ESSNET

USE OF ADMINISTRATIVE AND ACCOUNTS DATA IN BUSINESS STATISTICS

WP6
Quality Indicators when using Administrative Data in Statistical Outputs

Deliverable 2011 / 6.2:
Guidance on calculating composite quality indicators for outputs based on administrative data

October, 2012

Contents

Guidance on calculating composite quality indicators for outputs based on administrative data

1. Introduction

Work Package 6 (WP6) of the ESSnet AdminData aims to develop quality indicators for outputs based on administrative data. WP6 has already developed a list of 23 basic quality indicators for this purpose[1]. Whilst all of these indicators have been shown to be useful in assessing the quality of outputs using administrative data, it would be helpful for users to be able to see this information in summarised form. This report describes work to investigate methods for developing composite quality indicators to provide a more general overview of quality.

It was decided that the work would focus on developing separate composite indicators for a range of quality 'themes', based on the dimensions of the ESS quality framework. This project does not include development of a composite indicator to measure overall output quality, since it is widely considered that expressing overall quality as a single number is not likely to be meaningful. Furthermore, producing information on quality for separate dimensions allows for the possibility of examining trade-offs between the different dimensions; an important consideration, as set out in the European Handbook for Quality Reporting[2].

The aim of a composite quality indicator is to provide useful, summarised information to users on the quality of a particular output. To be effective, it is important for the composite indicators developed to reflect user requirements. For this reason, it will be necessary for any specific parameters needed in calculating a composite indicator to be set based on the needs of an output, rather than fixing them as standard across all outputs and statistical organisations. It is important to note that the aim of creating a composite quality indicator is to assist users and not to accommodate any comparison between organisations.

The first step in this work was to group the basic indicators into quality dimensions. In doing this, it was discovered that some of the indicators do not fit readily in the ESS quality dimensions and so extra quality themes were identified to cover all indicators. Following this, appropriate methods to calculate composite indicators for each of those themes were considered with reference to the literature. A general approach has been chosen and developed along with practical examples of its use for each relevant quality theme. These steps are described in the remainder of this report.

---

[1] See http://essnet.admindata.eu/Document/GetFile?objectId=5492

[2] See http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQR_FINAL.pdf

2. Grouping indicators into quality themes

The basic quality indicators developed by WP6 have been grouped into quality themes, based on the ESS quality dimensions and two extra groupings which are required to cover all of the basic indicators. The full list of grouped indicators can be found in Annex 1.

The basic quality indicators fit into the following quality themes:

- Accuracy
- Timeliness and punctuality
- Comparability
- Coherence
- Cost and efficiency
- Use of administrative data

There are two ESS quality dimensions which are not covered by the basic indicators: Accessibility and Clarity, and Relevance. This is because quality with regard to these dimensions is not impacted by whether the outputs are compiled using administrative data or survey data.

The indicators that fit in the themes Use of administrative data, and Cost and efficiency are mostly background information and all present information that is more easily understood separately. Therefore, it is not useful to develop composite indicators for these themes. Composite indicators will be developed for Accuracy, Timeliness and Punctuality, Comparability, and Coherence.

3. Methods for calculating composite indicators

Annex 2 contains a review of existing literature on calculating composite indicators. Two main approaches can be identified. The first approach is to normalise the component indicators in some way and aggregate them using a weighted or unweighted average. The second approach is to model the data in some way to assess the impact of each component indicator on quality.

The second approach includes methods such as Principal Component Analysis (PCA), Factor analysis and Structural equation models. Whilst this second approach is attractive from a theoretical point of view, it is often difficult to implement successfully in practice. For example, Brancato and Simeoni (2008) developed one structural equation model with reasonable results, but noted that the model was unable to properly represent the Accuracy dimension. Even assuming a successful model can be identified for a specific data set, it is unlikely that this model will be suitable for other data or in other organisations. Furthermore, it is likely the model will need to be continually re-specified to remain useful. For these reasons, it is considered that the simpler, first approach is more suitable for developing a generic method to calculate composite indicators for outputs based on administrative data.

For further information, see Annex 2. Note that much of the literature on this topic concentrates on indicators which compare performance across countries or regions. Methods specific to this context are not covered in the literature review.

4. Development of composite quality indicators

4.1 Normalisation of basic quality indicators

The basic quality indicators measure a range of different quality concepts. Where possible, the indicators have deliberately been expressed so that lower quality is reflected by a higher value (since most of the indicators measure errors and so are naturally in this direction). This removes one possible inconsistency, but it remains the case that the various indicators are on different scales. Superficially, it is apparent that many of the indicators are percentages. However, even the indicators that are expressed as percentages are not necessarily directly comparable. For example, a non-response rate of 20% is not of equivalent quality to 20% overcoverage.

A range of options were investigated for normalising the basic indicators, so that they are on the same scale and can be combined more easily. This work was a collaboration between Portugal and the UK. Many of the methods discussed in the literature (see, for example, Nardo et al (2008)) relate only to the situation where indicators are being compared across geographies and so are not appropriate for this purpose. There are two main methods that could be more generally applicable:

- Standardisation - converting the indicators to a common scale by subtracting a mean value for the indicator and dividing by a standard deviation.
- Min-Max – converting the indicators to a common scale by subtracting a minimum value and dividing by the difference between a maximum and minimum value for the indicator.

It is difficult to implement either of these methods to normalise the basic quality indicators, since it is not immediately obvious how to calculate mean, minimum, maximum or standard deviation. It may be possible to compare indicator values over time, but this will not always be practicable. It is therefore necessary to adapt the concept of normalisation. The following formula adapts a typical standardisation method so that it can be applied to quality indicators.

$$\text{Standardised value} = \frac{\text{Indicator value - Reference value}}{\text{Maximum - Minimum}}$$

The reference value in this formula is intended to denote the point at which the value of the quality indicator changes from being acceptable to unacceptable. For example, if a non-response rate of 20% is acceptable, but anything larger is unacceptable then the reference value for non-response rate would be 20%. This means that when the value of the basic indicator exceeds the reference value (which denotes unacceptable quality) the standardised value will be positive. Negative standardised values indicate that the quality is acceptable. It should be noted that this method will only work if it is possible to meaningfully define the reference values.

The maximum and minimum values in the formula are used to transform the different indicators onto the same scale. Many of the basic indicators are defined as percentages and have physical maximum and minimum values of 100% and 0% respectively. However, dividing all percentage indicators by 100 ignores the fact that a particular percentage value does not have the same quality implication for all indicators. To properly standardise the indicators, we need to divide by a quantity that reflects the range of likely values for the indicator.

In order to properly standardise the basic quality indicators, it is necessary to define reference, minimum and maximum values. The next section explores options for setting the reference values. Minimum and maximum values are considered afterwards.

