course, data have to meet certain requirements of detail and homogeneity. Since this method needs assumptions it should be considered that the requirement for a B method has been fulfilled.*

40. The Taskforce has the opinion that a method using a volume index of the output of hospitals based on numbers of contributory activities like nursing days, X-ray pictures and operations cannot fulfil the requirement for A nor B methods. The reason is that the effects of the complementarity of contributory activities are measured wrongly. E.g. a decrease in nursing days per patient caused by an increase in the quality of operations would be measured as a decrease in volume of output while in fact the number of treatments did not change. It appears from trial calculations that mistakes can be considerable.

###### **3.1.1.2 Stratification and homogeneity**

###### **• Cases of inpatient treatments**

41. Some treatments are a combination of different kinds of care with substitution effects. Most of them are represented by inpatient treatments with hospital stays.

42. An A method for the output of *hospitals and comparable institutions* asks for a classification of complete treatments (or treated patients) that takes into account differences between diseases and differences in treatment methods. Generally speaking the use of a DRG type of classification is an important factor to get an A method because each diagnosis related groups is defined to represent an identical level of consumption of resources (medical and structural) and to be medically consistent.

43. Whether a DRG type of classification meets the requirements for an A, B or C method depends on the extent of detail and degree of homogeneity of data results. This could be established in further research.

#### • *Old people's homes*

44. The output of old people's homes is the service delivered by the institutions to the occupants. A candidate indicator would be the average number of occupants during the year or occupant days. If the population of occupants is very homogeneous or the composition of the population is stable, this is an appropriate indicator. However, because of the steadily aging European population and rapid institutional changes in the social and medical sector, the indicator average number of occupants during the year seems to be too rough.

45. A solution is available if occupants of old people's homes can be classified into groups that require different degrees of care. E.g. in the Netherlands basic statistics of old people's homes classify occupants into four groups according to the intensity of care they need. If also the price (or cost) per occupant day for each care intensity group is available, this data can be applied as weights. Then the stratification requirement for an A method has been fulfilled.

46. If price or cost information is not available, the Taskforce recommends implementing a small sample survey to use as a last resort, with weights per intensity of care group based on "expert-guess" from specialists familiar with the sector. If this happens in a proper way, the requirement for a B method has been fulfilled.

#### • *Other cases*

47. A large part of output of medical activities of *general practitioners* are consultations of patients. Of course those patients have different diseases. Yet, the character of the consultations is more or less the same. A smaller part of output concerns special treatments of patients like dressing wounds.

When special treatments are delivered two cases can occur.

48. It could be a simple act such as immunization. That means that a normal appointment has to be made and the price of the immunization is the price of a consultation. That means also that the volume corresponding to the act of immunization can be considered the same as for a simple consultation and we don't need to know the number of immunizations since we consider an immunization equivalent to a consultation and the requirement for an A method has been fulfilled.

49. When a specific treatment more important than a simple consultation or a simple act is provided (mostly specialists, dentists or paramedics) it is certainly recorded and paid as a specific medical act. If in that case specific output indicators are available, the requirement for an A method has been fulfilled. If there is only one indicator for the total number of consultations (normal + specific) the requirement for a B method has been fulfilled. In Member States where

specific acts become numerous within general practitioners activity, surveys should be implemented to obtain output indicators by kind of treatments.

50. A special situation occurs when general practitioners have a collective agreement with a health insurance company or the Ministry of Health to care for the health of a certain group of people. If the number of consultations etc. is available, there is no special problem. However, if the number of consultations and specific acts general practitioners provide is unknown, the estimation of a volume index is a major problem. One possibility could be to use the total number of people in the group. This is a very rough approximation that is based on the dubious assumption that the composition of the group, the health of the group and the quality of the care provided by general practitioners has no noticeable change between two years. The Taskforce considers that the requirement for A or B methods has not been fulfilled.

51. The Taskforce recommends that, in Member States where collective agreements are widespread, annual sample surveys should be implemented in order to obtain the number of consultations and specific acts provided by general practitioners, dentists etc. Depending on the size and reliability of the sample the requirement for A or B methods has been fulfilled.

### **3.1.1.3 Representation**

52. Generally speaking, output indicators require large samples in order to be sufficiently representative of the whole population, both when used in A or B methods.

53. A special representation problem arises when for one part of output (e.g. 70%) there is integral observation and for the remaining 30% no observation exists. Then a potential solution could be a combination of output indicator methods and price indicator methods. The volume index of the former 70% will, generally speaking, not be representative for the other 30% of the output. However, the -implicit- price index of 70% could, indeed, be representative for the remaining 30%. This solution is based on the assumption that price indicators from an incomplete observation more often are representative than the corresponding volume indicators. (See document TFHTH/42).

54. The Taskforce observed that medical and social security statistics often give a lot of integral information e.g. about hospital care that is not used for national accounts purposes.

55. Whether such a method can fulfil the requirement for a B method depends on the validity of the assumptions to be made in a certain case.

### **3.1.1.4 Quality**

56. Accounting for quality change is a strict requirement for an A method. Quality should cover both changes in the physical characteristics of treatments and changes in treatment mix. If both quality changes are accounted for the requirement for an A method has been fulfilled. If only quality change resulting from treatments mix is accounted for then the requirement for a B method has been fulfilled. If there are no corrections for quality at all, as is the present situation in most Member States for medical services, only B or C methods are possible. Section 3.3 will give more detail on this point.

### **3.1.1.5 Weights**

57. Requirement for an A method:

- With non-market production weighting with the observed *costs* per product group in t-1.

- With market production weighting with the observed *value of output* per product group in t-1.

58. Requirement for a B method:

- With market production weighting with the observed costs per product group seems to be a good approximation and will give a B method.

- With non-market as well as market production weights for complete treatments in hospitals derived from costs of complementary activities could be an acceptable approximation. See section 3.1.1.1 above "Direct relation to output".

### 3.1.2 Price methods

59. Price methods play a part with *market production*, not with *non-market production*. Some articles suggest for countries where the market sector is significant that the price index for market health could be used to be extended to non-market health with proper adjustments. This solution seems difficult to implement mainly because the kind of care provided in such a non-market sector certainly will not be comparable with that provided in the market sector.

60. A very important point is the use of price data from consumer price statistics (CPI's) in national accounts. Generally speaking, the requirements for deflators in national accounts and for price indices in CPI's coincide. This means that price data and methods that are acceptable for the CPI might be also acceptable for national accounts. However, an important question is whether CPI price indices comply with national account concepts. At least two aspects have to be discussed.

61. The first point concerns the use of the CPI to deflate output where it is a representative deflator of private consumption. However, in the case of medical and social services, the major part of production coincides with private consumption and there are no complications such as trade and transport margins. But according to the recent Commission Regulation adopted in July 1998 for the HICP, Member States should record the care prices net of reimbursements. This means that using the CPI, any institutional arrangement for the delivery of health such as the change of rate of reimbursement will have an impact in the volume measure while there is no change in the number of acts nor quality of care. Hence it appears impossible to have the same price index for both health private consumption and output and adjustments should be made to cover the total price of the output.

62. In principle, another problem could be that mostly CPI's are calculated with fixed weights of a base year while in national accounts annually changing weights should be used.

63. Consequently the use of CPIs depends on the way they are compiled. In some countries CPI could be suitable. In that case and provided the requirements expounded in this section are met, it would be advisable to use as much CPI price data as possible for reasons of efficiency, because production price statistics (PPI's) for medical and social services appear to be very scarce and in order to advance mutual comparability.

64. Below an attempt will be made to describe the requirements for A methods and for B methods according to the six aspects discussed in section 2.5 and to give some examples. There seems to be much similarity between the requirements for acceptable price methods and acceptable output indicator methods. Therefore, the discussion below is less extensive than with output indicator methods.



### 3.1.2.1 *Direct relation to output*

#### • *A methods*

65. A strict requirement for an A method next to quality allowance, is that the price indicators used have a direct relation to output. For instance with hospitals output is the complete treatment of diseases classified into diseases and kind of treatment (e.g. DRG type of classification). Then cost changes caused by complementary activities are measured as a price change.

**66. PRICE INDICES BASED ON PRICES OF CONSULTATIONS (GENERAL PRACTITIONERS, SPECIALIST PRACTITIONERS) BY KIND OF CONSULTATION; PRICES OF DENTAL TREATMENTS BY KIND OF TREATMENT ETC ARE ALL INDICATORS SATISFYING THE REQUIREMENT FOR AN A METHOD.**

#### • *B methods*

67. The Taskforce has the opinion that a method using a price index of the output of hospitals based on separate price indices of complementary activities like nursing days, as in present CPI's cannot fulfil the requirement for A or B methods since all changes resulting from complementary activities are measured as volume changes.

68. If prices of complete treatments in hospitals are not available, an approximate deflator of hospital output could –indirectly- be derived from costs of complete treatments classified into diseases and kind of treatment (e.g. DRG type of classification).

*69. An example of an approximation is the method presented by Henk van Tuinen, Bram de Boo and Jaco van Rijn (1997). A prototypical price index is calculated using both basic Consumer Price Index (CPI) information and additional information on the average number of inpatient days. The total cost of treatments is the sum of the costs of inpatient days and specialist services. The methodology is applied to different levels of aggregation of treatments. Comparisons of the results are carried out between the prototypical price index and the hypothetical price index resulting from the standard CPI procedure from 1980 to 1994.*

70. The first experiment concerns appendectomy that is a relatively homogeneous commodity. The prototypical price index increases by an average rate of 1.5% a year as against a 4.4% increase in the hypothetical price index. The difference is due to the decrease in the average number of inpatient days for an appendectomy from almost 11 in 1980 to less than 7 in 1994. The study shows that a frequent reweighting of the hypothetical price index is not a solution to the complementarity problem.

71. A second experiment extends the methodology on most inpatient treatments according to the level of 17 sections of the International Classification of Diseases (ICD). In this case the 17 groups are considered as if they were homogeneous. The prototypical price index increases by an average rate of 2.5% a year as against a 4.8% increase in the hypothetical price index. The difference is due to the decrease of the average number of inpatient days in most groups of diseases. The study shows that it is quite impossible to assume that the decrease in the average number of inpatient days per treatment is caused mainly by changes in the composition of the group of diseases.

72. According to the Dutch figures the volume effects of a change from the hypothetical to the prototypical price index for inpatient medical services on some aggregates are the following (average per year):

Gross value added inpatient medical services: 3.28%

Gross value added health and veterinary services: 1.42%

Gross Domestic Product: 0.07%

73. *Since this method needs assumptions it should be considered that the requirement for a B method has been fulfilled.*

74. Note: The output method presented in De Boer and Zijlmans (Annex III and section 3.1) is based on the ideas of Van Tuinen c.s. concerning the CPI and is in fact the symmetric of their price method.

### **3.1.2.2 Stratification and homogeneity**

75. The result of deflation has to be an acceptable volume index. Therefore, the requirements for stratification and homogeneity for price methods are the same as for output indicators. So, we can refer to section 3.1.1.3 for a more extensive discussion.

76. The similarity of price and output methods also holds for collective agreements between practitioners and insurance companies. Suppose data are available on prices per patient that insurance companies pay to doctors. Deflation with such data does not fulfil the requirement for A or B methods. The reason is that the resulting volume indicator (number of patients) is too heterogeneous.

### **3.1.2.3 Representation**

77. Generally speaking, A as well as B methods do not require, for price indices, as large samples as for output indicators in order to be sufficiently representative for the whole population. The first underlying assumption is that there is a certain equality of price of a product on the market. Thus, price indicators observed in a small sample can be representative for the whole population. The second underlying assumption is that closely related products show comparable price changes. So, not all products need to be in observation. The conclusion is that from price data based on relatively small samples, provided that the observed units are chosen with care, reliable price index numbers can be derived.

#### **3.1.2.4 Quality**

78. Corrections of the price index for quality are a strict requirement for an A method. If there are no corrections for quality at all, as seems to be the present situation in CPI's in most Member States for medical services, only B or C methods are possible. The requirements as to quality changes for price methods are the same as for output indicator methods. So, we can refer to 3.1.1.4 for a more extensive discussion.

#### **3.1.2.5 Weights**

79. Weighting at two levels of aggregation has to be discussed here. Firstly, the lower level of aggregation of the current price data in national accounts. If deflation takes place at this level the price indices for higher levels of aggregation are -implicitly- weighted, according to the rules, with *values in t*.

80. Secondly, it is possible that the deflators used above are based on price data at lower levels of aggregation. At this detailed level weights can be very different: fixed base year CPI, expert-guess etc. Since this takes place at lower levels of aggregation, it seems to be acceptable.

## 3.2 INPUT METHODS

### 3.2.1 Input-based approach to measuring volume health output

81. In Member States the current way of measuring output of non-market health services at constant prices has been to use, as a proxy, input indicators. As said before input methodologies attempt to deflate all inputs, which are then summed and the total used as a proxy for total volume output (see para. 2.2). They fall into two broad categories.

82. The first, an input price approach, derives constant prices by deflating current price inputs. Ideally, a comprehensive breakdown of inputs is required and should be separately deflated with an appropriate deflator. In this way decomposition into price and volume effects is possible. For example wage increases in the same categories are in the price component and switches between categories are in the volume component. In practice this detail is seldom available or if available is not used and only the change in the general rate of salaries is included.

83. The second method, an input indicator approach, extrapolates inputs in the base year using quantity information. As said before all elements of input should be covered. Most often only employment in the health service is included, adjusted onto a standardised basis, or the number of hours worked. The unit of quantity for compensation of employees is considered to be an hour's work of a given type and level of skill. A quantity measure of work done would be a weighted average of quantity relatives for different types of work, weighted by the values of compensation of employees in the base year.

