



Reference Cost Data Quality

A Final Report for Monitor

23 April 2014

Contents

Important Notice from Deloitte	3
Executive Summary	4
1 Introduction	6
1.1 Background	6
1.2 Scope of this report	7
1.3 Report structure	7
2 Approach	9
2.1 Overall data quality assessment metrics	9
2.2 Screening stages	10
2.3 Operations	12
3 Outputs	14
3.1 Overall data quality assessment metrics	14
3.2 Screening stages	16
4 Recommendations	22
4.1 Approach to pre-processing steps	22
4.2 Proposed option	23
4.3 Next steps	25
Appendices	27
Appendix A Data	28
Appendix B Detailed statistics	29
Appendix C Original long list of screening stages	31
Appendix D Cost allocation metrics	36
Appendix E Analysis of spell-based reference costs	38
Appendix F Comparability of reference costs and HES data	40
Appendix G Illogical relativities	42
Appendix H Outliers methodology	45

Important Notice from Deloitte

This final report (the “Final Report”) has been prepared by Deloitte LLP (“Deloitte”) for Monitor in accordance with the contract with them dated 13 February 2014 (“the Contract”) and on the basis of the scope and limitations set out below.

The Final Report has been prepared solely for the purposes of providing the review under Work-stream 1, as set out in the Contract. It should not be used for any other purpose or in any other context, and Deloitte accepts no responsibility for its use in either regard.

The Final Report is provided exclusively for Monitor’s use under the terms of the Contract. No party other than Monitor is entitled to rely on the Final Report for any purpose whatsoever and Deloitte accepts no responsibility or liability or duty of care to any party other than Monitor in respect of the Final Report or any of its contents.

The information contained in the Final Report has been obtained from Monitor and third party sources that are clearly referenced in the appropriate sections of the Final Report. Deloitte has neither sought to corroborate this information nor to review its overall reasonableness. Further, any results from the analysis contained in the Final Report are reliant on the information available at the time of writing the Final Report and should not be relied upon in subsequent periods.

Accordingly, no representation or warranty, express or implied, is given and no responsibility or liability is or will be accepted by or on behalf of Deloitte or by any of its partners, employees or agents or any other person as to the accuracy, completeness or correctness of the information contained in this document or any oral information made available and any such liability is expressly disclaimed.

All copyright and other proprietary rights in the Report remain the property of Deloitte LLP and any rights not expressly granted in these terms or in the Contract are reserved.

This Report and its contents do not constitute financial or other professional advice, and specific advice should be sought about your specific circumstances. In particular, the Report does not constitute a recommendation or endorsement by Deloitte to invest or participate in, exit, or otherwise use any of the markets or companies referred to in it. To the fullest extent possible, both Deloitte and Monitor disclaim any liability arising out of the use (or non-use) of the Report and its contents, including any action or decision taken as a result of such use (or non-use).

Executive Summary

As part of its role as Sector Regulator, Monitor is responsible for the National Tariff Payment System, along with NHS England. This includes developing the Tariff Model which will be used to calculate the 2015/16 National Tariff and beyond. For the Tariff Model to be fit for purpose, it is necessary to ensure that the input data, including reference costs (RCs), is as reliable, consistent and relevant as possible. As such, Monitor has engaged Deloitte to assess the quality and integrity of the reference cost data relied upon in the Tariff Model. Specifically, this report focuses on the quality of reference cost data and the development of potential pre-processing steps that may be applied prior to inputting reference costs into the Tariff Model.

It is understood that currently, the only pre-processing step applied to reference costs is the removal of unit costs that are outliers, defined as either 20 times above or below the average for each healthcare resource group ('HRG') included in the RC database. As such, this study marks an initial investigation into whether it is possible to further improve RC data quality through enhanced pre-processing steps. It is anticipated that Monitor will continue to work in this area in the future.

This study is based on an analysis of reference cost data for three years (2010/11, 2011/12 and 2012/13), adjusted for the Market Forces Factor (MFF). The approach employed consists of four steps:

- **Overall data quality (1).** The first step is to assess the overall quality of the reference cost data. This is conducted by testing the RC dataset for normality, bimodality and volatility.
- **Screening stages.** The second step is to run a number of prioritised screening stages that have been agreed with Monitor, to identify providers and data points that are likely to exhibit poor data quality.
- **Operations.** The third step is to develop a set of high-level operations (pre-processing steps) which are used to remove certain observations or providers from the reference cost input data.
- **Overall data quality (2).** The final step is to assess the overall quality of the 'cleansed' reference cost input data. This enables a 'before and after' comparison and illustrates the impact of the proposed pre-processing steps on data quality.

An assessment of raw reference cost data shows that approximately 50% of unit costs are normally distributed and not bimodal. Additionally, it is found that around 15-25% of average unit costs exhibit year on year volatility of more than 20%.

Based on the approach outlined above, the proposed option (that is, the preferred pre-processing steps) to improve the quality of reference cost data input into the Tariff Model includes:

- Removing outliers from the raw reference cost data set based on the "maximum normed residual test";

- Removing providers with at least five unit cost submissions below £5 and at least 10 unit cost submissions above £50,000 (subject to an average unit cost check);
- Removing providers submitting RCs which are more than 50% lower than the national average for more than 25% of the HRGs submitted and who at the same time also submit RCs which are 50% higher than the national average for more than 25% of the HRGs submitted;
- Removing providers who report reference costs that include more than 75% duplicate costs across HRGs and departments; and
- Removing providers submitting reference costs containing more than 15% illogical relativities.

The proposed option set out above results in the removal of 15 providers (5.6%) in 2010/11, 7 providers (3.7%) in 2011/12 and 14 providers (7.2%) in 2012/13.

Overall data quality improves under the proposed option, when compared to the baseline scenario. Specifically:

- Volatility has reduced slightly in each year under the proposed option, in particular the number of volatile HRGs has fallen by between 1% and 2% respectively; and
- The number of HRG unit costs that are normally distributed and not bimodal has increased by 11 percentage points, from 53% to 64%.

The impact of the proposed option on weighted average unit costs is also considered. This gives an indication of the potential impact of the proposed option on tariff prices. For 2010/11, 2011/12 and 2012/13, 6.7%, 5.5% and 10.5%, average unit costs respectively have been affected by more than 10% as a result of implementing the proposed option.

In terms of next steps, prior to implementing the proposed option, it is recommended that Monitor:

- Tests the impact of the proposed option on the national tariff by running the input data through the Tariff Model;
- Tests the impact of the proposed option on provider sustainability; and
- Consults with the sector, giving stakeholders an opportunity to provide feedback on the proposed option.

1 Introduction

Monitor has new duties as the sector regulator for health services in England under the Health and Social Care Act ('HSCA'). Monitor's remit is to:

- Set prices for NHS-funded care (in partnership with NHS England);
- License providers;
- Enable integrated care;
- Safeguard choice and prevent anti-competitive behaviour; and
- Support the continuity of services should a provider encounter financial difficulties.

Monitor is now responsible for the National Tariff Payment System, along with NHS England, taking over from the Department of Health ('DH'). This includes developing the Tariff Model which will be used to calculate prices for the 2015/16 national tariff and beyond.

For the Tariff Model to be fit for purpose, it is necessary to ensure that the input data, including reference costs, is as reliable, consistent and relevant as possible. As such, Monitor has engaged Deloitte to assess the quality of reference cost data as an input in to the Tariff Model.

1.1 Background

Reference costs are one of the key inputs into the Tariff Model. This is because national tariff prices are based on the average of reference costs. The quality of reference costs affects the extent to which national tariff prices reflect the true underlying cost of delivering health care services. As such, it is critical that reference costs are as robust as possible. Robust reference costs play an important role in ensuring that prices place the appropriate incentives on providers and support the delivery of high quality patient care.

Submitted reference costs could vary across providers due to a number of factors, including:

- Differences in efficiency across providers;
- Differences in case-mix across providers;
- Differences in cost allocation methodologies across providers; and
- Differences in data quality across providers.

It is important to understand the drivers of such variations and distinguish between genuine variation in the cost base and variation driven by poor data quality.

This review of reference cost data is particularly important in light of previous studies that have highlighted significant problems related to the quality of reference costs:

- **An evaluation of the reimbursement system for NHS-funded care.** Monitor commissioned the review into the Payment by Results ('PbR') system, which identified a number of issues with the quality of reference cost data. For example, it found that in 2009/10, one in eight NHS providers submitted reference costs containing material errors; and
- **DH discussion papers.** DH has published a number of discussion papers on the volatility of reference cost submissions. They note, for example, that 40% of average unit costs exhibit year-on-year change exceeding 10%. Volatility in reference costs could generate volatile tariff prices, which could have a negative impact on financial planning for both providers and commissioners.

This review builds on the issues identified in previous work and develops an analytical framework for assessing the quality of reference cost data. Additionally, it puts forward recommendations to address some of the identified issues. Based on conversations with NHSE and Monitor, it is understood that currently, the only pre-processing step applied to reference costs is the removal of unit costs that are outliers, defined as unit costs reported by a provider that are either 20 times above or below the average for each healthcare resource group ('HRG').

1.2 Scope of this report

This report presents:

- The approach undertaken to assess the quality of reference cost data;
- Key findings on the quality of reference cost data; and
- Recommendations around pre-processing steps that can be applied by Monitor to improve the quality of reference cost input data for the 2015/16 national tariff.

