Fundamental review of the National Efficient Price (NEP)

FINAL REPORT

Independent Hospital Pricing Authority

26 August 2019

Contents

Inherent Limitations

This report has been prepared as outlined in the Scope Section. The services provided in connection with this engagement comprise an advisory engagement, which is not subject to assurance or other standards issued by the Australian Auditing and Assurance Standards Board and, consequently no opinions or conclusions intended to convey assurance have been expressed.

No warranty of completeness, accuracy or reliability is given in relation to the statements and representations made by, and the information and documentation provided by the Independent Pricing Authority personnel consulted as part of the process.

KPMG has indicated within this report the sources of the information provided. We have not sought to independently verify those sources unless otherwise noted within the report.

KPMG is under no obligation in any circumstance to update this report, in either oral or written form, for events occurring after the report has been issued in final form.

The findings in this report have been formed on the above basis.

Third Party Reliance

This report is solely for the purpose set out in the Scope Section and for the Independent Hospital Pricing Authority information, and is not to be used for any other purpose.

This report has been prepared at the request of the Independent Hospital Pricing Authority in accordance with the terms of KPMG's contract dated 17 September 2018 and varied 4 February 2019. Other than our responsibility to the Independent Hospital Pricing Authority, neither KPMG nor any member or employee of KPMG undertakes responsibility arising in any way from reliance placed by a third party, including but not limited to Independent Hospital Pricing Authority, on this report. Any reliance placed is that party's sole responsibility.

Executive Summary

Purpose and scope of the review

The Independent Hospital Pricing Authority (IHPA) engaged KPMG to review all current processes and statistical techniques used in the development of the National Efficient Price (NEP), drawing upon the literature and internal IHPA documentation (the fundamental review).

The objectives of the review are to:

- conduct a literature review of modern data analysis and statistical modelling techniques applicable to activity based funding of hospital services;
- review all current processes and statistical techniques used in the development of the NEP; and
- produce a list of recommendations of improvements to the current processes and statistical techniques used in the development of the NEP, quantified against an appropriate metric.

Conduct of the review

KPMG undertook the review as an iterative process. The starting point for the review was a comprehensive review of the literature. KPMG revisited the literature as new issues emerged during the course of testing potential solutions. This involved iterative discussions with IHPA staff and submitting progressive results and ideas to the Technical Advisory Committee (TAC) for feedback mid-stream through examining any one issue. KPMG focused on acute episodes of care however, many of the findings and suggested areas for improvement are applicable for all care types.

General findings

There are a number of general findings that had implications for what KPMG subsequently tested but that also have implications for the on-going management and development of the NEP. These are as follows:

Rarely are there universally recognised solutions: Through the literature review, KPMG identified considerable variability in the approaches taken by hospital price regulators in developing 'technical solutions' to address common issues. The key inference from this is that there are multiple options to tackle common problems, and there is no universal 'better practice'.

Interdependency of solutions: KPMG found from the literature review that in some cases, a technical approach adopted by a price regulator to address one issue is linked to an approach adopted for a related issue. The implications for the NEP include:

- Adjusting one aspect of the NEP that interacts with other NEP components requires consideration of the inter-dependencies before any adjustment is made to any one component (noting that IHPA staff are aware of this and it is evident in the NEP processes).
- When examining approaches used by other price regulators to address a common issue, a specific characteristic of other price models cannot simply be lifted and transplanted into the NEP without understanding how that component interacts with other components in the host model and how it might interact with other components in the NEP. This was taken into account by KPMG when for example, we examined the approach adopted in Ireland regarding inlier boundary points.

Rationale for an NEP process and/or component needs to be stated and documented: KPMG found that both in Australia and in other countries that have Activity Based Funding (ABF), often the rationale behind why a 'price regulator' has taken a specific approach is not documented or explicitly understood. Having a shared and documented understanding is critical for the transparency of the NEP (one of the objectives for the implementation of ABF) and to ensure that the solution is appropriate.

Practical solutions still need to be underpinned by principles: KPMG found that many of the data preparation processes are based on practical solutions which of themselves provided an efficient pathway during the formative stages of the NEP. While none of these practical solutions were found to be invalid or inappropriate, they should always be guided by a set of principles and/or statistical rationale. Note that KPMG has suggested an alternative solution to one of these practical approaches to data preparation (data trimming which is discussed in the specific findings below).

Statistical robust methods are preferred even if currently they do not deliver an improved performance for NEP: KPMG found examples of technical aspects of the NEP that were not technically optimum. The alternative technically superior methods tested (which tend to be more complex methods) often had the same level of statistical performance as the current NEP method. However, the more robust method should be adopted where it future proofs the NEP from data characteristics which currently seem not to impact on the NEP. KPMG found that other industries tend to adopt the more robust (and complex) methods. The areas where KPMG makes recommendations are discussed in the specific findings that follow this section.

Specific findings

The specific findings relate to each specific element of the NEP that was reviewed.

Trimming methods

In reviewing the current trimming methods, KPMG found that:

1. The current trimming method sets a standard threshold which does not take into account the variability in the underlying data and that it needs to be modified to better deal with episodes with improbable costs.
2. Internationally, the purpose of trimming in some cases differs to the Australian context. Thus, it is important to consider and understand the purpose of trimming in assessing the applicability of other methods to the Australian context.

A suggested change to the trimming process is to adopt a bottom up costing method to deem a minimally plausible cost value which should be set dynamically for different casemix end classes (e.g. DRGs). Specific options for a more dynamic approach are discussed in the body of this report.

Inlier Boundary Points

1. There are a number of key characteristics of the data that are not adequately dealt with by the current method. This includes:
   - The length of stay distributions are non-normal and highly variable in terms of the underlying statistical function.
   - While many DRG distributions could be described as log normal (skewed right), many DRGs have very dispersed distributions with some being potentially heteroscedastic in nature.
   - There are a substantial number of episodes with extreme length of stay values that potentially could push the boundary points ‘too wide’.
2. Internationally these issues are addressed in variable ways including the measure of central tendency and measure of dispersion used, and variable mechanisms to deal with extreme values.
3. While a core method may be suitable for DRGs with a typical log-normal distribution, a different or modified approach may be required for other DRGs with a more dispersed distribution.
4. The analysis of alternative methods found in the literature including a hybrid model developed by KPMG all delivered a similar level of statistical performance compared to the current method. However, two of the alternative international methods have the advantage of having features that aim to address the issues described above.

Suggested changes to the inlier boundary method:

1 There is a technical argument supporting the need to adopt a method that deals with the non-normality of length of stay distributions and that reduces the sensitivity of high boundary point methods to extremely high values. The two alternative international methods address the technical efficiencies of the current method and offer future proofing if the degree of heterogeneity in the data increases.

2 A more conservative approach is required to set the upper bound particularly for DRGs with high coefficient of variation (CV) values, such that the upper bound is 'closer' to the central tendency that the current method.

Price loadings for unavoidable costs

KPMG examined the extent to which the variables that have a price loading are correlated (a potential factor ignored in the current model) and its impact on the NEP. The key findings are:

1. There are material correlations across the patient, treatment and radiotherapy variables which should be taken into account in a modification to the current model.
2. The preferred alternative model would calculate all patient episode adjustments using a single multivariate regression model, with each adjustment factor as a covariate and no interaction effects among these factors.
3. Adopting this model would result in:

    a. Higher adjustment factors for:

    - Remote Patient;
    - Very Remote Patient; and
    - Radiotherapy
    - Specialist Psychiatric under 17, MDC 19 or 20, Non specialist paediatric hospital
    - Specialist Psychiatric under 17, MDC 19 or 20, Specialist paediatric hospital
    - Specialist Psychiatric under 17, Not MDC 19 or 20, Non specialist paediatric hospital
    - Specialist Psychiatric under 17, Not MDC 19 or 20, Specialist paediatric hospital.

    b. Lower adjustment factors for:

    - Remote Treatment; and
    - Very Remote Treatment.

4. The preferred alternative model performs as well as the current Model, as measured by overall Cost Ratio, R-squared and SMAPE and has minimal impact on individual hospital cost ratios, with all changes being less than 2.4% and most being less than 1%. Only two hospitals move from a cost ratio below 1 to above 1 and both change by less than 0.06. No hospitals move from a cost ratio above 1 to below 1.

A suggested change to the approach for determining price loadings for unavoidable costs are to adopt a multivariate regression approach as it is methodologically preferable to the current mixture of methods used to calculate the patient and episode adjustment factors given that it makes explicit the impact covariance has on the NEP regardless of the materiality of any such interactive effects.

Price loadings for private patients

KPMG examined solutions to avoid the potential for public and private patients with the same clinical presentation and treatment (in terms of ABF relevant data items) to generate different amounts of hospital revenue.

1. The distribution of the difference between the Private Patient Service Adjustment (PPSA) and the ratio of private benefits to episodes cost shows that PPSA would be likely to overestimate the

medical plus ancillary revenue for private episodes more often than it under estimates that revenue.
2. When averaged over all episodes, PPSA underestimates the net value of private revenue across private episodes.
3. A number of factors appear to be systematically associated with the difference between the benefits to cost ratio and the PPSA but any mechanism(s) underlying this finding could not be identified.

KPMG does not put forward any change to the current NEP approach as while the NEP Model appears to have scope for improved performance with respect to accuracy of PPSA, the mechanisms underlying the differences found through this review could not be determined. As a result, a suitable approach could not be designed to improve the Model's performance in this regard.

Price stabilisation

KPMG reviewed the constraints that are currently placed on year on year changes to price weights as a strategy to stabilise prices. KPMG found that there are multiple elements to price stabilisation, such as removal of spurious data, backcasting and loadings for unavoidable costs. The current approach to placing constraints on year on year changes to price weights as part of the overall price stabilisation approach is not sufficiently robust to ensure that changes to price weights are driven by underlying changes to actual cost drivers. This results in price weights changes in different directions from one year to the next requiring a more nuanced approach to price stabilisation. Different options have been set out in this report.

Indexation

KPMG reviewed the following aspects of the current approach to indexation:

- The time period used to develop the index value
- Whether there is a need to have a more dynamic approach to indexation that takes into account differential and underlying trends

KPMG found that the current method is based on an unstated assumption that the trends in costs for underlying cost drivers are relatively stable and similar across the major cost drivers. Whereas, KPMG's observations of health care costs and experience in other industries such as the insurance industry, suggests that this is unlikely to be the case.

A suggested change to the indexation method is for IHPA to consider adopting a more nuanced approach to indexation including:

- Undertaking sensitivity analysis of inflationary trends using different time periods so that, if there are material differences in trends between shorter and longer time periods, an adjusted index value would be used instead of the standard five year trend value.
- Identifying and adjusting for systemic or structural changes to the cost of services (or component cost); and
- Calculating separate index values for major cost drivers where there is evidence of differential trend.

Back-casting volume multipliers

KPMG assessed whether there is a suitable alternative to the current method that would reduce or eliminate the potential for error due to impact on hospitals' practice. This was assessed as KPMG found that the current method tends to overestimate the effects of any one back casting measure due to a failure to consider impact of price changes on purchasers' behaviour in the subsequent year.

However, KPMG's assessment is that the preferred method is the current method. While it suffers from the identified risk with regard to not accounting for NEP Model and Price induced changes in hospitals' behaviour and casemix, there is no basis on which to conclude this risk is larger than for alternative methods used in different industries other than one method (the Fisher Index method) which is not practical as it would delay the application of back casting by one to two years and thus substantially delay the reconciliation process.

The Reference cost

KPMG undertook a qualitative assessment of the current method to derive the reference cost and concluded that the principles and objectives of the current method – to produce a Reference Cost that maintains equivalency of NWAU values over time, independently of changes in the casemix present in the NHCDC data set – are reasonable. However, the method as currently applied has a number of weaknesses that should be addressed. The main weaknesses of the current method identified by our review are:

- Bias inherent in the choice of the chained Paasche Index;
- Risk of bias due to substantial changes in the AR DRG classification, activity counting rules and costing standards since the 2009 10 NHCDC data were collected; and
- Exclusion of subacute, non-admitted and emergency services' costs when calculating the Paasche Indices.

Based on the issues identified with the current method, this review makes the following suggestions:

1. That the index used for the calculation of the Reference Cost be changed to the (chained) Fisher Index.
2. That each year's index be calculated using activity data for all activity streams funded through the NEP.
3. That the index be rebased to a new base activity year. The specific year chosen should take into account the magnitude of changes in classifications, counting rules and costing standards since national ABF was introduced and changes planned for the near future.
4. That the rebased Reference Cost be set to the modelled average acute admitted episode cost for the chosen base year and cost weights be calculated using that new base Reference Cost, in that year.
5. That IHPA include the question of rebasing the Reference Cost in annual review of the NEP process and scope. This question should focus on whether there has been significant change to the activity classifications used for any of the activity streams, significant change to counting rules or related policies (such as admission criteria), and whether there has been significant change that would affect comparability of costing data over time.
6. That whenever IHPA rebases the Reference Cost, it publishes a time series of rebased historical Reference Costs and NEPs, to allow conversion of cost and price weights between NEP Determinations produced with different Reference Cost base years.

Final observation

The changes suggested in this review are likely to result in minor changes to the statistical performance of the current model. However, the changes are suggested as they address some of the technical issues found in the current model which are more likely to future proof the NEP against data volatility and changes in cost trends not evident in the historical data.