### 3.2.2 Strengths and weaknesses

84. The main advantage for using an input over an output-based approach for measuring health is that data are readily available (employment numbers, wages and salaries, intermediate consumption, etc.). In addition the methodology is usually fairly straightforward and it is consistent with what is generally done for other non-market services.

85. The main weaknesses with input methods are that, at constant prices, they are not consistent with the health output definition given in the paragraph 2.4. Say, the quantity of health care received by patients measured in term of treatments. Constant prices derived from input methods do not relate directly to quantity and quality of treatments. Comparisons with trial calculations mentioned in section 3. show that differences in the growth measure are significant. Measures with input methods imply that analysis of productivity is not possible.

### 3.2.3 The productivity issue

86. The demands on the Health Service in all countries continue to increase and Governments are always trying to balance, on the one hand, keeping costs down whilst at the same time satisfying increased demand for services offered. This has to be achieved whilst maintaining as far as possible the quality of service provided. That means that restrictions are made on input, mainly employment, whilst technological improvements as well as new treatments and medicines all contribute to the efficiency with which health care is delivered. In that case input methods interpret the reduction of the number of employees in hospitals as a reduction of output. The assumption that there is no change in productivity for health services is not defensible.

Is it possible to improve inputs methods to allow for productivity?

87. Some Member States do incorporate productivity adjustments to their input-based non-market output series. By definition, productivity calculations need both an input and output element, so any adjustments made will, of necessity, be indirect ones. There are three main categories used.

88. The first approach is to make adjustments based on productivity assumptions in other comparable areas. As these are inevitably going to be market activities, the assumption is that they are fairly comparable, but this is by no means certain. If productivity indicators in these areas are themselves fairly weak because of difficulties in estimating constant price output, then they should be avoided for estimating non-market productivity.

89. A second approach is to apply a simple flat rate, often arbitrary, adjustment per annum. Only Norway of the Taskforce Member States does this - at ½ per cent per year - for the non-market part of health. An arbitrary assumption cannot be easily defended and there is no reason to suppose that it would be the same for all Member States. It could be argued that zero productivity change is a safer assumption than anything positive but it is unlikely to reflect reality.

90. An alternative method assumes that, in the long run, wage-rate increases can be broken down as increases paid to compensate for inflation, and as increases paid for increased productivity. This at least makes an approximate and probably partial allowance for changes in productivity. Unlike the other two methods, it is based on statistical data in the non-market sectors. However this improvement is not sufficient to comply with the requirements for a proxy of an ideal volume index as mentioned in section 2.5.

#### *3.2.4 A, B or C criteria*

91. The Taskforce agrees that input-based approaches should be classified as C methods. As more data become available in Member States in order to monitor the cost of the health sector, the Taskforce thinks that it is preferable to implement output methods as soon as possible rather than to look for unlikely improvements of input methods.

#### **3.2.5. Recommendations**

92. The Taskforce agrees that it is desirable to move away from input methods to output methods for measuring healthcare. However, it recognizes that, in the short-term, those input methods may have to be used.

### 3.3 QUALITY CHANGES

93. Quality changes should cover both changes in the physical characteristics of products and changes in the product mix. The permanent technological improvements and health research advances as well as the rising trend of the demand and supply of care within a general context of cost control make the quality changes in health services an important issue. The Taskforce did not find documentation of explicit adjustment for quality within health national accounts.

94. Health improvement covers two cases. Changes mainly come from introduction of new treatments replacing or not older ones. Changes also come from improvement of existing treatments that refers to the way the medical rules are carried out and to the input mix involved. Of course there is a grey area between the two categories and the Taskforce did not have the expertise needed to define when the quality changes in an existing treatment become a new treatment. Further research should give rules for defining new treatments.

95. An important consideration is how larger changes in medical treatments are accounted for in the volume index. An example is the treatment of kidney stones. In the past the patient needed an operation whereas now kidney stones are eliminated by mechanical vibrations. Such a large change in quality obviously yields a new treatment and should be accounted for in the volume index by the introduction of the revised treatment as a different, new product. If this happens in the usual way for the introduction of new products in price and volume statistics the Taskforce thinks that at least the requirement for a B method has been fulfilled.

96. The next consideration is whether a method is acceptable if corrections for small quality changes in medical treatments are not accounted for. Generally, it is difficult to measure such changes in a proper way. Examples are small changes in medical techniques, changes in the number of infectious diseases contracted by patients in a hospital, changes in the number of medical errors and changes in additional facilities for patients. The Taskforce assumes that in these cases the year to year changes are small and partly counterbalancing, though over a number of years effects could cumulate. Nevertheless, the opinion of the Taskforce is that the requirement for a B method has been fulfilled.

#### **4. IMPLEMENTATION OF THE RECOMMENDATIONS**

97. The Taskforce recommends that output indicator methods should be used to estimate the output of human health activities at constant prices. Not only do they have the advantage of providing volume measures of output that are consistent between market and non-market sectors, they also provide a more appropriate decomposition between price and volume than do input methods. This is an important consideration for the interpretation of health figures in the national accounts. Initially, output indicator methods will only account for quality changes arising from changes in product mix. Defining and measuring quality changes in health care needs to be investigated.

98. The Taskforce considers that from a methodological point of view output indicator methods are feasible for human health activities. The immediate availability of suitable data is undoubtedly a problem, but the main difficulty is the availability of manpower and associated resources to evaluate, utilise and develop the data sources. To establish output indicator methods, national account services require the support of statistician working in the health sector. This, with all the liaison and cooperation involved, can be a time-consuming process.

99. Nonetheless, the Taskforce, aware of the wide set of data already available in Member States and of the new data that will becoming available in order to monitor the health sector, encourages national account services to take the opportunity to build new methods. Implementation is likely to be a step-by-step process.

100. Member States should be able to implement output indicator methods as soon as possible for simple treatments, such as those supplied by general practitioners, specialist practitioners, dentists, paramedics, etc., for which data are already available and easy to use.

101. For more complex treatments, such as those supplied by hospitals, Member States need more time for data collection and research. The introduction of a DRG type classification, which group inpatients by the homogeneity of diagnosis, treatment and cost, is decided by the Healthcare Administration not by the National Statistics Institute. The Taskforce encourages national account services to establish a dialogue with Healthcare Administrations with a view to using such administrative statistical sources.

102. In the medium term, Member States should develop approximative output indicator methods, such as the Dutch methods showed in Annex III. Though they need time to be put in place, they are based on data that are usually available.

## **5. RECOMMENDATIONS FOR FURTHER RESEARCH**

### **5.1 Social work activities and veterinary activities**

103. With the exception of old people's homes in section 3.1, the Taskforce did not deal with social work activities (NACE N Group 853). Neither did it deal with veterinary activities (NACE N Group 852). Further study is required to see whether the Taskforce recommendations for human health activities are applicable to these services.

### **5.2 Implementation of output methods**

104. There are a number of Member States that have just introduced, or are in the process of introducing, their own DRG type classification for health system and hospital administration purposes. The classification as implemented may or may not be suitable for national accounting purposes. The main issues are the characteristics of the classifications itself, the handling of changes in the classification and the construction of the volume index with the data provided by the classification and the calculation of cost weights. Member States with DRG type classifications should investigate these issues with a view to developing a methodology and exchange their experiences with other Member States.

105. There are other Member States that have a classification of treatments but not of the corresponding costs. Nevertheless, price and volume indices can be constructed by cross-classifying treatments and the costs of complementary activities as explained in Annex III. Member States with such classifications should test this methodology and exchange their findings with other Member States.

### **5.3 The quality issue**

106. This is a specific issue important to include in the research programme. As mentioned in the Section 3.3 investigation should be undertaken to establish rules in order to define new treatments.

### **5.4 The price issue**

107. The Task Force has not been able to review all the work currently being undertaken for consumer price indices or producer price indices in respect of human health activities. Different institutes have carried out studies and these should be evaluated to see whether their approach can be applied in national accounts.



## **ANNEX I**

**Price index numbers of complementary goods by Henk van Tuinen, Bram de Boo and Jaco van Rijn.**

## ANNEX II

### **Informative elements on Diagnosis Related Groups (DRGs) and its use in France to support hospital management**

These first elements are extracted from a French article on implementation of the “Programme de Medicalisation des Systèmes d’Information” which is used for the hospitals management and, from the article “Use of Diagnosis Related Groups (DRGs) to support hospital sector management in the European Community” (by M. Casas - in health services research G.N. Fracchia and Theofilatou-1993-) based on a study for the European Community.

**1- Definition of DRG :** Classification of hospital stays in a limited (500) number of homogeneous groups from a medical and economic point of view. These groups are defined to represent an identical level of consumption of resources (products such as laboratories analysis, X-ray, nursery cares, surgery operations...) and to be medically consistent. It has been established in USA (in 1970 with “data analysis “ methodology. It is annually revised (13 Th. revision in 1996). Firstly used for prospective payments of Medicare to old and disable people, it has been extended to all patients hospitalised (AP-DRG). There is a more complex version (1100) groups including relationship between primary and secondary diagnosis, outpatient care, psychiatric care, long stays...

**2- Definition of case mix :** is the range of services produced by a hospital (classified according to DRGs) giving the number of treatments for each DRG.

### **3- Uses of DRG classification in Member States**

The use of DRGs had spread in various European countries since 1984 (14 European countries; 11 Member States). Portugal, Belgium, Norway, Ireland, Italy fund hospitals according to their medical activity measured using DRGs. Information required for the DRG grouping system exists from several European countries patient information with the implementation of the Minimum Basic Data Set (MBDS) recommended by the European Community in 1982. It seems that other system has been adopted in Germany.

Using European data collected according to DRGs needs more harmonisation on many issues: countries do not use the same version of DRGs, nor the same classification for diagnosis, nor the same guidance to record the stay length, there are various approaches to obtain costs per products or DRG.

### **4- The French application**

In France this classification has been adapted to implement a new system for funding hospitals (“Programme de Medicalisation des Systèmes d’Information”: PMSI). This system has been used in a first time (1995) in order to compare the subsidies received by an hospital with the amount it could have had if its activity would have been paid according to a national average of costs of activities. Nowadays the system is applied to public hospitals (more than 100 beds) since 1997 for funding purposes and should be ready for use to private hospitals that have a convention with the Social Security, at the end of 1998. In a first step the application has been restricted to short length care (including less than 24 H) in medicine, surgery and obstetric services and has been extended in a second step to recovering care. Collection of information for medium length care is compulsory since the 1<sup>st</sup> July 1998 and the budgetary application will likely be ready to use in the year 2000. Long length care will never be covered. The application to psychiatric hospitals is being experimented.

**- Entrance-discharge report:** Each short hospital stay gives rise to a standardized report with administrative and medical information: identification of the hospital unit, age, sex, date of

entrance and discharge, transfer to other hospital units, main diagnosis, associated diagnosis, medical acts provided. That means there are as many discharged reports as weeks of care by kind of care. Diagnosis are encoded according to the international classification CIM-10 of the WHO and medical acts according to a national catalogue. Medical acts have a Relative Cost Index (ICR) which gives the share and relative costs of medical, nursery activities and material involved in this act.

Each report is classified within a DRG (using software).

**- Calculation of reference costs per DRG :** data from a sample of hospitals are compiled to establish a national scale of costs per DRG. The sample is adjusted representing the national structure of hospital stays by kind of hospitals and excluding hospital stays that have an extreme cost. The six kinds of hospitals are the following: little, medium and big hospitals, the university hospitals, the private hospitals that have a convention with the public hospital service and the cancer centers.

All the running costs of the hospital are split up for each medical and medico-technical unit (salaries of medical, para-medical administrative employees). In order to these units calculate the cost of elementary activities (x-ray, analysis, surgery...) and so on for all other costs. So that each hospital stay cost is calculated including expenditure directly affected to the inpatient, expenditures of medico-technical acts, expenditures of care, accommodation, administration and structure expenditures.

The reference cost of a DRG is the average cost in the sample for this DRG (it was previously the median cost). These costs by DRGs are converted in ISA points (Synthetic Activity Index). In the ISA scale, the cost of a birth without complication has been a reference with 1000 ISA points in 1995 and 1998. Each DRG reference cost in the hospital sample is divided by the cost (in the sample) of a birth without complication that gives the number of ISA points for each DRG. That means for example that the cost of heart transplantation (18697 ISA points) is 210 times the Radiation oncology seance cost (89 ISA points). In 1996 and 1997 the method operated as if the national output (say the number of ISA points) have been the same as in 1995 for an identical national case-mix.

Dividing the total cost in the sample by its total number of ISA points gives the value (cost) per ISA point which can be split up for each kind of expenditure (medical care, accommodation, structure...).

Given the number of ISA points by DRG from the sample it is possible to calculate the number of ISA points for each hospital for hospitals not belonging to the sample and the sum of ISA points shows the national hospital activity.

Dividing the total cost of any hospital by its total number of ISA points gives the value of its own ISA point. Comparisons between hospitals (and with the national standard) are carried out according to the value of their ISA point.

ISA points are used to determine the annual amount to be granted to the hospital and the regional health budget (CF table for a case mix).

This detailed presentation shows that many statistics are collected and could be involved in output indicators method for national account purposes. But several difficulty should be got over because reference costs and ISA scales are annually revised and DRG classification often changes too. Data are not available at the moment for national account purposes.

## ANNEX III



**Statistics Netherlands**

Division Presentation and Integration

**A method for the derivation of volume indices and deflators of complete treatments in hospitals.**

Sake de Boer and Gerrit Zijlmans

version: 1 September 1998

## 1. Introduction.

In this note we discuss a method for the estimation of price indices and volume indices for health services in the national accounts. This paper focuses on the problem of complementarity of activities within hospitals.

