

## **ANNEX 6**

### **GROSSING-UP PROCEDURE**

# GROSSING-UP PROCEDURE

There are broadly three areas to be addressed through this survey:

- Single cost estimates of expenditure within a SIC code, e.g., in-house operating costs;
- Percentage estimates for expenditure related to each media, i.e. air, water, land; and
- Total expenditure summaries both across SIC codes and for all UK industry.

The estimation of industry wide costs was achieved through grossing up the sample totals. The grossing-up procedure to derive final population estimates was similar to previous years, ensuring that every cell in the database had a value (0 or above). This was accomplished through the strengthened validation procedures described in Annex 9 and through the use of imputing (infilling of blank or incorrect values). Section 1 describes the method of imputing whilst section 2 describes the method of grossing up cost estimates. The method for analysing expenditure for each media (using percentage allocations) is described under section 3 and total grossed-up environmental expenditure is described under section 4.

## 1. METHODS OF IMPUTING AND MEDIA SPLITS FOR EXTERNAL OPERATING COSTS

Respondents to the questionnaire did not always answer every question or provided answers that were inconsistent. It was, therefore, necessary to develop a process for imputing these missing answers in order that the expenditure totals were always internally consistent. This meant, for example, that the total expenditure for a given category was equal to the sum of the expenditures by media for that category.

The method for ascribing an answer depended upon the type of answer required (percentage or value) and the other data available from both the respondent and within the category. The process is outlined in **Figure 1**.

If more than one value was missing on one questionnaire response the missing values were determined through the imputing procedure so that the most aggregated totals were calculated first, progressing down to the least aggregated. For example: If a respondent answered 'not known' to total environmental operating costs, total external costs and total in-house operating costs then the total environmental operating costs would be calculated and this value split between the other two answers depending upon other responses within the cell. An exception to this is where the respondent answered 'not known' to an item of external operating costs, though provided a total external operating cost summing those items where the costs were known and provided. In these

instances, the total external operating cost was re-calculated with the unknown quantity added and Total Environmental Operating Costs (internal plus external) similarly recalculated.

An ordering of the calculations was needed to ensure that the internal consistency remained with each questionnaire where imputing had taken place.

The media splits for in-house, end-of-pipe and integrated expenditure were simply calculated by multiplying the related total by the relevant percentage split. However, the questions on external operating costs asked for specific figures rather than a total and media percentage splits. Hence it was necessary to adopt a different approach for the calculation of spending by media within external operating costs.

Three categories related directly to the general media splits used for the other totals, these were wastewater, solid waste and soil/ground water. This meant that anything in the category "other" could be classified as air, noise, nature protection or other. To determine what proportion of the external other total to attribute to each media the answer to the question "please specify" was used. The descriptive answers were used to classify "other" expenditure to one or more of the possible media categories. A large number of items were wrongly classified as other rather than waste. Hence it was necessary to include waste as possible media category when splitting up external other spending. For any missing answers values were imputed in the same manner as other percentage splits. Regulatory charges were included within the other media category.

**Figure 1. Process for Calculating Missing Values**

Condition 1	Condition 2	Condition 3	Condition 4	Action 1	Action 2	Value to enter
Answer is n/a						0
Answer is blank or 'not known'	Missing entry is a monetary value (total) *	Value can be derived from other entries by respondent		Derive value		derived value
		Value cannot be derived from other entries by respondent		Calculate 'spend per employee' for category.	Multiply by number of employees for respondent.	calculated value rounded to 2 decimal places
	Missing entry is a percentage **	Related percentages sum to 100%				0
		Related percentages sum to 0%	Total related to percentages is £0			0
			Category contains data	Calculate categories media splits weighted by related totals (rounded to whole percents)		-
				Calculated media splits sum to 100%		calculated values
				Calculated media splits sum to >100%	reduce smallest of values rounded up by 1% (Check again)	calculated values
				Calculated media splits sum to <100%	Increase largest of values rounded down by 1% (Check again)	calculated values
			Category does not contain data	Ascribe equal media splits.		equal values
		Related percentages sum =/ 0 or 100		Use available data from respondent and category to ascribe values.		ascribed values
	Missing entry is a value (subsection of other value) ***			Calculate average percentage attributed to sub-section	Multiply percentage found by main total	calculated value rounded to 2 decimal places