It is important to note that this is the first attempt to apply systematic pre-processing steps to reference cost data. Given the limited precedent in this area, the aim of this project is to make a first step on a longer journey towards improved data quality. The initial aim is to identify and agree with the sector a set of screening stages that can be applied to raw reference cost data in order to remove providers or observations that exhibit poor data quality. Additionally, the aim is to create a framework for assessing data quality, which could help improve the quality of reference cost returns; it is intended that the recommendations put forward in this report, with Monitor approval, will be subject to consultation with the sector. Given the time frame for this project, the focus is on identifying the most material data quality issues; Monitor may consider undertaking further work in this area in future.

The mapping of reference costs between years has also been considered as part of this project. The methodology and key findings from this work-stream are presented in a separate report.

1.3 Report structure

The remainder of this report is set out as follows:

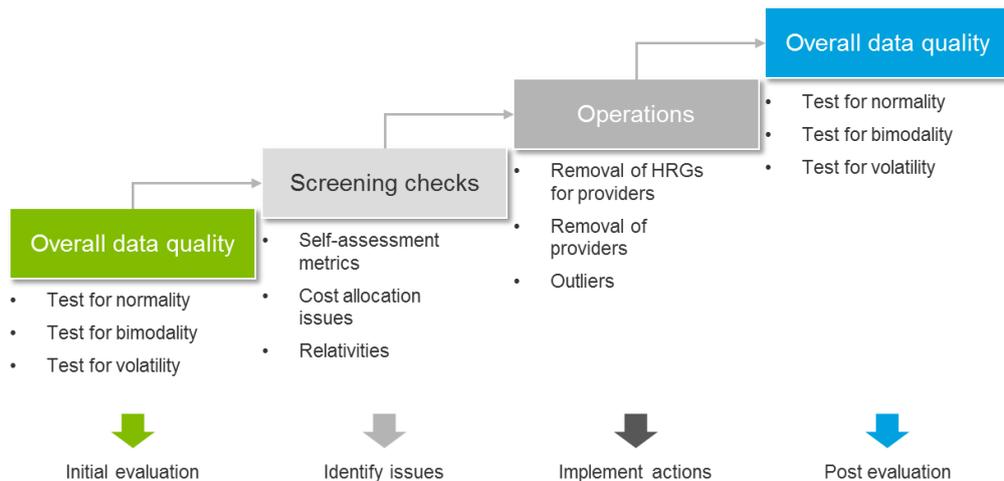
- Section 2 details the analytical approach used to assess the quality of reference cost data;
- Section 3 summarises the key outputs from this analysis; and
- Section 4 provides recommendations on potential pre-processing methodologies.

Appendices set out technical details and further relevant information.

2 Approach

This section presents the approach used to assess the quality of reference cost data. This approach has been developed following discussions with Monitor, and is designed to identify potential data quality issues and pre-processing steps that may be applied to address these issues. Figure 1 outlines the overall approach.

Figure 1: Overall approach to data quality assessment



The approach consists of four steps:

- **Overall data quality (1).** The first step is to assess the overall quality of the reference cost data. This is conducted by testing for normality, bimodality and volatility.
- **Screening stages.** The second step is to run a number of prioritised screening stages that have been agreed with Monitor, to identify providers and data points that exhibit poor data quality.
- **Operations.** The third step is to develop a set of high-level operations (pre-processing steps) which are used to remove certain observations or providers from the reference cost input data.
- **Overall data quality (2).** The final step is to assess the overall quality of the 'cleansed' reference cost input data. This enables a 'before and after' comparison and illustrates the impact of the proposed pre-processing steps on data quality.

Each of these steps is discussed in more detail below.

2.1 Overall data quality assessment metrics

Three metrics were selected to assess the overall quality of the reference cost data. These metrics were also used to assess potential improvements in the quality of reference cost data, following application of pre-processing steps. These metrics are:

- **Tests for normality.** The distributions of submitted reference costs are assessed for each HRG and department to understand whether the underlying data is normally distributed.¹ A normal distribution of unit costs for HRGs across providers implies that provider costs are evenly distributed around the average, and the average cost therefore represents a reliable approximation of submitted reference costs.
- **Tests for bimodality.** The distributions of unit costs are assessed for each HRG and department for evidence of dual peaks which potentially indicate bimodality. Bimodality suggests that the average cost is potentially unrepresentative of the true cost for certain providers. Bimodality also has potential implications for currency design; however this is beyond the scope of this report.
- **Tests for volatility.** The unit cost for each HRG and department is compared across 2010/11, 2011/12 and 2012/13 to assess the level of volatility, taking into account valid reasons for changes in reported unit costs, such as changes in currency design or costing guidance.²

Further details on the statistical tests for normality and bimodality are provided in Appendix B.

2.2 Screening stages

A number of prioritised screening stages have been agreed with Monitor and are designed to identify providers that exhibit poor reference cost data quality. Poor quality may be identified across certain HRGs and departments, or for the provider as a whole. For example, providers submitting reference costs for a number of HRGs that are both above and below the national average are identified as having potential cost allocation issues. A long list of screening stages has been developed based on discussions with Monitor (further details are presented in Appendix C), which were refined into three categories of screening checks:

- **Self-assessment metrics.** This category relates to checks that providers should undertake prior to submitting reference cost data.
- **Cost allocation issues.** This category seeks to capture poor cost allocation approaches employed by particular providers.
- **Relativities.** This category relates to illogical unit costs observed for particular providers.

Each of these is discussed in further detail below.

¹ The tests are for log normality, that is, whether the natural log transformation of unit costs is normally distributed. This approach is commonly adopted in the statistics literature in health care, given the typically high number of low cost patients and the low number of very high cost patients. See, for example, William H. Greene, *Econometric Analysis*, 2011.

² Volatility in HRG unit costs is limited to those HRGs where no significant design changes have occurred across 2010/11, 2011/12 and 2012/13, as identified by the Health and Social Care Information Centre (in particular the Casemix team) in the Codebook. (<http://www.hscic.gov.uk/casemix/costing>)

2.2.1 Self-assessment metrics

This category of tests is designed to identify fundamentally poor data quality by verifying whether self-assessment checks have been undertaken by providers. The 2012/13 reference cost collection required Trusts to complete for the first time a self-assessment quality checklist embedded in the collection templates, as well as mandatory and non-mandatory validations.³ These stages identify issues such as:

- Negative costs and activity;
- Duplicate values across a complete row of data;
- Unit costs above £50,000 and below £5; and
- Missing unit costs and activity.

These metrics are included as screening stages within the approach.

2.2.2 Cost allocation issues

When submitting reference costs, providers are required to allocate shared costs to individual HRGs. The following methods have been developed to identify providers that may be adopting poor cost allocation principles:

- Identifying providers with a number of repeated unit costs across a single department and HRG; and
- Identifying providers with a significant number of unit costs both greater and less than the national average.

These metrics aim to identify providers that may either be inconsistently applying top-down cost allocation approaches or use considerable judgement in allocating costs. Monitor's Approved Costing Guidance emphasises the importance of activity based costing.

2.2.3 Relativities

Tests regarding relativity are used to identify providers with a significant number of unit costs that could be illogical compared to others. Monitor's 'Costing Patient Care' publication highlighted that reference costs sometimes generate cost relativities that are inconsistent with the clinical design of HRGs (an illustrative example is presented in Box 1).⁴ A *relativity index* is constructed from national reference cost schedules based on HRG complexities. Providers' individual relativity indices are then generated for each HRG and each HRG root is identified as exhibiting illogical relativities if the indices do not match those generated from the national hypotheses (see Appendix G for further

³ <https://www.gov.uk/government/publications/nhs-reference-costs-2012-to-2013>

⁴ <http://www.monitor-nhsft.gov.uk/costingpatientcare>

details). A high number of illogical relativities potentially indicates issues with cost allocation, and as such is considered to be an indicator of poor data quality.

Box 1: Example of an illogical relativity

If there are two HRGs for the same age group that are based on the same underlying procedure, with one relating to higher complexity cases, then providers submitting higher unit costs for the HRG associated with less complex cases would exhibit an illogical relativity for that HRG root.

The illustrative example presented below relates to appendectomy procedures for patients aged 19 years and over. The appendectomy procedures with complications and co-morbidities ('CC') – FZ20B – should logically have a higher unit cost than the same procedure without major CCs – FZ20A. In this case, the national average unit costs are consistent with this relativity; however the difference in unit costs for Hospital A indicates that there is an illogical relativity for this particular HRG.

Provider	HRG code	HRG description	Unit cost	Relativity
Hospital A	FZ20A	Appendectomy Procedures, 19 years and over with Major CC	£3,000	FZ20A < FZ20B
Hospital A	FZ20B	Appendectomy Procedures, 19 years and over without Major CC	£3,250	
National average	FZ20A	Appendectomy Procedures, 19 years and over with Major CC	£3,815	FZ20A > FZ20B
National average	FZ20B	Appendectomy Procedures, 19 years and over without Major CC	£2,672	

Source: National average unit cost data from 2011/12 reference cost national schedules. Cost for Hospital A is illustrative only.

2.3 Operations

Having identified providers that exhibit poor data quality, either for particular HRGs or as a whole, the next step is to design 'operations' that can be applied to the reference cost data set. Two types of operation are considered:

- The removal of a provider submission for a particular HRG and department (that is, the removal of an observation that may reflect a data error, for example); and
- The removal of a provider across all HRGs (that is, the removal of the provider from the reference cost data set) as the provider's data quality may be poor in a number of areas.

The impact of the recommended operations on the quality of the reference cost data is evaluated against the overall data quality assessment metrics, as outlined in Section 2.1.