1 Introduction

1.1 Purpose and scope of the review

The Independent Hospital Pricing Authority (IHPA) engaged KPMG to review all current processes and statistical techniques used in the development of the NEP, drawing upon the literature and internal IHPA documentation (the fundamental review).

The objectives of the review are to:

- conduct a literature review of modern data analysis and statistical modelling techniques applicable to activity based funding of hospital services;
- review all current processes and statistical techniques used in the development of the NEP; and
- produce a list of recommendations of improvements to the current processes and statistical techniques used in the development of the NEP, quantified against an appropriate metric.

1.2 Conduct of the review

KPMG undertook the review as an iterative process. While the starting point for the review was a comprehensive review of the literature (Appendix 1), KPMG revisited the literature as new issues emerged during the course of examining issues and testing potential solutions. This often involved iterative discussions with IHPA staff and submitting progressive results and ideas to the Technical Advisory Committee (TAC) for feedback mid-stream through examining any one issue.

KPMG focused on acute episodes of care however, many of the findings and suggested areas for improvement are applicable for all care types.

1.3 Assessment metrics

A three level assessment framework (refer Appendix 2 for full details) has been adopted for this review noting that this is predominantly applied to testing alternative methods. In some cases, the review has also looked at specific issues other than alternative methods, in which case, the assessment process is specific to that issue. The assessment framework comprises a set of criteria at three levels as follows:

Level 1: method is consistent with core principles

1 The method is effective in achieving ABF objectives
2 It is feasible to apply the method in practice
3 The risks associated with the method are manageable.

Level 2: NEP attribute specific criteria (when applicable)[1]

Level 2 criteria are specific to the aspect of the NEP that is being tested and are listed in the relevant sections in this report.

Level 3: macro impacts

The method improves the analytical performance of the NEP in terms of:

- Explanation of costs (measured by R^2)
- Cost ratio (ratio of modelled total cost to actual total cost for a LHN)
- Cost model costs' accuracy (measured by Symmetric Mean Absolute Percent Error (SMAPE)).

[1] Method specific criteria are applied if there is a specific nuance that the macro metrics are not sufficiently sensitive to.

The objectives of national ABF as stated in the National Health Reform Agreement[2] that relate specifically to ABF are:

- improve patient access to services and public hospital efficiency through the use of activity based funding (ABF) based on a national efficient price (Schedule A);
- ensure the sustainability of funding for public hospitals by increasing the Commonwealth's share of public hospital funding through an increased contribution to the costs of growth (Schedule A);
- improve the transparency of public hospital funding through a National Health Funding Pool and a nationally consistent approach to ABF;

The first and third of the objectives above relate specifically to the functional form of ABF (i.e. methods used to derive the NEP and the NWAU model). The second of the objectives above is not dependent on the functional form of the model and this is not used to assess potential changes to the NEP methodology. The first of these objectives has two distinct components namely, patient access and hospital efficiency. Therefore, three distinct criteria have been formed that relate to the effectiveness of the funding model in terms of meeting the national ABF objectives:

- Patient access to services: does the change impact on the capacity of hospitals to enhance patient access to services?
- Public Hospital efficiency: does the change better reflect the efficient cost of service?
- Transparency of public hospital funding: does the change lead to a more transparent link between funding and cost of services?

---

[2] The NHRA has nine objectives of which three relate specifically to ABF. http://www.federalfinancialrelations.gov.au/content/npa/health/_archive/national-agreement.pdf.

2 Data trimming

2.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches in the following aspects of the NEP for *acute admitted care type only*:

- Trimming methods where the focus of the testing was on the following:
  - Overlap of the current trimming business rules to assess whether there is any redundancy in the rules
  - Using more dynamic approach to setting thresholds for trimming.

The testing used 2015/16 acute admitted episodes from hospitals in scope of activity based funding that were used by IHPA to construct the 2018/19 NEP using AR-DRG Version 9.

2.2 Purpose of data trimming

The purpose of data trimming is to identify and remove from the costed activity data, episodes with a spurious total cost value. Spurious values are removed because they are considered likely to be in error and thus should not influence the calculation of the NEP. Note that the episode is excluded from the calculation of the NEP but included in NEP processes that do not use the total cost variable (e.g. calculation of inlier boundary points).

2.3 Current approach to data trimming

The current method used by IHPA applies a series of data tests as follows:

- Stage 0 Trimmed: Jurisdictional Advice (no patient level cost data)
- Stage 1 Trimmed: Hospital-DRG extreme costs (cost-funding ratio)
- Stage 2 Trimmed: Cost lower than $23
- Stage 3 Trimmed: extreme cost increase over 300% or decrease less than 25% between previous and current episode
- Stage 4 Trimmed: Extreme cost ratio between in scope cost and modelled cost.

2.4 Alternatives tested

The following specific issues were considered for testing:

Research question #1: What is the overlap between the current trimming methods?

- This research question was considered as any substantial overlap would adversely impact the efficiency of the actual NEP process although it would not affect the NEP outcome.
- A specific criterion was used to assess this issue rather than the overarching assessment framework given this is an exploratory issue, namely, the number of episodes that overlap any two of the trimming criteria currently used to derive the NEP.

Research question #2: What is the impact of alternative trimming methods?

- Two alternative methods were selected. One was as an alternative for one of the current criteria namely total cost is less than $23. The other method is an alternative for the full set of criteria used to trim costed episodes. The specific reasons for their selection are discussed in the next section.

- The overarching assessment framework described in 1.3 is used to assess the alternative methods.

2.5 Results and discussion

2.5.1 Overlap of current methods

Research question #1: What is the overlap between the current trimming methods? We analysed the issue of overlap between the four trimming stages by systematically excluding a stage from the trimming process. The results are shown in Table 1.

Table 1: Overlap in the current trimming process

	Current 4-stage approach	Stage 1 excluded	Stage 2 excluded	Stage 3 excluded	Stage 4 excluded
Total acute episodes	5,612,366	5,612,366	5,612,366	5,612,366	5,612,366
Working in progress (WIP) or not costed	270,006	270,006	270,006	270,006	270,006
Admitted acute episodes	5,342,360	5,342,360	5,342,360	5,342,360	5,342,360
Stage 0 Trimmed: Jurisdictional Advice (no patient level cost data)	125	125	125	125	125
Stage 1 Trimmed: Hospital-DRG extreme costs (cost-funding ratio)	6,073	Excluded	6,073	6,073	6,073
Stage 2 Trimmed: Cost lower than $23	18,992	20,232	Excluded	18,992	18,992
Stage 3 Trimmed: extreme cost increase over 300% or decrease less than 25% between previous and current episode	76	89	284	Excluded	76
Stage 4 Trimmed: Extreme cost ratio between in scope cost and modelled cost.	457	364	7,731	481	Excluded
Total remaining episodes	5,316,637	5,321,550	5,328,147	5,316,919	5,317,094
Trimming share	0.5%	0.4%	0.3%	0.5%	0.5%

The results from this analysis indicate some overlap in business trimming rules but not to the point where there is material advantage to rationalise the rules. The main overlap is between stage 2 and 4 namely, 7,367 cases trimmed from stage 2 also meet the criterion for trimming in stage 4 (when stage 2 is excluded 7,731 cases are trimmed from stage 4 compared to 364 cases when stage 2 is not excluded).

Key findings

- The overlap in the current trimming stages is not sufficient to warrant a change.

2.5.2 Impact of alternative methods

Research question #2: What is the impact of alternative trimming methods?
We analysed the following alternative trimming methods:

- Universal minimum threshold based on a bottom-up costing of a bare minimum episode, calculated to be $275 for this review.
- Irish method of trimming: log transformation of the data and then trim episodes 3 standard deviations above and below the mean of transformed data[3]. Note that the Irish model then uses two standard deviations of the trimmed data to set boundary points.

The first method was selected as an alternative to one of the data exclusion criteria namely, total cost is less than $23. This current criterion was considered inadequate because it suggests that the minimal plausible total cost value is $23 which is unlikely to be the case for acute admitted patients. The alternative approach takes a standard costing approach to set the minimum plausible total cost value based on the following scenario:

- A patient is transferred from the ED and admitted to a short stay observation unit; the patient self-discharges after two hours. During this time, the patient undergoes nursing observation but does not receive treatment from any clinician.
- The nominal cost[4] of this episode comprises administrative time to admit the patient - $50, orderly transportation - $50, bed preparation - $50, nursing observations - $100, and administrative time to discharge - $25, resulting in a total cost of $275.

The second method was selected as it was the only data trimming method identified in the literature reviewed regarding hospital funding models[5]. This method however, as noted in the discussion that follows, has a broader purpose than just trimming spurious values (which is the NEP purpose of trimming). The results are included in this section as part of a broad discussion of data trimming methods.

The summary results are shown in Table 2.

Table 2: Alternative trimming methods

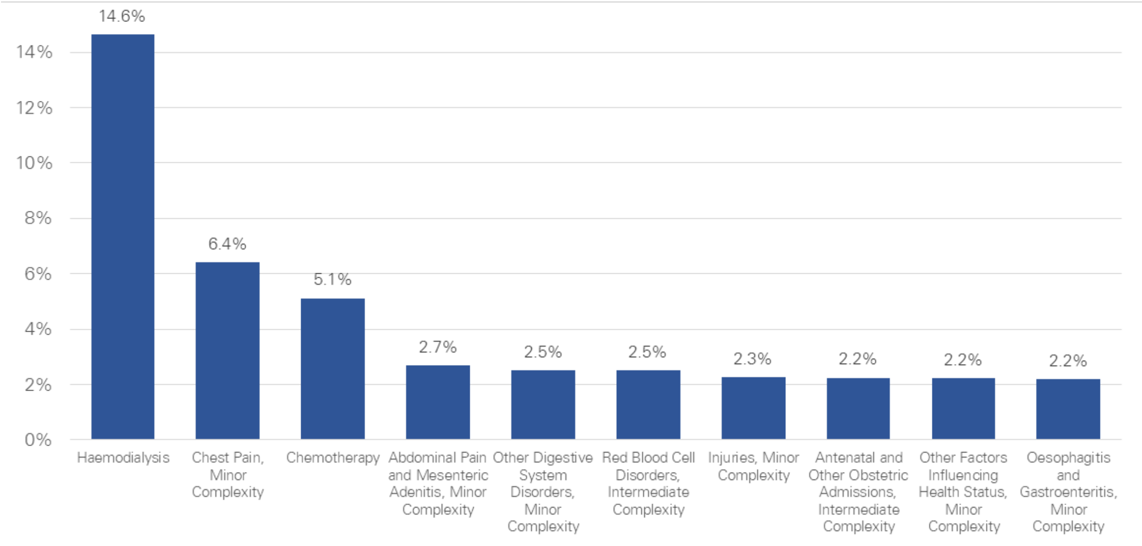
	Current 4-stage approach	Universal bottom-up costing minimum threshold $275	Irish method
Trimmed acute admitted episodes	25,723	496,238	371,349
Share of total	0.5%	8.8%	6.6%
NEP[6]	$5,012	$5,011	$5,011

[3] Fiachra Bane (20165). *Introduction to the Price Setting Process for Admitted Patients V1.0 26 May 2015*. Healthcare Pricing Office, HSE.
[4] The nominal cost is what KPMG considers to be a plausible direct cost for the cost component.
[5] Other international jurisdictions may also have processes for identifying and removing episodes with spurious cost values. However, these processes were not discussed in the literature reviewed for this project.
[6] Note that this comparison holds constant all other aspects of the NEP process. Thus for example, the indexation value has not been recalculated as a result of excluding a different sub-set of episodes which ordinarily would occur as a result of a different trimming method.

Universal bottom-up costing minimum threshold

The results show that a universal minimum threshold at $275 would result in a large number of trimmed episodes – almost 9% of all episodes. This suggests that there are a range of episode costs that may legitimately fall under the $275 minimum. Inspection of the data identified the top ten DRGs that accounted for 45% of all these trimmed cases, with three accounting for 25% (haemodialysis, chest pain with minor complexity and chemotherapy). The analysis was undertaken on the data set of in-scope costs and thus excludes e.g., high cost drugs and PBS reform related costs. Consequently, this may have skewed the results for DRGs where these costs are a material proportion of total costs.

Figure 1: DRGs with acute admitted episode costs less than $275



Other trimmed cases were spread over a larger number of DRGs.

Some of these low cost episodes[7] may reflect the low cost of home care (e.g. dialysis in the home programs). Some may reflect the low cost of holding patients in short stay observation units (e.g. chest pain),while others simply may represent poor quality costing data.

Key findings

- The 'one size fits all' nature of a universal threshold is a limitation given the heterogeneity across DRGs. While a universal threshold may be relevant for a large cross section of DRGs, such a threshold may need to be modified for specific case types.
- Conversely, there are potential benefits of dynamic or DRG-specific thresholds.

Irish boundary point method

The Irish boundary point method (Irish method) also trims a substantially larger number of episodes than the current approach but it is important to note that the purpose of trimming method used in the Irish payment model is fundamentally different to trimming in the Australian context although there is some overlap in purpose.

The Australian purpose of trimming is to identify spurious or problematic values of episode cost. The aim of the Irish trimming method is to remove 'extreme values' regardless of whether they are valid values in order to reduce the sensitivity of the boundary point method to extreme values. The Irish method uses the standard deviation to set boundary points. Extreme values would overly influence

[7] The episodic cost is based on the costs in scope of the national funding arrangements and thus for example, excludes some cost items such as capital costs, blood costs and high cost drugs.

the standard deviation which would then set the boundary points wider that what otherwise would be the case. Extreme values would include spurious data as well as likely to include valid values. The Irish method by its design includes a large number of cases not trimmed by the bottom up costing as, the former trims extreme high and low value cases, whereas, the bottom up method only trims low value cases.