Starting point are the ideas presented in Nordhaus [1] and Van Tuinen, De Boo and Van Rijn [2]. They discuss the complementarity of certain household consumption goods and its consequences for the estimation of the price index of household consumption. Van Tuinen c.s. present an improved method for the estimation of the price index of household consumption. As an example treatments in hospitals consist of a combination of two activities, an operation and a number of inpatient days. Van Tuinen c.s. argue that as a result of better medical techniques the number of inpatient days per treatment decreases, implying a decrease in the total costs and thus the price index per treatment. However, currently no consumer price information is available for complete treatments, but only for the separate activities that contribute to them. Using this kind of price information leads to overestimation of the all-over price index if the effect of the complementarity of operations and inpatient days on the weighting coefficients is not accounted for. Without this adjustment the cost decrease caused by the decrease in the number of inpatient days is treated as a volume effect and not as a price effect.

The problem of complementarity does not only affect private consumption price statistics but has also an impact on the measurement of volume and price indices in national accounts. Ignoring complementarity can result in underestimation or overestimation of the growth rate of important national accounts indicators. In this note we illustrate in which way volume indices for complete treatments in short-stay hospitals can be derived from information on the separate activities that are part of these treatments.

If prices and production values of groups of complete treatments could be obtained directly from the bookkeeping accounts of the hospitals, deflators of the production values of treatments could easily be derived. However, this kind of information is not (yet) available. Our method gives an *approximation* of this ideal situation applying a statistical method adopted from input-output analysis. It is based on the cross tabulation of complete treatments with the hospital activities contributing to them. This method meets two requirements: firstly, it accounts for complementarity and secondly, it encompasses the complete short-stay hospital industry in the national accounts. The latter implies that all treatments carried out in hospitals and all the complementary activities are included in the model.

## 2. Classification of complete treatments and complementary activities.

We illustrate the method with a numerical example. For reasons of simplicity the number of distinguished complete treatments and complementary activities is small.

Output of hospitals is very heterogeneous. We tackle this problem by distinguishing as a first step treatments with an operation and treatments without an operation and between procedures with day and night nursing care and procedures with only day care. In addition we distinguish policlinic visits to hospitals. This results in five product groups (complete treatments):

- treatments with an operation combined with day and night nursing care (TODN)
- other treatments combined with day and night nursing care (TXDN)
- treatments with an operation with only day nursing care (TOD)
- other treatments with only day nursing care (TXD)
- policlinic visits (TPV)

In this example seven complementary activities are distinguished:

■ operations combined with day and night nursing care	(AODN)
■ operations combined with only day nursing care	(AOD)
■ other procedures combined with day and night nursing care	(AXDN)
■ other procedures combined with only day nursing care	(AXD)
■ day and night nursing care	(ACDN)
■ day nursing care	(ACD)
■ polyclinic services	(APS)

It is assumed that all activities of hospital doctors and nurses are included in these seven activities.

### 3. Indices without accounting for complementarity.

Since we are interested in the quantitative consequences of complementarity we first derive in this section

price and volume indices *ignoring* the complementarity of activities.

The figures in table 1 on the number of operations, other procedures, nursing days and polyclinic services (first two columns) somewhat resemble the Dutch situation in 1994 and 1995. However, the cost per unit figures in column 3 and 4 are fictitious and are made up for purposes of illustration. So, the resulting indices from this example have no relation to reality.

We assume that the costs per nursing day increase by 5% between year  $t-1$  and year  $t$ , whereas the cost increase of the other activities is about 3%. From data on quantities and data on costs per unit, total costs are calculated (see last three columns of table 1) in current prices as well as in constant prices, i.e. prices of  $t-1$ .

**Table 1. Quantities and costs of activities in short stay hospitals**

	Number (*1000)		Costs per unit (Hfl)		Total costs (Million Hfl)		
	$t-1$	$t$	$t-1$	$t$	$t-1$	$t$	$t$ (price $t-1$ )
AODN	700	710	6000	6180	4200	4388	4260
AOD	350	370	3000	3090	1050	1143	1110
AXDN	200	210	1000	1030	200	216	210
ACD	150	160	500	515	75	82	80
ACDN	16000	15700	600	630	9600	9891	9420
ACD	600	650	300	315	180	205	195
APS	23000	23000	100	103	2300	2369	2300
Total					17605	18294	17575

From the last three columns of table 1 indices can be derived for the total of short-stay hospital activities:

Value index:  $18294/17605 = 103.91$

Laspeyres volume index:  $17575/17605 = 99.83$

Paasche price index:  $18294/17575 = 104.09$

In this example, the volume of hospital activities shows a slight decrease despite the increase in the number of operations and other therapeutic procedures. This decrease is caused by the fact that this deflation method treats the decrease of nursing days as a decrease of volume and not as a decrease in costs.

#### 4. Indices accounting for complementarity.

The essential drawback of the approach in the previous section is that *complementary activities* and not *complete treatments* are analysed as the output of hospitals. Van Tuinen c.s. [2] argue that for correct measurement of price and volume changes complete treatments must be seen as the output of hospitals. This is also the view of Eurostat's *Task Force on Health in constant prices*.

In the Netherlands the only available data are the number of complete treatments (represented by the number of patients treated) and the prices and quantities of the complementary activities. The prices and costs of complete treatments are not available. However, they can be derived by a cross tabulation of complete treatments and complementary activities.

Table 2 gives the number of complete treatments (indicated by the number of patients treated) for the years t and t-1, again somewhat resembling the Dutch figures for 1994 and 1995.

**Table 2. Complete treatments in short stay hospitals**

	t-1	t	quantity index
	<i>Patients treated (*1000)</i>		
TODN	700	710	101.43
TXDN	900	900	100
TOD	350	370	105.71
TXD	250	280	112
TPV	23000	23000	100

Table 3 connects the complete treatments of table 2 with the complementary activities of table 1. This cross tabulation is the heart of the method and essential for the resulting indices. Underlying table 3 are estimates of the number of units of an activity that contributes to one unit of a certain complete treatment. These estimates sometimes can be derived from medical statistics, but are to be based on assumptions in other cases. An essential restrictive condition is that eventually every unit of activity has to be allocated to a treatment.

For reasons of simplicity we assume in table 3 that every patient is operated only once (in reality the number of operations exceeds the number of operated patients). As a result AODN equals TODN and AOD equals TOD. Further on it is assumed that the average number of nursing days (day and night care) is 9.5 in t-1. This gives 6650 nursing days (ACND) for treatments with an operation (TODN). For every operation with only day nursing (TOD) we assume one nursing day (ACD). The other figures in table 3 have been estimated residuary.

**Table 3. Cross tabulation of complete treatments and complementary activities for year t-1 (quantities)**

	TODN	TXDN	TOD	TXD	TPV	Total
	<i>number</i>	<i>of treated</i>	<i>patients</i>	<i>(*1000)</i>		
	700	900	350	250	23000	-
	<i>number of units of complementary activities (*1000)</i>					
AODN	700	-	-	-	-	700
AOD	-	-	350	-	-	350
AXDN	-	200	-	-	-	200
AXD	-	-	-	150	-	150
ACND	6650	9350	-	-	-	16000
ACD	-	-	350	250	-	600
APS	-	-	-	-	23000	23000

Table 4 “translates” for year t-1 costs of hospital activities (rows) into costs of complete treatments (columns). Table 4 is derived from table 3 by multiplying each row with the corresponding costs per unit (see table 1). The last row of table 4 gives the total costs per treatment in year t-1.

**Table 4. Cross tabulation of complete treatments and complementary activities for year t-1 (costs)**

	TODN	TXDN	TOD	TXD	TPV	Total
	<i>number</i>	<i>of treated</i>	<i>patients</i>	<i>(*1000)</i>		
	700	900	350	250	23000	
	<i>Million Hfl</i>					
AODN	4200	-	-	-	-	4200
AOD	-	-	1050	-	-	1050
AXDN	-	200	-	-	-	200
ACD	-	-	-	75	-	75
ACND	3990	5610	-	-	-	9600
ACD	-	-	105	75	-	180
APS	-	-	-	-	2300	2300
Total costs	8190	5810	1155	150	2300	17605

Next, applying Laspeyres volume index formula, the total costs per treatment category are used as weights in the computation of the volume index of total output. Table 5 gives the results.



**Table 5. Volume index t/t-1 of total output of short-stay hospitals**

	total costs t-1 (table 4)	quantity index t/t-1 (table 2)	total costs t prices t-1
	<i>Million Hfl</i>		<i>Million Hfl</i>
TODN	8190	101.43	8307
TXDN	5810	100	5810
TOD	1155	105.71	1221
TXD	150	112	168
TPV	2300	100	2300
Total	17605	<i>result:</i> 101.14	17806

The Laspeyres volume index of total output is  $17806/17605 = 101.14$ . The -implicit- price index of total output can be derived from total costs in current prices of year t (table 1) and total costs of t in prices of t-1 (table 5), which yields  $18294/17806 = 102.74$ .

## 5. Comparison of indices ignoring and indices accounting for complementarity

Table 6 compares the indices for total output of hospitals based on complete treatments on the one hand and the indices based on complementary activities (not accounting for complementarity) on the other hand.

**Table 6. Indices t/t-1: comparison of methods**

	based on activities	based on treatments	difference
value index	103.91	103.91	0
volume index	99.83	101.14	+ 1.3
price index	104.09	102.74	- 1.3

Table 6 shows that the differences in the indices derived with both methods are considerable. Most striking is that the slight decrease of volume from section 2 appears to be an increase of more than 1% if complementarity is accounted for.

Of course, again, a warning must be made. Our example is a simplification of reality and some data, e.g. those for the costs (weights) of operations and nursing days, are fictitious. For that reason definitive conclusions on the influence of complementarity cannot yet be made.

## 6. Some problems and possible solutions.

- a. Replacing fictitious data.

In the example the costs per unit of activity are fictitious. To make things more realistic, observed data on costs per unit per (type of) operation, per (type of) nursing day and per (type of) polyclinic visit should be used. It is to be hoped that data collected for consumer price statistics can be used for this purpose.

b. Heterogeneity of output.

In our example five types of output are distinguished. Each of those five product groups (complete treatments) is still heterogeneous. The results can be improved by distinguishing more types of output.

However, not all heterogeneity can be removed by using more detailed data. Reducing the number of inpatient days in hospitals by sending patients away earlier than before can still cause heterogeneity if those patients require additional care and further treatment in special rehabilitation hospitals or at home. This is a source of heterogeneity because it represents a quality decrease of treatments in short-stay hospitals. The method we have presented in the previous sections does not account for such quality changes.

A solution for this problem may be to extend the model by adding the activities of rehabilitation hospitals, general practitioners and physiotherapists and home care by district nurse services and home help services. However, still a solution is to be found for any heterogeneity e.g. as a result of an increase of volunteer help.

c. Cross tabulation of complementary activities and complete treatments.

Cross tabulation of treatments and activities is the heart of the method presented. It is necessary to utilise as much statistical sources as possible and to make assumptions -if necessary- as realistic as possible. In the Dutch situation for instance a problem that has to be solved is the allocation of second, third etc. operations (and other treatments) to complete treatments. Published medical statistics do not give that information.

## **7. Recapitulation of required data.**

### **7.1 Number of complete treatments (see table 2)**

A good indicator of the number of complete treatments is the number of treated patients. Such information should be provided by medical statistics. A requirement is that patients are classified into homogeneous groups according to disease and treatment. An example is a classification into Diagnosis Related Groups (DRG).

The Vademecum of health statistics of the Netherlands (1997, table 9.3.1) gives the number of “Discharged patients by group of diagnosis”, further distinguished into “in nursing day care” and “admitted for day and night”.

The Vademecum gives also the number of visits to outpatients departments of hospitals (table 8.1.4).

### **7.2 Number of complementary activities (see table 1).**

The total number of operations, other procedures, nursing days etc., if possible classified into different categories, should be provided by medical statistics.

The Vademecum of health statistics of the Netherlands (table 8.3.20) gives the number of operations (44 groups), other therapeutic procedures (5 groups) and diagnostic procedures (11 groups). Table 10.1.1 provides numbers of nursing days.

### **7.3 Cross-tabulation of treatments and complementary activities (see table 3)**

Table 3 (and table 4) is the heart of the presented method. Here is embodied which activity contributes to which treatment and to which degree. A part of the necessary data could be derived from medical statistics. It is especially important to have statistical data on average nursing day care per treatment group.

The Vademecum of health statistics of the Netherlands (table 9.3.1) gives the average length of stay in hospital per patient per diagnosis group. Other information can be derived by a combination of table 8.3.20 and table 9.3.1. Table 8.3.20 gives for instance the number of operations on 44 parts of the body. Sometimes these data can easily be related to the disease groups of table 9.3.1. E.g. “Operations on ears” can be directly related to “Diseases of the ears”. In other cases a combination of common sense and assumptions is required to fix the relation.

#### 7.4 Prices/costs per unit of complementary activity (table 1)

Table 4 realises the aim of the method: the weights for the volume indices of the disease and treatment groups. It is derived by multiplying the number of activity units in table 3 with their unit price or unit costs. This information could be derived from consumer price statistics or, perhaps, from medical cost data.

### 8 Some concluding remarks

In this short note we have discussed an approximation method for the estimation of volume and price indices of hospital services in the national accounts. Prices and costs of complete treatments are derived from prices and costs of complementary activities in hospitals. The results are used as weighting coefficients with the aggregation of volume indices for treatment groups into a volume index for total hospital production.