2.3.1 Outliers

When assessing operations and the removal of observations and providers, it is important to consider outliers, that is, potential extreme cost values within the reference cost data. Based on conversations with DH and Monitor, it is understood that currently, the only pre-processing step applied to reference costs is the removal of unit costs that are outliers, defined as either 20 times above or below the average for each HRG.

There are a number of statistical methods that can be used to identify and remove outliers from a data set. Following a review of these methods, it is recommended that an outlier methodology termed the “maximum normed residual test” is implemented by Monitor. This approach has been applied in the literature.⁵ The maximum normed residual test is based on the largest absolute deviations from the average of each HRG. This test is potentially less sensitive to the assumption that unit cost distributions for HRGs are statistically normal, than the current approach to outlier detection (although the test does assume that cost distributions are approximately normally distributed). This is potentially important, given the relatively high proportion of HRG distributions which are classified as non-normal; as outlined in Section 3.

This test is undertaken across each HRG and department to identify observations within HRGs and departments that are extremely different from other observations in those HRGs and departments. It is important to apply this general removal of outliers before pre-processing steps are applied to the reference cost data. This outlier test is described in more detail in Appendix H.

⁵ See, for example: http://projecteuclid.org/download/pdf_1/euclid.aoms/1177693492 and <http://www.jstor.org/discover/10.2307/2346808?uid=3738032&uid=2&uid=4&sid=21103706692927>

3 Outputs

This section presents key findings in relation to:

- Overall data quality metrics; and
- Screening checks.

The analysis presented in this section is based on reference cost data for 2010/11, 2011/12 and 2012/13, adjusted for MFF. Pre-processing steps that have been developed to improve the quality of reference cost data are presented in Section 4.

3.1 Overall data quality assessment metrics

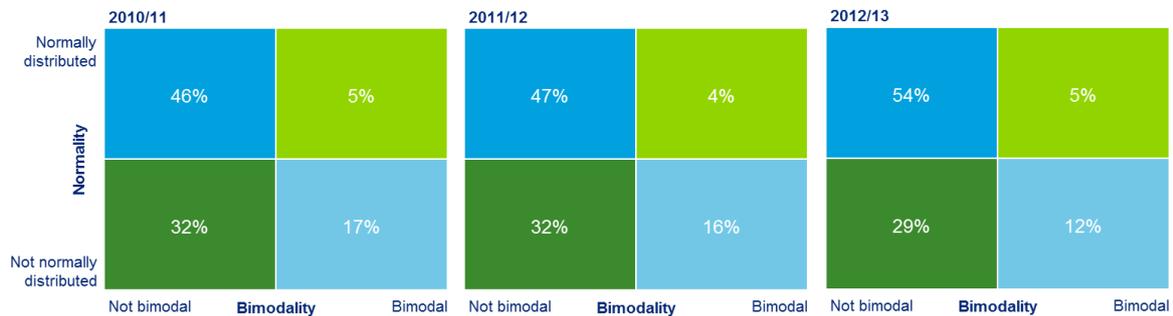
Across MFF adjusted reference cost data between 2010/11 to 2012/13, unit costs across providers are statistically normally distributed and do not show evidence of bimodality for approximately 50% of HRGs. Additionally, there is evidence of year-on-year volatility in unit costs for around 40% of HRGs, where such change would not be expected (for example, due to changes in currency design or costing guidance).

3.1.1 Normality and bimodality

Figure 2 summarises the findings of the tests for normality and bimodality for reference costs across 2010/11, 2011/12 and 2012/13. Specifically, the boxes illustrate the potential combinations of normality and bi-modality; that is, HRG distributions can be:

- Normal and not bimodal;
- Normal and bimodal;
- Not-normal and not bimodal; or
- Not-normal and bimodal.⁶

⁶ Note that the tests for normality and bimodality are run as two separate tests which is why the unit cost distribution for an HRG can be classified as normal and bimodal. Clearly this is an unusual case and appears counterintuitive in some respects; however it only applies to a small number of HRGs and is not an impactful result.

Figure 2: Normality and bimodality

In summary, it is found that:

- The statistics are consistent across 2010/11 and 2011/12, with evidence to suggest improved data quality for 2012/13;
- In 2010/11, unit costs across providers for 46% of HRGs and departments were statistically normally distributed without evidence of bimodality; whereas 17% of HRGs and departments are statistically non-normal and show evidence of bimodality in their cost distribution; and
- In 2012/13, the number of HRGs and departments with unit costs that are statistically normally distributed and do not show evidence of bimodality increases to 54%, while the HRGs and departments with non-normal and bimodal unit cost distributions decreases to 12%.

3.1.2 Volatility

Table 1 shows the level of volatility in reference cost data across the three years, based on various thresholds. The threshold identifies the maximum year-on-year percentage change in unit costs (either above or below) before a HRG and department is classified as volatile. Only HRGs where volatility is not expected, that is, HRGs which have not had a change in currency design or costing guidance, are considered.⁷ The mapping of reference costs between years that has also been considered as part of this project assesses in detail currency design, re-mapping and costing guidance changes related to HRGs in order to identify which HRGs are comparable between years. However, for the purpose of this report, volatility in HRG unit costs is limited to those HRGs where no significant design changes have occurred across 2010/11, 2011/12 and 2012/13, as identified by the Health and Social Care Information Centre in the Codebook.⁸

⁷ This comprises around 750 HRGs.

⁸ <http://www.hscic.gov.uk/casemix/costing>

Table 1: Volatility

Threshold	% HRGs exceeding threshold		
	2010/11 – 2011/12	2011/12 – 2012/13	2010/11 – 2012/13
10%	40%	33%	48%
20%	21%	16%	26%
30%	13%	11%	15%

In summary, it is found that:

- Between 2010/11 and 2011/12, 40% of HRGs and departments considered show a unit cost change of more than 10%; and
- If the acceptable level of annual change in unit costs is increased to 30%, far fewer HRGs and departments (13%) are categorised as volatile.

This summary shows that the number of HRGs classified as 'volatile' is sensitive to the volatility threshold that is applied. It is noted that DH previously used a threshold of 10% when assessing volatility.

3.2 Screening stages

This section presents the results of screening stages that have been applied to the reference cost data. Specifically, this section identifies the number of providers exhibiting poor data quality for certain HRGs and departments, or poor data quality as a whole.

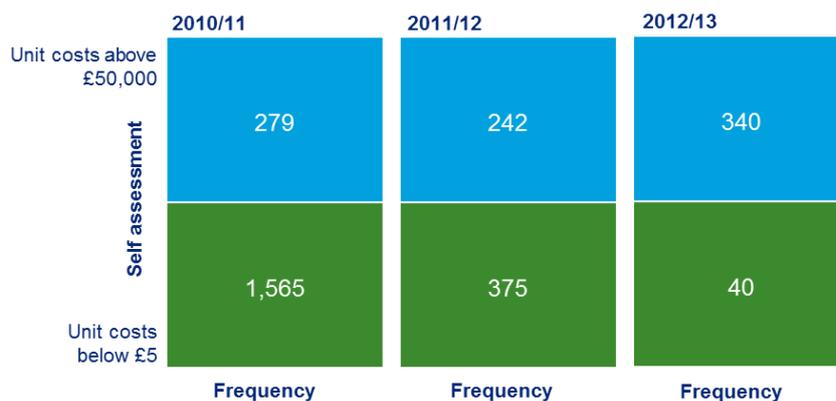
3.2.1 Self-assessment metrics

The self-assessment metrics considered are:

- **Negative costs and activity.** Across all three years, there were no instances of negative costs or activity;
- **Duplicate values across a complete row of data.** There were no instances of duplicate values across all data;
- **Missing unit costs and activity.** There were no instances of missing unit costs and activity; and
- **Unit costs above £50,000 and below £5.** There were various instances of unit costs below £5 or above £50,000 across the three years, as illustrated in Figure 3, adjusted to

include only the data that feeds into the Tariff Model.⁹ The number of unit costs above £50,000 has remained relatively constant between the three years; however, the number of unit costs reported below £5 has decreased over the three years, in line with the increased emphasis on providers' validation of their data. This potentially suggests improved data quality over time.

Figure 3: Self-assessment metrics



3.2.2 Cost allocation issues

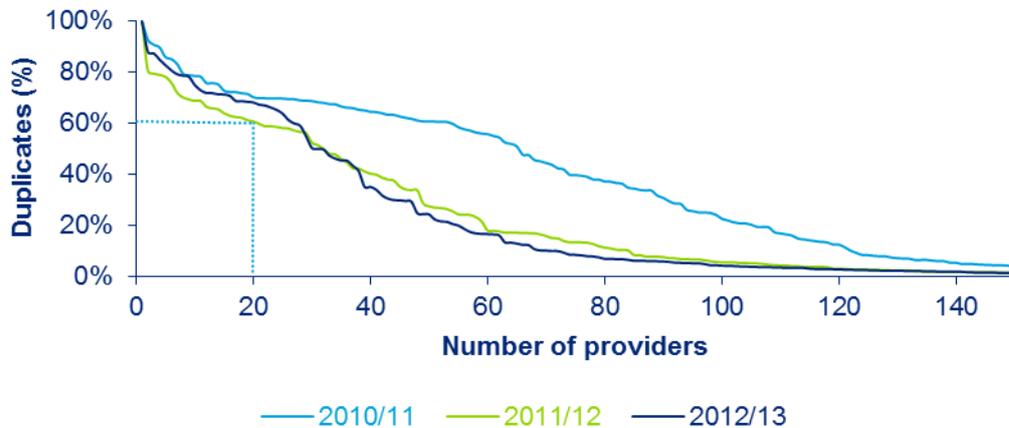
Cost allocation issues are determined through the following checks:

- Identifying providers with a number of repeated unit costs across a single department and HRG.** Figure 4 shows the percentage of costs that are repeated across at least one HRG for the same HRG and department, against the number of providers to which this applies. For example, for 20 providers, around 60% of submitted unit costs are repeated at least once across the same HRG and department in 2011/12. The shape of the curve is similar across all of the three years, and there are a limited number of providers submitting a high percentage of duplicates. This graph also shows that the number of providers with a high percentage of duplicate costs reduces slightly from 2010/11 to 2012/13.