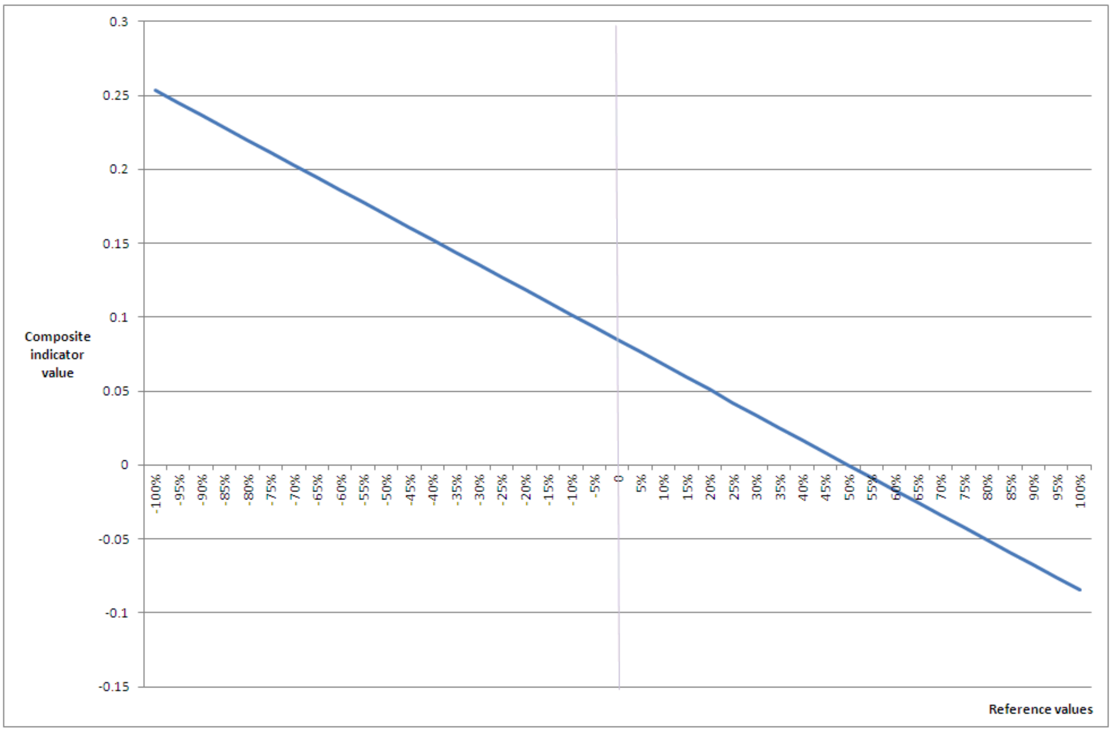
4.2 Setting reference values

The reference value denotes the point at which the value of a particular basic indicator changes from unacceptable to acceptable quality. For some indicators, it may be possible to make an educated guess at where this happens from a theoretical point of view. However, it is important to remember that acceptable quality for an output is driven by the uses of the output and the quality requirements of those uses. Reference values should therefore ideally be developed in consultation with users. In some cases, survey managers might already have a good idea of user needs and be able to set suitable reference values for the indicators. Once set, reference values should be kept constant unless there is a genuine change in user needs. Reference values should never be altered to mask any deterioration in the quality of outputs.

Even with appropriate input from users and survey managers, the setting of reference values is likely to be subjective to some degree. It is important to consider how sensitive the final composite indicator is to the reference value. One way to do this is to calculate the composite indicator using a range of different reference values and examine the effect.

Figure 1 shows values of an example composite indicator for the Accuracy dimension based on standardising each of the basic quality indicators relating to that dimension and calculating the mean of those values. Minimum and maximum values were set based on the likely range of acceptable values for the indicators. The point 0 on the x-axis denotes the value of the composite indicator for the best estimate of the reference values. The other points on the line show what happens to the value of the composite indicator when the reference values are decreased or increased by up to 100%. The line is straight because the reference values were changed uniformly across the different basic indicators.

Figure 1: Sensitivity of reference values for an example composite indicator



In this example, the graph shows us that the composite indicator continues to be positive (implying unacceptable quality) until we increase the reference values by around 50%. If we believe that the reference values genuinely denote the point at which the quality changes from unacceptable to acceptable within a tolerance of 50% then we can be confident in saying that the accuracy of this output is unacceptable. If it is not possible to define the reference values that precisely, then we would have to conclude that it is not possible to make a definitive statement about the quality of the output.

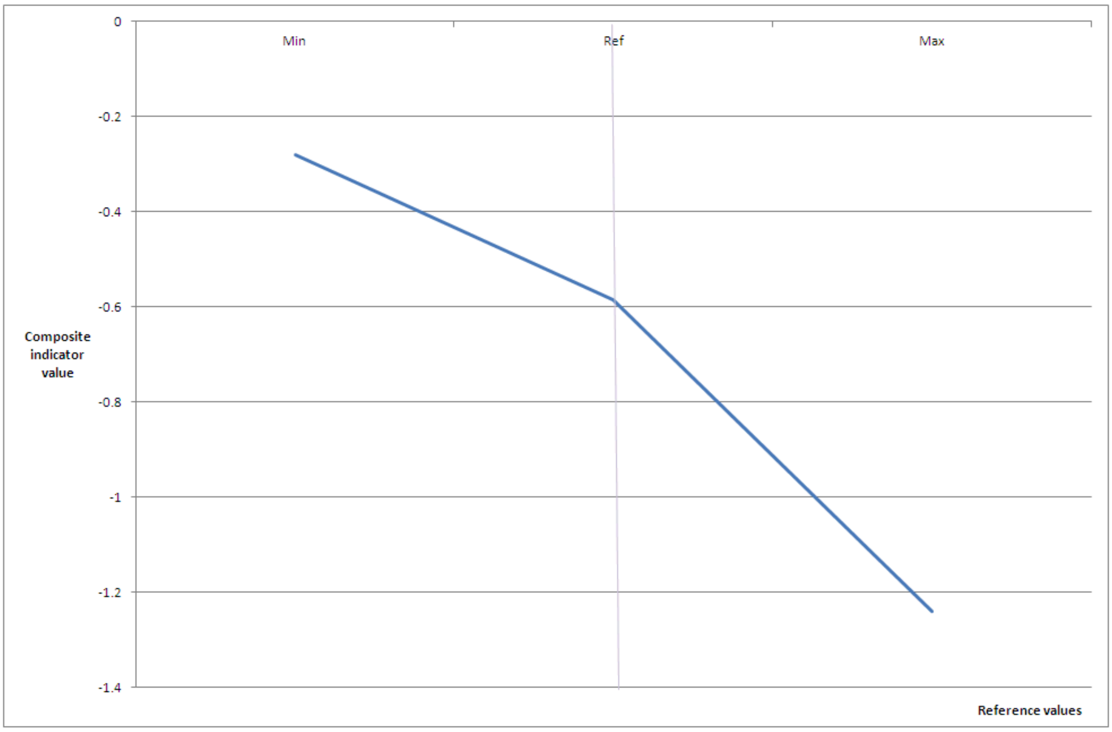
4.3 Setting minimum and maximum values

We can refine both the sensitivity analysis and the standardisation of basic indicators by thinking about the lowest and highest values we would realistically expect for the reference value (the point at which quality changes from being acceptable to unacceptable). The likely range of values may be easier to define than the reference value itself. For example we might be confident that the true reference value for non-response rate is somewhere between 10% and 40%, but only be able to make an educated guess at where in that range it lies.

The lowest and highest values of the reference value will differ between the basic indicators and give an indication of the expected spread of those indicators. For this reason, this range of values can be used in place of an educated guess for the minimum and maximum indicator values in the denominator of the normalisation formula. This allows us to take account of the fact that percentage values have different quality implications for different basic indicators.

We can plot the composite indicators that result from using the minimum and maximum reference values along with the best estimates of those reference values (denoted as “Ref” in the graphs below) to better understand the meaningfulness of the composite indicator. Figure 2 shows an example composite indicator derived from the basic indicators in the Accuracy dimension, using best estimate, minimum and maximum reference values (that is, calculating the composite indicator using each of these three sets of reference values, and joining those points together). In each case, the normalised indicators were combined using a simple mean.

Figure 2: Example composite indicator using minimum, mean and maximum reference values



In this example, the composite indicator has negative values for the whole range of likely reference values. We can therefore confidently say that the output has an acceptable level of accuracy.

4.4 Use of weighted and unweighted versions of the indicators

For most of the basic quality indicators, it is possible to calculate weighted and unweighted versions. For example, an unweighted non-response rate indicates how many businesses have missing values. A weighted non-response rate, using register Turnover as the weight, indicates the proportion of Turnover that is missing.

As part of a Principal Component Analysis of quality indicators, Smith and Weir (2000) found that weighted indicators contain different information to unweighted indicators. It is therefore necessary to decide whether it is more useful to include the weighted or unweighted version of each indicator, or both.