We expect that a first implementation of this method will be time and capacity consuming. Since data from medical statistics, price statistics and national accounts are combined, the first implementation will require a lot of co-operation and consultations between the statistical experts of the related departments. A lot of problems (further working out of the method, data collection, choice of classifications, fixing relations between treatments and activities, etc.) have to be solved. However, at the moment the implementation has been finished gains are high. Considerable misestimations caused by complementarity are avoided. Further on, the results not only can be used in the national accounts but also in consumer price statistics and in health statistics (calculation of costs of health care in constant prices).

### References.

[1] W.D. Nordhaus: *Do real output and real wage measures capture reality? The history of lighting suggests not.* Cowles Foundation discussion paper 1078. New Haven USA: Yale University, 1994.

[2] Henk van Tuinen, Bram de Boo and Jaco van Rijn: *Price index numbers of complementary goods. A novel treatment of quality changes and new goods, experimentally applied to inpatient medical care.* Paper prepared for discussion at the 1997 meeting of the International Working Group on Price Statistics (Ottawa group).

## ANNEX IV



Statistics Norway

*Ann Lisbet Brathaug*

### DRG USED IN NORWAY - AN EXAMPLE

#### 1. *The DRG system*

##### 1.1 HISTORY

The development of the DRG (Diagnosis Related Groups) system was initiated at the end of the 1960s by researchers at the University of Yale in USA. The purpose was to find a system for solving problems concerning quality, efficiency and costs in hospitals by using established methods from the manufacturing industry. The DRG system is now in use in USA and in several European countries as one of the basic sources for reimbursing hospitals.

In Norway the DRG system has been used in somatic hospitals since 1988. From July 1 1997 the DRG system is used as one of the financing sources of hospitals<sup>3</sup>. The Norwegian DRG version is the same as utilised by Health Care Financing Administration (HCFA) in the USA for Medicare patients. The version is updated yearly to cover both developments in medical technology and changes in diagnosis. At the time being, the Norwegian DRG system is not developed for outpatient treatments in hospitals nor for mental hospitals. Of about 80 hospitals in Norway 60 are classified as DRG hospitals. These hospitals cover approximately 95 per cent of all hospitalisation.

##### 1.2. COST WEIGHTS

From 1996 the HCFA-12 version is adopted. The version consists of approximately 500 DRGs. Each DRG is related to a cost weight. The cost weight defines the average cost of the specific DRG relatively to average cost at the national level (for the average patient). The cost weights are estimated on the basis of costs related to the specific DRGs. The specific DRGs are made up of four different components; costs related to average length of a stay, x-ray costs, laboratory costs and operation costs. So far, only inpatient treatments are included in the calculation of the cost weights. On average the costs related to the average length of a stay contribute to 67 per cent of the DRG's cost weight.

The present Norwegian cost weights are derived in 1992 and based on information from 9 hospitals. The weights are under revision, and the new weights based on cost information from 10 hospitals will be implemented from January 1 1999.

##### 1.3 DEFINITIONS IN THE NORWEGIAN DRG SYSTEM

- A hospital's **DRG points** are calculated as the sum of the DRG weights (cost weights) for all hospital stays (discharges).

---

<sup>3</sup> The DRG reimbursement is made up as: DRG points \* 0,45 \* the price for one DRG. The DRG points are calculated as the sum of DRG weights for all hospital stays (discharges). The unit price, which is 27 136 n.kr. for 1998, is calculated as the hospitals average expenses (costs) per inpatient. Source for the expenses: The hospitals accounts

- **DRG index:** The cost weights will tell how much resources are needed for the patients within a DRG. The cost weights summed up for the hospital will describe the “heaviness” of the mix of patients in the hospital. The DRG index is calculated as the sum of DRG points divided by the hospital stays (discharges), and, hence, the DRG index describes the average “heaviness” of the inpatients.
- The number of **index stays** at a hospital is a theoretic number which is used to correct for differences in the mix of patients in hospitals. Index stays are equal to DRG points. If all patients at the hospital have DRG-weights equal to 1, the number of index stays will be equal to the number of discharges. Hospitals with a high share of “heavy” inpatient treatments will have a higher number of index stays than discharges, and the opposite will be the case if the hospitals have a high share of “easy” inpatients. In this way, the DRG system can, better than the discharges, be used to describe the output from hospitals, and output from different hospitals will be comparable. However, it must be underlined that quality changes in the output are not taken into consideration. The number of index stays at hospitals can be summed up to form a global number of index stays.

## *2. Example: Output method compared to input method*

### *2.1 GENERALLY*

At time being, the DRG system is not being used for national accounts purposes in Norway.

In the following example the global number of index stays (DRG points) is used to calculate output of outpatient treatment in somatic hospitals at constant prices. This is done by calculating a volume index based on the index stays.

The output method is compared with the input method used in the Norwegian National Accounts. The present method can shortly be described as: The annual central and local government accounts are used to estimate output from hospitals at current prices. The output is non-market and, therefore, valued by the production costs as the sum of compensation of employees, intermediate consumption, consumption of fixed capital and other taxes on production. The output at current prices is deflated using a weighted cost price index based on price indices for the components of input (price indices for the components of intermediate consumption, consumption of fixed capital and wage rates) and assuming a fixed productivity growth per year.

Since the output method is covering only inpatients in somatic hospitals, the National Account figures for hospital services are corrected to cover the same population. This is done by excluding mental hospitals and outpatient treatments from the total output from hospitals. The method required some assumptions which may influence the figures.

The Norwegian input method is implying a fixed annual productivity growth, and, therefore, in the example is also included results from the input method when no productivity growth is implemented. Hence, three methods are compared; the output method based on the DRG system, the input method with fixed productivity growth included, and the input method without any productivity growth.

The following exercise must be looked upon as an illustration of the results coming from an output method based on genuine output statistics (the DRG system) compared to the results coming from an input method. It must be underlined that the output method is not corrected for quality changes in the products, and, therefore, not an A method as described in the report. The index used will, however, fulfil the requirements for an B method.

## 2.2 RESULTS

Using Norwegian figures, it was impossible to cover a longer period than 1990 to 1995. The first calculation of DRG points started in 1988, but due to uncoverage of hospitals the time series are not consistent back to 1988.

**Table 1. Output from hospitals, excluded mental hospitals and outpatient treatments.**  
**Annual volume changes**

Output method	4,2	1,6	4,1	0,7	2,7
Input method (incl. fixed productivity growth)	3,9	2,0	2,2	0,2	3,6
Input method (no productivity growth)	3,5	1,7	1,8	-0,1	3,2

**Table 2. Output from hospitals, excluded mental hospitals and outpatient treatments.**  
**Volume index. 1990 = 100**

	1991	1992	1993	1994	1995
Output method	104,2	105,9	110,3	111,1	114,2
Input method (incl. fixed productivity growth)	103,9	106,0	108,3	108,6	112,5
Input method (no productivity growth)	103,5	105,3	107,2	107,1	110,5

Normally one would predict that the volume changes should be higher using an output method than using an input method. This is also confirmed in the exercise based on the Norwegian data. The results show that the accumulated growth from 1990 to 1995 is 14,2 per cent for the output method and 10,5 per cent for the input method (no imputation for productivity growth), see table 2. Hence, the average growth rates (geometric) for the years from 1990 to 1995 are 2,7 and 2,0 per cent respectively. The impact on GDP by using the output method, is very small in the Norwegian case, and hardly noticeable on the growth rates.

Table 1 shows that even though the annual growth rates from the output method normally are higher than from the input method, this is not the case for the year 1995. From 1994 to 1995 the input method (with no productivity growth) is 0,5 points higher than the output method. This result is not as predicted, but it is hard to explain why it becomes like this. Part of it can be explained by assumptions done, see paragraph two under 2.1 Generally. There has been no time to investigate this case further on.

### References

SINTEF Unimed (1996): *Samdata Sykehus. Sammenligningsdata for somatiske fylkeshelsetjeneste 1996*. Raport 6/97

Ministry of Health and Social Affairs: *Prisliste DRG 1998 med kodeveiledning HCFA-12* (From January 1 1998)

Statistics Norway (1997): *National Accounts 1978 -1996*, Official Statistics of Norway (NOS) C 426

## ANNEX V

### LIST OF DOCUMENTS

#### FIRST MEETING (12-13 February 1998)

- 1TFHTH/1 : Agenda and elements of discussion for the first meeting on Health (N. Dufour)
- 1TFHTH/2 : Task force mandate (Eurostat Document)
- 1TFHTH/3 : EUROSTAT DOCUMENT CN 333 REV.2 ABOUT HARMONIZATION OF CONSTANT PRICE DATA
- 1TFHTH/4 : Draft of Decision (Eurostat Document)
- 1TFHTH/5 : Report of Task Force on NACE L (P. Konijn)
- 1TFHTH/6 : Practice in national accounts at constant price in Belgium and issue on delimitation of the health non-market sector (C. Mathys)
- 1TFHTH/7 : Price index numbers of complementary goods by Henk van Tuinen, Bram de Boo and Jaco van Rijn
- 1TFHTH/8 : Deflating health expenditures by A.J. de Boo and M.M. Smit
- 1TFHTH/9 : Practice in national accounts at constant price in United Kingdom (D. Harper)
- 1TFHTH/10 : Practice in national accounts at constant price in Spain (G. Mejias)
- 1TFHTH/11 : Practice in national accounts at constant price in France (G. Houriez)
- 1TFHTH/12 : “Analyse Medico-économique de l’activité hospitalière (PMSI)”
- 1TFHTH/13 : Overview of Dutch deflation methods in national accounts (P. Verbiest)

#### SECOND MEETING (2-3 April 1998)

- 2TFHTH/14 : Council Regulation (n° 58/97) of December 1996 concerning Structural Business Statistics (Eurostat Document)
- 2TFHTH/15 : Presentation of Eurostat Task Forces concerning Structural Business Statistics on NACE sections M, N and O (K. Hayes)
- 2TFHTH/16 : Draft report on “The national accounts’ needs and links to Structural Business Statistics in the section M, N and O” (Eurostat Document)
- 2TFHTH/17 : Report of the first meeting of the TF on Education (K. Hayes)
- 2TFHTH/18 : Report of the first meeting of the TF on Health (N. Dufour)
- 2TFHTH/19 : Agenda, elements of discussion for the second meeting on Health (N. Dufour)
- 2TFHTH/20 : New method to be implemented in Spain (G. Mejias)
- 2TFHTH/21 : New method to be implemented in United Kingdom (D. Harper)
- 2TFHTH/22 : Current practice in national accounts at constant price for health in Norway (A. L. Brathaug)
- 2TFHTH/23 : - Variables contained in the OECD health data base (D. Roberts)

- David Roberts transmitted also the OECD Publication on principles of health accounting for international data collections

2TFHTH/24 : The main characteristics of the data base about health expenses in Belgium C. Mathys)

2TFHTH/25 : Draft outline of the final report for Education (K. Hayes)

### THIRD MEETING (4-5 June 1998)

3TFHTH/26 : Report of the second meeting of the TF on Health (N. Dufour)

3TFHTH/27 : Agenda, elements of discussion and list of documents for the third meeting of the TF on prices and volumes for Health (N. Dufour)

3TFHTH/28 : Report of the second meeting of the TF on NACE L (P. Konijn)

3TFHTH/29 : Report of the second meeting of the TF on Education (K. Hayes)

3TFHTH/30 : Classifications used in health sector (D. Roberts)

3TFHTH/31 : Presentation on Diagnosis Related Groups –DRGs- (N. Dufour)

3TFHTH/32 : Use of Diagnosis Related Groups to support hospital sector management in the European Community (Study from M. Casas 1993)

3TFHTH/33 : Measuring output of non-market hospital services (Statistics New Zealand-Service of National accounts)

3TFHTH/34 : Definition of market/other non-market producers and output (Extract from a note of E. Dalgard for OECD)

3TFHTH/35 : Elements about quality change in health (N. Dufour)

3TFHTH/36 : Output of non-market (K. Hayes)

3TFHTH/37 : A method for the derivation of volume indices and deflators of complete treatments in hospitals (S. De Boer and G. Zijlmans)

3TFHTH/38 : Trial calculation in Norway using output volume indicator based on DRGs and comparisons with input methods (A.-L. Brathaug)

3TFHTH/39 : Does the French Consumer Price Index Overstate Inflation ? (F. Lequiller-INSEE)

### FOURTH MEETING (7-8 September 1998)

4TFHTH/40 : Report of the third meeting of the TF on NACE L (P. Konijn)

4TFHTH/41 : Report of the third meeting of the TF on Health (N. Dufour)

4TFHTH/42 : Estimation of volume indices for output from incomplete quantity data (S. De Boer)

4TFHTH/43 : New health methodology implemented in the UK (D. Harper)

4TFHTH/44 : DRG used in Norway - An example- (A.-L. Brathaug)

4TFHTH/45 : Report of the third meeting of the TF on Education (K. Hayes)

4TFHTH/46 : Five documents on HCPI : Council Regulations (EC) No 1687 and 1688 /98 of July 1998, Commission Regulation (EC) No 1749/96, Council Regulation



**(EC) No 2494/95, Report to the Council of Ministers on the HCPI in European Union**

**4TFHTH/47 : Revision of the CPI hospital services component. Elaine M. Cardenas (BLS)**

**4TFHTH/48 : Four Documents of the OECD meeting (September 98) concerning output method on health from Polish Central Statistical Office, Australian Bureau of Statistics, Statistics New Zealand and Israel Central Bureau of Statistics**