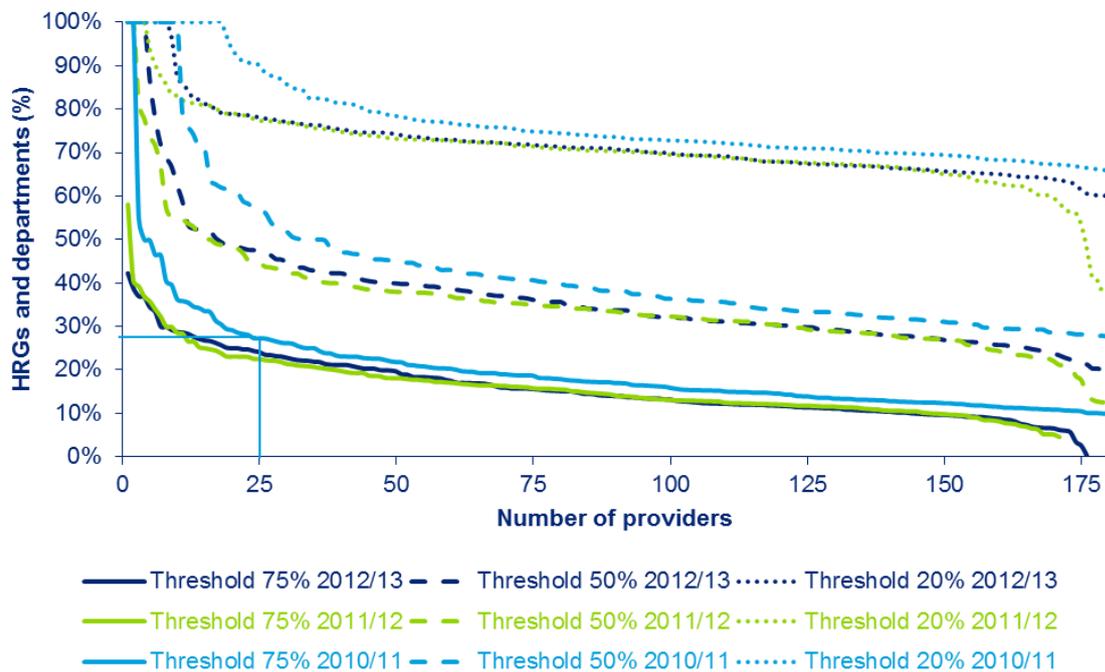
⁹ Data adjusted for the market forces factor ('MFF') for each year according to the MFF provider index for that year.

An example of how this screen stage works is included in Appendix D.

Figure 4: Repeated costs across years



- Identifying providers with a significant number of unit costs either greater or less than the national average.** Figure 5 shows the percentage of HRGs and departments with unit costs either greater than or less than the average by varying thresholds of permitted cost variation (20%, 50% and 75% of the average). For example, for 25 providers, at least 27% of their submitted HRGs are at least 75% above or below the average in 2010/11. For a small number of providers, a very high percentage of HRG unit costs submitted are clearly either above or below average. However, for the majority of providers, far fewer submitted HRG unit costs are clearly above or below average. When applying this screening method, it is important to only apply this screening to those providers who submit reference costs for a number of HRGs that are both above and below the national average. This is because some providers may be consistently above the average if they are, for example inefficient compared to the average provider.

Figure 5: High and low unit costs across years

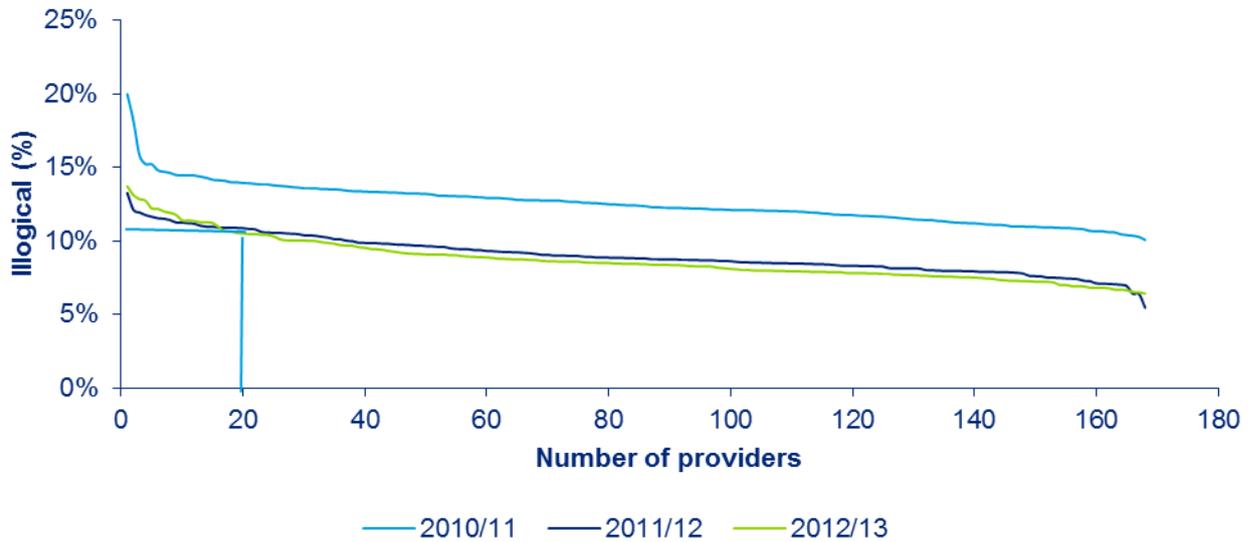
This graph demonstrates:

- As the threshold for permitted cost variation increases, the number of unit costs above and below the average decreases;
- The pattern of providers with higher costs is similar across the three years, with a small number of providers submitting a high number of costs either above or below the average; and
- The higher curve for each threshold in 2010/11 compared with other years suggests that providers' cost variation has decreased over time, potentially indicating improved cost allocation methods over time.

If Monitor implements a pre-processing step based on this screening check, there will be an element of subjectivity in order to set the appropriate threshold. This may therefore need to be subject to consultation across the sector.

3.2.3 Relativities

Figure 6 shows the number of HRG roots (as a percentage of providers' activity across each department) that exhibit at least one illogical HRG unit cost. This shows that in 2011/12, 20 providers exhibited signs of illogical relativities across approximately 10% of the HRG roots in which they deliver care.

Figure 6: Illogical relativities over time

In summary, it is found that:

- The shape of the curve is similar across all three years;
- There are a relatively small number of providers submitting a high number of illogical relativities (above 15%);¹⁰ and
- In 2010/11, a high number of providers exhibit illogical relativities across at least 10% of the HRG roots in which they deliver care.

The fact that most providers exhibit illogical relativities across 5-10% of their activity indicates that there could be genuine variation in resource utilisation between HRGs. This may not necessarily be counterintuitive.

For example, in certain instances at the provider level, a more complex procedure could have a lower cost than its less complex equivalent, although this would not be expected across all providers. Specifically, for a procedure conducted in a District General Hospital (DGH), pre- and post-procedure diagnostics could be conducted at a specialised centre for the complex cases, and conducted within the DGH for the non-complex cases. This could lead the DGH to correctly report lower costs for the more complex cases where the pre- and post-diagnostics are delivered externally, and higher costs for the lower complexity cases where pre- and post-procedure diagnostics are not delivered externally.

¹⁰ Note that there is a step change in the percentage of illogical relativities for providers between 10% and 15%. That is, a high number of providers with around 10% illogical relativities, but very few providers with more than 15% illogical relativities (only around five providers in 2010/11). This indicates that around 10% illogical relativities could be a reasonable number for single providers, but that over 15% may be consistent with poor data quality.

3.2.4 Summary

The screening stages have identified some potential issues in the underlying quality of reference cost data. Having assessed these issues, the next step is to use this information to develop operations to transform the data that is inputted into the Tariff Model.