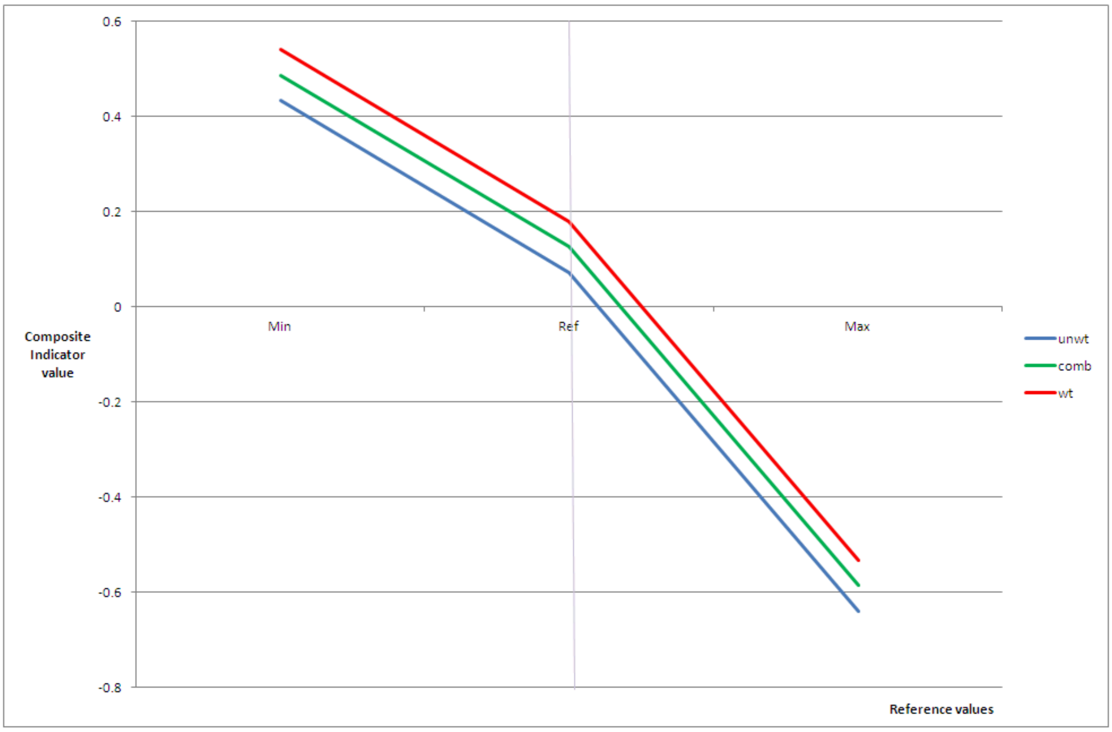
4.5 Combining and weighting the indicators

The final step in calculating a composite indicator is to decide how to combine and weight the basic indicators. This is related to the choice of which indicators should be included. The simplest option is to take a simple mean of all of the indicators – either the unweighted or weighted versions, or both. However, it may be necessary to use a weighted mean to produce a more meaningful composite indicator. Higher weights should be given to any indicators that are more important to the quality needs of the outputs. This could include giving higher weights to weighted versions of the indicators, for example, if they are more important than the unweighted versions (for some or all of the basic indicators). Weighting can also be used to ensure that each aspect of the quality dimension gets equal consideration in the composite indicator, since it may be the case that some of the basic indicators are related to each other.

The choice of appropriate weights needs to be handled carefully. Cecconi et al (2004) prefer using an unweighted average, since it removes the necessity to make a judgement on weights. However, Nardo et al (2008) suggest a practical method to develop suitable weights by asking relevant experts to allocate a budget of 100 points to the set of indicators and derive weights by taking the average of those allocations.

Figure 3 shows an example of the range of values for a composite indicator (using best estimate, minimum and maximum reference values) for a particular output, again using basic indicators from the Accuracy dimension. Three versions of the composite indicator are plotted; one using unweighted versions of the indicators (“unwt”), one using weighted versions of the indicators (“wt”), and one using both unweighted and weighted versions (“comb”). In each case, the indicators are combined using a weighted mean, with higher weights given to indicators which are considered to be more important. The “comb” indicator is an unweighted mean of the “unwt” and “wt” indicators. It would be possible to use a weighted mean if either of the unweighted or weighted indicators were considered more important.

Figure 3: Example composite indicators using three options for the versions of indicators included



For these examples, the lines cross from being positive to negative fairly near to the best estimate ("Ref") values. This suggests that the realised values of the indicators are too close to the reference values to be able to make definitive conclusions about the quality of the output. Note that, in this example, the composite indicator using weighted versions of the indicators is the most clearly positive. We would be slightly more confident concluding that the quality is unacceptable if the weighted indicators were more important to users. However, for any of these examples, quality statements should be presented very carefully, making note of the uncertainty in the composite indicator. It is recommended to avoid using composite indicators when the result is ambiguous and to concentrate on the constituent basic indicators instead.

The final weighting of the individual indicators should be decided based on the importance of different aspects of quality to the users of the output. In the same way, the final choice on whether to use unweighted indicators, weighted indicators or some combination of both should be addressed with reference to users.

4.6 Conclusions

It is possible to derive a composite quality indicator for a quality theme by standardising the values of the basic indicators and combining them using a mean (weighted or unweighted).

The standardisation relies on defining a reference value for each indicator, the point at which quality becomes unacceptable. It is important to consider the quality needs of the output for users in defining these reference values. If the resulting composite indicator is positive, that implies that the level of quality for that dimension is unacceptable for the output. Negative values imply acceptable quality.

The sensitivity of the composite indicator can be tested by defining minimum and maximum values for the reference values and plotting the range of resulting composite indicators between these extremes. If the values are either all positive or all negative, the outcome of the composite indicator will be meaningful. If part of the range of values is positive and part negative then it may not be possible to comment on the quality with complete confidence and care should be taken. In some cases, the only reliable outcome may be to publish the individual basic quality indicators separately (or those that are considered to be of greatest importance to users). Note that the plots are intended to assist producers in deciding whether it is meaningful to denote an aspect of quality as being acceptable or unacceptable for an output. When it is meaningful, the published composite indicator should simply state that the Accuracy, for example, is of an acceptable level based on a range of indicators. The plots themselves are not intended to accompany published outputs.

When combining the standardised indicators, weighting can be used to give the correct emphasis to the indicators, based on user needs for quality and ensuring that no aspects of quality are given disproportionate emphasis in the composite indicator.

The analysis above shows that it can be difficult to derive a meaningful composite indicator even when the only gradation is between acceptable and unacceptable quality. It is therefore not recommended to define composite indicators that attempt to grade the quality in any more detail than this. For example, trying to distinguish between acceptable and good quality will add further complications and is likely to lead to spurious results.

The following sections of this report consider how composite indicators can be developed in practice for the different quality themes: Accuracy, Timeliness and Punctuality, Comparability, and Coherence.

5. Developing a composite indicator for Accuracy

5.1 Choice of indicators

The first step in creating a composite indicator is to decide which of the basic quality indicators are useful or important for the particular quality theme. Table 1 lists the nine basic quality indicators that relate to Accuracy.

Table 1: List of basic quality indicators in the Accuracy dimension

	Indicator
9	Item non-response (% of units with missing values for key variables)
10	Misclassification rate
11	Undercoverage
12	Overcoverage
13	% of units in the admin source for which reference period differs from the required reference period
14	Size of revisions from the different versions of admin data - RMAR (Relative Mean Absolute Revisions)
15	% of units in admin data which fail checks
16	% of units for which data have been adjusted
17	% of imputed values (items) in the admin data

It is important to consider whether all of these indicators are needed for the Accuracy composite indicator and also whether there are any important concepts of Accuracy missing. The formula descriptions in the list of basic indicators note that it is possible to weight eight of these indicators ("Size of revisions" is the only one for which this would not make sense). Therefore, we also need to consider whether weighted, unweighted or both versions of the indicators should be used in the composite indicator.

Some of the Accuracy indicators are related to each other: "% of imputed values (items) in the admin data" is directly related to "% of units for which data have been adjusted" and "Item non-response", since adjusting suspect data and dealing with non-response in administrative data are both commonly done using imputation. The indicator "% of units in admin data which fail checks" is also related to "% of units for which data have been adjusted", since the data adjustments will generally be a consequence of failing checks. Including all four indicators in the composite indicator with the same weighting as the others will give disproportionate emphasis to this aspect of accuracy. Therefore, it will probably be necessary to combine the normalised indicators using a weighted mean to produce a meaningful composite indicator.

5.2 Example construction of composite indicator for Accuracy

Table 2 contains examples of values for unweighted and weighted (where appropriate) values for each of the basic indicators belonging to the Accuracy dimension. The figures are illustrative only, but based on values that could typically be expected, for example when estimating annual Turnover using VAT data. Note that the values for indicators 15 ("% of units in admin data which fail checks") and 16 ("% of units for which data have been adjusted") are identical. This reflects the fact that in many statistical offices it is not possible to re-contact businesses to confirm suspicious values, so that the natural action for businesses which fail checks is to automatically adjust them. However, there are other options for dealing with businesses that fail checks, so it will not always be the case that these values are the same.