\* Answers for Questions such as 1.1a, 1.2vi, 1.3, 1.4, 2.1a, 2.2a \*\* Answers for Questions such as 1.2b, 2.1c, and 2.2c

\*\*\* Answers for Questions such as 1.2i - v

## 2. METHOD OF GROSSING-UP COST ESTIMATES

Cost estimates, (e.g. responses to Questions 1.1a, 1.3, 1.4, 3.1, 3.2, & 3.3), were grossed-up using numbers of employees. This was the same method used in the 1999 to 2011 surveys and therefore provides a consistent basis for comparison.

Normally responses would be grossed-up using the numbers of businesses. However, because of the range of employee numbers within the grossing-up sample frame (1 to 249 and 250+) the number of employees was used instead (see 1999 report for a more detailed explanation). Each cell of the grossing-up frame was grossed-up separately and the grossed-up totals were added to determine the overall total expenditures. It should be noted that, although the sample frame only considered businesses with greater than ten employees, the population of employees (E) included businesses with fewer than ten employees.

The grossing-up uses Equations 1 and 2:

$$\text{Equation 1 } T = t * M$$

$$\text{Equation 2 } M = E/e$$

where: T = Total grossed-up U.K. expenditure for that cell of the grossing-up frame

t = sum of expenditure for the responding businesses in that cell

E = population of employees in that cell for the U.K.

e = number of employees for the responding businesses in that cell

## 3. METHOD OF ANALYSING EXPENDITURE FOR EACH MEDIA

Questions 1.1b, 2.1c, & 2.2c asked respondents to estimate the percentage split of expenditure between different media, e.g., waste water, air, etc.

The following method was used to calculate the expenditures by environmental media:

The grossing-up used the following sequence of equations:

$$\text{Equation 3 } C_{ai} = x_i * \%_{ai}$$

where:  $C_{ai}$  = cost estimate for media

$x_i$  = cost response by respondent 'i' to earlier part of the question

$\%_{ai}$  = percentage for media 'a' by respondent 'i'

Equation 4

$$t_a = \sum_{i=1}^n C_{ai}$$

where:  $n$  = number of valid responses in cell of grossing-up frame

Equation 5

$$P_a = \frac{t_a}{\sum_a^g t}$$

where:  $P_a$  = estimate of proportion of expenditure on media for cell of grossing-up frame for

each of the seven environmental media a to g

Equation 6

$$T_a = T \times P_a$$

where:  $T_a$  = Estimate of UK expenditure for media 'a' for cell of grossing-up frame

$T$  = Total grossed-up U.K. expenditure, (from Equation 4.5)

(Note that  $T_a$  is directly proportional to  $t_a$ . Hence, an alternative method of grossing-up  $C_{ai}$  independently, and then factoring  $T_a, T_b, \dots, T_g$  so that their grossed-up totals sum to  $T$  would produce the same results as the above simpler method. This is because  $T_a$  would still be directly proportional to  $t_a$  since  $M$  is constant for each media).

#### 4. TOTAL GROSSED-UP ENVIRONMENTAL EXPENDITURE

Total expenditure for a particular question and SIC sector:

Equation 7

$$T = \sum T_b$$

Where:  $T$  = Total grossed-up expenditure for all businesses in the relevant row of the grossing-up frame

$T_b$  = total grossed-up expenditure for businesses in size band 'b' in the row.

Totals for different SIC sectors are obtained by adding the relevant disaggregate totals in each column of the grossing-up frame.

## **5. IDENTIFICATION AND TREATMENT OF OUTLIERS**

Identification: During the data analysis phase, all provisional totals by SIC sector and expenditure type were looked at to see how much each company's total varied from the mean of its respective grossing group. Any company with a total of more than five standard deviations away from the mean value of that group was considered to be an outlier. This strategy identified eight companies that were then formally identified as an outlier for one or more expenditure items.

Treatment: These were then viewed in the context of spend within each individual SIC category against spend in previous years. On the basis of this assessment, none of the outliers were omitted from the grossing procedure.