**4TFHTH/49 : Agenda and list of document for the fourth meeting (N. Dufour)**

**4TFHTH/50 : Draft of the final report**

**4TFHTH/51 : Hospital price inflation : What does the new PPI tell us? B. Catron and B. Murphy (BLS)**

**4TFHTH/52 : Technique, quality and productivity trends within medical care- Extracts from Productivity trends in the public sector in Sweden. Report to the expert group on public finance. Ministry of finance**

**LIST OF PARTICIPANTS TO THE TASK FORCE ON HEALTH**

SPAIN	Mrs. Gloria MEJIAS	Tel : 00-34-91-583 91 44 Fax : 00-34-91-583 93 53 E-mail : glmejias@strogoff.ine.es
UNITED KINGDOM	Mr. David HARPER	Tel : 00-44-171-533-60 18 Fax : 00-44-171-533-59 03 E-mail : david.harper@ons.gov.uk
NETHERLANDS	Mr. Sake DE BOER	Tel : 00-31-70-337 47 57 Fax : 00-31-70-337 59 81/31 70 387 74 29 E-mail : sbor@cbs.nl
NORWAY	Mrs. Ann-Lisbet BRATHAUG	Tel : 00-47-22-86 48 52 Fax : 00-47-22-864996 E-mail : abr@ssb.no
FRANCE	Mr. Guillaume HOURIEZ	Tel : 00-33-1-41 17 37 12 Fax : 00-33-1-41 17 78 48 E-mail : Guillaume.Houriez@dg75-g412.Insee.atlas.fr
OECD	Mr. David ROBERTS	Tel. : 00-33-1-45 24 88 44 Fax : 00-33-1-45 24 96 57 E-mail : david.roberts@oecd.org
BELGIUM	Mrs. Claude MATHYS	Tel. :00-32-2-221 22 07 Fax : 00-32-2-221 31 43 Email : claudemathys@nbb.be

EUROSTAT	Mrs. Nicole DUFOUR	Tel : 00-352-43-01-35-045 Fax : 00-352-43-01-33-029 Email: <a href="mailto:Nicole.dufour@eurostat.cec.be">Nicole.dufour@eurostat.cec.be</a>
----------	--------------------	---

**REPORT OF THE TASK FORCE HEALTH II**

**VOLUMES MEASURES FOR HEALTH**

November 2000

## EXECUTIVE SUMMARY

The aim of Task Force has been to systematically present the recommendations existing so far, to develop further the ideas discussed by the first TF on Health, to make proposals for parts of CPA N that have not yet been covered, and to exchange more recent experience on specific issues.

The focus of the Task Force's deliberations has been on the most important CPA classes of section N: 85.11 hospital services, 85.12 medical practice services, 85.13 dental practice services and 85.31 social work services with accommodation.

The first Task Force defined health output as: *"The health output is the quantity of health care received by patients, adjusted to allow for the qualities of services provided, for each type of health care. The quantities should be weighted together using data on the costs or prices of the health care provided"*. Two elements of this definition were considered particularly important for the classification of methods for constant price estimation: adjustment for quality and the notion of complete treatment. Another strict requirement for a first best method is that the indicators and the weighting coefficients used have a direct relation to output. All the requirements, some of these with somewhat less importance and relevance than those mentioned above, appear in the set of criteria used for the evaluation of methods. The Task Force recommends that the examination of the criteria is done domain by domain.

Regarding complete treatments, the Task Force will use a pragmatic notion which aims at capturing at least full treatments for each of the CPA services distinguished. An aspect of the complete treatment issue which continues to be relevant is the re-admission problem (in the case of hospitals) or the first visit problem (as the Task Force labelled the same phenomenon in the case of medical and dental practice services). The Task Force did not have the possibility to investigate deeply into this issue, and recommends that trial calculations should be carried out in countries where such data is available. If evidence from these calculations shows that re-admissions into hospitals should have no significant impact on the indicator for complete treatments, the Task Force thinks that the indicator can qualify for an A method even without corrections for re-admissions.

When giving recommendations on whether first visits or the total number of visits should be the appropriate indicator in the case of medical and dental practice services, it is the Task Force's view that it is necessary to distinguish between the different services (general practitioners, the various specialists). The Task Force thinks that there is a difference between consultations and treatments provided by a specialist and those provided by a general practitioner. Each visit to the general practitioner can be looked upon as a treatment in itself, and the treatments are normally rather homogeneous, while this is not the case regarding visits to specialists. This leads to the conclusion that for general practitioners the recommended output indicator is the total number of consultations by treatment, while for specialists the recommended indicator is the number of first visits broken down by type of specialist and type of treatment. Similarly, the output indicator for dental practice services is the number of first visits broken down by type of treatments. Nevertheless, the Task Force thinks that the first visit issue should also be explored by non-TF countries.

The Task Force's investigations showed that the availability of weights is a problem in general. In many fields one will have satisfactory output indicators, but the cost data needed for the weighting are not available. If no direct and representative weights are available, the Task Force recommends that indirect weights should be derived. Expert judgements can also be taken into account. Indirect weights and expert guesses can be accepted as B methods if sensitivity analysis show that the resulting volume indices are stable/not strongly influenced by changes in the weights.

To be a first best method, quality changes must be taken into account. Quality changes should cover both changes in physical characteristics of products and changes in the product mix. This will be a strict requirement for an A method in all areas, while the B methods should cover partial quality changes (product mix). In the Task Force's opinion partial quality changes related to product mix is captured by a sufficiently detailed product classification, for example by using DRG type of classifications for hospital services. Regarding service-related quality changes the Task Force did not have the expertise to investigate this further, but gives some advises on quality indicators that should be further explored. For example, the Task Force recommends to test whether the ratio qualified staff (man year)/occupant days could be an acceptable indicator of quality changes in the field of hospital psychiatric services, rehabilitation services, nursing services etc. For inpatient treatments in hospitals it is recommended that the MS should examine i) the number of infectious diseases contracted by patients in hospitals, ii) changes in the number of medical errors within the hospital, and iii) changes in additional facilities for patients. Further research is also necessary with respect to indicators of changes in the skills/qualifications of the staff.

The Task Force has classified output indicators into A/B/C methods for the following services: i) hospital services distinguished into inpatient treatments in general and specialised hospitals, hospital psychiatric services, rehabilitation services provided by specific rehabilitation institutions and nursing services, ii) medical practice services distinguished into consultations and treatments by general practitioners and by medical and surgical specialists, iii) dental practice services, and iv) social work services without accommodation.

For all hospital services, except nursing services, the recommended output indicator that fulfil the requirements for an A method will be a DRG type indicator if appropriate cost weights and full quality adjustments are taken into account. For rehabilitation services provided by specific rehabilitation institutions the number of occupant days by level of care is also acceptable for an A method. Regarding nursing services, the Task Force's view is that occupant days by type of institution (proxy for level of care) are sufficiently homogeneous to meet the requirement for an A method. The same approach is recommended for social work services with accommodation.

The Task Force thinks that the development and implementation of new methods in the field of health services will be dependent on the liaison and co-operation between national accountants, health statisticians and health administration.

## Contents

### Executive summary 116

### 1. Introduction 119

### 2. Task Force I: Definitions, main recommendations and underlying criteria 120

### 3. Key issues 122

#### *3.1. Definition of services: Inpatient and outpatient treatments 122*

##### 3.1.1 Inpatient treatments 122

##### 3.1.2 Outpatient treatments 122

#### *3.2. Complete treatments 123*

#### *3.3. Diagnosis Related Groups (DRG) 125*

#### *3.4. Weights 126*

### 4. Health and social work activities: discussion and recommendations 128

#### *4.1 Hospital services (CPA 85.11) 128*

##### 4.1.1 Inpatient treatments in general and specialised hospitals 128

##### 4.1.2 Hospital psychiatric services 130

##### 4.1.3 Rehabilitation services in rehabilitation centres/hospitals 131

##### 4.1.4 Nursing services 131

#### *4.2 Medical practice services (outpatients - CPA 85.12) 132*

##### 4.2.1 Consultations and treatments by general practitioners 132

##### 4.2.2 Consultations and treatments by medical and surgical specialists 134

#### *4.3 Dental practice service (CPA 85.13) 134*

#### *4.4 Social work services with accommodation (CPA 85.31) 135*

#### *4.5 Other services under CPA section N 136*

### List of documents 137

## 1. INTRODUCTION

1. At the meeting of the Working Party on National Accounts in February 1999 it was agreed that there should be a follow-up to the reports of the three Task Forces on non-market services. Eurostat had proposed a step-by-step process to address the difficulties concerning the implementation of the new methodology in the member states. Regarding health services (CPA N) two types of work should be undertaken. Firstly, all member states should describe and assess the available statistical sources. In particular, they should investigate information derived from hospital administration, social security and insurance records. Secondly, an expert group (Task Force Health II) should be set up to continue the methodological work on volume measures for health.

2. The aim of Task Force II (TF II) has been to (i) systematically present the recommendations existing so far, (ii) develop further the ideas discussed by the first TF, (iii) make proposals for parts of CPA N that have not yet been covered, and (iv) exchange more recent experience on specific issues.

3. The following countries were represented in the Task Force in addition to Eurostat: the Netherlands, France, Italy and Norway. Norway chaired the group. The Task Force met three times: in March and in May 2000 in Luxembourg, and in October 2000 in Voorburg.

4. All TF countries made trial calculations. The detailed results are presented in separate papers (see the attached list of documents) which served as the basis for the TF's work. The ISTAT paper concentrates on the problem of collective agreements for general medical practice services. INSEE's contribution discusses output methods for private health services, hospitals (including psychiatry and rehabilitation care) and medical care in old people's homes. The CBS analyses volume measurement of hospital cure and care. Special attention is given to outpatient treatment. Statistics Norway describes the system of Diagnosis Related Groups (DRG) and looks into the problem of re-admissions. Trial calculations are made for hospital output and for nursing and care services for the old and disabled.

5. The focus of the Task Force's deliberations has been on the most important CPA classes of section N: 85.11 hospital services, 85.12 medical practice services, 85.13 dental practice services and 85.31 social work services with accommodation. There was not enough time to have a closer look at CPA classes 85.14 other human health services, 85.20 veterinary services and 85.32 social work services without accommodation.

6. The outline of the report is as follows: Section 2 presents the core definitions and recommendations of the first Task Force (TF I). Section 3 describes several key issues. Section 4 presents the TF's recommendations for the various classes of CPA N 'health and social work services'.

7. Of the Task Force countries, direct volume output indicators is not implemented in the National Accounts except for Italy, where output indicators are being used to measure the volume of inpatient treatment in hospitals. So far the volume indicator does not include any quality adjustment. In the Netherlands it will be decided next year if they should start implementing direct volume measures for health. In Norway, a revision of the National Account's time series will take place in 2001. The plan is to implement volume indicators for the health services during this revision process. In the context of the TF work INSEE had calculated a direct volume index for the year 1998. Calculations for 1999 will be made as soon as basic data are available. Based on that experience a decision will be made whether this method will be used in national accounts.

## 2. TASK FORCE I: DEFINITIONS, MAIN RECOMMENDATIONS AND UNDERLYING CRITERIA

8. The aim of the Task Force I was to investigate existent and embryonic output methods which provide for consistent constant price estimates for both market and non-market health services. The Task Force I gave the following definition of health output: *The health output is the quantity of health care received by patients, adjusted to allow for the qualities of services provided, for each type of health care. The quantities should be weighted together using data on the costs or prices of the health care provided* (para. 24). TF I underlined also that the quantity of health care received by patients should be measured in terms of complete treatments.

9. Two elements of this definition were considered particularly important for the classification of methods for constant price estimation: adjustment for quality and the notion of complete treatment. Together with other requirements, which are listed below, they appear in the set of criteria used for the evaluation of methods.

10. A condition underlying the recommendations is that the constant price estimate of market output and non-market output should be consistent. This means that an output price or output indicator method should be used for market output and a unit cost or output indicator method should be used for non-market output. Input methods were considered not acceptable. It must be underlined that consistency does not require the methods to be the same providing the definitions of output is the same. However, consistency is lost when an output method is used for market output and an input method is used for non-market output (para. 20).

11. Like other Task Forces on price and volume measures TF I had to come up with a classification of the different methods into A (most appropriate), B (the alternatives if A methods cannot be applied) or C methods (those that should no longer be used). For the various sub-levels of health services possible methods were judged against a set of criteria. This allowed, in a second step, the mapping into the A/B/C classification.

12. The criteria were the following (TF I Report para. 32):

- prices/quantities should relate directly to output; i.e. they should refer to complete end products and not to contributory activities or to contributory intermediate or primary inputs,
- prices/quantities should have sufficient stratification; i.e. different prices/quantities should be available for all different product groups comprising output,
- product groups should have sufficient homogeneity. This requirement will be met if there is only one product group. If there is more than one product within a product group, an additional requirement is that the composition of the product group does not change over time,
- prices/quantities should be sufficiently representative for the product group. Ideally, this requires an integral observation of the prices/quantities of all items in the product group, but this may not be possible. Prices/quantities may not be available for the product group; or the prices/quantities do not cover all products of the group; or the prices/quantities for the products are based on a sample. In these circumstances, changes in the prices/quantities that are observed should be representative of changes in the prices/quantities that are not observed.
- prices/quantities for a product group should account for changes in quality of products; changes of quality should be included in the volume component. Changes in quality cover two dimensions: changes in physical characteristics and changes in product mix,



- prices should be combined using values of output in t as weights; quantities should be combined using the value of output in t-1 as weights for market output and the total costs in t-1 as weights for non-market output.