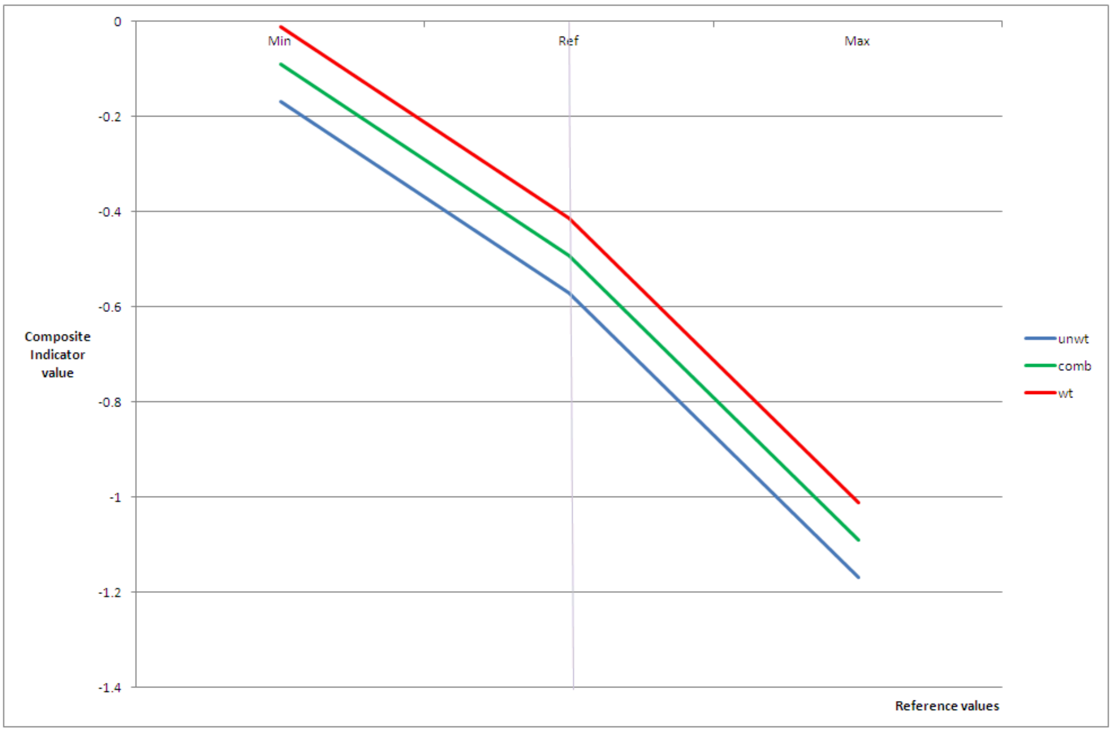
Table 2 also contains reference values (best estimate (Ref), minimum and maximum) for each of the indicators. For the purposes of the example, a set of values have been chosen for purely illustrative purposes. These example reference values should not be used in practice. Reference values should always be set based on consultation with survey managers and users.

Table 2: Example values for basic indicators in the Accuracy dimension

	Indicator	Indicator value (%)		Reference value (%)		
		Unweighted	**Weighted**	**Min**	**Ref**	**Max**
9	Item non-response	15	12	20	25	30
10	Misclassification rate	5	8	2.5	5	15
11	Undercoverage	10	15	20	25	30
12	Overcoverage	5	10	20	25	30
13	% of units with different reference period	20	7	20	30	40
14	Size of revisions	1	*n/a*	0.5	2	5
15	% units failing checks	7	11	2.5	5	10
16	% units with data adjusted	7	11	2.5	5	10
17	% imputed values (items)	22	23	22.5	30	40

Figure 4 shows the values of composite indicators for the range of minimum to maximum reference values using the three choices "unwt", "wt" and "comb". The indicator values are combined using a simple mean. Because there is no weighted version of indicator 14 ("Size of revisions"), the unweighted value is used when compiling the "wt" indicator and is used twice in the "comb" indicator.

Figure 4: Example composite indicators for the Accuracy dimension

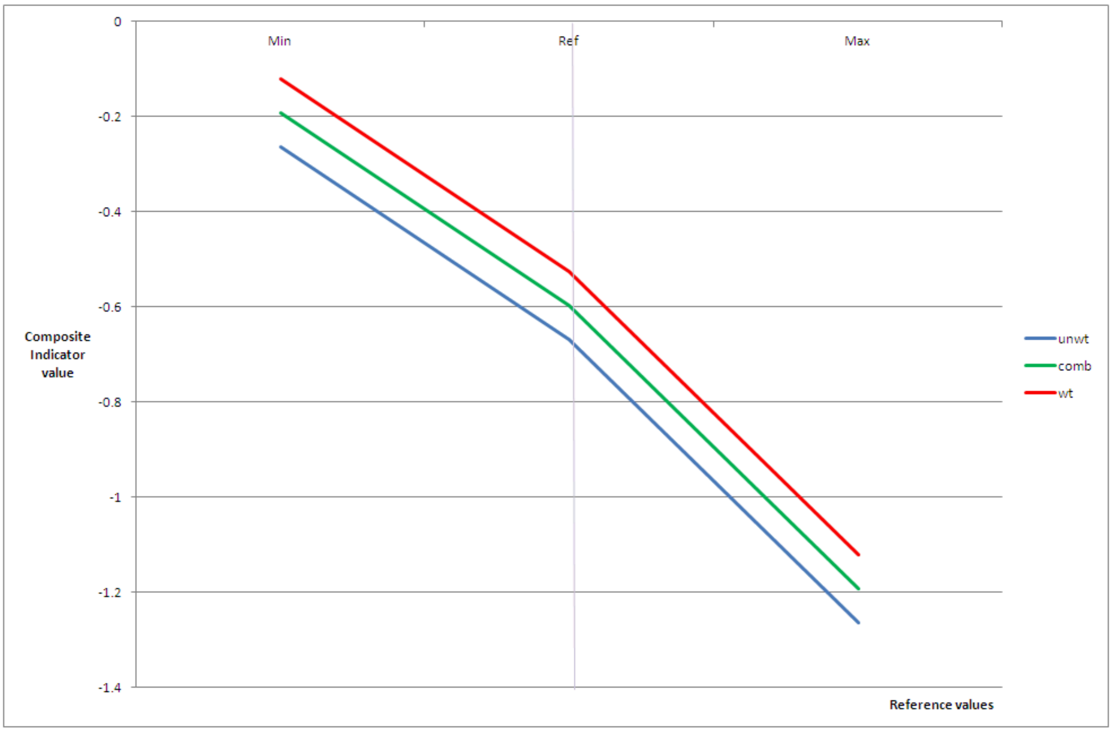


As previously mentioned, some of the basic indicators in the Accuracy dimension are related to each other. To produce a more representative composite indicator, it makes sense to reduce the weights of these indicators. Figure 5 shows composite indicators derived from the same data, but using a weighted mean where indicators 9, 15, 16 and 17 are given half the weight of the other indicators. That is, composite indicators are calculated as:

$$\text{Composite indicator} = \frac{(0.5\times \mathrm{I}_9) + \mathrm{I}_{10} + \mathrm{I}_{11} + \mathrm{I}_{12} + \mathrm{I}_{13} + \mathrm{I}_{14} + (0.5\times \mathrm{I}_{15}) + (0.5\times \mathrm{I}_{16}) + (0.5\times \mathrm{I}_{17})}{7}$$

where I_9 to I_{17} are the normalised values of indicators 9 to 17 respectively.