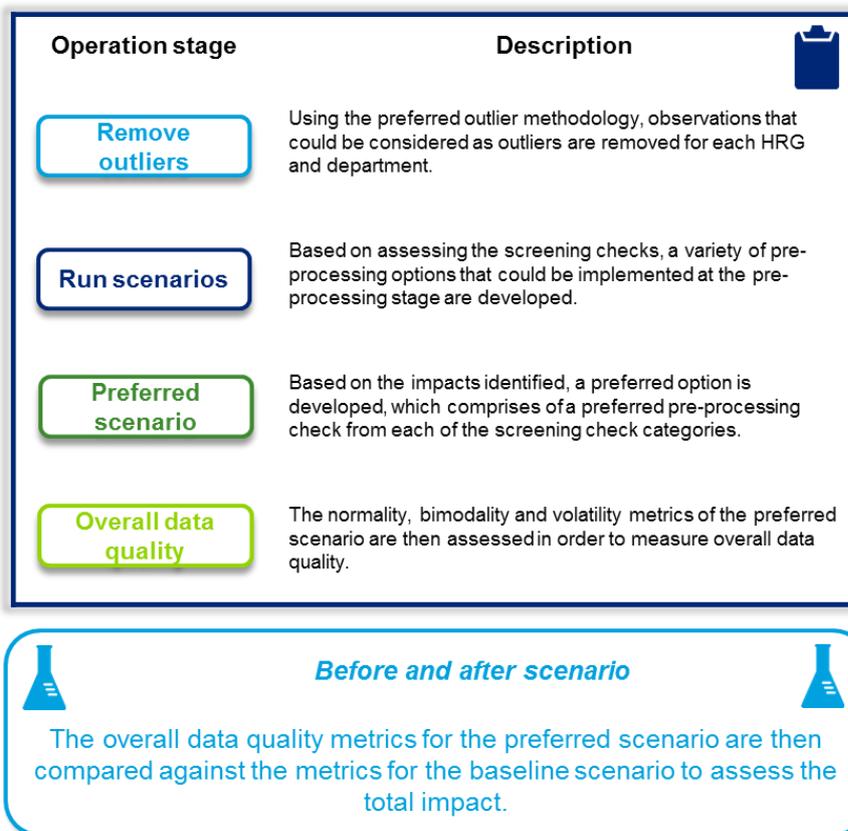
4 Recommendations

This section sets out recommendations for the pre-processing steps that could be implemented to enhance the quality of reference cost data used as an input into the Tariff Model. The recommendations relate to 2015/16; however Monitor may wish to develop this framework and implement further pre-processing steps in future years.

4.1 Approach to pre-processing steps

Figure 7 presents the approach to identifying the preferred pre-processing option.

Figure 7: Approach to pre-processing steps



As outlined above, a number of options have been considered in developing the recommended pre-processing option. The preferred option is based on identifying the screening stages that are most material, in terms of impact. This approach is favoured given the limited time-frame for conducting the analysis. Appendix C presents an analysis of a longer list of screening checks and their associated impacts.

4.2 Proposed option

This section presents the proposed pre-processing option (preferred scenario). In summary, this option entails:

- Removing outliers from the raw reference cost data set based on the “maximum normed residual test”;
- Removing providers with at least five unit cost submissions below £5 and at least 10 unit cost submissions above £50,000;
- Removing providers submitting more than 25% of costs above and 25% of costs below the average by 50% or more;
- Removing providers with more than 75% duplicate costs across HRGs and departments; and
- Removing providers submitting more than 15% illogical costs, based on relativities.

Each of these steps is outlined in further detail below, along with the estimated impact of the proposed pre-processing option. Impact is assessed against the baseline scenario. This counterfactual scenario is defined as how reference cost data is currently pre-processed before it is further adjusted in the Tariff Model; that is, the raw data set, adjusted by the MFF with outliers removed based on the approach historically used by DH.

4.2.1 Operations based on self-assessment metrics

It is recommended that all providers with more than five unit cost submissions below £5 are removed, for HRGs with a sufficiently high average unit cost. It is also recommended that all providers with more than 10 unit cost submissions above £50,000 are removed, for HRGs where the average unit cost is sufficiently low.¹¹ Providers reporting unit costs above £50,000 or below £5 – even for a very small number of HRGs – where these extreme costs would not be expected, could have poor data quality overall, and on this basis, are removed. Whilst it is acknowledged that these removals are on the basis of a small number of observations, the presence of these costs are considered as predictive of poor data quality, due to the self assessment process which requires providers to review these figures.

4.2.2 Operations based on cost allocation metrics

1. It is recommended that providers submitting more than 25% of costs above and 25% of costs below the average by 50% or more are removed. For providers meeting this condition, more than 50% of their unit cost submissions vary materially from the national average, with a significant number of submitted unit costs both above and below the average. This indicates that data quality could be poor across a range of HRGs for these providers. As such, the preferred option is to remove providers rather than individual data points.

¹¹ The thresholds for these operations have been selected based on assessing the distribution of the number of unit costs below £5 or above £50,000 for providers, and identifying clear changes. The threshold average cost for unit costs exceeding £50,000 is £40,000; and the threshold for costs below £5 is £100.

- It is recommended that providers with more than 75% duplicate costs across HRGs and departments are removed. Providers may be expected to submit costs that are the same across two or more HRGs in a number of cases. However, numerous providers submit over three quarters of each of their unit costs at least once, potentially indicating that providers apply significant judgement when allocating costs to HRGs.

4.2.3 Operations based on relativity

It is recommended that providers submitting more than 15% illogical relativities are removed. Providers submitting a high number of unit costs that could be illogical potentially exhibit poor data quality across all of the HRGs for which they submit costs. As such, the preferred approach is to remove these providers rather than individual observations.

4.2.4 Summary

Table 2 summarises the impact of the proposed option on overall data quality. The preferred option is compared with the baseline scenario. This allows a comparison of the current situation, and the future scenario, if the recommended pre-processing option is implemented by Monitor.

Table 2: Proposed option summary

	Preferred option	Number of observations (providers) remaining 2010/11	Number of observations (providers) remaining 2011/12	Number of observations (providers) remaining 2012/13	Volatility 1011-1112 (% change on baseline)	Volatility 1112-1213 (% change on baseline)	Volatility 1011-1213 (% change on baseline)	% HRGs statistically normal and not bimodal (2011/12)
Before	 Baseline	1,174,867 (266)	1,223,924 (181)	1,450,604 (195)	1,412	1,332	1,328	53%
After	 Preferred option*	1,150,150 (251)	1,198,895 (174)	1,372,953 (181)	1,400 (-1%)	1,303 (-2%)	1,317 (-1%)	64% (+11%)

*Note: in this assessment volatility is defined as the number of HRGs with unit costs in excess of 20% change. The HRGs compared are restricted to those where no obvious volatility is expected.¹²

In the preferred scenario, the number of data points has decreased in each year compared to the baseline, reflecting the removal of 15 providers (5.6%) in 2010/11, 7 providers (3.7%) in 2011/12 and 14 providers (7.2%) in 2012/13.

Overall data quality has improved under the preferred scenario:

- Volatility has reduced slightly in each year under the preferred scenario, in particular the number of volatile HRGs has fallen by between 1 and 2 percentage points respectively; and

¹² 20% has been identified as the threshold level of volatility here due to the fact that a very high number of HRGs have unit cost volatility of up to 10% and between 10% and 20%. The level of volatility of which room for improvement therefore exists is where unit cost volatility exceeds 20%.

- The number of HRG unit costs that are normally distributed and not bimodal has increased by 11 percentage points, from 53% to 64%.¹³ As noted above, setting prices based on average costs may be more appropriate for HRGs that are normally distributed. As such, an increase in the proportion of HRG unit costs that are normally distributed potentially leads to an improvement in data quality in this respect.

The impact of the recommended pre-processing option on average unit costs is also considered. This gives an indication of the potential impact of the proposed option on tariff prices. For 2010/11, 2011/12 and 2012/13, 515 (6.7%), 440 (5.5%) and 1,124 (10.5%), average unit costs respectively across HRGs and departments have been affected by more than 10% by implementing the preferred pre-processing steps. These impacts are summarised in Table 3. This potentially illustrates a material impact on tariffs for these HRGs. It is noted that tariff prices are based on average unit costs. As such, the removal of providers constituting a high proportion of activity for a particular HRG could have a material impact on tariff prices.

Table 3: average cost impacts

Unit cost impacts	2010/11	2011/12	2012/13
Number of unit costs with more than 10% increase (%)	139 (1.8%)	145 (1.8%)	382 (3.6%)
Number of unit costs with more than 10% decrease (%)	376 (4.9%)	295 (3.7%)	742 (7%)
Total number of unit costs with more than 10% increase or decrease (%)	515 (6.7%)	440 (5.5%)	1,124 (10.5%)

Finally, the impact on total reference cost activity has also been considered. In particular, implementing the preferred option leads to a decrease in total reference cost activity by 2.7%, 2.5% and 5.4% in 2010/11, 2011/12 and 2012/13, respectively, compared with the baseline option.

4.3 Next steps

Prior to implementing the proposed option, it is recommended that Monitor undertakes the following steps:

- **Test impact on tariff.** Whilst this study has attempted to capture this impact by considering average unit costs, it is recommended that Monitor runs the proposed reference cost input data set through the Tariff Model to understand the true impact on tariff prices.
- **Test impact on provider sustainability.** It is recommended that Monitor considers the impact of the proposed option on provider stability. Whilst Monitor would like to improve the quality of reference cost data, it needs to balance this objective against short term provider stability.

¹³ These figures are based on 2011/12 data.

- **Consult with the sector.** As noted at the outset, this is the first time pre-processing steps have been implemented to improve the quality of the reference cost data. As such, it will be important to engage with the sector and consult on the proposed option. This will give stakeholders an opportunity to comment on and critique the proposed methodology; Monitor may wish to update its approach in future years based on this feedback.

Additionally, Monitor may consider developing its methodology over time. Monitor could, for example, consider removing providers from the reference cost data set based on discrepancies between HES and reference cost activity data. This issue is considered in further detail in Appendix F.