13. As stated in paragraph 56 of the first report a strict requirement for an A method is accounting for quality changes. If both quality dimensions are accounted for the requirement for an A method is fulfilled. If only quality changes from treatments mix is accounted for the requirement for a B method has been fulfilled. If there are no corrections for quality at all, only B or C methods are possible.

14. A strict requirement for an A method next to the adjustments for quality changes is that the indicators and the weighting coefficients used have a direct relation to output (para. 37). The importance and the relevance of the other criteria listed in paragraph 11 above are somewhat less high. However, a method should fulfil all the requirements at a minimum level. The combination of the "scores" regarding each requirement is decisive when it comes to classifying a method as A, B or C. TF II recommends that this examination is done domain by domain. In section 4 the Task Force will discuss to what extent a suggested method will meet the requirements, and thus classify the method into A, B or C.

### 3. KEY ISSUES

#### 3.1. Definition of services: Inpatient and outpatient treatments

15. The recommended methods require indicators (prices or quantities) at the product level. Hence, the further discussion will refer to the services as defined in section N of the CPA (health and social work services). Section N covers:

- Hospital services (CPA 85.11) - inpatient treatments (cover surgical services, medical services, gynaecological and obstetrical services, rehabilitation services etc.)
- Medical practice services (CPA 85.12) - outpatient treatments (cover services by general practitioners and by medical and surgical services)
- Dental practice services (CPA 85.13)
- Other human health services (CPA 85.14) (cover services provided by midwives, nurses, physiotherapists etc, ambulance services, medical laboratories)
- Social work services with accommodation (CPA 85.31) (cover welfare services to old and disabled persons delivered through institutions etc)
- Social work without accommodation (CPA 85.32).

16. A well-established distinction is the one between inpatient and outpatient treatment. Inpatient treatment is by definition provided by hospitals only, whereas outpatient treatments are provided both by hospitals and by general practitioners and specialists outside hospitals. CPA classifies all outpatient treatments under CPA 85.12 (medical practice services).

##### *3.1.1 Inpatient treatments*

17. Hospital activities cover general and specialised somatic hospital activities, mental hospital activities, specialised somatic nursing homes activities, rehabilitation, maternity home activities etc. Normally, a general hospital will cover several kinds of specialised services such as psychiatry, obstetrics, surgery, medicine and rehabilitation. In addition there will be specialised hospitals for psychiatry, cancer, heart diseases etc. and institutions for long term stays providing nursing and rehabilitation services. Hospital services include services delivered under the direction of medical doctors chiefly to inpatients, aimed at curing, restoring and/or maintaining the health of a patient (medical and paramedical services, nursing services, laboratory and technical services, radiological and anaesthesiological services, etc.).

18. Inpatient treatments cover both day treatment and overnight stays in hospitals. A day treatment is defined as less than 24 hours of hospitalisation. In this report inpatient treatments will be split into hospital services in general and specialised hospitals (medicine, surgery, obstetrics etc.) rehabilitation services (mostly long term care), psychiatric services (both short and long term care), see table 1.

##### *3.1.2 Outpatient treatments*

19. Most hospitals also provide outpatient treatments; i.e. patients have consultations with a specialist at the hospital or receive a dialysis, chemotherapy etc treatment. The outpatient treatment in hospitals is similar to consultations and treatments by general practitioners and by private medical and surgical specialist (outside hospitals). Consultations and treatments by general practitioners (GPs) cover services consisting of prevention, diagnosis and treatment by doctors of medicine of physical and/or mental diseases of a general nature. The services can be provided in general practitioners' practices and also be delivered by outpatient clinics etc. Consultations and treatments by medical and surgical specialists (in or outside hospitals) include specialised medical services like consultation services in paediatrics, gynaecology-obstetrics, neurology, psychiatry, etc. and surgical consultation services, treatments as dialysis, chemotherapy, insulin therapy, X-ray treatments etc. The discussion in section 4 will distinguish between outpatient

treatments by GP and outpatient treatments by medical and surgical specialists (in and outside hospitals).

20. An overview on inpatient and outpatient treatments on the basis of a cross-classification by provider is given in table 1.

**Table 1 Inpatient and outpatient treatments by provider**

Medical services	General and specialised hospitals (excluded psychiatric hospitals)	Psychiatric hospitals	Rehabilitation centres and hospitals	Nursing homes	General practitioners or medical and surgical specialists
Inpatient treatment: • day treatment • overnight stays • long term care	X  X	X  X  X	X  X  X	   X	
Outpatient treatments	X	X	X		X

### 3.2. Complete treatments

21. In the TF I report para. 25 it is said that the quantity of health care should be measured in terms of complete treatments. A complete treatment in the comprehensive sense may consist of a great number of complementary services and a number of providers will be involved. In most cases, patients visit their general practitioner first, and the practitioner may refer them to a specialist who will perhaps transfer them to hospital either as an outpatient or an inpatient. After the inpatient treatment the person may be followed up by outpatient treatments or he/she might be sent to a nursery or rehabilitation centre (see figure 1). In an ideal world the patient should be monitored during his long march through the health sector. In practice, however, most countries do not have the possibility to do this because of a serious lack of data, and besides there will be a need to preserve confidentiality of the data. TF I underlined that the feasibility of measuring complete treatments depends on the degree of fragmentation of the services making up a treatment (para. 26) *"Outpatients consume sets of discrete services that are provided by independent producers. Inpatients consume bundles of services while they are hospitalised, but usually these services can not be linked to outpatient services they receive before and after hospitalisation. Even treatment received as an inpatient may be fragmented if, for example, the patient is transferred from a hospital to a nursing home."* Separate measures of treatment are required for the different producers of health services. The TF II agrees with the view of TF I and will use a pragmatic notion of a complete treatment which aims at capturing at least full treatments for each of the CPA services distinguished later on (see figure 1).

**Figure 1. Link between a treatment in a the comprehensive and the narrow sense**

GP => specialist => hospital (inpatient) => rehabilitation centre => GP: one complete



### 3.3. Diagnosis Related Groups (DRG)

26. The DRG system is a system for classifying hospital stays in general and specialised hospitals into groups that are medically meaningful and as homogeneous as possible regarding resources used. Based on medical and administrative information about the discharges, each hospital stay will be placed in one and only one DRG. In countries, the DRG system is normally used by health administrators, mainly in the Ministries, and the system is developed to finance hospitals. The investigation made by the Task Force shows that DRG system is now implemented in many countries. Originally, the DRG system was not designed for outpatient treatments or for mental hospitals/psychiatric services within a general hospital, nor was it designed for treatments in rehabilitation centres and nursing homes. Of the TF countries this is the case in Italy and Norway, while in France the DRG system is at present used for both inpatients and outpatients in medicine, surgery, obstetrics, and rehabilitation, and psychiatry will be included in the system from 2001. It is also important to underline that since the DRG system is made for administrative purposes, the statisticians' needs and requests are not necessarily well taken into account. Since the system is relatively new, the Task Force has not had the possibility to clarify if new treatments are introduced in the system in a proper way and if the weighting is done according to the requirements.

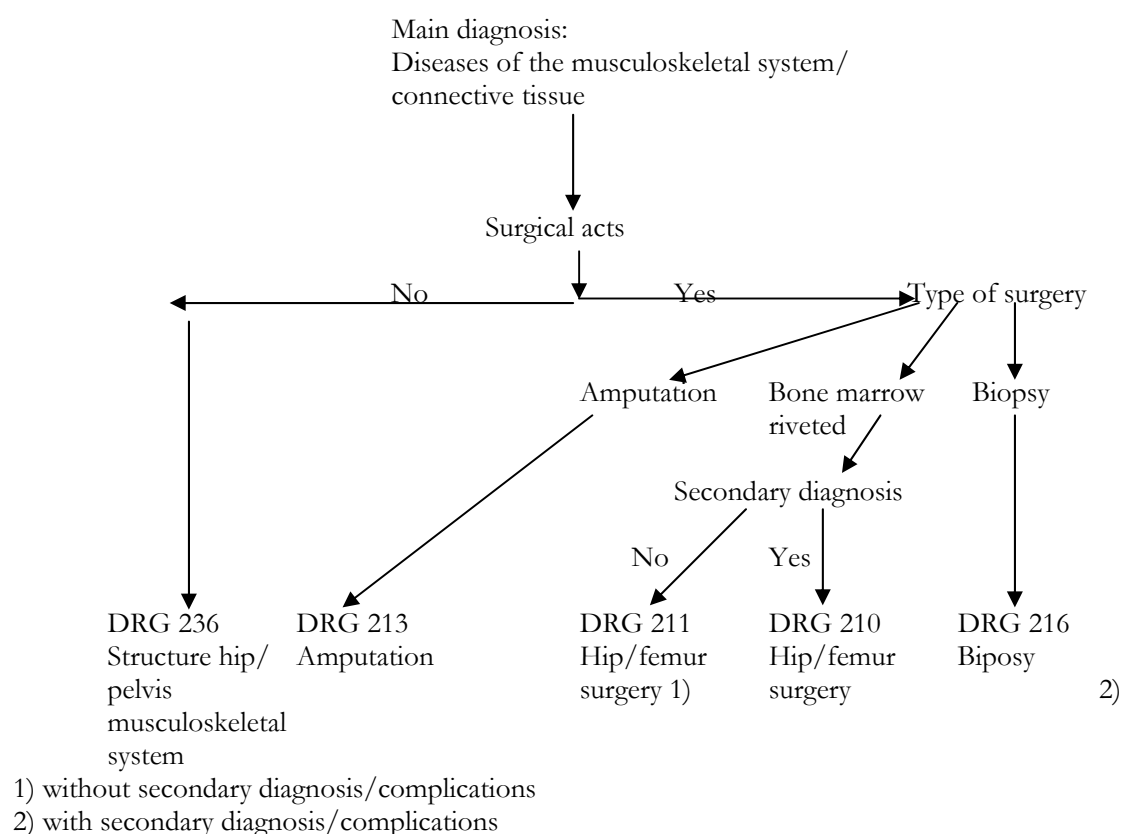
27. The first DRG system was initiated at the end of the 1960s by researchers at the University of Yale in the USA. Several DRG versions were developed in the USA in the 1970s and 1980s. All DRG versions were based on the classification system originally developed at the Yale University, but dependent on the patient population various versions were developed. In the USA there are now two main systems. One is the HCFA group (Health Care Financing Administration), which is used by Medicare to finance older patients outside the public financing system. The second is the All Patient group, which is used by New Jersey and New York to finance all hospital services. It was in the state of New Jersey, that the DRG system for the first time in 1980 was taken into use in a broader scale, with the purpose of financing all hospital services (unit cost financing). It is these American DRG systems that many countries have applied and to some extent developed further. For instance, the Nordic countries have introduced a Nordic version of the DRG system (NordDRG). The version follows the same structure as HCFA, but has some additional features. Further, a Norwegian version of NordDRG has been developed, called NorskDRG. This version includes a separate group for paediatrics and new-borns. This is also the case in Italy and France for new-borns.

28. The DRG classification is rather detailed. Those versions which are based on the HCFA groups, consist of approximately 25 main diagnosis groups and nearly 500 Diagnosis Related Groups.

29. To place a patient in a DRG, it is necessary to know the main diagnosis, the secondary diagnosis, procedures (medical and surgical acts), sex, age, length of stay, complications and associated diseases, and discharge information (date, discharged to etc). The DRG grouping of the discharged patients is thus based on information from the hospitals' administrative registers. To get correct DRGs, it is essential that the coding at the hospitals is done according to standard classifications, both regarding diagnosis, procedures and administrative information. In most countries, the patients are normally diagnosed according to the international classification of diseases. The classification in use is the World Health Organisation's (WHO) International Statistical Classification of Diseases and Health Related Problems (ICD). The Tenth Revision (1993) is the latest in a series that was formalised in 1983 as the Bertillon Classification or International List of Causes of Death. In the updated classification, conditions have been grouped in a way most suitable for general epidemiological purposes and the evaluation of health care. The ICD itself thus meets the requirement for diagnostic information for general purposes, while a variety of other classifications would be used in conjunction with it and would deal either with different approaches to the same information or with different information (notably medical and surgical procedures and disablement).

30. As mentioned in the previous paragraph information on procedures (medical and surgical acts) is necessary to get the correct DRGs. As far as the Task Force knows, the classification for coding surgical procedures will differ from country to country. For instance, in the Nordic countries the “NOMESCO Classification of Surgical Procedures” is introduced. At the hospital several procedures can be registered for each patient. A hierarchy exists between the procedures. The programme for grouping the data into DRGs will always choose the procedure that is most recourse demanding.

**Figure 2. Allocating patients to DRGs: An example**



### 3.4. Weights

31. Regarding the weighting criterion, the requirement for an output indicator method to be classified as A method is (para. 57 in TF I Report):

- *With non-market production weighting with the observed costs per product group in t-1*
- *With market production weighting with the observed value of output per product group in t-1.*

The requirement for a B method is (para. 58):

- *With market production weighting with the observed costs per product group seems to be a good approximations and will give a B method.*

32. Especially in hospitals with non-market as well as market-production weights for complete treatments derived from costs of complementary activities could also be accepted as a B method (para. 58).

33. The Task Force’s investigations showed that the availability of weights is a problem in general. In many fields one will have satisfactory output indicators, but the cost data needed for the weighting are not available. In addition the Task Force recognises that proper cost weights

will hardly be available every year, but as long as the weights are updated at regular intervals it is the Task Force's opinion that this is acceptable also for an A method.

34. If no direct and representative weights are available, the Task Force recommends that indirect weights should be derived. Expert judgements can also be taken into account. The TF does not recommend to use cost weights from other countries, as sometimes are suggested. In all cases where indirect weights or expert guesses are used, it is necessary to make sensitivity analyses to investigate whether or not the choices will result in volume indices that are stable (i.e. not significantly influenced by the choice).