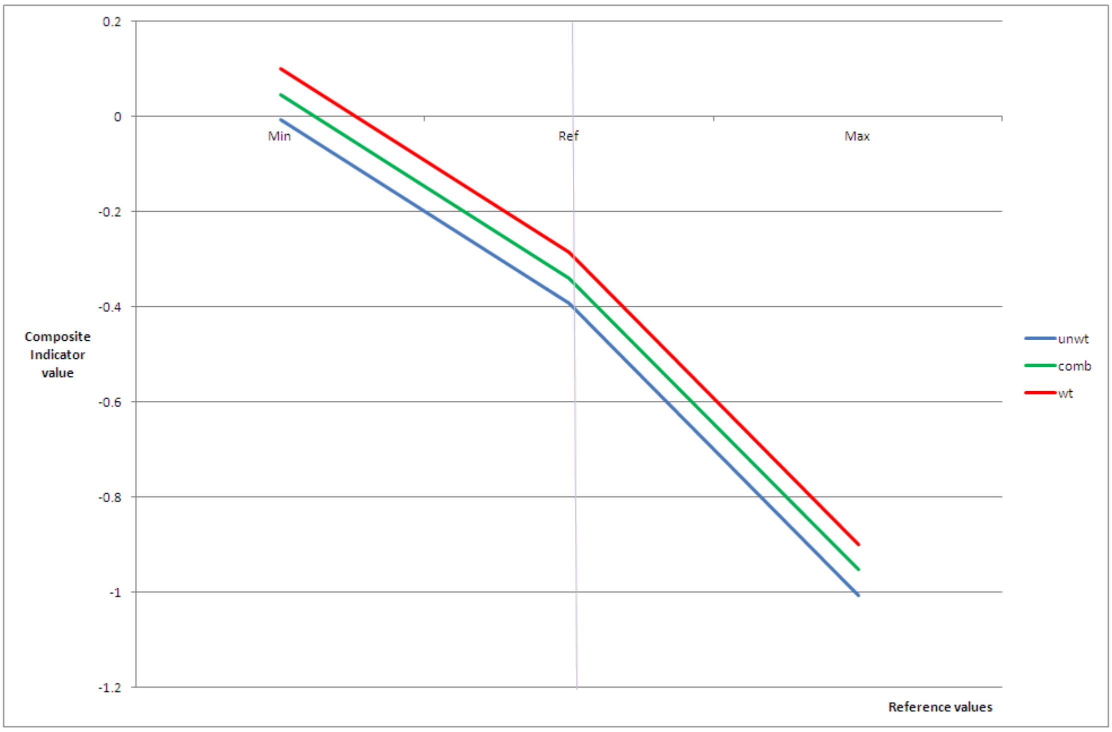
Figure 5: Example composite indicators for the Accuracy dimension, with weighting to reduce impact of related indicators



Weighting can also be used to give a more useful composite indicator if there are some of the basic indicators that are of more importance to users. Figure 6 shows composite indicators using a weighting where "Overcoverage" is given little importance (since it can be dealt with easily if it is identified), "Size of revisions" and "% units failing checks" are given higher importance. "% units with data adjusted" is excluded from the composite indicator (or, equivalently, given zero weight), since the same information is contained in "% units failing checks". The composite indicator is calculated as:

$$\text{Composite indicator} = \frac{I_9 + I_{10} + I_{11} + (0.1 \times I_{12}) + I_{13} + (2 \times I_{14}) + (2 \times I_{15}) + I_{17}}{9.1}$$

Figure 6: Example composite indicators for the Accuracy dimension, with weighting to reflect importance of basic indicators to users



Figures 4 to 6 show that the choice of weights can affect the values of composite indicators. In this example, the outcome changes from being clearly acceptable quality (figures 4 and 5) to having some doubt for the unweighted and combined versions of the composite indicator (figure 6). More extreme cases are of course possible.

These illustrations show how it is possible to construct a composite indicator for Accuracy based on the proposed method. The final choice on weighting and choice of indicators should be made based on consultation with users.

6. Developing a composite indicator for Timeliness and punctuality

Table 3 lists the basic quality indicators relating to Timeliness and punctuality.

Table 3: List of basic quality indicators in the Timeliness and punctuality dimension

Number	Indicator
4	Periodicity (frequency of arrival of the admin data)
18	Delay to accessing / receiving data from admin source

The ultimate decision on whether an output is timely and punctual is when it is published, compared to when the output was due to be published. Bearing this in mind, it is relatively straightforward to set reference values for these two indicators.

“Periodicity” measures the frequency of arrival of administrative data and a natural reference value would therefore be the frequency required by the statistical output.

“Delay to accessing / receiving data from admin source” is calculated as:

$$\frac{\text{Time from the end of reference period to receiving Admin data}}{\text{Time from the end of reference period to publication date}} \times 100\%$$

Quality in relation to this indicator is clearly unacceptable when data are received too late to be able to publish the output to schedule. The change from acceptable to unacceptable quality therefore happens when the time from the end of the reference period to receiving Admin data is the same as the time from the end of the reference period to publication date. In the formula above, this implies a reference value of 100%.

For both of these indicators, it is more difficult to define minimum and maximum values. Because the reference value is so clear cut, it is not meaningful to create upper and lower bounds for its value. One plausible option would be to put the minimum and maximum equal to the reference values described above. This would result in a denominator of zero in the normalisation formula, which illustrates the difficulty.

However, using these indicators it is possible to create a simpler composite indicator describing whether the Timeliness and punctuality is acceptable or not. If the data do not arrive with the desired frequency or on time to be used in the output, then the consequences for the output are serious. For either of these indicators, a failure to meet the minimum requirement would result in an output of unacceptable quality.

A composite indicator for Timeliness and punctuality can therefore be calculated by comparing each of the basic indicators to their reference value. If either of the indicators have unacceptable quality, then the composite indicator should state that the output has unacceptable Timeliness and punctuality. If both indicators are acceptable, then the output can be said to have acceptable Timeliness and punctuality.

Table 4 contains example basic indicator values and accompanying reference values for Timeliness and punctuality. The example is fictitious, but based on the concept of using quarterly administrative data to estimate a quarterly output.

Table 4: Example values for basic indicators in the Timeliness and punctuality dimension

Indicator		Indicator value	Reference value
4	Periodicity	4 times a year	4 times a year
18	Delay to accessing data	106.7% (delay of 32 days)	100% (delay of 30 days)

In this example, we would conclude that the output is of unacceptable Timeliness and punctuality, since the administrative data are not available on time. However, if a method was developed to estimate the output using forecast data from the previous quarter, it might be possible to change the reference value to 140 days, say. This would allow us to create an output of acceptable Timeliness and punctuality but may raise complications in terms of the other dimensions (e.g. accuracy) because of the estimate being based on the model, not the raw data.

7. Developing a composite indicator for Comparability

Table 5 lists the basic quality indicators relating to Comparability.

Table 5: List of basic quality indicators in the Comparability dimension

Number	Indicator
19	Discontinuity in estimate when moving from a survey-based output to admin data-based output

There is only one basic indicator in the Comparability dimension, so it is not necessary to calculate a composite indicator to gain an overall measure of the Comparability. However, it could still be useful to normalise the indicator, by comparison with a reference value, to determine whether the quality is acceptable or unacceptable.

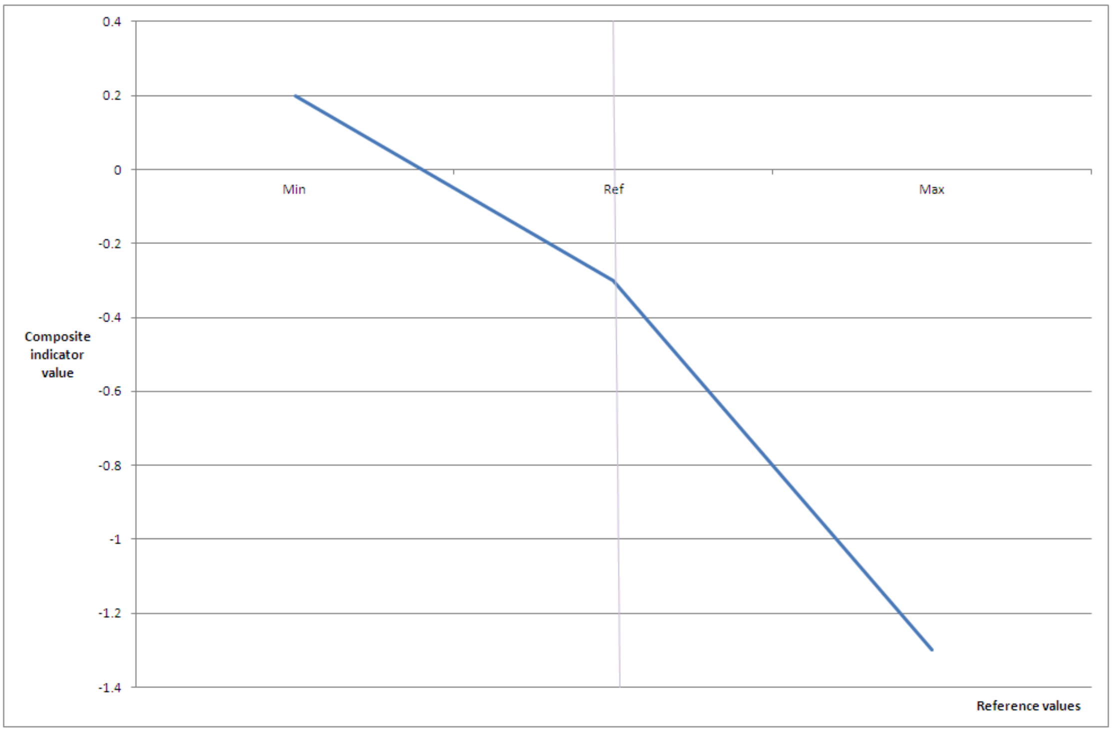
Table 6 contains example indicator and reference values for the Comparability dimension.