Key findings

- The Irish method of trimming results in a larger share of trimmed episodes by design as it seeks to remove extreme values regardless of whether they are valid or not which has implications for setting high boundary points in particular. Therefore, the Irish method is not suitable when the only objective of trimming is to remove spurious values from the actual calculation of the NEP.

2.5.3 Assessment

The following tables summarises the assessment of the alternative method against the assessment criteria. Note that formal assessment of the Irish trimming method has not been undertaken given that the purpose underlying the Irish method is fundamentally different to the purpose of trimming in the NEP context. However, further discussion of the Irish method is provided in the section on boundary points as the Irish trimming method interplays with their boundary point methodology.

Table 3: Assessment of the current data trimming method

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Meets this requirement noting that trimming of spurious data is unlikely to have an impact on this requirement.
Public Hospital efficiency	Partially meets this requirement noting that current method is likely to leave in the final data set, episodes with improbable low costs which then has the potential to skew boundary points and impact on incentive efficiency drivers (marginal).
Funding transparency	Meets this requirement in that the business rules are specific and easy to understand.
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement noting that trimming as defined aims is to identify and remove poor quality data.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	Meets this requirement noting that trimming methods are unlikely on their own to create perverse incentives.
Gaming	Meets this requirement noting that trimming methods are unlikely on their own to be gamed.
Level 2: specific to NEP feature	
Reliability in identifying episodes with spurious values	One of the four business rules sets the threshold value too low namely ‘Cost lower than $23’; the probable minimum cost of an acute episode is likely to be far greater than $23. Thus the current method is likely to include episodes with improbable cost values which increases risk of distorted cost weights and skewed NEP.
Level 3: macro criteria	
R^2	79.26%
Cost Ratio	1.0
SMAPE	9.39%

Table 4: Assessment of the bottom up costing data trimming method

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Meets this requirement noting that trimming of spurious data is unlikely to have an impact on this requirement.
Public Hospital efficiency	Meets this requirement noting that what actually constitutes the minimum plausible value needs to be validated.
Funding transparency	Meets this requirement in that the business rules are specific and easy to understand.
Level 1: feasibility	
Data availability	Meets this requirement.
Data Quality	Meets this requirement noting that trimming as defined aims is to identify and remove poor quality data.
Data Volume	Meets this requirement.
Level 1: risks	
Perverse incentives	Meets this requirement noting that trimming methods are unlikely on their own to create perverse incentives.
Gaming	Meets this requirement noting that trimming methods are unlikely on their own to be gamed.
Level 2: specific to NEP feature	
Reliability in identifying episodes with spurious values	This method is more robust in setting a minimum threshold value as it is based on the minimum plausible cost. The value used in this testing could be revised through consultation with jurisdictions and hospitals. The decision then becomes whether the threshold is set at the agreed plausible minimum cost of a fraction (e.g., 80%) of this value. For the purposes of illustrating this concept, the actual derived minimum plausible value was used.
Level 3: macro criteria	
R^2	79.02%
Cost Ratio	1.0
SMAPE	12.69%

2.5.4 Conclusions and implications

The current approach to trimming needs to be revised to introduce a more realistic minimally plausible value. The bottom up approach testing provides one such option. The value used in this review was chosen to illustrate the alternative method and needs to be validated with the potential of setting it dynamically to take into account the variable nature of distributions across DRGs. There are three a number of options to setting the minimally plausible cost dynamically such as:

1 Set it for each DRG taking into account the length of stay and significant consumable costs.
2 Set it for each DRG taking into account length of stay only.
3 Set it as a multiplicative value of the unadjusted mean cost (e.g. 10 per cent).

Based on the data used in the analysis, the implications of changing the approach to setting a realistic minimum plausible cost value are:

1 Significantly more episodes would be removed from the calculation of the NEP unless there is a commensurate improvement in the quality of costing data.
2 There are likely to be DRGs with a relatively high volume of spurious data (episodes with a cost greater than $23 but less than the agreed minimally plausible value) which would be impacted (cost weights) by the application of the suggested method.
3 The proposed approach is marginally more complex than the current method particularly if a dynamic approach to setting the minimum value is implemented as this will result in the values varying across DRGs.

3 Inlier Boundary points

3.1 Aspects of NEP tested

The testing of the inlier boundary points focused on the following:

a. Assessing whether the key assumption underpinning the current method is still valid
b. Testing the sensitivity of the linier boundary methods to the underlying nature of the distributions
c. Assessing alternative methods to deal with a technical deficiency of the current method.

3.2 Purpose of setting inlier boundary points

While not explicitly stated in IHPA documentation, KPMG understands that outlier mechanisms are used to deal with a number of policy related issues:

- *Payment/funding risk*: without outliers, ABF has a greater level of payment risk for both payer and provider in that some providers could be unfairly underfunded and others unfairly overfunded. Outliers reduce this risk for both providers and payers.
- *Pricing stability:* setting the price for an inlier population removes the sensitivity of the price to changes in outlier populations which tend to be more volatile than the inlier population.
- *Promoting efficiency:* without an outlier mechanism, the price will be materially influence by long stay outliers and thus moving the price towards the inefficient providers undermining the national objective of efficiency.

Note that there are also other mechanisms in the NEP process that deal with pricing stability which will be examined in future stages of this review.

3.3 Current approach to setting inlier boundary points

The current approach used by IHPA uses a multiplicative value of the mean length of stay to set inlier boundary points. This includes:

- For mental health AR-DRGs (MDCs 19, 20) and AR-DRGs with highly variable length of stay distributions (21 AR-DRGs), the low inlier boundary point is (mean length of stay/1.5) and the high boundary point is set at (mean length of stay*1.5).
- For all other AR-DRGs, the low inlier boundary point is (mean length of stay/3) and the high boundary point is set at (mean length of stay*3).

For simplicity, the term 'L3H3' is used in this report to refer to the current IHPA method.

It is also important to note that the method excludes same-day episodes occurring in AR-DRGs designated for a separate same-day payment, and uses LOS adjusted to remove ICU days for ICU-unbundled AR-DRGs.

3.4 Test Plan

The following specific issues were considered for testing:

Research question #1: Does the original assumption underpinning the decision to use L3H3 still hold, namely that the coefficient of variation (CV) lies between 0.25 and 1.5?

- This research question was included as one of the key rationales/assumptions for choosing the L3H3 method was that the CV for length of stay lies between 0.25 and 1.5. The underlying rationale was more driven by consideration of funding of long stay rather than short stay outliers and specifically that the H3 method would deliver too many long stay outliers if the CV exceeded

1.5 and would deliver too few if it fell below 0.25. If this is no longer the case, then it brings into question as to whether the current method is still appropriate.

- A specific criterion was used to assess this issue rather than the overarching assessment framework given this is an exploratory issue.

Research question #2: How sensitive is the current method (predominantly L3H3) to variation in CV?

- This research question was included as given the influence of extreme values on measures of central tendency and dispersion (which form most boundary point methods), then the sensitivity or otherwise to the CV becomes an aspect in deciding on the optimum boundary point method.
- A specific criterion was used to assess this issue rather than the overarching assessment framework given this is an exploratory issue.

Research question #3: What is the impact of estimating the boundary points using alternative approaches?

- The methods chosen for testing were selected because they address the technical deficiencies of the current method (discussed further later on in this section) and the methods are used internationally. The actual methods are described later on in this section).
- The overarching assessment framework described in 1.3 is used to assess the alternative methods.

3.5 Results and discussion

3.5.1 Validity of assumption regarding CV

Research question #1: Does the original assumption underpinning the decision to use the L3H3 still hold, namely that the coefficient of variation (CV) lies between 0.25 and 1.5?

We analysed 5,316,637 episodes across 798 DRGs from acute admitted services for 2015/16. We found that 681 (85.3%) of DRGs covering 3,588,658 (67.5%) episodes were within the 0.25-1.5 range. Nine (1.1%) DRGs covering 1,285,410 (24.2%) episodes had CVs below 0.25, while 100 (12.5%) of DRGs accounting for 372,145 (7.0%) episodes had CVs above 1.5.

Table 5: Summary of length of stay distributions

	Coefficient of variation (length of stay)		
	<0.25	**Within 0.25-1.5 range**	**>1.5**
DRG (count) #	9	681	100
DRG (%)	1.1%	85.3%	12.5%
Episode (count)	1,285,410	3,588,658	372,145
Episode (%)	24.2%	67.5%	7.0%

It was not relevant to calculate a coefficient of variation for 8 DRGs (same day procedures) covering 70,424 episodes as these had standard deviations equal to zero (see Appendix 3 for list).

The profiles of key DRGs with CV values below 0.25 (Table 6) and above 1.5 (Table 7 ND Table 8) are provided below. Haemodialysis (DRG L61Z) has a CV of 0.06 and accounts for 1,072,909 (20.1%) of all episodes followed by chemotherapy that has a CV of 0.0 and accounts for 3.5% of all episodes. These two AR-DRGs are essentially sameday DRGs and thus the low CV values are not surprising. The remaining AR-DRGs with low CVs account for a very small proportion of the entire population (less than 0.2%).

Table 6: DRGs with coefficient of variation less than 0.25

DRG code	DRG description	CV	Freq	Share
R63Z	Chemotherapy	0.00	184,274	3.5%
M63Z	Male Sterilisation Procedures	0.02	2,981	0.1%
N11B	Other Female Reproductive System GIs, Minor Complexity	0.06	2,147	0.0%

DRG code	DRG description	CV	Freq	Share
L61Z	Haemodialysis	0.07	1,072,909	20.2%
B62Z	Apheresis	0.08	451	0.0%
C03B	Retinal Procedures, Minor Complexity	0.14	4,033	0.1%
N07B	Other Uterus and Adnexa Procedures for Non-Malignancy, Minor Complexity	0.14	10,931	0.2%
Q61C	Red Blood Cell Disorders, Minor Complexity	0.16	6022	0.1%
C10Z	Strabismus Procedures	0.20	1,662	0.0%
O66C	Antenatal and Other Obstetric Admissions, Minor Complexity	0.26	17,931	0.3%

Table 7: DRGs with coefficient of variation greater than 1.5 (top ten ranked by CV)

DRG code	DRG description	CV	Freq	Share
D15Z	Mastoid Procedures	2.62	1,042	0.0%
L08A	Urethral Procedures, Major Complexity	2.65	167	0.0%
E75B	Other Respiratory System Disorders, Minor Complexity	2.82	16,565	0.3%
U67B	Personality Disorders and Acute Reactions, Minor Complexity	3.24	14,157	0.3%
Z65Z	Congenital Anomalies and Problems Arising from Neonatal Period	3.35	286	0.0%
A13A	Ventilation >=336hours, Major Complexity	3.52	829	0.0%
I23A	Local Excision & Removal of Internal Fixation Device, Except Hip & Fmr, Maj Comp	3.54	1,542	0.0%
B67B	Degenerative Nervous System Disorders, Intermediate Complexity	3.72	4,215	0.1%
K12Z	Other Bariatric Procedures	4.77	205	0.0%
B82A	Chronic & Unspec Para/Quadriplegia, Major Complexity	5.80	637	0.0%
Z64B	Other Factors Influencing Health Status, Minor Complexity	6.36	46,975	0.9%

Table 8: Top 10 DRGs with coefficient of variation greater than (top ten ranked by frequency)

DRG code	DRG description	CV	Freq	Share
Z64B	Other Factors Influencing Health Status, Minor Complexity	6.36	46,975	0.9%
E65A	Chronic Obstructive Airways Disease, Major Complexity	1.80	18,923	0.4%
O66A	Antenatal and Other Obstetric Admissions, Major Complexity	1.50	17,965	0.3%
U61B	Schizophrenia Disorders, Minor Complexity	2.55	16,667	0.3%
E75B	Other Respiratory System Disorders, Minor Complexity	2.82	16,565	0.3%
I75B	Injuries to Shoulder, Arm, Elbow, Knee, Leg and Ankle, Intermediate Complexity	1.72	15,703	0.3%
U67B	Personality Disorders and Acute Reactions, Minor Complexity	3.24	14,157	0.3%
B81B	Other Disorders of the Nervous System, Minor Complexity	2.09	14,030	0.3%
X62A	Poisoning/Toxic Effects of Drugs and Other Substances, Major Complexity	1.90	13,193	0.2%
X63B	Sequelae of Treatment, Minor Complexity	1.70	13,090	0.2%

Key findings

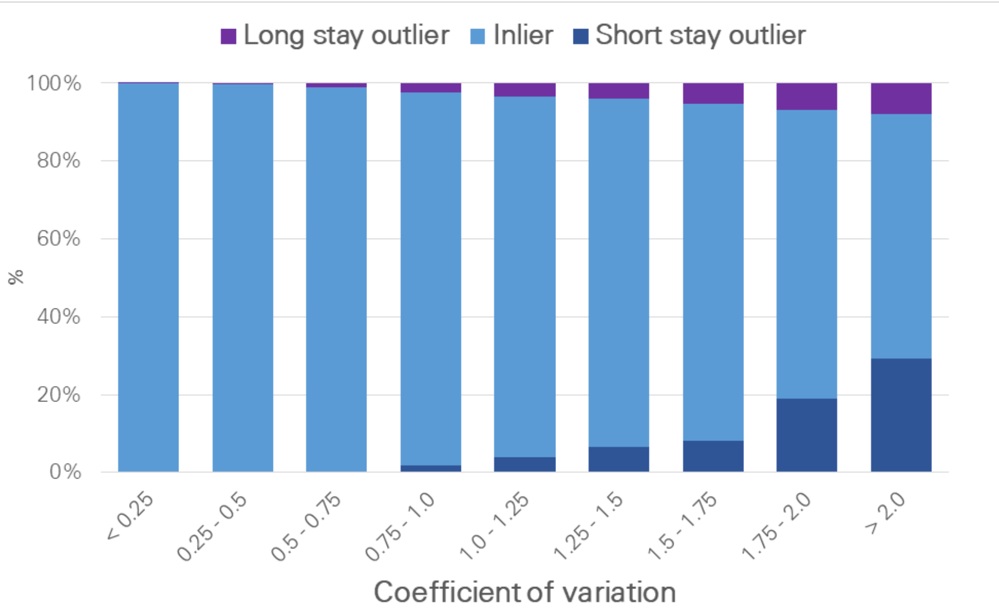
- The original assumption that led to the use of early use of the L3H3 method in Australia largely stands particularly if sameday dialysis episodes and chemotherapy (also a sameday DRG) are removed from the analysis.
- DRGs that have a CV value less than the low threshold CV of 0.25 are largely short stay episodes, while DRGs with a CV value that exceeds the upper threshold value of 1.5 are largely DRGs that by their construction, are heterogeneous e.g. 'other' DRGs and tend to have small sample sizes.