35. The Task Force's recommendations regarding weights are the following:

1. *Best solution*: proper output values or cost weights (compatible with output indicator) based on full (exhaustive) coverage.
2. *Second best*: proper cost weights based on representative samples. This will be the case in France and Norway where the cost weights for DRGs are given by a survey on costs in a sample of hospitals.
3. *Third best*: indirect cost weights. Examples on indirect weights are given in section 4.
4. *Fourth best*: expert guesses. This may be acceptable given that weights do not have to be one hundred percent right.

In general, the Task Force thinks that both the best and the second best solutions are acceptable for an A method. Indirect cost weights and expert guesses can be accepted as B methods if sensitivity analysis show that the resulting volume indices are stable/not strongly influenced by changes in the weights.

#### 4. HEALTH AND SOCIAL WORK ACTIVITIES: DISCUSSION AND RECOMMENDATIONS

36. The focus in this chapter is on services where output price methods are impossible; i.e. non-market production, or pose specific problems so that output indicator methods become relevant.

##### 4.1 Hospital services (CPA 85.11)

37. Hospital services cover services delivered under the direction of medical doctors chiefly to inpatients (services to outpatients are covered by CPA 85.12) aimed at curing, restoring and/or monitoring the health of a patient. They cover both medical and paramedical services, nursing services, laboratory and technical services, radiological and anaesthesiological services etc. In the CPA classification hospital services are broken down at the 6-digit level into hospital surgical services, hospital medical services, hospital gynaecological and obstetrical services, hospital rehabilitation services, hospital psychiatric services, and other hospital services. The further discussion will relate to the following inpatient services:

1. Inpatients: overnight stays and day treatment (less than 24 hours of hospitalisation) in general and specialised hospitals (excluding psychiatric hospitals). This item is covering both hospital surgical, medical, gynaecological and obstetrical services, rehabilitation services (if provided in general hospitals) etc.
2. Psychiatric hospital services (both short and long term care, see table 1)
3. Rehabilitation services in rehabilitation centres/hospitals (covering mostly long term care and in this discussion not covering rehabilitation services within general hospitals, see table 1)
4. Nursing services provided in nursing homes.

38. The reason for this product split is that the DRG system in several countries is designed only for item 1. Up to now the services in rehabilitation centres and psychiatric hospitals have not been included in the DRG. An exception is that rehabilitation services in general hospitals are included. The Task Force has not been able to get any information whether there exist plans to include into an extended DRG system these institutions providing mostly long term care services. The exception to this is France where rehabilitation services in all institutions are covered and psychiatric treatment will be included in the DRGs from 2001.

##### 4.1.1 Inpatient treatments in general and specialised hospitals

39. It is the Task Force's view that it is not acceptable to just count the overall number of discharges to obtain an A or B method. The treatments are too heterogeneous. To meet the requirement for complete treatments it is necessary to take the whole package of complementary medical, paramedical, diagnostic, analytical and non-medical services comprising treatments into account. The appropriate method will be to use a classification that assigns inpatients to homogeneous groups according to their diagnosis, treatment and cost.

40. The TF I Report (para. 42 and 43) says "*An A method for the output of hospitals and comparable institutions asks for a classification of complete treatments (or treated patients) that takes into account differences between diseases and differences in treatment methods. Generally speaking the use of DRG type of classification is an important factor to get an A method because each diagnosis related groups is defined to represent an identical level of consumption of resources (medical and structural) and to be consistent.*" The investigations done by the Task Force show that the American based DRG classification is rather detailed. It consists of approximately 500 DRGs, in France even more (Wagner 2000). At this level of detail it can be supposed that the items are as homogeneous as possible and the Task Force finds this a sufficient level of detail to meet the requirements for *stratification and homogeneous product groups*.

41. To be a first best method, quality changes must be taken into account. Quality changes should cover both changes in physical characteristics of products and changes in the product mix. The permanent technological improvements and progress in health research as well as the rising trend of demand and supply of care within a general context of cost control make the



quality changes in health services an important issue. Improvements cover both changes from the introduction of new treatments replacing older ones, and improvements of existing treatments that refer to the way the medical rules are carried out, and to the input mix involved. Product-related changes in quality include:

- changes in quality due to better performing equipment/new techniques
- changes in quality due to better performing doctors/nurses (qualifications)
- changes in quality "related to the environment" (e.g. patients getting a infectious disease in the hospital, medical errors and changes in additional facilities for patients)

42. Changes in quality due to better performing equipment/new techniques (evolution) will sooner or later be captured by the creation of a new DRG, while the changes in quality due to qualifications and environment cannot be captured in this way. The TF underlines that it is difficult to define when the quality change in an existing treatment becomes a new treatment. Therefore, the TF will argue that to capture changes in product mix and new treatments it is important to have a sufficiently detailed classification. Regarding the latter two quality aspects mentioned in the previous paragraph, the Task Force thinks they can be of importance, but it did not have the possibility to investigate this further. The TF recommends that the MS should examine i) the number of infectious diseases contracted by patients in hospitals, ii) changes in the number of medical errors within the hospital, and iii) changes in additional facilities for patients. If the year to year changes is significant, adjustments must be carried out to obtain an A method. Further research is also necessary with respect to indicators of changes in the skills/qualifications of the staff.

#### A method

43. The Task Force's conclusion is that that DRGs fulfil the requirement for an A-method if appropriate weights are used and adjustments are made for all changes in quality. The DRG type of classification will normally be sufficiently detailed to meet the requirements for appropriate homogeneity, stratification and representativity. As long as the DRG system is used as a financing system for hospitals, cost weights will be available, mostly based on a representative survey of costs in a samples of hospitals. These cost weights will normally meet the requirement for an A method, see section 3.4. Using DRGs means that quality changes related to product mix (including the emergence of new treatments that are allocated to a new DRG) are fully taken into account. Changes in the quality of the basic services, however are not completely covered, and must be taken into account separately (see paragraph 42). Re-admissions into hospitals should be implemented in the method if the level of re-admissions changes significantly from year to year (see paragraph 23).

#### B method

44. DRGs with partial quality adjustments (only product mix) will fulfil the requirements for a B method.

45. Methods based on the ICD classification can be B methods if they fulfil the following requirements: i) detailed main diagnosis and ii) costs related to the ICD-diagnosis. In addition adjustments for quality changes resulting from treatment mix should be taken into account. This aspect is well taken care of if the diagnoses are sufficiently detailed. The diagnosis can be seen as a proxy for the output (treatment). The problem in this case will be to identify the cost weights since the costs for each diagnosis will depend on the treatment given (also depending on age and sex etc.). Indirect cost weights can be derived by cross-classifying the expenditure related to inpatients by sub-sectors (type of care), diagnosis groups and age groups as indicated in a Dutch paper (Eefting, Kleima et al 2000, see also Boer and Zijlmans 1998). Their analysis clearly shows that the concept of complete treatments requires detailed information. The level of diagnosis groups is important and the age dimension as well.

46. The number of inpatients grouped by detailed hospital disciplines (general medicine, general surgery, obstetrics and gynaecology etc.) will also qualify for a B method if appropriate weights and partial quality adjustments are taken into account (see Collesi 2000).

## C method

47. The Task Force considered the following methods to be C methods:

1. input methods
2. output indicators like the simple number of discharges (not taking into account the diseases and treatments etc.)

### *4.1.2 Hospital psychiatric services*

48. As mentioned in previous paragraphs (26 and 38), the present DRG system in many countries is not designed for psychiatric hospitals. The only exception the Task Force knows of is France, where psychiatric services will be included in the system from 2001. Data available in countries, will normally be the number of patients, admissions, day treatments, discharges, days of hospitalisation (occupant days). In some countries also diagnosis (by ICD) is known and can be cross-classified by type of psychiatric institution. What is lacking is data on the level of care received and diagnosis-related costs.

49. The Task Force agrees that the care received/treatment is more homogeneous for psychiatric services than for services in general hospitals (surgical, medical etc.). The reason for this is that the psychiatric patients normally do not have to go through surgical, radiological etc. procedures. However, there will of course be different types of care provided to different patients, some demanding long stays ("heavy" cases) and some demanding only short stays ("easy" cases). In some countries heavy and easy cases are treated in different institutions. If this is the case, an approximation for the level of care might therefore be to use information on days of hospitalisation per type of institution, assuming that the level of care per inpatient per overnight stay is equal within each of the different types of institutions. The Task Force considers this assumption to be acceptable to meet the requirements for at least a B method. Whether an A method is possible depends on the degree of specialisation of the institution and whether the weights are appropriate (see section 3.4) and quality adjustments are carried out.

## A method

50. The A method is a DRG approach with full quality adjustments and appropriate cost weights. There are two ways of capturing the care level in that case; either, the levels are systematically recorded in each institution via DRG type classifications or the various care levels can be related to specific institutions.

## B method

51. DRGs with only partial quality adjustments will be classified as a B method.

52. Occupant days (days of hospitalisation) by per type of institution (approximation for level of care) will be classified as a B method given appropriate cost weights. The weights will refer to the total costs of the different type of institutions. Regarding cost weights, for a B method indirect weights can be accepted. In Norway about 80 percentage of the costs in institutions are related to compensation of employees. If the number of personnel by occupation and by type of institution is known compensation of employees could be estimated. Hence, indirect weight can be derived using these estimates. For a B method also weights derived from expert judgements can be accepted.

53. Another output indicator that will qualify for a B method is the number of treatments by diagnosis (if appropriate level of details). The problem using this indicator will be related to non-available costs. The Task Force thinks that indirect weights or expert judgements could be used also in this field, and the method would still classify as a B method.

54. To be a B method the indicator used should be adjusted for at least partial quality changes. If the breakdown of treatments by diagnosis or the breakdown of occupant days by level of care is sufficiently detailed, the product mix aspect will be covered. Regarding quality changes relating to

service characteristics the Task Force has discussed several indicators but has had no time to test them in practice. The Task Force recommends that the ratio qualified staff (man year) /occupant days should be further investigated as a quality indicator in this fields. Studies done in Norway show that this ratio on average has increased significantly during the 1990s (Brathaug and Nørgaard 2000).

C method

55. Input methods will be classified as C methods. An output method based only on days of hospitalisation (not broken down by care of level) will also be classified as a C method.

#### *4.1.3 Rehabilitation services in rehabilitation centres/hospitals*

56. Rehabilitation services provided by general hospitals will be covered by the DRG system, and thus discussed under heading 4.1.1. What is discussed in this section is therefore rehabilitation services provided by specific rehabilitation institutions. These services will normally be related to longer stays, but can also be short. It can be argued that the rehabilitation services (treatments) are relatively homogeneous and even more homogeneous than psychiatric services. For this reason, the recommendations for rehabilitation services may differ slightly from the recommendations for psychiatric services.

57. Statistics that are currently available for rehabilitation institutions include the number of occupants, discharges and occupant days. In addition there will be information on the personnel.

A method

58. The best A method is a DRG approach with full quality adjustments and appropriate cost weights.

59. For rehabilitation services provided by rehabilitation institutions where no DRG classification is available, the number of occupant days (days of hospitalisation) by level of care is acceptable for an A method provided appropriate weights and full quality adjustments are made. The reason why this method is suggested as an A method for rehabilitation services, but not for psychiatric services, is that the services are more homogeneous. The Task Force recommends to test whether the ratio qualified staff (man year)/occupant days could be an acceptable indicator of quality changes related to service characteristics.

B method

60. DRGs with only partial quality adjustments will be classified as a B method, as will the number of occupant days by level of care if only partial quality adjustments are carried out.

61. In the field of rehabilitation services, the Task Force thinks that total number of occupant days can be an acceptable indicator for a weak B method, the reason being that the services are rather homogeneous. Also for a weak B method adjustments for quality changes should be carried out; in this case represented by adjusting for quality changes related to service characteristics.

C method

62. Input methods will be classified as C methods. If investigations in the MS show that rehabilitation services are more heterogeneous than the TF thinks, the simple number of occupant days will not meet the requirement for a B method.

#### *4.1.4 Nursing services*

63. Nursing services cover services provided by nursing homes/institutions (mostly long term medical and nursing care under medical supervision). The medical care element will be of less importance, since more serious illnesses will be treated in general or specialised hospitals. The services are provided to old or disabled people (mentally or physically disabled). In some

countries there will be separate nursing homes for somatical and psychogeriatric patients and there will be combined institutions providing care to both somatical and psychogeriatric patients. Even though the nursing homes provide mostly long term care, they will also provide some short term care. In Norway the short term care accounts for about 12 percent of total occupant days during a year. This percentage has been stable during the 1990s. The distinction between short and long term care is of less importance since the care provided is more or less the same and provided by the same type of personnel.

64. The Task Force concludes that the complete treatment is equal to the care provided. The care will consist of nursing hours, the nursing hours being homogeneous at least within an institution. However, the care/treatment can be heterogeneous between institutions in the respect that care is provided by different types of personnel and there can be more or less qualified staff per patient in the different institutions.

65. Available information concerning nursing homes will be the number of patients (at the beginning and end of the year) and the number of occupant days. In addition there will normally be information about personnel. In some countries administrative registration systems exist for patients in nursing homes and patients receiving home care (Kleima 2000, Brathaug/Nørgaard 2000). These systems will give detailed information about the patients' need for care and the care they actually receive, normally defined as hours of care.

#### A methods

66. In the Task Force's view occupant days by type of institution (indirect measurement of level of care) and adjusted for quality changes will meet the requirement for an A method. The Task Force thinks that the nursing services within an institution are sufficiently homogeneous. However, proper weighting and quality adjustments are necessary to obtain the A method. Proper weighting will require costs for the different types of institutions. As possible quality indicator is recommended the ratio qualified staff (man year)/occupant days by type of institution.