Appendices

Appendix A Data

This appendix describes the data used for the analysis and some of the overarching processing steps. Specifically, the analysis has used:

- Publicly available Reference cost data for years 2010/11, 2011/12 and 2012/13.¹⁴
- For 2010/11 and 2011/12, unit costs have been adjusted for MFF using provider level adjustments contained in the relevant years' final national tariff.
- For 2012/13, data has already been adjusted for MFF in the published reference costs.
- Tests are then primarily conducted on the resulting MFF adjusted unit costs.
- All data not directly used in the Tariff Model has been discarded from the analysis, for example renal and mental health departments. This results in c.7% of data being removed from the analysis each year. In particular, only the following departments are kept in the analysis in order to match the granularity of the Tariff Model:
 - Day case; elective inpatient; elective inpatient with excess bed days; non-elective short stay; non-elective inpatient long stay; non-elective inpatient long stay with excess bed days; outpatient procedures; and A&E.
- No further adjustments are made to the data prior to undertaking the analysis.

¹⁴ For 2010/11: <https://www.gov.uk/government/publications/2010-11-reference-costs-publication> ;

For 2011/12: <https://www.gov.uk/government/news/financial-year-2011-to-2012-reference-costs-published> ;

For 2012/13: <https://www.gov.uk/government/news/financial-year-2011-to-2012-reference-costs-published>

Appendix B Detailed statistics

Overall data quality assessment metrics

This appendix describes the technical detail involved in implementing the overall data quality assessment metrics.

Normality

For the average unit cost of each HRG and department across providers, a Skewness and Kurtosis (“SK”) test for normality is calculated based on skewness and another based on kurtosis and then combines the two tests into an overall test statistic.¹⁵ The significance level is defined at 5% such that p-values above 5% are considered statistically normally distributed whilst those below 5% indicate non-normality. Sample sizes of less than 25 are excluded from this analysis.

Bimodality

For the average unit cost of each HRG and department across providers, a bimodality coefficient is calculated based on skewness and kurtosis. In particular the coefficient is calculated as:

$$\beta_{ij} = \frac{\gamma_{ij}^2 + 1}{\kappa_{ij}} \in (0,1)$$

Where;

β_{ij} : bimodality coefficient for HRG i in department j

γ_{ij}^2 : squared skewness for HRG i in department j

κ_{ij} : kurtosis for HRG i in department j

The higher the value of the coefficient, the more likely a distribution is said to be bimodal. High values of skewness and/or low values for kurtosis increase this coefficient and therefore the likelihood of bimodality.

In the literature, a value of 5/9 has been used as a threshold to identify distributions that could be bimodal.¹⁶

¹⁵ References on Skewness Kurtosis test: D’Agostino, R. B., A. Balanger, and R. B. D’Agostino, Jr. 1990. A suggestion for using powerful and informative tests of normality, *American Statistician* 44: 316-321.

Royston, P. 1991. sg3.5: Comment on sg3.4 and an improved D’Agostino test. *Stata Technical Bulletin* 3:19. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 110-112.

¹⁶ See, for example, Knapp, T.R., “Bimodality Revisited”, *Journal of Applied Statistical Methods*, 6 (1), Article 3.

It is noted that ideally a representative sample of the distributions of HRGs would be plotted manually and the data analysed visually to support a more complete and holistic understanding of the features of each cost distribution. However, in this case the requirement has been to propose automated methods to assess data quality. The automated methods are easier to implement and potentially require fewer subjective judgements; however they do potentially trade off valuable information in some instances.

Appendix C Original long list of screening stages

A long list of screening stages was developed and agreed with Monitor. This list was then refined according to Monitor's assigned priorities and feasibility requirements.

Table 4: Original long list of screening stages

ID	Item	Priority (1=high priority)	Where used/reason to exclusion
A	Unit costs reported by a provider that exhibit volatility	8	Although of limited value as a screening stage to a number of issues, this measure is a key metric upon which data quality impacts are estimated.
B	RCs reported by a provider that exhibit illogical relativities (for example reporting lower cost for more complex HRGs)	1	A relativity index has been constructed as part of the screening stage in this study.
C	Series of self-assessment and provider level data quality indicators	4	A selection of important self-assessment metrics have been selected as part of the screening stage for this study.
D	Providers that report significantly higher than average costs for certain HRGs and lower than average for other HRGs (indicating poor cost allocation)	3	This is a key screening check that has been used to identify providers with potential data quality issues.
E	Where possible, a high level reconciliation of HES and reference costs	5	This impact is considered at a high level in an appendix to this report.
F	Provider costs above average cost one year and below average in other years (indicating poor cost allocation)	9	This metric is linked to the volatility measures. In particular, year-on-year volatility in unit costs has been used to measure overall data quality. Therefore assessing costs over time has been employed as an overall data quality metric, rather than an individual screening check to identify providers or data points with data quality issues.
G	Where possible, instances where total cost in RCs don't reconcile with financial statements (data potentially supplied by Monitor's Provider Regulation team)	6	Difficulty in obtaining providers' financial data within the necessary timescales has restricted the use of this metric. It is recommended that Monitor revisits this stage and conducts a feasibility assessment around the usability of this as a potential indicator.
H	Estimated RCs for providers from the mapping exercise that exhibit volatility	7	Mapping reference costs between three years with a view to potentially setting prices based on multiple years of reference cost data is a separate work stream from this report. Due to the complexity of this exercise, an analysis of this volatility metric would test only a dual hypothesis (data quality <i>and</i> accuracy of mapped cost estimated). Due to the lack of power in this regard, this metric

			has not been considered as part of the data quality assessment.
I	Providers submitting the same costs for different currencies	2	This is a key screening stage that has been used to identify providers with potential data quality issues.

4.3.1 Further self-assessment metrics

The long list of self- assessment metrics considered is detailed here.

Table 5: Long list of self-assessment metrics

Item	Priority (high/low)	Where used/reason to exclusion
Invalid HRG/TFC codes (from HES) implying poor RC quality	Low	The intuition would be that poor quality HES data is linked to poor reference cost data quality. Due to the dual hypotheses involved and the more direct testing approaches on reference costs adopted, this metric has not been assessed.
MAQS	Low	The Materiality and Quality Score (MAQS) is a self-assessment check performed by providers submitting patient-level costs to measure the extent to which their cost allocation is appropriate. However, the MAQS is subject to bias as it is self-submitted whilst also the design is evolving which suggests the hypothesis could be a dual proposal assessing the appropriateness of MAQS as well as the quality of data. This data is also only available for 2012/13 for 66 trusts, rather than the full number of trusts across three years.
Provider submitted unit costs over £50k and below £5	High	This is a key screening check that has been used to identify providers with potential data quality issues.
Submissions where the reported day case unit cost greater than elective unit cost	Medium	This metric has not been considered as the assertion that day case unit costs should not exceed elective unit costs is not necessarily true across all providers. Therefore identifying these cases would not necessarily flag providers with poor data quality as the unit costs could be genuine.
Instances where a provider's market share of cost or activity is greater than 5%	Low	There is no appropriate market share level which indicates data error, therefore this metric has not been considered.
Where possible, a high level reconciliation of the Audit Commission's national benchmarker	Low	The Audit Commissioner's national benchmarker has been accessed through Monitor's servers and discussed with Monitor's KIM team around the appropriateness of this tool in feeding in to the reference cost input data work stream. While the tool is useful for benchmarking a small number of providers in a local setting across a variety of care aspects, there is no function for extracting overall data quality metrics at a national level and the underlying data cannot be accessed from the tool. As a result, this metric has not been employed.
Negative activity reported	Medium	Although this screening check has been used to identify providers with potential data quality issues, there was no material impact and so this indicator was not considered for further analysis.
Whole number activity not reported	Medium	This metric has been assessed however there was no material impact and so this indicator was not considered for further analysis.
Missing unit costs	Medium	Although this screening check has been used to identify providers with potential data quality issues, there was no material impact and so this indicator was not considered for further analysis.

Negative costs reported	Medium	Although this screening check has been used to identify providers with potential data quality issues, there was no material impact and so this indicator was not considered for further analysis.
Duplicate items (department, HRG, provider) provided	High	This is a key screening check that has been used to identify providers with potential cost allocation issues and therefore poor data quality.
Providers' reported Information Governance Toolkit (IGT) score, which is an external measure of data quality.	Low	The Information Governance Toolkit (IGT) is an online system which allows NHS organisations and partners to assess themselves against Department of Health Information Governance policies and standards. The IGT has been accessed and the appropriateness of this tool in feeding in to the reference cost input data work stream has been assessed. While the toolkit is useful for benchmarking a particular provider and accessing a suit of different metrics, overall data quality scores can only be accessed for each individual provider in turn, therefore there is no function for extracting overall data quality metrics at a national level. In addition, some of the elements that construct the measures for overall data quality are not made explicit, and the underlying data cannot be accessed from the toolkit. As a result, this metric has not been employed at this stage.

Options assessment and development of preferred option

This section presents a number of scenarios which could be implemented as pre-processing steps to improve the quality of reference costs. Each scenario is tested and the impact of the proposed operation on volatility is estimated across reference costs in 2010/11, 2011/12 and 2012/13. Table 6 summarises the impact of each of the scenarios.