3.5.2 Sensitivity of L3H3 to CV

Research question #2: How sensitive is the current predominantly L3H3 method to variation in CV?

We grouped the DRGs by their CV in bands of 0.25 from 0 to 2, and calculated the share of short stay and long stay outliers.

Figure 2: Outliers versus coefficient of variation for current NEP inlier method



The key results using actual data show that:

- The share of DRG outliers increases with CV. These results are intuitive in that DRGs with low CV values have few outliers and the proportion of episodes within a DRG that are deemed outliers generally increases as CV increases.
- Short stay outliers increased more markedly than long stay outliers as CV increases particularly once CV is more than 1.75.

Further analysis highlights that there are certain DRGs with high CVs that have the majority of short-stay outliers. Schizophrenia disorders, both minor and major complexity, have a CV of 2.6, lower boundary points of 13 and 31 days, and short-stay outliers of almost 10,000 and 4,000 episodes respectively; minor personality disorders have a CV of 3.2, a lower boundary point of 3 days, and almost 7,000 short-stay outliers. Collectively, these three DRGs make up 92% of the short-stay inliers for DRGs with a CV greater than 2. These three DRGs are mental health DRGs for which currently L1.5H1.5 approach is used to set boundary points because of the length of stay characteristics discussed above.

More generally, this suggests that the non-normal nature of the distribution of DRGs can be problematic particularly for DRGs with high values of CV. The current method uses the arithmetic mean as the measure of central tendency which is highly sensitive to skewed distributions and to distributions which are more volatile. The large majority of DRGs have a non-normal length of stay distribution. Many DRG distributions tend to be skewed to the right with long length of stay tails resulting in the arithmetic mean generating an 'inflated' value of the central tendency.

There are a number of options of dealing with this matter which have been adopted in other countries that have ABF arrangements in place. This matter is considered in the next section. There are clear implications for outlier policy.

The NEP needs to be set so that it encourages efficiency without penalising providers whose high length of stay has more to do with patient variation than inefficiency. There is an argument that long stay boundary point should be set such that it retains the inefficient length of stay cases within the inlier population leaving natural patient variation as the outliers. This is of course difficult to achieve through a pure statistical method.

Key findings

- A preferred boundary setting method would address the following:
    1. Deal with the non-normality of the distributions
    2. Deal with extreme values
    3. Minimise volatility
    4. Set the high boundary point high enough that captures inefficient practices within the inlier population.
    5. Rewards efficiency.
    6. Promote equity.
- The first three of these matters suggests the use of a method different to the current approach (four options are considered in the next section). The other matters are more a matter of judgement than pure technical considerations.

3.5.3 Alternative methods

Research question #3: What is the impact of estimating the boundary points using alternative approaches?

Three methods were chosen as each of these deals with the non-normality issues albeit in different ways and are currently used in other countries:

- Log transformation of the data using standard deviation as the measure of dispersion (the Irish method[8]): boundary points are set 2 standard deviations above and below the mean of transformed data. Note that the Irish trimming method was also applied as it is integral to the method used to derive boundary points (refer to the earlier discussion in section 2.4). The influence of extremely high values on boundary points was one of the reasons for using the average length of stay in Australia as the additive factor rather than standard deviation (which had been previously used in Australia)[9].
- Function of the inter quartile range (northern European method[10]): Q75+(Q75-Q25)*1.5.
- Dynamic model that sets the inlier boundary points based on type of length of stay distribution and includes rules for eliminating 'extreme' values from the calculation of the boundary points.

---

[8] Fiachra Bane (20165). *Introduction to the Price Setting Process for Admitted Patients V1.0 26 May 2015*. Healthcare Pricing Office, HSE.
[9] In the early years of what was then referred to as casemix funding, states such as South Australia and Victoria used standard deviation to set boundary points e.g. ALOS +/- a multiplicative value of the SD. Work commissioned by the Commonwealth in the early 1990's led to a recommendation that eventually led to the use of the current method or L3H3 that replaced SD with ALOS.
[10] Stephani V, Quentin W, Van den Heede K, Van de Voorde C, Geissler A. Payment methods for hospital stays with a large variability in the care process. Health Services Research (HSR) Brussels: Belgian Health Care Knowledge Centre (KCE). 2018. KCE Reports 302. D/2018/10.273/36.

Note that this last method first identifies whether the length of stay distribution fits one of four statistical distributions but if not, then the current boundary point method is applied. Where the distribution matches one of these nominated distributions a boundary point method specific to the distribution is then applied. The full technical description of the dynamic model is contained in Appendix 4.

One other method was assessed namely the Portuguese method[11] which uses the geometric mean. This method sets high boundary points and at 2 standard deviations above the geometric mean. However, it does not consider low boundary points and thus was not considered in the final testing. The results of the exploratory analysis of this method are presented in an Appendix 5. Some exploratory analysis was also undertaken comparing the Irish method with and without its associated trimming method the results of which are shown in Appendix 6.

The results for the dynamic model are not provided in detail because testing found that it was indistinguishable from the current model. Across the 798 DRGs, the dynamic model identified 2 normal distributions, 6 log-normal distributions and 2 multi-modal distributions. Further these DRGs have small populations which account for less than 0.1 % of all episodes. Thus the majority of DRGs and the overwhelming majority of episodes have the same boundary points as the current method. The threshold critical values set to test for the nature of the distributions were varied but the overall results did not improve. Specifically, in the 'PROC Univariate' test of distribution, when the p-value is less than the critical value, you reject the null hypothesis and conclude that the data did not come from the specified distribution. Even at a critical value of 0.01, less than 2% of all DRGs were found to be normal, log-normal or multi-modal.

The dynamic model also incorporated a mechanism for excluding highly influential episodes, based on extremely high lengths of stay. In essence, any episode meeting the both of the following two criteria were excluded:

- length of stay is higher than the 98th percentile for the DRG; and
- excluding the observation would decrease the DRG average length of stay by at least 5%.

Note that in almost all DRGs, episodes meeting the second criterion already satisfied the first.

This influential outlier trimming mechanism had no material impact on the results, excluding only two episodes.

For reference, the DRGs identified through the dynamic method, and the adjusted boundary points derived via the dynamic method, are shown in the Table below.

Table 9: DRGs identified by the dynamic model as normal, log normal or multi-modal.

DRG	Episodes	Distribution	Current method inlier boundary		Dynamic method inlier boundary	
			Lower	Upper	Lower	Upper
B42A	334	Log normal	6	48	3	58
F11A	294	Log normal	13	114	9	104
F40A	271	Log normal	3	31	2	39
H06A	265	Log normal	14	29	5	49
K01A	420	Log normal	11	92	7	82
P05A	79	Log normal	37	71	8	128
P08Z	25	Normal	35	263	12	162
P62A	425	Normal	24	217	16	122
P63A	126	Multi modal	12	101	10	61
P64A	207	Multi modal	11	101	10	58
Total	2,446					

[11] Alberto et. al. (2012). *Factors influencing hospital high length of stay outliers.* British medical Journal Health Services Research (2012), 12:265.

While many of the DRG length of stay distributions 'appear' to be log normal, they do not fulfil the technical requirements that define this statistical distribution. Notwithstanding this, many of the international methods assume that the distributions are more log normal than normal and consequently adopt inlier methods consistent with this assumption.

It is evident from the literature that most ABF payers other than Australia have adopted a method to directly deal with the non-normal nature of the distribution and to deal with extremely high values of length of stay in a systemic manner. It is also evident in the literature that the methods used to deal with these issues vary across other countries. We applied the Irish and the interquartile range methods on 5,316,637 episodes across 798 DRGs for acute admitted episodes from the 2015/16 dataset. Table 10 provides a summary of the comparison between the methods for lower and upper boundary limits.

Table 10: Comparison of boundary limits using current NEP trimming method

	Current	Irish	IQR
Low			
Mean (days)	3	2	2
Median (days)	1	1	1
Min (days)	1	1	1
Max (days)	45	22	30
Relative to the current method, share of DRGs where boundary:			
Decreased	[redacted]	35%	43%
Stayed the same		61%	57%
Increased		4%	0%
High			
Mean (days)	22	21	16
Median (days)	13	13	11
Min (days)	2	2	2
Max (days)	363	289	155
Relative to the current method, share of DRGs where boundary:			
Decreased	[redacted]	59%	91%
Stayed the same		10%	5%
Increased		31%	4%

The underlying nature of the length of stay distributions influences the impact of these methods on whether the high boundary point increases or decreases.

Table 11: mean CV values by direction of change in high boundary point

Direction of change in boundary value	Mean CV value	
	Irish	IQR
Decreased	0.93	1.09
Increased	1.20	1.13

The above statistics indicate that DRGs with less dispersed length of stay distributions have a lower high boundary point whereas DRGs with more dispersed distributions have an increased value of the high boundary point with the Irish method having the most marked differentiation namely, the mean CV value of DRGs that had a decrease in high boundary point is much less than the mean CV value for DRGs that had an increase in high boundary point compared to the other two methods. Thus, the Irish method appears to be more sensitive to the underlying distribution compared to the IQR method and the current NEP method.

To test this, we re-ran the analysis shown Figure 2 (page 12 of this report) using the full range of methods considered in this review, the results of which are shown in Appendix 7. The results are very similar: the alternative methods follow a very similar trend of increasing outliers with increasing CV, however the Portuguese and Irish methods have less short-stay outliers (19% and 22% for the

Portuguese and Irish methods respectively, relative to 29% for the current method); the Portuguese method has more long stay outliers than the current method (19% compared to 8%) while the Irish method has less (4% to compared 8%) in DRGs with high CV.

While both the Irish and the IQR method deals with the non-normality, they do so in different ways and then set the multiplier values differently. Thus, there are a series of issues to consider:

1. How to deal with the non-normal nature of the distributions.
2. The multiplicative factor that to use when deriving the boundary points
3. Whether a single method is sufficient given that while many of the distributions are skewed (right) many others display some form of heteroscedasticity and other distributions are relatively flat in nature.

The common outcome of both the Irish and the IQR method is that the high boundary point remains the same in very few DRGs. The Irish method tends to set the high boundary point higher than the IQR and the current method which is possibly more a result of using 2 standard deviations than log transforming the data. The IQR method results mostly in a reduction in high boundary points (91%) of DRGs compared to the other two methods.

Key findings

- The IQR method decreases the low and high boundary points relative to the current method in the majority of cases; while the Irish method has a mixed result (increase and decrease) for high boundary points and decreases the low boundary point (based on NEP trimming method).
- These alternative methods are also broadly resilient to changes in underlying distributions, but will generate high volumes of outliers when the underlying distribution is particularly spread (i.e. has a high coefficient of variation) if the current trimming method is maintained.
- There is a critical interdependency between trimming and setting boundary points. This is a caveat of the current analysis that requires further investigation.
- A dynamic method is theoretically possible and arguably technically superior, however, in practice it is indistinguishable from the current NEP method.

The Irish, the IQR and dynamic method deal with the technical deficiencies discussed earlier in this section, namely, non-normality and episodes with extreme length of stay values. Which method is 'better' is first driven by decision as to whether or not to trim extreme values or use some other method to deal with extreme values and then policy considerations regarding outliers. Regardless of which core method is used (i.e. log transforming or different measures of central tendency), the two critical issues are:

- Should a different method be used for DRGs which display heteroscedasticity?
- How conservative or aggressive should the boundary points be set (i.e. the value of the multiplicative parameter)?

There are examples of ABF payers recognising that a universal method of determining outliers has its limitations. For example:

- France selectively removes low boundary points based on a policy decision to encourage alternative care models[12].
- Victoria introduced L2H2 for DRGs dispersed length of stay distributions similar to what currently occurs with the NEP.