Table 6: Example values for basic indicator in the Comparability dimension

Indicator		Indicator value (%)	Reference value (%)		
			Min	Ref	Max
19	Discontinuity	0.7	0.5	1.0	2.0

Figure 7 displays the resulting “composite” indicator for the range of reference values. Since there is only one indicator in the Comparability dimension and it is an indicator which already takes account of survey weights, there is only one version of the indicator to plot.

Figure 7: Example composite indicator for the Comparability dimension



In this example, it is not quite clear that the output is of acceptable quality with respect to Comparability. Since there is only one basic indicator, it is possible to make this deduction directly from Table 6.

8. Developing a composite indicator for Coherence

Table 7 lists the basic quality indicators relating to Coherence.

Table 7: List of basic quality indicators in the Coherence dimension

Number	Indicator
5	% of common units across two or more admin sources
6	% of common units when combining admin and survey data
20	% of consistent items for common variables in more than one source
21	% of relevant units in admin data which have to be adjusted to create statistical units

Indicators 5 ("% of common units across two or more admin sources") and 6 ("% of common units when combining admin and survey data") both give useful background information, but neither directly measures the Coherence. However, there are particular situations where a higher proportion of common units across different sources would lead to higher quality; for example, where the multiple sources are used to validate data. When setting reference values for these indicators, this context should be taken into account. For many outputs, the quality may not be directly impacted by the values of indicators 5 and 6. In these cases, it would be sensible to not include those indicators when compiling the composite indicator (or equivalently to give them a weight of zero).

Note that for indicators 5, 6 and 20 a higher indicator value implies higher quality. When creating composite indicators, we are assuming that higher basic indicator values imply lower quality. It is possible to deal with this by careful choice of the minimum and maximum reference values. The minimum reference values should be larger than the maximum reference values, to reflect the fact that a higher reference value is a tighter restriction. This will result in a negative value for the denominator of the normalised indicator (since the maximum value minus the minimum value will be negative). The negative denominator will have the effect of converting the normalised indicator to the correct scale. For example if the basic indicator value for "% of consistent items for common variables in more than one source" is 50% and the (best estimate) reference value is 60% then the indicator value minus the reference value is -10%. The negative value would imply acceptable quality, despite the fact that the indicator is below the reference value and a higher proportion of consistent values would be expected to give higher quality. Dividing this -10% by a negative denominator has the effect of converting this -10% into a positive normalised value, to reflect the unacceptable quality. By switching the direction of the minimum and maximum reference values, we can appropriately normalise indicators for which a higher value implies higher quality.

Table 8 contains example indicator and reference values for the basic indicators in the Coherence dimension.

Table 8: Example values for basic indicators in the Coherence dimension

Indicator		Indicator value (%)		Reference value (%)		
		Unweighted	Weighted	Min	Ref	Max
5	Common units across sources	48	60	60	50	30
6	Common units combining admin and survey data	71	92	85	80	60
20	Consistent items	50	75	80	70	60
21	Units needing adjusting	32	8	5	10	20

Figures 8 and 9 show resulting composite indicators, both combine the normalised indicators using a simple mean. The composite indicators in Figure 8 use all four basic indicators, whereas those in Figure 9 only use indicators 20 and 21.

Figure 8: Example composite indicator for the Comparability dimension, using all basic indicators

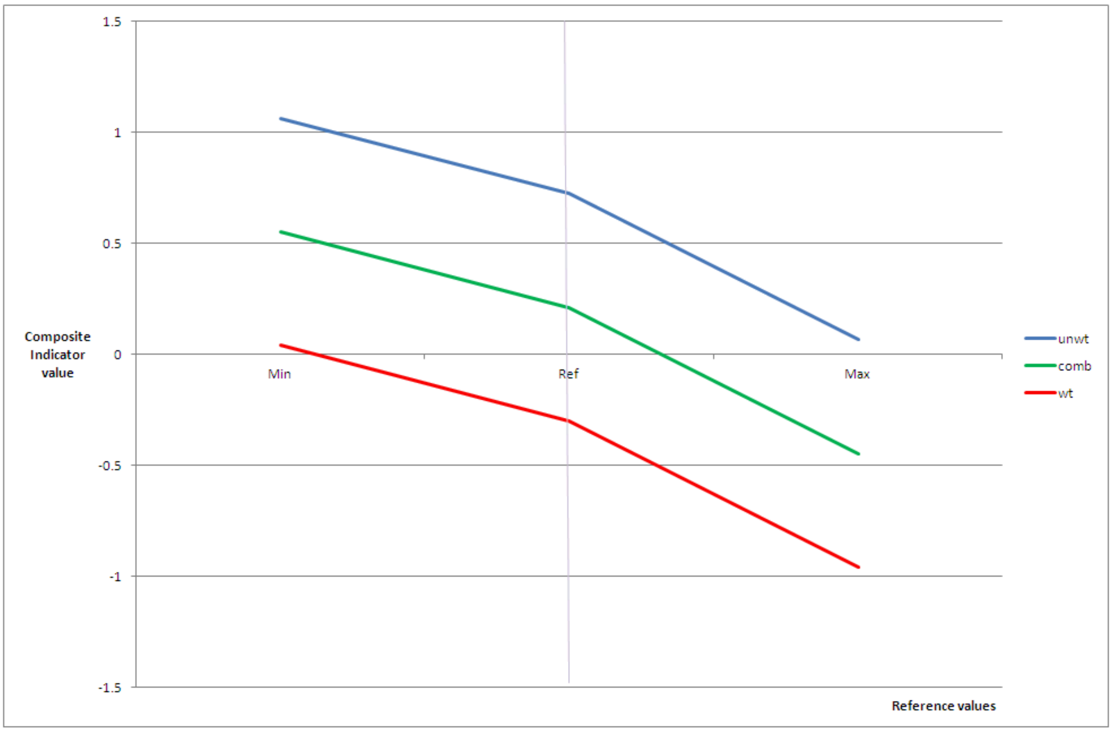
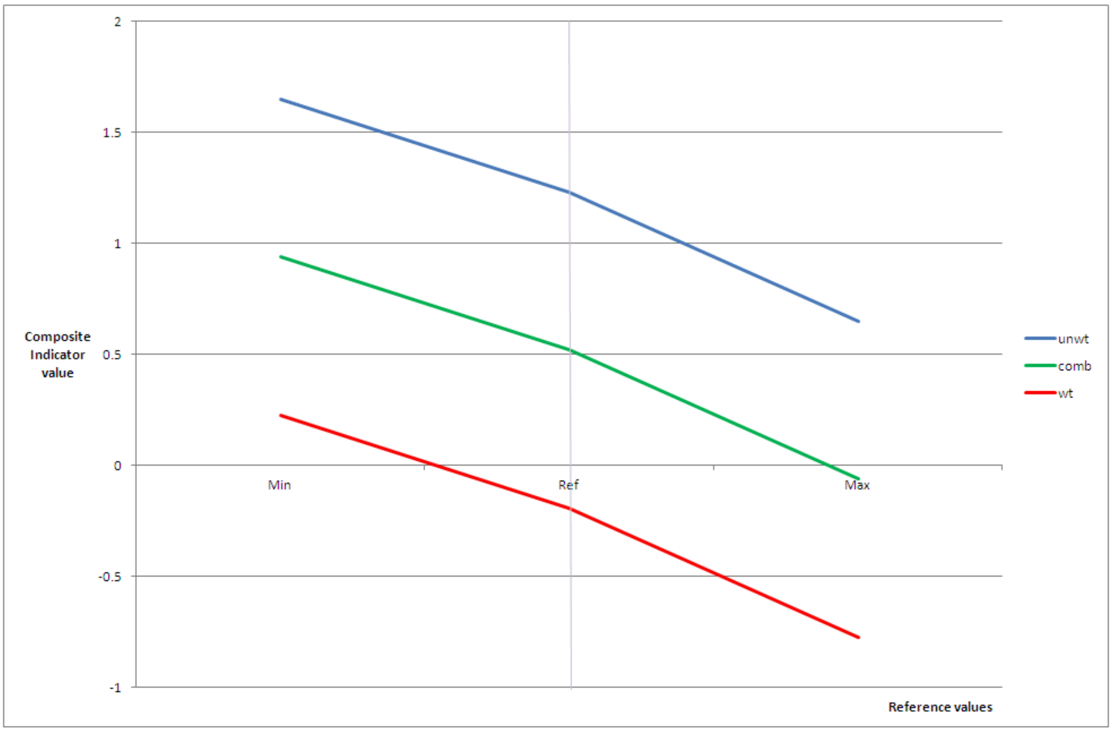