Table 6: Individual options assessment

Screen category	Number	Pre-processing recommendation	Number of observations (providers) remaining 2010/11	Number of observations (providers) remaining 2011/12	Number of observations (providers) remaining 2012/13	Volatility 1011-1112 (% change on baseline)	Volatility 1112-1213 (% change on baseline)	Volatility 1011-1213 (% change on baseline)
Baseline		N/A	1,180,256 (268)	1,229,062 (181)	1,453,581 (195)	1,420	1,334	1,335
Self assessment metrics	1	Remove all observations where unit costs are below £5 or above £50,000	1,178,412	1,228,445	1,453,201	1,399 (-1.5%)	1,308 (-1.9%)	1,313 (-1.6%)
	2	Remove all providers where unit costs are below £5 or above £50,000 across more than 10 submissions	1,051,680 (252)	1,089,690 (164)	1,339,970 (185)	1,438 (+1.3%)	1,330 (-0.3%)	1,318 (-1.3%)
	3	Eliminate providers submitting more than 10% of costs above and 10% below average by 50% or more	292,083 (133)	261,665 (54)	321,350 (70)	1,610 (+13.4%)	1,601 (+20%)	1,654 (+23.9%)
	4	Eliminate providers submitting more than 25% of costs above and 25% below average by 75% or more	1,180,256 (268)	1,229,062 (181)	1,453,581 (195)	No change	No change	No change
	5	Eliminate providers submitting more than 25% of costs above and 25% below average by 50% or more	1,178,811 (266)	1,229,055 (180)	1,429,263 (194)	1,417 (-0.2%)	1,317 (-1.3%)	1,321 (-1.0%)
Cost allocation	6	Eliminate providers submitting more than 20% of costs above and 20% below average by 50% or more	1,121,552 (254)	1,224,423 (179)	1,400,875 (188)	1,435 (+1.1%)	1,312 (-1.6%)	1,331 (-0.3%)
	7	Eliminate providers submitting more than 15% of costs above and 15% below average by 50% or more	852,096 (220)	904,695 (140)	1,046,079 (155)	1,447 (+1.9%)	1,317 (-1.3%)	1,356 (+1.6%)
	8	Eliminate providers with more than 75% duplicate costs across services and departments	1,152,976 (256)	1,222,530 (175)	1,415,159 (185)	1,427 (+0.5%)	1,304 (-2.2%)	1,313 (-1.6%)
	9	Eliminate providers with more than 50% duplicate costs across services and departments	846,860 (204)	1,041,645 (149)	1,262,371 (164)	1,478 (+4.1%)	1,337 (+0.2%)	1,329 (-0.4%)
	10	Eliminate providers submitting more than 10% illogical HRGs	75,112 (94)	1,085,899 (143)	1,345,509 (163)	2,322 (+60%)	1,276 (-4%)	1,468 (+10%)
Relativities	11	Eliminate providers submitting more than 15% illogical HRGs	1,175,800 (263)	1,229,062 (181)	1,453,581 (195)	1,413 (-0.4%)	1,334 (0%)	1,338 (0%)

Note: in this assessment volatility is defined as the number of HRGs with unit costs in excess of 20% change. The HRGs compared are restricted to those where no obvious volatility is expected. No outlier rules have been applied in this analysis as the options are considered independently.

Appendix D Cost allocation metrics

This section describes in more detail the cost allocation metrics used in this report.

Duplicates

Analysing duplicate cost submissions involves identifying providers with a number of repeated unit costs across a single department and HRG. In particular, duplicate costs are identified as the number of submitted costs by each provider (as a percentage of each provider's submitted HRGs) that are repeated across at least one HRG for the same HRG and department. An example of duplicate costs for a single provider is presented in Table 7.

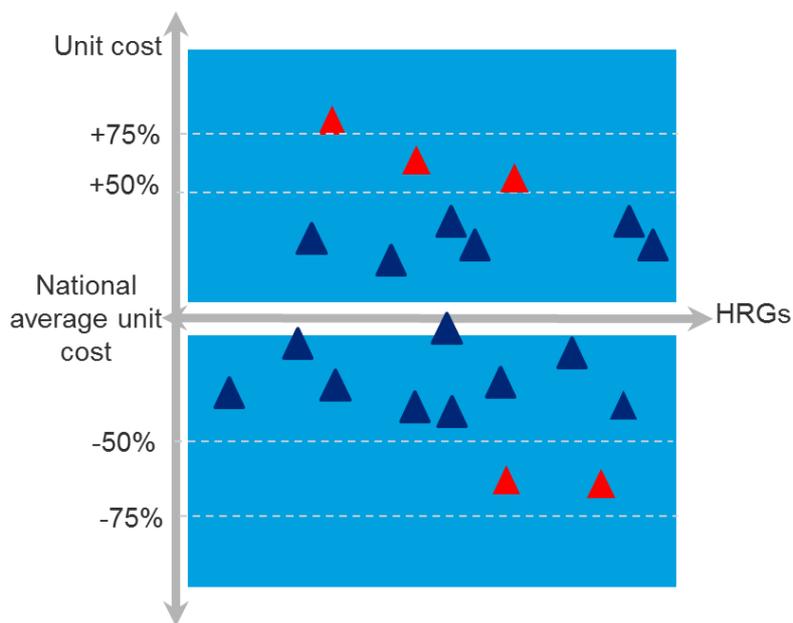
Table 7: Duplicates example

HRG	Department code	Unit cost	Duplicate submission	Duplicates tag
FZ20A	EL	£100	3	1
FZ21A	NEI_L	£100	3	1
FZ22A	EL	£150	0	0
CZ21V	NEI_L	£100	3	1

Each duplicate cost submission is identified and the number of duplicates is counted. In this case there are three occurrences of unit costs of £100 for this provider. The number of duplicates is then counted once each (rather than 3+3+3), therefore in this example the total number of duplicates would be 3, and the number of duplicate cost submissions would be $\frac{3}{4} = 75\%$ (3 duplicates from a total of 4 submitted costs).

4.4 High and low costs

This section describes the approach to detecting high and low provider costs in more detail. Figure 8 illustrates unit costs for a single provider.

Figure 8: High and low costs for a provider

In this case, the provider submits five of its twenty HRGs either above or below the average by 50% or more, equating to 25% of its submitted unit costs. However, in the preferred scenario, for example, there is also the requirement that a number of submitted costs must be both above and below average for a provider to be identified as having poor data quality. Consider the preferred scenario: remove providers submitting more than 25% of costs above and 25% below of average by 50% or more. In this case two submitted costs (10%) are more than 50% below the average and 3 submitted costs (15%) are more than 50% above the average. Therefore the provider would not be removed under this metric. This provider submits a single unit cost (5%) with a value more than 75% greater than the average and does not submit any costs which are more than 75% below the average.

Appendix E Analysis of spell-based reference costs

Background and approach

Monitor would like to understand whether reference costs collected at a spell level could be used in the Tariff Model. This section contains an initial assessment of potential issues related to spell based reference costs.

Both the 2011/12 and 2012/13 Reference cost datasets support spell based grouping, unlike earlier currency designs. The process of converting episode costs into spell costs is complex and the collection of spell-based reference costs has been mandated by DH (and now Monitor) since 2011-12. Therefore reference costs at a spell level have been collected since 2011/12 only.

An initial analysis of 2011/12 spell based reference costs has been conducted in order to investigate potential data quality issues with the spell based reference cost data. This analysis has been undertaken using 2011/12 data only, but findings potentially also apply to 2012/13 data which has also been collected at a spell level.

Several simple checks have been identified based on discussions with Monitor that could indicate poor data quality. This is not an exhaustive list and the findings do not reach clear conclusions around the appropriateness of spell based reference costs data as an input in to the Tariff Model; rather this is a preliminary investigation that could be used as the basis for further work in this area.

Details of the analysis undertaken are presented below:

- **The difference between unit costs between FCE and spell level data has been assessed.** Each spell includes at least one FCE, implying that its unit cost will be higher than for any given HRG (although it is noted that there could be grouping issues that lead to FCE level costs being higher than spell level costs).
- **Tests for normality.** The distributions of submitted spell based unit costs are assessed for each HRG and department to understand whether the underlying data is normally distributed.¹⁷ A normal distribution of unit costs for HRGs across providers would imply that provider costs are evenly distributed around the average, and therefore the average cost represents a reliable approximation of submitted reference costs.
- **Tests for bimodality.** The distributions of submitted spell based unit costs are assessed for each HRG and department for evidence of dual peaks which potentially indicate

¹⁷ The tests are for log normality, that is, whether the natural log transformation of unit costs is normally distributed. This approach is commonly adopted in the statistics literature in health care, given the typically high number of low cost patients and the low number of very high cost patients. See, for example, William H. Greene, *Econometrics Analysis*, 2011.

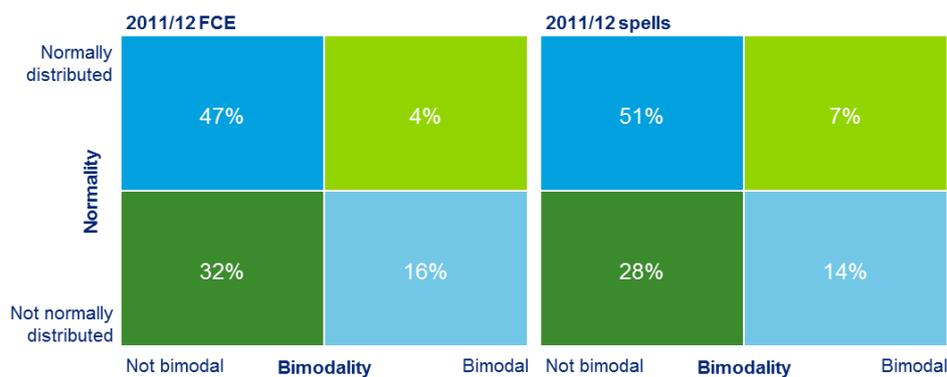
bimodality. Bimodality would suggest that the average cost is not representative of the true costs for certain providers.