[12]Stephani V, Quentin W, Van den Heede K, Van de Voorde C, Geissler A. Payment methods for hospital stays with a large variability in the care process. Health Services Research (HSR) Brussels: Belgian Health Care Knowledge Centre (KCE). 2018. KCE Reports 302. D/2018/10.273/36.

- Germany and Ireland set a maximum value for the high boundary point as a way of desensitising boundary points to extreme values and volatile nature of the length of stay distributions in some DRGs[13].

Many countries have taken a more conservative approach to setting the high boundary point to 'bring in' the upper bound closer to the selected measure of central tendency. Examples, include France uses 2.5 times the arithmetic mean and Germany and Ireland set a maximum value for their high boundary point.

3.5.4 Hospital level impact

A number of additional analyses were undertaken to assess the impact of the alternative methods on the cost ratio at the individual hospital level:

- Various cost ratio metrics such as share of hospitals with cost ratios greater than 1.05 or less than 0.95
- Various change in cost ratio metrics such as share of hospitals that moved from a cost ratio less than one (actual costs lower than predicted costs) to a cost ratio greater than one (actual costs greater than predicted costs)
- Scatter plots of showing the hospital cost ratios under the current and alternative methods
- Regression analysis to identify if any particular types or peer groups of hospitals were specifically impacted by the alternative methods.

Key findings

- Both the IQR and Irish methods cause meaningful changes to cost ratios at the hospital level with a relatively small proportion of hospitals, less than 5%, have cost ratios that change signs.
- Regression analysis does not find any evidence that any specific groups of hospitals are more or less impacted by the alternative methods.

The analytical results that support these findings are presented in the remainder of this section.

The following table illustrates that The IQR method more so that the Irish method impacts on the cost ratio for individual hospitals at the aggregate level with less hospitals having a cost ratio greater than 1.05 but more hospitals having a ratio less than 0.95.[14]

[13] Stephani op cit.
[14] A .05 threshold was selected as being indicative of a material impact on funding i.e., cost ratio greater than 1.05 or less than 0.95.

Table 12: summary cost ratio results

Statistical parameter	Current NEP (Actual value)	IQR method	Irish method
Net cost ratio	1.000	1.000	1.000
Hospitals with cost ratio greater than 1.02	43%	39%	42%
Hospitals with cost ratio less than 0.98	48%	50%	47%
Hospitals with cost ratio greater than 1.05	35%	31%	35%
Hospitals with cost ratio less than 0.95	38%	40%	39%
Maximum cost ratio	2.534	2.605	2.598
Median cost ratio	0.988	0.983	0.987

Both methods however, result in a material number of hospitals whose cost ratio changed by more than 0.05 namely, 13% of hospitals under the IQR method and 10% of hospitals under the Irish method (Table 13).

Table 13: impact of alternative models on cost ratio

Statistical parameter	IQR method	Irish method
Hospitals with a cost ratio that has changed from >1 to <1	4%	2%
Hospitals with a cost ratio that has changed from <1 to >1	1%	2%
Hospitals whose cost ratio has changed by more than |0.05|	13%	10%
Hospitals whose cost ratio has changed by more than |0.05| and increased in value	3%	5%
Hospitals whose cost ratio has changed by more than |0.05| and decreased in value	10%	5%

The figures below provide a scatter plot comparing the results for individual hospitals which was then subjected to regression analysis to assess whether the impact on individual hospitals was associated with any characteristic of the hospital.

Figure 3: Cost ratio scatter plot for IQR compared to current method

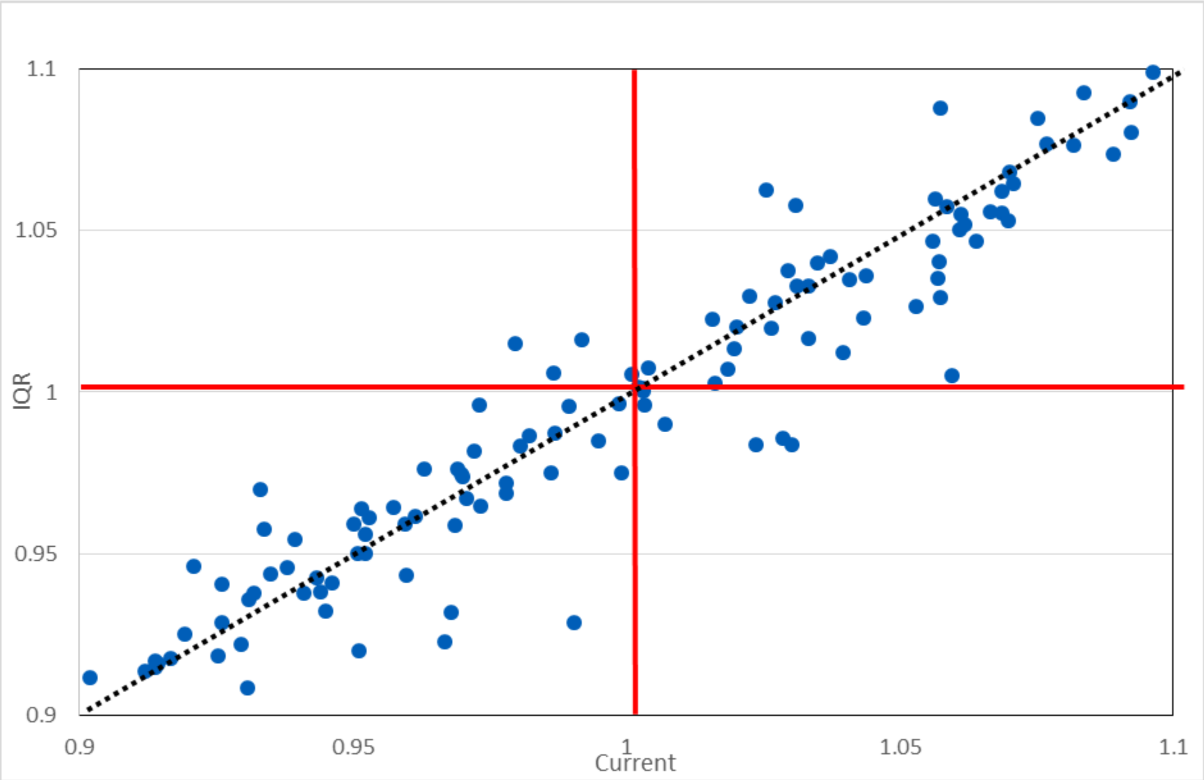
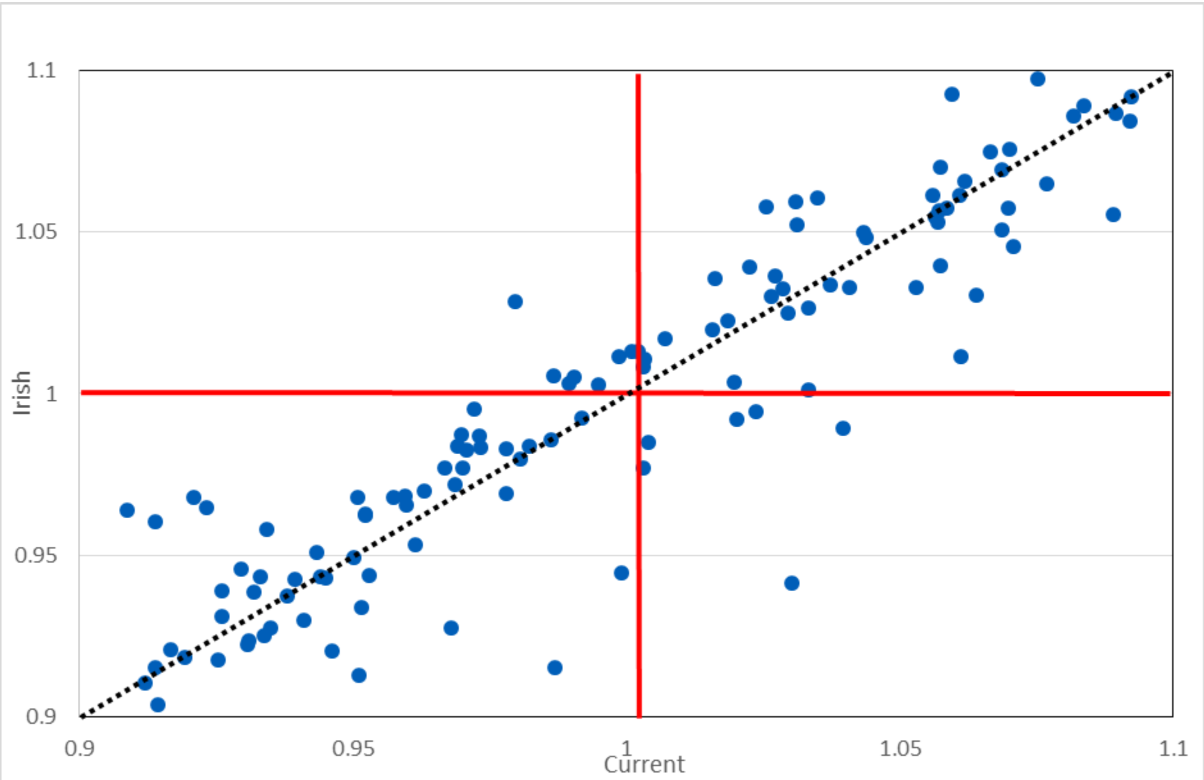


Figure 4: Cost ratio scatter plot for Irish compared to current method



A series of regression analyses were conducted to test if certain hospital types were more or less impacted by the alternative methods. Below are coefficients for each peer group from a simple linear regression with cost ratio. The coefficient highlights how the cost ratio for each peer group varies relative to the average ratio of one; if there was no difference in cost ratios between hospitals, all the coefficients would be zero.

Across the board, the alternative methods have a relatively small impact with the exception of the IQR method impact on E4 (rehabilitation) and D1 (small regional) hospitals. These hospitals have relatively high cost ratios under the current method that are tempered somewhat under the IQR method.

Table 14: Coefficients from a linear regression of hospital peer group to cost ratio[15]

Peer group	Current NEP (Actual value)	IQR method	Irish method
A1	-0.030	-0.040	-0.043
A2	0.006	-0.011	-0.004
B1	-0.029	-0.038	-0.046
B2	0.033	0.032	0.027
C1	-0.005	-0.015	-0.021
C2	-0.064	-0.085	-0.083
D1	0.448	0.385	0.454
D2	-0.127	-0.128	-0.111
D3	0.027	0.018	0.021
E4	0.335	0.100	0.238
E9	-0.012	-0.093	0.000
F	-0.098	-0.084	-0.065
G	-0.058	-0.183	-0.087

3.5.5 Conclusions and implications

Summary

The main conclusions for the NEP arising from this review are:

1 There is a technical argument supporting the need to adopt a method that deals with the non-normality of length of stay distributions.

2 There is a technical need to reduce the sensitivity of high boundary point methods to extremely high values.

3 A more conservative approach is required to set the upper bound particularly for DRGs with high CV values, such that the upper bound is 'closer' to the central tendency than the current method. One option is to extend the application of the alternative factor used for MDCs 19 and 20 ((mean length of stay/1.5) and the high boundary point is set at (mean length of stay*1.5)) to AR-DRGs that have display a relatively high CV value (such as 1.5 given that this the value used in the assumption that led to the original decision to use L3H3 or a higher value such as a CV of 2 simply to constrain the number of DRGs where the alternative method is applied). This approach would also deal with the second point above.

4 While none of the alternative methods provide an overall better statistical result compared to the current NEP method, they address the above technical requirements and thus are more likely to impact on the effectiveness of the NEP at the hospital level.

[15] All the hospitals in the data used for the analysis all receive activity based payments under NEP including the rehabilitation and small rural hospitals.

Assessment against the criteria

The following tables summarises the assessment of the alternative methods against assessment criteria.

Table 15: assessment of the current inlier method

Level/Criterion	Assessment
Level 1: effectiveness	
• Patient access to services	Direct impact unlikely. However, the extent to which the current method does not adequately deal with non-normality and extreme values would result in potential skewing of boundary points that could influence patient admission decisions because of perceived (un) profitability of some AR-DRGs.
• Public Hospital efficiency	Direct impact unlikely. However, the extent to which the current method does not adequately deal with non-normality and extreme values would result in potential skewing of boundary points that could mask the true efficiency of a hospital that could then influence hospital practices and hospital efficiency.
• Funding transparency	Meets this requirement as the process for setting boundary points is clear and easily understood.
Level 1: feasibility	
• Data availability	Meets this requirement.
• Data quality	Meets this requirement.
• Data volume	Meets this requirement.
Level 1: risks	
• Perverse incentives	The extent to which the current method does not adequately deal with non-normality and extreme values would result in potential skewing of boundary points that could influence hospital admission and patient care practices because of hospitals' perceptions of efficiency is distorted because of the impact on boundary points and subsequently the NEP.
• Gaming	No specific risks associated with this method over and above the potential for gaming directly associated with setting boundary points.
Level 2: specific to NEP feature	
• No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.26%
Cost Ratio	1.0
SMAPE	9.39%

Table 16: assessment of the Irish inlier method

Level/Criterion	Assessment
Level 1: effectiveness	
• Patient access to services	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any patient access risk associated with the current method.
• Public Hospital efficiency	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any hospital efficiency access risk associated with the current method.
• Funding transparency	Meets this requirement as the process for setting boundary points is clear and easily understood.