Figure 9: Example composite indicator for the Comparability dimension, using only indicators 20 and 21



There is a large difference between using the unweighted and weighted versions of the indicators when constructing these composite indicators. Using the weighted versions, the Coherence is of acceptable quality for most of the range of reference values. However, using the unweighted versions, the Coherence is clearly of unacceptable quality. As with the other composite indicators, the choice of which versions to use depends on the needs of the users and producers of the data. These graphs demonstrate how important it is to get that choice right. In this example, there is relatively little difference whether including or excluding the background information indicators, 5 and 6. This will not be the case for all outputs.

9. Conclusion

This report has considered methods for calculating composite quality indicators for outputs based on administrative data. Whilst various methods are discussed in the literature, none of them produce easily interpretable results that are relevant for this purpose. Therefore, a simple method has been developed and described to create composite quality indicators for four separate quality dimensions: Accuracy, Timeliness and punctuality, Comparability, and Coherence. The other two quality themes covered in the list of basic quality indicators for outputs based on administrative data are more related to background information and would not benefit from being summarised in composite indicators.

It has been decided not to attempt to produce a single composite indicator covering all aspects of quality. Whilst this is mathematically possible, there is significant doubt that such an indicator would be meaningful. It is important to note that the composite indicators described in this report are intended to assist users in understanding whether the quality attributes of particular outputs are acceptable or unacceptable. The composite indicators have not been developed with the purpose of allowing comparison between countries or outputs and are not designed to enable such comparisons.

This report gives details of the recommended method for calculating composite quality indicators and examples for each of the four quality dimensions covered. In every case, the setting of parameters for the composite indicators should be based on user requirements for the quality of the particular output.

10. References

Brancato G. and Simeoni G. “Modelling Survey Quality by Structural Equation Models”. Proceedings of Q2008 European Conference on Quality in Survey Statistics, Rome, July 2008: Web.

Cecconi C., Polidoro F. and Ricci R. “Indicators to define a territorial quality profile for the Italian consumer price survey”. Proceedings of Q2004 European Conference on Quality in Survey Statistics, Mainz, May 2004: CD-ROM.

Munda G. and Nardo M. “Weighting and Aggregation for Composite Indictors: A Non-compensatory Approach”. Proceedings of Q2006 European Conference on Quality in Survey Statistics, Cardiff, 2006: Web.

Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffman A. and Giovannini E. “Handbook on constructing composite indicators: methodology and user guide”, OECD (2008): Web.

Smith P. and Weir P. "Characterisation of quality in sample surveys using principal components analysis". Proceedings of UNECE Work session on Statistical Data Editing, Cardiff, October 2000: Web.

Annex 1: Grouping of basic quality indicators into quality themes

The tables below list the basic quality indicators matched to each of the quality themes. For reference, the tables include the indicator numbers as shown in the WP6 list of quality indicators[3].

Accuracy

Number	Indicator
9	Item non-response (% of units with missing values for key variables)
10	Misclassification rate
11	Undercoverage
12	Overcoverage
13	% of units in the admin source for which reference period differs from the required reference period
14	Size of revisions from the different versions of admin data – RMAR (Relative Mean Absolute Revisions)
15	% of units in admin data which fail checks
16	% of units for which data have been adjusted
17	% of imputed values (items) in the admin data

Timeliness and punctuality

Number	Indicator
4	Periodicity (frequency of arrival of the admin data)
18	Delay to accessing / receiving data from admin source

Comparability

Number	Indicator
19	Discontinuity in estimate when moving from a survey-based output to admin data-based output

Coherence

Number	Indicator
5	% of common units across two or more admin sources
6	% of common units when combining admin and survey data
20	% of consistent items for common variables in more than one source

[3] See: http://essnet.admindata.eu/WikiEntity?objectId=5452

21	% of relevant units in admin data which have to be adjusted to create statistical units

Cost and efficiency

Number	Indicator
7	% of items obtained from admin source and also collected by survey
8	% reduction of sample size when moving from survey to admin data
22	Cost of converting admin data to statistical data
23	Efficiency gain in using admin data

Use of administrative data

Number	Indicator
1	Number of admin sources used
2	% of items obtained exclusively from admin data
3	% of required variables derived from admin data that are used as a proxy

Annex 2: Literature review on methods for developing composite indicators

Prepared by Carys Davies, UK.

Brancato G. and Simeoni G. "Modelling Survey Quality by Structural Equation Models". Proceedings of Q2008 European Conference on Quality in Survey Statistics, Rome, July 2008: Web
http://q2008.istat.it/sessions/paper/09Brancato.pdf

This paper investigates the capacity of standard quality indicators to reflect quality components and overall quality, using structural equation models. The paper applies confirmatory factor analysis first-order and second-order models.
Structural equation models provide measures of the impact of each manifest variable (e.g. quality indicators) on the relative latent factor (e.g. quality or quality components) as well as measures of reliability, such as the Squared Multiple Correlation.

The paper evaluates the goodness of fit of the models using the Santorra-Bentler scaled X^2 statistic, instead of the standard X^2 statistic, since the standard X^2 statistic tends to be erroneously too high in the case of non-normality. In cases of unfavourable indicators of fit, inspection of modification indices can help guide model re-specification.

Section 4 presents theoretical structural equation models. The paper evaluates overall quality as a second order latent factor, where no relationships among quality components are assumed. Two different theoretical first order models are also considered. The first model evaluates quality components as latent factors, where correlations between quality components can be assumed. The second model considers quality as a general latent dimension, which derives from all quality indicators; no quality components are included in the latent structure.

The three theoretical models described in the paper were tested with real data. The models were then evaluated using the Goodness of fit statistics and Squared multiple correlations, to identify the best measurements of the common factor and loadings, to evaluate relationships in the model. The analysis showed that the second-order model did not converge and the simple first-order quality model did not produce interpretable results. The more reasonable model was the first order latent factor model on quality components. However, this model was not able to represent more complex quality components, such as Accuracy.

Cecconi C., Polidoro F. and Ricci R. "Indicators to define a territorial quality profile for the Italian consumer price survey". Proceedings of Q2004 European Conference on Quality in Survey Statistics, Mainz, May 2004: CD-ROM.

This paper details a methodological approach to synthesising basic indicators in order to compare territorial data collection quality, for the Italian consumer price survey. Section 4 examines four main standardisation methods. Standardising the basic indicators helps to eliminate the influence of the unit of measure, making them more comparable.

The main standardisation methods which were evaluated are:

Method 1 – the ratio between the indicators and the mean of the series

Method 2 – the ratio between the indicators and the maximum of the series

Method 3 – the ratio between the differences of the indicators with respect to the average of the distribution and the standard deviation

Method 4 – the ratio between the indicators with respect to the minimum of the distribution and its range

Method 2 was chosen for the analysis as it offers easy interpretation of results since the range varies between 0 and 1 or 0 and 100. The method also provides the possibility to evaluate the classification of the areas in cardinal and ordinal views.