#### B methods

67. Occupant days by type of institution without any adjustments for quality changes in the basic services will meet the requirements for a B method. If proper cost weights are unavailable, indirect cost weights or expert guesses can be accepted for the B method.

68. Since the nursing services are rather homogeneous, occupant days adjusted for quality changes might be considered acceptable for a B method.

#### C method

69. All input methods are classified as C methods.

## **4.2 Medical practice services (outpatients - CPA 85.12)**

70. The following discussion will distinguish between i) consultation and treatment by general practitioners (GPs) and ii) consultation and treatment by medical and surgical specialists (in and outside hospitals). The first service consists of prevention, diagnosis and treatment by doctors of medicine of physical and/or mental diseases of a general nature. The latter consists of consultation services in paediatrics, gynaecology-obstetrics, neurology, psychiatry etc, surgical consultation services, treatment in outpatient clinics such as dialysis, chemotherapy, x-ray treatment etc.

### *4.2.1 Consultations and treatments by general practitioners*

71. The services of GPs are differently organised in countries. In some countries they produce market services, in other countries non-market services. In some countries GPs have a collective

agreement with a health insurance company or with the general or regional government to care for the health of a certain group of people. In such cases the doctors are paid to be available and to provide their service at request. Services under collective agreements can be market production or non-market production.

72. General practitioners who do not work under a collective agreement, will normally be reimbursed (or directly paid) according to patients treated and the tests etc. carried out. Under these circumstances the number of consultations by kind of treatment is normally available. As argued in paragraph 25, the nature of one visit to the GP will normally be looked upon as a complete treatment in itself, and the treatments can be considered as reasonably homogeneous. In the Task Force's view this leads to the conclusion that the overall number of visits (consultations) should be the appropriate indicator for the total number of complete treatments.

73. In the case of collective agreements, the number of consultations will not always be known which creates a problem for the estimation of a volume index. In Italy, for example, this is the case. The problem was discussed by TF I (para. 50 and 51) and the conclusion was *"one possibility could be to use the total number of people in the group. This is a very rough approximation that is based on the dubious assumption that the composition of the group and the quality of care provided by general practitioners has no noticeable change between two years. The Taskforce considers that requirement for A or B methods has not been fulfilled."* TF I recommended that in MSs where collective agreements are widespread (and no information on the number of visits is available), annual sample surveys should be implemented in order to obtain the number of consultations and specific acts provided by general practitioners. Whether this method will be an A or B method will have to be judged depending on size and reliability of the survey.

#### A method

74. The recommended output indicator is the total number of consultations by treatment. As weights one should ideally have the costs related to each treatment. To be the first best method adjustments have to be made also for all quality changes. The TF did not feel competent enough to make a suggestion on quality indicators that cover the physical characteristics of treatments provided by general practitioners. The proposed A method is equivalent to the DRG type of indicator for inpatients in hospitals.

75. The TF's view is that the number of consultations by diagnosis will meet the requirements for an A method if proper weights are used and quality changes taken into account.

76. *Collective agreements:* If consultations by treatments (or diagnosis) are available, this will under the same conditions as mentioned above, fulfil the requirement for an A method. Nevertheless, this information will in some countries not be available, only the total number of (potential) patients (by age). In this case, an A method will be hard to obtain. The TF thinks that an A method is obtained if annual sample surveys are carried out in order to obtain the number of consultations and specific acts provided and the size and reliability of the sample meet the necessary requirements for an A method. Also quality changes have to be taken into account.

#### B method

77. Regarding the indicators classified under A methods, the direct cost weights might be hardly available and indirect weights will have to be estimated. Since the GP normally get reimbursed according to the number of consultations (including continuation visits) by tests etc carried out etc., an acceptable proxy for the cost weights are the reimbursement fees which normally can be recorded from administrative sources. This cost weight indicator will fulfil the requirement for representativity, sufficient stratification and homogeneity. The total number of consultations by treatment or diagnosis with indirect cost weighting will be classified as a B method.

78. If only partial quality changes are taken into account, the indicator 'total number of consultations by treatment/diagnosis' will only meet the requirement for a B method. The TF thinks that since the indicator is consultations broken down by treatments or diagnosis, the

quality element following from product mix (and new treatments) is taken into account in an appropriate way.

79. The TF thinks that also the total number of consultations (without further breakdown) is acceptable for a B method, the reason being that the consultations will be rather homogeneous. Quality changes arising from service characteristics should be taken into account.

80. *Collective agreements:* In cases where collective agreements concur with services provided to patients who pay the bill themselves (market production) a possibility to receive a B method for non-market production is to deflate output at current prices with the market price index. However, this solution needs a plausibility check by someone familiar with the relevant market.

C methods

81. Input methods are classified as C methods.

#### *4.2.2 Consultations and treatments by medical and surgical specialists*

82. Normally, the consultations and treatments by medical and surgical specialists will be heterogeneous, depending on type of medical/surgical specialists. This will lead to slightly different recommendations than for general practitioners; the type of medical/surgical specialist providing the service has to be taken into account.

83. For specialist consultations and treatment the complete treatment distinction (i.e. first visits) is more important than for services provided by general practitioners, the reason being that the services are more heterogeneous and the nature of the specific treatment often implies that there will be continuation visits. The Task Force recommends that complete treatments should be the a strict requirement for an A method in the case of specialists, this leading to the conclusion that regarding specialists first visits are useful output indicator. Nevertheless, this issue should also be explored by non-TF countries.

A method

84. The number of first visits broken down by type of specialist and type of medical act (treatment), with proper weighing and adjusting for quality changes, will meet the requirements for an A method.

B method

85. As an acceptable indicator for a B method the TF recommends to use number of first visits by specialist. Proper weighing is necessary.

C method

86. All input methods will be classified as C methods. The same will be the case with a method based only on the total number of consultation or first visits.

### **4.3 Dental practice service (CPA 85.13)**

87. Dental practice services include orthodontic services and dental practice service. Orthodontic services include treatment of protruding teeth, crossbite, overbite, including dental surgery even when given in hospitals to inpatients, services in the field of oral surgery and other specialised dental services. Dental practice services include diagnosis and treatment of diseases affecting the patient or aberrations in the cavity of the mouth and services aimed at prevention of dental diseases. These dental services can be delivered in health clinics, such as attached to schools, firms, homes for the aged, as well as in consulting rooms. The services also cover services in the fields of general dentistry, such as routine examinations, preventive dental care, treatment of caries, etc.

88. It should be noted that dental services are mostly market services. However, in all the TF countries there will be a minor part which are non-market. In Norway all children less than 20 as well as old and disabled people living in nursing homes, will be provided public dental services free of charge. What is most commonly available information on dental services are the number of consultations and the number of patients who have received the service during a year.

89. The TF had no time to have a closer look into dental services. TF I says (para 49 in TF I Report): *"When a specific treatment more important than a simple consultation or a simple act is provided (mostly specialists, dentists or paramedics) it is certainly recorded and paid off as a specific medical act. If in that case specific output indicators are available, the requirement for an A method has been fulfilled. If there is only one indicator for the total number of consultations (normal + specific) the requirement for a B method is fulfilled."* The TF II underlines that a method can only be classified as an A method if quality changes are taken into account.

#### A methods

90. In the case of market production there are two possible methods and in the case of non-market production one method (the latter):

1. deflation of the output per dental treatment by a price index (adjusted for quality change) and weighted with output values or costs in t
2. output indicators per treatment adjusted for quality changes and weighted with output values or costs in t-1.

91. The output indicator must cover the complete treatment (medical act). A proxy for the complete treatment is the number of first visits. To meet the requirement regarding representativity and homogeneity the first visits should be broken down by type of treatment. Further, the indicator has to meet the requirement for proper weighting and full quality adjustments to be classified as an A method.

#### B method

92. Number of first visits by type of treatment adjusted for partial quality changes will be classified as a B method.

93. According to TF I also the total number of consultations will fulfil the requirement for a B method. TF II discussed whether, alternatively, the first visit approach should be taken into account. The view is that if a breakdown by type of treatment is impossible, it will be rather meaningless to count only first visits. The problem with using first visits (if the treatment dimension is unavailable) is that rather heterogeneous treatments will be grouped together; i.e. those consisting of only one visit and those requiring a number of continuation visits. The TF thinks that the number of consultations will be more homogeneous than the first visits. Consequently, the TF will support the recommendation from TF I that total number of consultations will fulfil the requirement for a B method.

#### C method

94. Input methods will be classified as C methods.

### **4.4 Social work services with accommodation (CPA 85.31)**

95. Social work services with accommodation cover welfare services delivered through residential institutions to old persons and to physically or mentally handicapped persons including social assistance services involving round-the clock care services. CPA 85.31, in addition, includes services delivered through residential institutions to children and young people (orphanages, homes for children in need of protection etc.) to other clients (homes for single mothers and other social rehabilitation services also for people addicted to drugs and alcohol).

96. The Task Force only discussed this item in connection with welfare institutions for old, physically or mentally handicapped persons, recognising that these services, economically,

constitute the major part of social work services. In the classification of services it is underlined that combined lodging and medical services under the direction of a medical doctor is classified under hospital services (CPA 85.11). In many countries there may be unclear borderlines and a possible overlap between the social work services and hospital services (nursing services).

97. The discussion regarding social work services for old, physically or mentally handicapped persons will be similar to the discussion under section 4.1.4 Nursing services. The complete treatment is equal to the care provided and the care will be homogeneous within an institution, but might be heterogeneous between institutions. Even though the discussion in the TF was related to institutions for old and disabled, the same conclusions can be drawn for the other social work services with accommodation.

A method

98. Occupant days by type of institution (if this breakdown adequately reflects the different levels of care) and adjusted for quality changes will meet the requirement for an A method. Appropriate weights are also necessary to obtain the A method.

B method

99. As B methods the TF recommends occupant days and necessarily adjusted for service-related quality changes. The recommended quality indicator is the ratio staff (man year) per occupant days.

C method

100. All input methods are classified as C methods.

#### **4.5 Other services under CPA section N**

101. As mentioned under the introduction (para. 5) there was not enough time to have a closer look at CPA classes 85.14 Other human health services, 85.20 Veterinary services and 85.32 Social work services without accommodation. These services are including:

- Other human health services (CPA 85.14): services provided by midwives, services provided by nurses, services provided by physiotherapists and other para-medical persons, ambulance services, services provided by laboratories etc.
- Veterinary service (CPA 85.20): services for pets and animals other than pets (animal and veterinary hospital and non-hospital medical, surgical and dental services to animals)
- Social work services without accommodation (CPA 85.32): child day-care services, also for handicapped children and young people, vocational rehabilitation services, day care centre services for old people, etc.

102. All the services mentioned in the previous paragraph, with the exception of child day care services, are rather unimportant compared to hospital services and social work services with accommodation. In addition a major part of these services will be market services. In Norway this will be the case for veterinary services (nearly 80 percent of the output is market), laboratory services, physiotherapists, other para-medical persons (nearly 100 percent of output is market) etc, while ambulance services and vocational rehabilitation services etc will be non-market. For the non-market services under this heading the same techniques as described for the other health and social work services can be used.



## TASK FORCE HEALTH II

### LIST OF DOCUMENTS

TFHealthII/1	Volume measures of non-market services (= document Eurostat B1/CN 398)
TFHealthII/2	Minutes of the meeting of the NAWP on 1-2 February 1999 (= document Eurostat B1/CN406)
TFHealthII/3	Statistics Denmark: Quantity indicators for non-market services
TFHealthII/4	Statistics Finland: Measuring the volume of non-market services; inventory of data on education and health
TFHealthII/5	Australian Bureau of Statistics: Measuring non-market sector output (paper presented at the 1998 OECD National Accounts meeting)
TFHealthII/6	Statistics New Zealand: Measuring output of non-market hospital services (paper presented at the 1998 OECD National Accounts meeting)
TFHealthII/7	Australian Bureau of Statistics: Non-market output - Recent work by the ABS (paper presented at the 1999 OECD National Accounts meeting)
TFHealthII/8	Australian Bureau of Statistics: Excerpts from internal documentation on health services
TFHealthII/9	Eurostat: Overview on the first Task Force's recommendations
TFHealthII/10	Note of the first meeting of the TF Health II
TFHealthII/11	Note of the second meeting of the TF Health II
TFHealthII/12	Note of the third meeting of the TF Health II
TFHealthII/13	D. Collesi (ISTAT): Volume measurement of health services - Institutional arrangements for the provision of health services in Italy
TFHealthII/14	D. Collesi (ISTAT): Output indicator for hospital care: the Italian situation regarding data on the DRG classification and future programs
TFHealthII/15	D. Collesi (ISTAT): Basic health care: A proposal for the deflation of the accredited general medical care services
TFHealthII/16	D. Collesi (ISTAT): Non-market output at constant prices: Methodology and application in the Italian national accounts

(paper presented at the 1999 OECD National Accounts meeting)

TFHealthII/17	Cl. Wagner (INSEE): Volume measures of human health services in national accounts - the present French method and propositions for a new method
TFHealthII/18	A. Brathaug/E. Nørgaard: Norway - volume indicators for health services
TFHealthII/19	J. Eefting/G. Gringhuis/F. Kleima/P. Warns (CBS) Volume measurement of health services - hospital cure and care
TFHealthII/20	CBS: Note on nursing homes and psychiatric hospitals
TFHealthII/21	ONS: Developments in the measurement of general government output (paper presented at the 2000 OECD National Accounts meeting)