Emerging outputs

Normality and bimodality

Figure 9 summarises the findings of the tests for normality and bimodality for reference costs across 2010/11, 2011/12 and 2012/13. The data has been MFF adjusted only.

Figure 9: Normality and bimodality in spells based reference costs



Comparing the normality and bimodality features of the spell based reference costs collected in 2011/12 with the FCE data previously identified, it appears that a greater number of the spell based unit costs are statistically normally distributed and do not show signs of bimodality (51% compared with 47%).

Next steps

Despite the potential improvement in statistical features of the HRGs costed at a spell level, this result would need to be investigated and refined by Monitor.

The high number of spell based unit costs that are not greater than their FCE equivalent is an overarching concern.

It is recommended that Monitor conducts a more detailed data quality assessment of spell based reference costs for 2011/12 using the framework to assess data quality outlined in this report. Monitor should also seek to full develop a sector impact analysis in this regard, to estimate the resource impact on providers of submitting additional cost information.

Appendix F Comparability of reference costs and HES data

This section assesses to what extent (episode) activity (counts) within the HES data set reconcile with those in reference costs. This analysis is undertaken at a provider level and separately at an HRG level.

Background and methodology

One of the primary uses of HES data in the Tariff Model is to provide activity counts for individual HRGs at the episode and spell level. As a result, an additional investigation was requested to analyse the activity counts between the 2012/13 HES data and the 2012/13 reference cost (RC) data at the episode level.¹⁸ This was conducted in two ways:

- Comparison between the activity in HES and RC at the individual HRG level; and
- Comparison between the activity in HES and RC at a provider level.

The two data sets used for this analysis were:

- The 2012/13 grouped HES data set provided to Monitor by the Casemix team; and
- The 2012/13 RC raw data set.

Emerging findings

Activity comparison at the HRG level

The analysis compared 1,825 HRGs with activity in both data sets:

- Based on 1,825 HRGs, the total activity for the HES data set was 15,656,683 compared to 15,549,728 in the RC data, reflecting a difference of ~0.69%.
- 1,598 HRGs (~87.6%) had less than 10% difference in activity between the HES and RC data sets
- 1,419 HRGs (~77.8%) had less than 5% difference in activity between the HES and RC data sets
- 567 HRGs (~31.1%) had less than 1% difference in activity between the HES and RC data sets

Based on this analysis, 227 (12%) of HRGs do not reconcile by more than 10%.

¹⁸ Activity reported within Reference Costs is at the episode level

Comparison between the activity in HES and RC at a provider level

The analysis compared 178 NHS trusts:

- 153 NHS Trusts (~86.0%) had less than 10% difference in activity between the HES and RC data sets
- 134 NHS Trusts (~75.3 %) had less than 5% difference in activity between the HES and RC data sets
- 55 NHS Trusts (~30.9%) had less than 1% difference in activity between the HES and RC data sets

Based on this analysis 25 (14%) providers' activity does not reconcile by more than 10%. Monitor may wish to consider removing providers from the reference cost data set if activity does not reconcile between reference costs and HES by more than a given threshold.

Appendix G Illogical relativities

A relativity index is constructed from national reference cost schedules based on HRG complexities. This is used to generate hypotheses of expected relativities. Providers' individual relativities indices are then generated for each department and each HRG root is identified as exhibiting illogical relativities if the indices do not match those from the national hypotheses. A high number of illogical relativities potentially indicate poor data quality.

This appendix contains further details on the approach to developing the relativity indices for each year.

In order to construct the relativity index, for each year, a three step process is undertaken:

- **Data sorting.** The reference cost data is sorted to generate an ordering of complexities and co-morbidities
- **Age factors.** The relativity index is based on identifying HRGs that are the same except only for differing complexities. Aside from complications and comorbidities, age is a key variant of HRGs that are otherwise driven by the same procedure. There is an adjustment to ensure comparisons are generated only for those HRGs which relate to the same age group.
- **Root comparison.** Relativities are analysed at the level of the HRG root. Cost differences between HRGs are only identified and subject to indexing if they belong to the same root. Therefore, if HRGs have the same root and refer to the same age group, differing only by complexity, then an indexation is generated for these HRGs such that the more complex HRG is given a higher value than the less complex HRG.

In the development of the index, lower values are assigned to lower complexity HRGs and higher values are assigned to more complex HRGs, with increasing complexity HRGs given higher numbers in single integer increments. For example, an appendectomy in 2012/13 is described in the following way:

Table 8: Example relativity index

Currency code	Currency description	Relativity index
FZ20F	Appendicectomy Procedures, 19 years and over with CC Score 5+	7
FZ20G	Appendicectomy Procedures, 19 years and over with CC Score 3-4	6
FZ20H	Appendicectomy Procedures, 19 years and over with CC Score 1-2	5
FZ20J	Appendicectomy Procedures, 19 years and over with CC Score 0	4
FZ20K	Appendicectomy Procedures, 18 years and under with CC Score 3+	7
FZ20L	Appendicectomy Procedures, 18 years and under with CC Score 1-2	6
FZ20M	Appendicectomy Procedures, 18 years and under with CC Score 0	5

Excerpt taken from 2012/13 data

In the above example, the HRGs with higher CC scores are assigned higher values. The index is also separate for the two separate age bandings within the HRG root.

In particular, in the above example, there are 7 values for the index, reflecting 7 different HRGs in the HRG root, FZ20. This is then split in to two sub-categories, reflecting procedures on different age groups. Once the index is split in to these two parts, each index starts at 7 with the highest complexity HRG within the age group taking the value 7. Lower complexity HRGs are then assigned appropriately lower values descending from 7. This data is then combined with the provider level submissions in the reference cost data for the relevant year analysing cost differences. A single flag for an illogical relativity is produced when a provider has submitted one of more inconsistent unit costs in a single root, compared with the other unit cost submissions in that HRG root. Table 9 illustrates how the index is applied to provider submitted unit costs.

Table 9: Example illogical relativities for a single provider

Currency code	Currency description	Relativity index (national)	Unit cost	Provider index	Illogical relativity count
FZ20F	Appendicectomy Procedures, 19 years and over with CC Score 5+	7	£3,000	6	1
FZ20G	Appendicectomy Procedures, 19 years and over with CC Score 3-4	6	£3,200	7	
FZ20H	Appendicectomy Procedures, 19 years and over with CC Score 1-2	5	£2,000	5	
FZ20J	Appendicectomy Procedures, 19 years and over with CC Score 0	4	£1,700	4	
FZ20K	Appendicectomy Procedures, 18 years and under with CC Score 3+	7	£5,000	7	
FZ20L	Appendicectomy Procedures, 18 years and under with CC Score 1-2	6	£4,500	6	
FZ20M	Appendicectomy Procedures, 18 years and under with CC Score 0	5	£4,000	5	

Costs are completely fictitious and illustrative only

In the above example, the higher cost for FZ20G than FZ20F for the provider indicates an illogical relativity as the lower complexity HRG has a higher cost, therefore producing an inconsistency with the national relativity index. This generates a single count for an illogical relativity for this HRG root. Note that even if two other HRGs exhibited illogical relativities, or there were further inconsistencies within the root, there would still only be one count for this HRG root. The number of roots exhibiting illogical relativities are then summed for each provider and expressed as a share of the total number of HRG roots submitted.

Appendix H Outliers methodology

When assessing operations and the removal of observations and providers, it is important to consider outliers, that is, potential extreme cost values within HRGs. Based on conversations with DH and Monitor, it is understood that currently, the only pre-processing step applied to reference costs is the removal of unit costs that are outliers, defined as either 20 times above or below the average for each HRG.

There are a number of methods which can be used to detect observations that could be considered outliers in a data set. A range of different methods used to detect outliers in reference costs have been considered and an outlier methodology termed the “maximum normed residual test” has been selected, which has been widely applied in the literature.

Maximum normed residual test

The maximum normed residual test is a statistical test used to detect outliers in a data set and is based on the largest absolute deviations from the average of each HRG. The two-sided test statistic is defined as:

$$t = \frac{\max_{i=1, \dots, N} |Y_i - \bar{Y}|}{\sigma}$$

Where:

σ = standard deviation

\bar{Y} = sample mean

The test detects one outlier at a time. This outlier is temporarily deleted from the dataset and the test is iterated until no outliers are detected.

This test is undertaken across the natural logarithm of cost for each HRG and department to identify observations within HRGs and departments which are extremely different from other observations in that HRG and department.

The test does assume that the underlying data is approximately shaped in a similar way to a normal distribution, that is, HRG log cost distributions should be approximately “bell shaped”. This is distinct from an HRG being defined as statistically normal after undergoing rigorous testing (as in the overall data quality metrics section). In particular, most tests for normality – those employed in the overall data quality section – often reject the null hypothesis of normality quite frequently, when the underlying data is reasonably bell-shaped, and outlier tests such as this one, could reasonably be employed. Tests that do not assume a specific distributional form were considered as part of this analysis but were rejected based on their lack of power in identifying outlying observations.

In this process outliers have been considered through assessing raw data series. An alternative approach that could be used is regression analysis to identify extreme costs for providers, which exist after controlling for a number of factors, for example size and case mix. It is also noted that

outliers have necessarily been identified through an automated process, however, there are a number of limitations to automation alone and, ideally, individual cost distributions would be assessed visually in order to further understand potential outliers from a more holistic perspective.