Of particular interest is Section 5, which details the synthesis of the basic indicators. Since the basic indicators have been normalised and standardised they can be grouped. A non-weighted average was preferred to group the indicators since a weighted average introduces a judgemental criterion in selecting the system of weights. Due to the limited number of basic indicators, a geometric mean was used to calculate the synthetic indicators. Whereas, an arithmetic mean was used to group the indicators for regions and macro areas. The synthetic measures were transformed into spatial indices in order to rank and compare chief towns, regions and macro areas.

Munda G. and Nardo M. "Weighting and Aggregation for Composite Indictors: A Non-compensatory Approach". Proceedings of Q2006 European Conference on Quality in Survey Statistics, Cardiff, 2006: Web http://www.ons.gov.uk/ons/media-centre/events/past-events/q2006---european-conference-on-quality-in-survey-statistics-24-26-april-2006/agenda/index.html

This paper evaluates the consistency between the mathematical aggregation rule, used to construct composite indicators and the meaning of weights.

Section 2 formally proves that equal importance is incompatible with linear aggregation; since in a linear aggregation weights have the meaning of a trade-off ratio.

The paper states that when using a linear aggregation rule, the only method which computes weights as scaling constants, with no ambiguous interpretation, is the trade-off method. Consider two countries differing only for the scores of two variables. The problem is then to adjust one of the scores for one of the countries so the two countries become indifferent. In order to compute N weights as trade-offs, it is necessary to assess N-1 equivalence relations. However, operationally this method is very complex. The assumption that the variable scores are measured on an interval or ratio scale of measurement must always hold. However, this is rarely the case in practice.

It is concluded that whenever weights have the meaning of importance coefficients, it is essential to use non-compensatory aggregation rules to construct composite indicators.

Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffman A. and Giovannini E. "Handbook on constructing composite indicators: methodology and user guide", OECD (2008): Web
http://www.oecd.org/std/42495745.pdf

This handbook provides a guide on constructing and using composite indicators, with a focus on composite indicators which compare and rank countries' performances.

Part 1 focuses on methodology for constructing composite indicators. Of particular interest are Sections 1.5 and 1.6 which detail normalisation, weighting and aggregation methods. Section 1.5 details nine different normalisation methods and provides formulas in table 3. Some of the methods included in this section are; standardisation, min-max and distance to reference.

- **Standardisation** converts indicators to a common scale, with a mean of zero and standard deviation of one.
- **Min-max** normalises indicators to have an identical range, by subtracting the minimum value and dividing by the range of the indicator values.
- **Distance to reference** measures the relative position of a given indicator to a reference point i.e. a target or benchmark.

Of the nine methods described, some may only be suitable for composite indicators which compare/rank countries' performances.

Section 1.6 presents methods for weighting and aggregation, including a table detailing compatibility of aggregation and weighting methods. This section also briefly describes some of the pros and cons of the methods. Further details and practical applications are given in Part 2, Step 6. The paper mostly focuses on the weighting and aggregation methods in terms of composite indicators which compare countries' performance. However, the methodology for some of these methods could be applicable to other types of indicators.

- For **principle components** or **factor analysis** weights are only introduced to correct for overlapping information between correlated indicators, they are not used to measure theoretical importance. If there is no correlation, weights cannot be estimated with this method.
- In the **unobserved components model**, individual indicators are assumed to depend on an unobserved variable plus an error term. The weight obtained is set to minimise the error and depends on the variance of an indicator, say q and the sum of the variances of all other indicators including q. This method resembles regression analysis.
- For the **budget allocation process**, experts allocate a 'budget' of 100 points to a set of indicators. The weights are calculated as the average budgets.
- Weights for the **analytic hierarchy process** represent the trade-off across indicators. The process compares pairs of indicators and assigns a preference. The relative weights of the individual indicators are calculated using an eigenvector.
- **Conjoint analysis** asks for an evaluation of a set of alternative scenarios e.g. a given set of values for the individual indicators. The preference is then decomposed. A preference function is then estimated using the information emerging from the different scenarios. The derivatives with respect to the individual indicators of the preference function are used as weights.

The aggregation methods discussed in Part 2, Step 6 are geometric methods, Non-compensatory multi-criteria approach and additive methods; the difference between the number of indicators above and below a threshold (around the mean), summation of weighted and normalised indicators. More information on the non-compensatory multi-criteria approach can be found in Munda and Nardo (2006).

Part 2, Step 4, looks at multivariate analysis techniques. It is noted that the methods are mostly for data expressed in an interval or ratio scale. However, some of the methods are suitable for ordinal data, for example, principle components analysis. Four main methods are considered, including; principal components analysis, factor analysis, Cronbach coefficient alpha and cluster analysis, as well as a few others.

- **Principal components analysis** aims to explain the variance of observed data through a few linear combinations of the original data.
- **Factor analysis** is similar to principal components analysis. The aim of the method is to describe a set of variables in terms of a smaller number of factors and to highlight the relationships between variables.
- The **Cronbach coefficient alpha** (c-alpha) assesses how well a set of items (individual indicators) measures a single uni-dimensional object (e.g. attitude, phenomenon). C-alpha is a coefficient of reliability based on the correlation between individual indicators.
- **Cluster analysis** uses algorithms to group items (individual indicators) into clusters, where items in the same cluster are more similar to each other than to those in other clusters.

Polidoro F., Ricci R. and Sgamba A.M. "The relationship between Data Quality and Quality Profile of the Process of Territorial Data Collection in Italian Consumer Price Survey". Proceedings of Q2006 European Conference on Quality in Survey Statistics, Cardiff, October 2006: Web
http://www.ons.gov.uk/ons/media-centre/events/past-events/q2006---european-conference-on-quality-in-survey-statistics-24-26-april-2006/agenda/index.html

The methodology discussed in this paper expands on the methods detailed in Cecconi et al (2004). The paper details the methodology used to synthesise the indicators for sample coverage, data collection infrastructure and micro data accuracy as well as creating an overall synthetic indicator.

Section 3.2 provides the methodology for standardising and synthesising the basic indicators. The standardisation method detailed in the paper is the one which was chosen in Cecconi et al (2004), the ratio between the indicator and the maximum value. In addition this paper has developed some mathematical notation for the chosen method.

This paper examines the methods used for synthesising the basic indicators in more detail than in Cecconi et al (2004) and also provides notation and formulas. Firstly the basic indicators are grouped by town, for each component e.g. sample coverage and then for all the basic indicators (overall), using a geometric mean. However, regional and geographic synthetic indicators are calculated using a weighted arithmetic mean. Again, these are calculated for each component and then for all the basic indicators.

Smith P. and Weir P. "Characterisation of quality in sample surveys using principal components analysis". Proceedings of UNECE Work

session on Statistical Data Editing, Cardiff, October 2000: Web
http://www.unece.org/fileadmin/DAM/stats/documents/2000/10/sde/4.e.pdf

This paper describes how to obtain some overall measure of quality by considering quality as a multivariate measure for any dataset, where each quality indicator represents one dimension of quality. This is an alternative approach to evaluating the total survey error, since total survey error evaluates quality in terms of overall accuracy but is very costly.

The paper focuses on the use of principal components analysis to find the measures which best capture the underlying variation in the data quality measures. The analysis is used to try and obtain a small number of indicators which provide the most data quality information, in order to make the assessment of data quality more straight forward.

Variables from the UK Monthly Inquiry into the distribution and Services Sector survey were used for the analysis. A relatively wide-ranging set of indicators were included covering sampling, response rates and data editing. The indicators were also calculated by stratum. The method was also applied to data from the U.S. Energy Information Administration's Annual Fuel Oil and Kerosene Sales report, using the same indicators where possible.

Before principal components analysis can be performed, the variables need to be standardised by subtracting the mean and dividing by the standard deviation. This removes the variability in the measures.

The results detail the proportions of variation in the data, explained by the principle components and also provide the loadings (coefficients to derive the principle components) for the first five principle components. The larger coefficients highlight which variables are most important, in each principle component.

The paper concludes that, for this set of indicators:

- Most of the variation is explained by response rates.
- Weighted indicators contain different information to unweighted indicators.
- Some of the related indicators (e.g. sampling fraction and sampling errors) contain very similar information.