Level/Criterion	Assessment
Level 1: feasibility	
• Data availability	Meets this requirement.
• Data quality	Meets this requirement.
• Data volume	Meets this requirement.
Level 1: risks	
• Perverse incentives	This method has less perverse incentives to the extent that it deals with the technical deficiencies of the current method and thus moderates any associated risks.
• Gaming	No specific risks associated with this method over and above the potential for gaming directly associated with setting boundary points.
Level 2: specific to NEP feature	
• No additional metrics identified	n/a
Level 3: macro criteria	
R^2	78.81%
Cost Ratio	1.0
SMAPE	9.50%

Table 17: assessment of the IQR inlier method

Level/Criterion	Assessment
Level 1: effectiveness	
• Patient access to services	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any patient access risk associated with the current method.
• Public Hospital efficiency	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any hospital efficiency access risk associated with the current method.
• Funding transparency	Meets this requirement as the process for setting boundary points is clear and easily understood.
Level 1: feasibility	
• Data availability	Meets this requirement.
• Data quality	Meets this requirement.
• Data volume	Meets this requirement.
Level 1: risks	
• Perverse incentives	This method has less perverse incentives to the extent that it deals with the technical deficiencies of the current method and thus moderates any associated risks.
• Gaming	No specific risks associated with this method over and above the potential for gaming directly associated with setting boundary points.
Level 2: specific to NEP feature	
• No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.06%
Cost Ratio	1.0
SMAPE	9.45%

Table 18: assessment of the dynamic inlier method

Level/Criterion	Assessment
Level 1: Effectiveness	
• Patient access to services	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any patient access risk associated with the current method.
• Public Hospital efficiency	Conceptually, this method technically addresses the two key deficiencies of the current method (that relate to non-normality and extreme values of length of stay) and thus technically moderates any hospital efficiency access risk associated with the current method.
• Funding transparency	The process is substantially more complex and thus not as easily understood which could impact on the transparency of the NEP.
Level 1: feasibility	
• Data availability	Meets this requirement.
• Data Quality	Meets this requirement.
• Data Volume	Meets this requirement.
Level 1: risks	
• Perverse incentives	This method reduces the potential for perverse incentives to the extent that it deals with the technical deficiencies of the current method and thus moderates any associated risks.
• Gaming	No specific risks associated with this method over and above the potential for gaming directly associated with setting boundary points.
Level 2: specific to NEP feature	
• No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.18%
Cost Ratio	1.0
SMAPE	9.39%

There is little difference between the four methods in terms of the macro level results. The current method has a slightly greater explanatory power – the three alternative methods are within a range of less than 0.5% (absolute) value of current R-Square. Similarly, the current method generates a lower value of residual error compared to the other methods but there is marginal difference across the three methods.

Key findings

- There is minimal difference in the statistical performance of the three alternative methods compared to the current inlier boundary point method.
- There is a modest impact on cost ratios for individual hospitals with no systemic pattern in terms of the characteristics of the hospitals impacted.

4 Price loadings for unavoidable costs

4.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches for price loadings for unavoidable costs in the following aspects of the NEP for *acute admitted care type only*. Testing focused on the following aspects:

a. Profiling episodic cost data to see if there are significant variations for relative frequencies of relevant adjustment factor status variables.
b. Assess the likely materiality of any such variations in terms of potential impact on NEP modelling.
c. Modify the NEP Model to replace the current steps for calculating the Indigenous, Patient Remoteness, Dialysis and Radiotherapy adjustments with three multivariate linear regressions.
d. Run the modified model and assess performance using the agreed metrics.

The testing used 2015/16 acute admitted episodes from hospitals in scope of activity based funding that were used by IHPA to construct the 2018/19 NEP using AR-DRG Version 9.

4.2 Purpose of setting price loadings for unavoidable costs

A legitimate and unavoidable cost is defined as a cost "associated with providing clinically appropriate public hospital services, where there are legitimate differences between patient and for provider groups, and which are unavoidable"[16] and by inference which are not reflected or cannot be accommodated in patient classifications that are used in the NEP. Currently, there are price loadings for the following:

- Admitted Acute Model:
  - Paediatric Adjustment;
  - Specialist Psychiatric Age Adjustment;
  - Patient Residential Remoteness Area Adjustment;
  - Indigenous Adjustment;
  - Radiotherapy Adjustment;
  - Dialysis Adjustment;
  - Patient Treatment Remoteness Area Adjustment; and
  - Intensive Care Unit (ICU) Adjustment.
- Subacute Admitted Model:
  - Patient Residential Remoteness Area Adjustment; and
  - Indigenous Adjustment.
- Non-admitted Model:
  - Indigenous Adjustment; and
  - Multidisciplinary Clinic Adjustment.
- Emergency Department (ED) Model:
  - Patient Residential Remoteness Area Adjustment;
  - Indigenous Adjustment; and

[16] Assessment of Legitimate and Unavoidable Cost Variations Framework. Version 2.2 May 2018, IHPA, pp4.

- Emergency Care Age Adjustment.

This paper focuses on the loadings in the Acute Model, which has the most complex loading arrangements and calculation methods. We note also that some of the loadings calculated for the Acute Model are used in other models.

4.3 Current approach to setting price loadings

IHPA has a framework and a prescribed process that governs its assessment of potential areas of legitimate and unavoidable costs. The administrative process is as follows:

Stage	Process Details
Stage 1:	(1a) Jurisdiction determines that it meets the criteria for assessment
Request for assessment	(1b) Jurisdiction requests an assessment by IHPA
Stage 2:	(2a) IHPA reviews the request and evidence provided
Assessment	(2b) IHPA provides notification of the request to all jurisdictions
	(2c) IHPA undertakes the assessment
Stage 3:	(3a) IHPA determines the decision
Draft decision	(3b) IHPA drafts the decision and provides to the jurisdictions
	(3c) IHPA reviews the written comments by the jurisdictions with regards to the draft decision
Stage 4:	(4a) IHPA drafts the final decision and provides to the jurisdictions
Final decision	(4b) IHPA refines the Pricing Framework if there are legitimate and unavoidable variations in costs

The actual method for determining the loading factor involves a sequential statistical process (using a combination of ratios of average costs and linear regressions) to calculate the price loadings. In our initial review[17] we concluded:

- The sequential process for calculating adjustments effectively ignores potentially material interactions between status indicators for the paediatric adjustment and subsequent adjustments (e.g. between dialysis status and Indigenous status).
- This risks some adjustments being inflated or deflated for significant subgroups of the episode population. e.g., if there is a higher proportion of Indigenous status episodes among dialysis episodes than non-dialysis episodes, then the Dialysis Adjustment may be inflated.
- The risk is focused on the following adjustments:
  - Indigenous Adjustment;
  - Radiotherapy Adjustment;
  - Dialysis Adjustment; and
  - Patient Remoteness Adjustment.

The Paediatric Adjustment and the Treatment Remoteness Adjustment were excluded, based on the multiplicative manner of their application in the NEP. This differs from the additive application of the included adjustments listed above.

[17] This formed part of the exploratory work reported in the literature review report.

These findings indicated there is value in confirming whether or not material interactive effects are present in the data. If they are present, then an alternative method(s) for calculating unavoidable costs' adjustments should be explored. Any alternative(s) should directly address the identified interactions.

4.4 Test plan

Given the identified risk of uncontrolled interactions in the NEP Model, further consideration of potential to improve this aspect of the NEP Model's performance should be based on more detailed diagnostic analysis.

Research question #1: Are there material relationships between the patient, episode and hospital attributes used to calculate adjustments for unavoidable costs?

The following approach was adopted:

- Profile episodic cost data from the NEP Acute Model to see if there are significant variations in relative frequencies of relevant pairs of adjustment factor variables. Indigenous status and Patient Remoteness status were not profiled against each other as the current NEP Model already controls for variability in these variables when calculating the relevant adjustments.
- Where (if) this profiling shows such variation(s), assess its materiality in terms of potential impact on NEP modelling.

Research question #2: Do the material relationships among the profiled status variables affect the NEP model results?

The following approach was adopted:

- Modify the NEP Model in Step 10.1, replacing the current steps for calculating the Paediatric, Specialist Psychiatric Age, Indigenous, Remoteness, Dialysis and Radiotherapy adjustments with different mixtures of existing steps and multivariate linear regressions. The four scenarios tested were:
    1. Paediatric Adjustment and Specialist Psychiatric Age Adjustment steps unchanged, with the Radiotherapy, Dialysis, Patient Residential Remoteness and Indigenous Adjustments calculated using three related multivariate regression models.
    2. Paediatric Adjustment and Specialist Psychiatric Age Adjustment steps unchanged, with the Radiotherapy, Dialysis, Patient Residential Remoteness and Indigenous Adjustments calculated using a single multivariate regression model.
    3. Paediatric Adjustment step unchanged, with the Specialist Psychiatric Age, Radiotherapy, Dialysis, Patient Residential Remoteness and Indigenous Adjustments calculated using a single multivariate regression model.
    4. All of the adjustments calculated using a single multivariate regression model.
- Run the modified model and assess performance using the agreed metrics.

4.5 Results and discussion

4.5.1 Relationship testing results

Research question #1: Are there material relationships between the patient, episode and hospital attributes used to calculate adjustments for unavoidable costs?

We analysed cross two-way tabulations of the following status variables:

- Indigenous status;

- Radiotherapy status;
- Dialysis status;
- Patient Remoteness – Outer Regional;
- Patient Remoteness – Remote; and
- Patient Remoteness – Very Remote;

In each case the potential for material relationships was assessed by comparing the distributions of one status variable for the two values of the cross tabulated status variable. Table 19 to Table 23 show the cross tabulations used to make this assessment.

Table 19: Number of episodes by Indigenous status and Dialysis status

	Non Dialysis		Dialysis		Total	
Non Indigenous	1,278,916	99.4%	8,056	0.6%	**1,286,972**	**100.0%**
Indigenous	79,848	98.6%	1,156	1.4%	**81,004**	**100.0%**
Total	**1,358,764**	**99.3%**	**9,212**	**0.7%**	**1,367,976**	**100.0%**

Table 20: Number of episodes by Indigenous status and Radiotherapy status

	Non Radiotherapy		Radiotherapy		Total	
Non Indigenous	1,282,545	99.7%	4,427	0.3%	**1,286,972**	**100.0%**
Indigenous	80,900	99.9%	104	0.1%	**81,004**	**100.0%**
Total	**1,363,445**	**99.7%**	**4,531**	**0.3%**	**1,367,976**	**100.0%**

Table 21: Number of episodes by Dialysis status and Radiotherapy status

	Non Radiotherapy		Radiotherapy		Total	
Non Dialysis	1,354,262	99.7%	4,502	0.3%	**1,358,764**	**100.0%**
Dialysis	9,183	99.7%	29	0.3%	**9,212**	**100.0%**
Total	**1,363,445**	**99.7%**	**4,531**	**0.3%**	**1,367,976**	**100.0%**

Table 22: Number of episodes by Patient Remoteness and Dialysis status

	Non Dialysis		Dialysis		Total	
Metropolitan and Inner Regional	1,214,399	99.3%	7,956	0.7%	**1,222,355**	**100.0%**
Outer Regional	123,483	99.2%	1,003	0.8%	**124,486**	**100.0%**
Remote	14,442	99.1%	124	0.9%	**14,566**	**100.0%**
Very Remote	6,440	98.0%	129	2.0%	**6,569**	**100.0%**
Total	**1,358,764**	**99.1%**	**9,212**	**0.9%**	**1,367,976**	**100.0%**

Table 23: Number of episodes by Patient Remoteness and Radiotherapy status

	Non Radiotherapy		Radiotherapy		Total	
Metropolitan and Inner Regional	1,218,379	99.7%	3,976	0.3%	**1,222,355**	**100.0%**
Outer Regional	124,024	99.6%	462	0.4%	**124,486**	**100.0%**
Remote	14,510	99.6%	56	0.4%	**14,566**	**100.0%**
Very Remote	6,532	99.4%	37	0.6%	**6,569**	**100.0%**
Total	**1,363,445**	**99.6%**	**4,531**	**0.4%**	**1,367,976**	**100.0%**

Examination of the tables show:

- Indigenous episodes are more than twice as likely to be Dialysis episodes than Non-Indigenous, although the proportions involved are small;
- Indigenous episodes are three times as likely to be Radiotherapy episodes as Non Indigenous but the proportions involved are very small;
- Dialysis episodes are equally likely to be Radiotherapy episodes as Non Dialysis episodes, and the proportions involved are very small;
- Very Remote patient episodes are more than twice as likely to be Dialysis episodes than other patient episodes, which is likely to be related to the corresponding finding for Indigenous status and Dialysis status;
- Other Patient Remoteness areas are similarly likely to be Dialysis episodes as each other;
- Very Remote patient episodes are a little more likely to be Radiotherapy episodes than other patient episodes, and the proportions involved are small; and
- Other Patient Remoteness areas are similarly likely to be Radiotherapy episodes as each other.

Key findings

- There is some scope for interactions between:
  - Indigenous status with each of Dialysis status and Radiotherapy status;
  - Patient Remoteness with each of Dialysis status and Radiotherapy status.
- It is likely that any interactions involving Patient Remoteness will be limited to Very Remote patient episodes.
- It is unlikely any interactions will be material, given the small numbers of episodes involved.

4.5.2 Impact of the alternative methods

Research question #2: Do the material relationships among the profiled status variables affect the NEP model results?

We modified the NEP Model by replacing current adjustment calculation steps with multivariate linear regression models. Four alternative models were tested, with each alternative integrating more of the adjustments' calculation steps into a single step. The four alternative models are described in Table 24.

Note that none of the regression models used in these alternatives included interaction effects for the adjustments concerned. This was intentional and was to ensure transparency of the resulting NEP Model formulae for calculating NWAU. The inclusion of any interaction effects would significantly complicate the formulae by adding in multiple new adjustments whose purpose and derivation would be difficult for non-technical audiences to interpret and accept.

Table 24: Alternative models tested

Model	Description
Alternative Model 1	Paediatric Adjustment step as in the current NEP Model. Specialist Psychiatric Age Adjustment step as in the current NEP Model. Three consecutive linear regression modelling steps replaced the current steps for calculating Indigenous, Patient Remoteness, Radiotherapy and Dialysis adjustments. The three regressions comprised: 1. Model cost ratio against covariates of : - Indigenous status; - Radiotherapy status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; for inlier episodes in hospitals with at least 100 radiotherapy episodes, and excluding high cost outliers. Calculate Radiotherapy Adjustment from model results. 2. Model cost ratio against covariates of : - Indigenous status; - Dialysis status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; for episodes in hospitals with at least 100 dialysis episodes, and excluding high cost outliers. Calculate Dialysis Adjustment from model results. 3. Model cost ratio against covariates of : - Indigenous status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; - Radiotherapy status; - Dialysis status; for all hospitals, and including high cost outliers. Calculate Indigenous and Patient Remoteness adjustments from model results. The equalisation and calibration steps were retained.

Model	Description
Alternative Model 2	Paediatric Adjustment step as in the current NEP Model. Specialist Psychiatric Age Adjustment step as in the current NEP Model. A single regression modelling step replaced the current steps for calculating Indigenous, Patient Remoteness, Radiotherapy and Dialysis adjustments. The single regression modelled cost ratio against covariates of : - Indigenous status; - Radiotherapy status; - Dialysis status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; for all hospitals, and including high cost outliers. Calculate Radiotherapy, Dialysis, Indigenous and Patient Remoteness adjustments from model results. The equalisation and calibration steps were retained.
Alternative Model 3	Paediatric Adjustment step as in the current NEP Model. A single regression modelling step replaced the current steps for calculating Specialist Psychiatric Age, Indigenous, Patient Remoteness, Radiotherapy and Dialysis adjustments. The single regression modelled cost ratio against covariates of : - Mental health episode, MDC 19 or 20, aged under 18, non-specialist paediatric hospital; - Mental health episode, MDC 19 or 20, aged under 18, specialist paediatric hospital; - Mental health episode, not MDC 19 or 20, aged under 18, non-specialist paediatric hospital; - Mental health episode, not MDC 19 or 20, aged under 18, specialist paediatric hospital; - Age status variable equals 1.1Indigenous status; - Radiotherapy status; - Dialysis status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; for all hospitals, and including high cost outliers. Calculate Specialist Psychiatric Age Adjustments, Radiotherapy, Dialysis, Indigenous and Patient Remoteness adjustments from model results. The equalisation and calibration steps were retained.

Model	Description
Alternative Model 4	A single regression modelling step replaced the current steps for calculating Paediatric, Specialist Psychiatric Age, Indigenous, Patient Remoteness, Radiotherapy and Dialysis adjustments. The single regression modelled cost ratio against covariates of : - Paediatric episode in a specialist paediatric episode hospital, crossed with DRG; - Mental health episode, MDC 19 or 20, aged under 18, non-specialist paediatric hospital; - Mental health episode, MDC 19 or 20, aged under 18, specialist paediatric hospital; - Mental health episode, not MDC 19 or 20, aged under 18, non-specialist paediatric hospital; - Mental health episode, not MDC 19 or 20, aged under 18, specialist paediatric hospital; - Mental health episode, MDC 19 or 20, aged 18 or older; - Age status variable equals 1.1Indigenous status; - Radiotherapy status; - Dialysis status; - Patient Outer Regional status; - Patient Remote status; and - Patient Very Remote status; For all hospitals, and including high cost outliers. Calculate Paediatric, Specialist Psychiatric Age, Radiotherapy, Dialysis, Indigenous and Patient Remoteness adjustments from model results. The equalisation and calibration steps were retained.

Impact on model performance

The alternative models had little or no impact on the adjustments for Patient Remoteness, Treatment Remoteness, Dialysis and Indigenous status (Table 25). However, Alternative Model 1 increase the Radiotherapy Adjustment, while Alternative Model 3 markedly increased the Specialist Psychiatric Age Adjustments. Alternative Model 4 produced further substantial increases for some of these last adjustments.

In considering the changes in Table 25, it is important to note that:

- Alternative Model 1 - Specialist Psychiatric Age Adjustments are multiplied by the sum of the other adjustments and multiplied by the Paediatric Adjustment, in the NEP Model formula.
- Alternative Model 2 - Specialist Psychiatric Age Adjustments are multiplied by the sum of the other adjustments and multiplied by the Paediatric Adjustment, in the NEP Model formula.
- Alternative Model 3 - Specialist Psychiatric Age Adjustments are added to the sum of the other adjustments and multiplied by the Paediatric Adjustment, in the NEP Model formula.
- Alternative Model 4 - all of the adjustments are added together, in the NEP Model formula.

Table 25: Comparison of adjustment values – current Model and alternative models

Adjustment	Current NEP	Alternative Model 1	Alternative Model 2	Alternative Model 3	Alternative Model 4
Specialist Psychiatric Age Adjustment:					
MDC 19/20, Non specialist paediatric hospital	41%	41%	41%	52%	49%
MDC 19/20, Specialist paediatric hospital	15%	15%	15%	25%	56%
Not MDC 19/20, Non specialist paediatric hospital	71%	71%	71%	85%	97%
Not MDC 19/20, Specialist paediatric hospital	58%	58%	58%	99%	142%
Indigenous Adjustment	4%	4%	4%	4%	4%
Patient Remoteness Adjustment:					
Outer Regional	8%	8%	8%	8%	8%
Remote	25%	27%	27%	27%	27%
Very Remote	29%	30%	30%	30%	30%
Radiotherapy Adjustment	32%	38%	41%	41%	40%
Dialysis Adjustment	27%	27%	28%	28%	27%
Treatment Remoteness Adjustment:					
Outer Regional	0%	0%	0%	0%	0%
Remote	8%	6%	6%	6%	6%
Very Remote	12%	11%	11%	11%	11%

The increase in Radiotherapy Adjustment, together with the reductions for Treatment Remoteness Adjustments suggest the Alternative Model 1 works to attribute more cost variation to patient remoteness and radiotherapy, with less unaccounted for variation remaining for attribution to treatment remoteness.

This is likely to be a consequence of the inclusion of the Radiotherapy and Dialysis status variables as covariates in the regression model used to calculate the Indigenous and Patient Remoteness adjustments. By including these additional covariates, the model controlled for variation due to these status variables, reducing the amount of unaccounted for variation in the regression model. It appears this led to slightly higher coefficients for the Remote and Very Remote patient status variables.

The large increases to the Specialist Psychiatric Age Adjustments produced by Alternative Model 3 and 4 were not accompanied by material changes in the other adjustments. At the episode level, the impacts of these large changes will vary, depending on whether other adjustments are applicable or not. This is due to the additive nature of these adjustments in Alternative Models 3 and 4, vis á vis their multiplicative nature in the current NEP Model and Alternative Models 1 and 2.

In terms of overall Model performance, the current Model and each of the alternative models are virtually indistinguishable (see Table 26). The only measures with any differences are the R^2 value and the SMAPE. Both differ only at the second decimal place. These differences are immaterial.

Table 26: Comparison of model performance measures – current Model and Alternative Model 1

Measure	Current NEP	Alternative Model 1	Alternative Model 2	Alternative Model 3	Alternative Model 4
Total cost ratio	0.9998	0.9998	0.9998	0.9998	0.9998
R^2	79.26%	79.25%	79.24%	79.25%	79.18%
MAPE	12.86%	12.86%	12.86%	12.86%	12.86%
SMAPE	9.39%	9.37%	9.37%	9.37%	9.37%

Impact at hospital level

Additional analyses were undertaken to assess the impact of the alternative methods on the cost ratio at the individual hospital level:

- Various cost ratio metrics such as share of hospitals with cost ratios greater than 1.05 or less than 0.95
- Various change in cost ratio metrics such as share of hospitals that moved from a cost ratio less than one (actual costs lower than predicted costs) to a cost ratio greater than one (actual costs greater than predicted costs)
- Scatter plots of showing the hospital cost ratios under the current and alternative methods
- Regression analysis to identify if any particular types or peer groups of hospitals were systematically affected by the Alternative Model 1.

Table 27 summarises the distributions of hospital cost ratios for the current Model and for the alternative models. As with overall impact on the model results, there is very little difference in the hospital level distributions of cost ratios.

Table 27: Summary cost ratio results – comparison of current Model and alternative models

Statistical parameter	Current NEP (Actual value)	Alternative Model 1	Alternative Model 2	Alternative Model 3	Alternative Model 4
Hospitals with cost ratio greater than 1.02	105	106	106	105	107
Hospitals with cost ratio less than 0.98	118	118	117	118	117
Hospitals with cost ratio greater than 1.05	105	106	106	105	107
Hospitals with cost ratio less than 0.95	118	118	117	118	117
Maximum cost ratio	2.5332	2.5343	2.5351	2.5363	2.5383
Median cost ratio	0.9880	0.9884	0.9897	0.9891	0.9896

Investigating the hospital level changes in cost ratios further reinforces the degree of congruence among the results from the current model and the alternative models. Alternative Models 3 and 4 produce the largest individual changes in cost ratios for any hospitals among the alternative models.

Nevertheless, no hospital's cost ratio changes by more than 2.4% across all of the alternative models. Only two hospitals have cost ratio results that move from above 1 to below 1 or vice versa. One hospital goes in each direction for Alternative Model 3, while two hospitals go from below to above in Alternative Model 4. In all four cases, the change in cost ratio is less than 0.005 in magnitude.

Table 28: impact of alternative models on cost ratio

Statistical parameter	Alternative Model 1	Alternative Model 2	Alternative Model 3	Alternative Model 4
Hospitals with a cost ratio that has changed from >1 to <1	0	0	1	0
Hospitals with a cost ratio that has changed from <1 to >1	0	0	1	2
Hospitals whose cost ratio has changed by more than |0.05|	0	0	0	0
Hospitals whose cost ratio has changed by more than |0.05| and increased in value	0	0	0	0
Hospitals whose cost ratio has changed by more than |0.05| and decreased in value	0	0	0	0
Maximum change in cost ratio	0.0091	0.0109	0.0151	0.0239
Largest increase in cost ratio	0.0091	0.0093	0.0100	0.0096
Largest decrease in cost ratio	-0.0073	-0.0109	-0.0151	-0.0239

Viewing scatter plots comparing the results for individual hospitals add to this overall picture of near congruence (see Figure 5 to Figure 8).

Figure 5: Cost ratio scatter plot for Alternative Model 1 compared to current Model

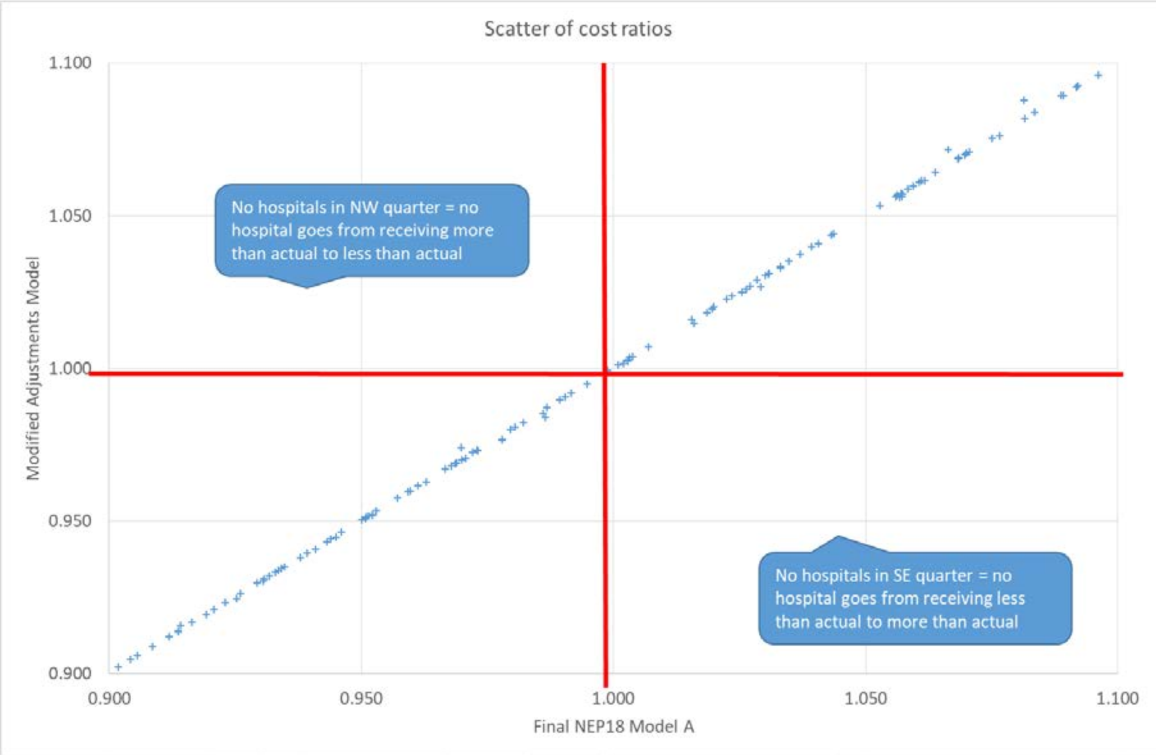


Figure 6: Cost ratio scatter plot for Alternative Model 2 compared to current Model

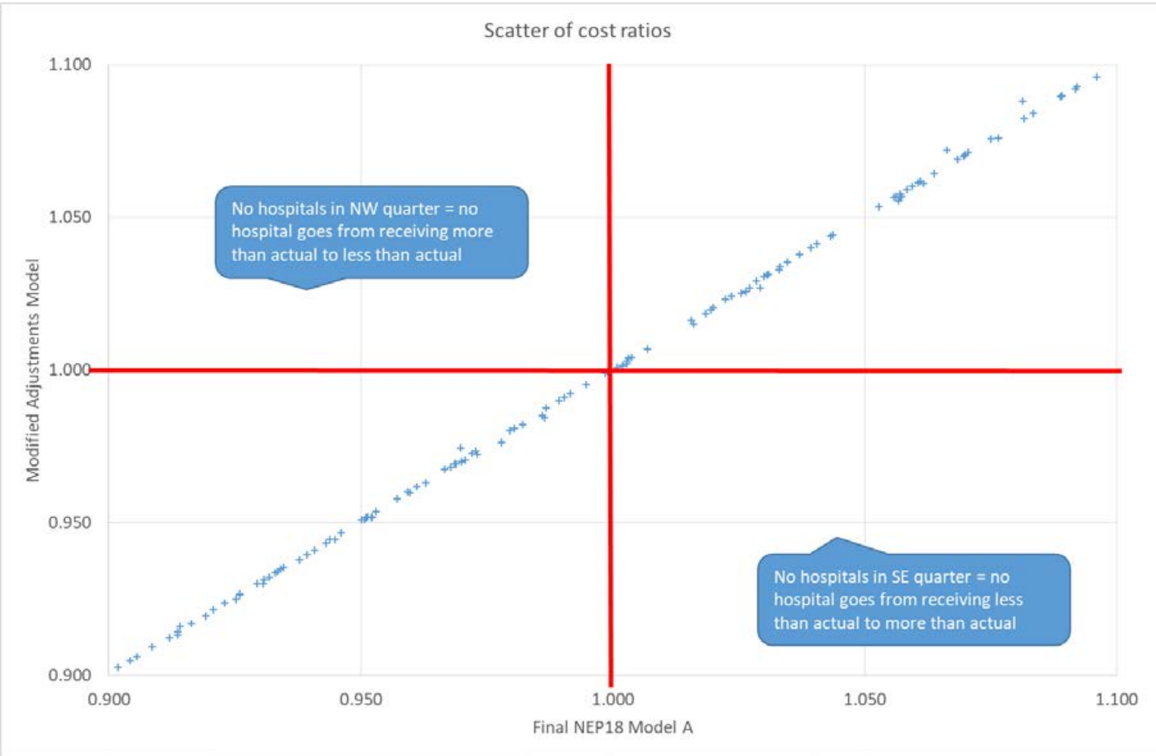


Figure 7: Cost ratio scatter plot for Alternative Model 3 compared to current Model

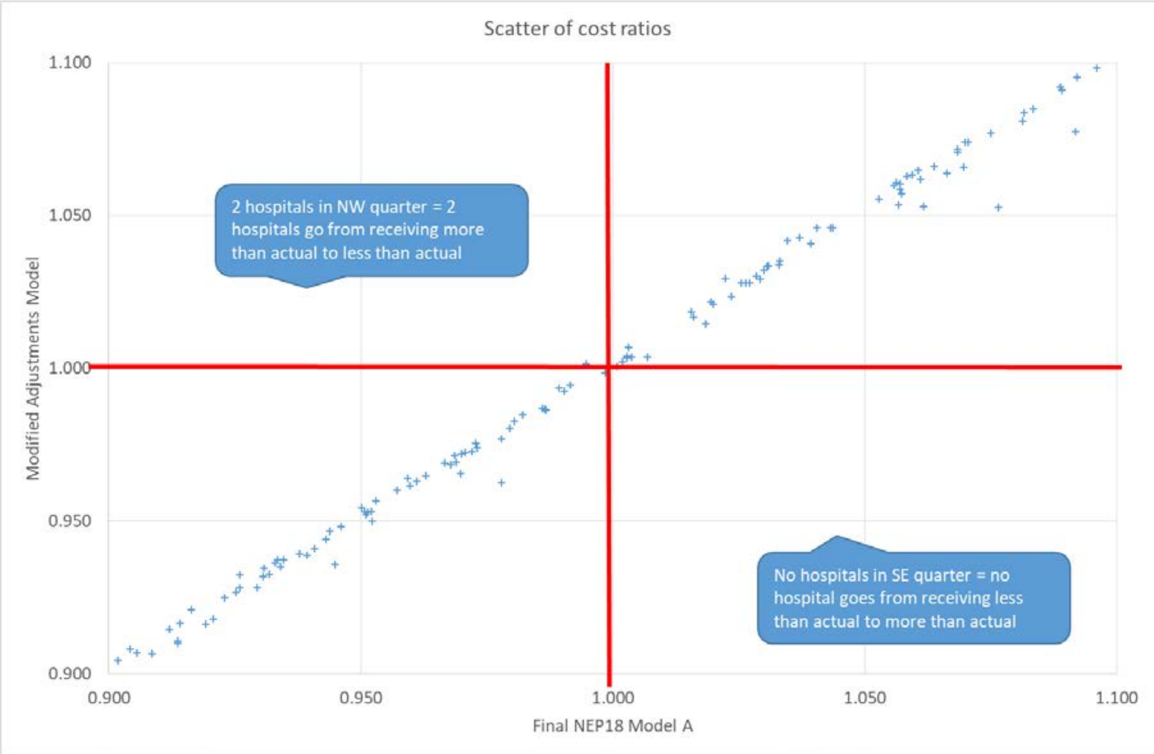


Figure 8: Cost ratio scatter plot for Alternative Model 4 compared to current Model

For Alternative Model 1, the distributions of changes in cost ratios were plotted for different hospital characteristics. These plots are shown in Appendix 8. These plots showed evidence of a higher change for hospitals with the following attributes:

- located in the Northern Territory;
- located in a remote area; and
- peer group D3.

Further investigation revealed that the remote area attribute was the important one, as it drives the relatively high changes for Northern Territory hospitals and for peer group D3.

To quantify the scale of this apparently biased impact of the Alternative Model 1 on the remote hospitals, a regression analysis of change in cost ratio against remoteness was carried out. The results of that regression are shown in Table 29.

All of the coefficients are smaller than 0.005 in absolute value. This is such a small magnitude as to be immaterial.

Table 29: Remoteness coefficients from regression of change in hospital cost ratio with Alternative Model 1

Remoteness category	Coefficient
Metropolitan	-0.0017
Inner Regional	-0.0014
Outer Regional	-0.0022

Remoteness category	Coefficient
Remote	0.0044
Very Remote	0.0017

This distribution and regression analysis was not replicated for the other alternative models as the high degree of congruence in results for the four alternative models indicated that such analyses would be likely to produce similar results to those for Alternative Model 1.

Key findings

- The alternative models generated higher adjustment factors for:
  - Remote Patient;
  - Very Remote Patient; and
  - Radiotherapy.
- The alternative models generated lower adjustment factors for:
  - Remote Treatment; and
  - Very Remote Treatment.
- The alternative models generated effectively unchanged adjustment factors for:
  - Indigenous patient;
  - Outer Regional patient;
  - Dialysis; and
  - Outer Regional treatment.
- Alternative Models 2 and 3 increase Specialist Psychiatric Age Adjustments substantially.
- The alternative models perform as well as the current Model, as measured by overall Cost Ratio, R-squared and SMAPE.
- The alternative models have minimal impact on individual hospital cost ratios, with all changes being less than 2.4% and most being less than 1%, for each model.
- There is a tendency for Remote hospitals to increase their cost ratios by a negligible amount, under the alternative models.

4.5.3 Conclusions and implications

Summary

The main conclusions for the NEP arising from this review are:

1. The use of a Model using a single multivariate regression model that include all of the patient and episode status variables as covariates (Alternative Model 4), with cost ratio as the response variable, performs as well as the current Model.
2. This Alternative Model 4, incorporating multivariate regression, would have negligible impact on hospital level funding outcomes.
3. The use of multivariate regression is methodologically preferable to the current mixture of methods used to calculate the patient and episode adjustment factors as it makes explicit the impact covariance has on the NEP regardless of the materiality of any such interactive effects.

Assessment against the criteria

Table 30 and Table 31 summarise the assessment of the alternative methods against assessment criteria.

Table 30: Assessment of the current method

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Direct impact unlikely. However, the extent to which the current method does not adequately control for different patient characteristics when calculating factors, there may be skewing of predicted costs. This could influence patient admission decisions because of perceived (un)profitability of some AR-DRGs.
Public Hospital efficiency	Direct impact is very unlikely. However, the extent to which the current method does not adequately control for different patient characteristics when calculating factors, there may be skewing of predicted costs. This could mask the true efficiency of a hospital in treating patients where those characteristics come into play.
Funding transparency	The use of different methods for different adjustments introduces a risk of confusion and reduces clarity around this aspect of the Model.
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	No specific risks identified.
Gaming	No specific risks identified.
Level 2: specific to NEP feature	
No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.26%
Cost Ratio	0.9998
SMAPE	9.39%

Table 31: assessment of Alternative Model 1

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Compared to the current method there is a reduced risk associated with dealing explicitly with any interaction between remoteness variables and other adjustment factors.

Level/Criterion	Assessment
Public Hospital efficiency	To the extent this Model actively controls for different patient and episode characteristics, it would improve the allocation of costs to those characteristics, through the relevant adjustment factors. This supports better pricing of hospital services based on those characteristics, allowing for better management of patient treatment costs.
Funding transparency	As this Model actively controls for different hospital characteristics and uses the same method to calculate each adjustment factor, there is improved transparency around the derivation and application of those factors.
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	No specific risk identified.
Gaming	No specific risk identified.
Level 2: specific to NEP feature	
No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.25%
Cost Ratio	0.9998
SMAPE	9.37%

Table 32: assessment of Alternative Model 2

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Compared to the current method there is a reduced risk associated with dealing explicitly with any interaction between remoteness variables and other adjustment factors.
Public Hospital efficiency	To the extent this Model actively controls for different patient and episode characteristics, it would improve the allocation of costs to those characteristics, through the relevant adjustment factors. This supports better pricing of hospital services based on those characteristics, allowing for better management of patient treatment costs.
Funding transparency	As this Model actively controls for different hospital characteristics and uses the same method to calculate each adjustment factor, there is improved transparency around the derivation and application of those factors.

Level/Criterion	Assessment
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	No specific risk identified.
Gaming	No specific risk identified.
Level 2: specific to NEP feature	
No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.24%
Cost Ratio	0.9998
SMAPE	9.37%

Table 33: assessment of Alternative Model 3

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Compared to the current method there is a reduced risk associated with dealing explicitly with any interaction between remoteness variables and other adjustment factors.
Public Hospital efficiency	To the extent this Model actively controls for different patient and episode characteristics, it would improve the allocation of costs to those characteristics, through the relevant adjustment factors. This supports better pricing of hospital services based on those characteristics, allowing for better management of patient treatment costs.
Funding transparency	As this Model actively controls for different hospital characteristics and uses the same method to calculate each adjustment factor, there is improved transparency around the derivation and application of those factors.
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	No specific risk identified.
Gaming	No specific risk identified.

Level/Criterion	Assessment
Level 2: specific to NEP feature	
No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.25%
Cost Ratio	0.9998
SMAPE	9.37%

Table 34: assessment of Alternative Model 4

Level/Criterion	Assessment
Level 1: effectiveness	
Patient access to services	Compared to the current method there is a reduced risk associated with dealing explicitly with any interaction between remoteness variables and other adjustment factors.
Public Hospital efficiency	To the extent this Model actively controls for different patient and episode characteristics, it would improve the allocation of costs to those characteristics, through the relevant adjustment factors. This supports better pricing of hospital services based on those characteristics, allowing for better management of patient treatment costs.
Funding transparency	As this Model actively controls for different hospital characteristics and uses the same method to calculate each adjustment factor, there is improved transparency around the derivation and application of those factors.
Level 1: feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: risks	
Perverse incentives	No specific risk identified.
Gaming	No specific risk identified.
Level 2: specific to NEP feature	
No additional metrics identified	n/a
Level 3: macro criteria	
R^2	79.18%
Cost Ratio	0.9998
SMAPE	9.37%

5 Private patient adjustment

5.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches for price loadings for private patients (private patient adjustments) on the following aspects of the NEP for *acute admitted care type only*. Testing focused on the following aspects:

a. Performance of the Private Patient Service Adjustment (PPSA) relative to the ratio of actual benefits to episode cost for private, insured episodes with data in both the Hospital Casemix Protocol (HCP) and National Hospital Cost Data collection (NHCDC).

b. Test for systematic relationships between the ratio and selected hospital, patient and episode attributes.

The testing used 2015/16 acute admitted episodes from hospitals in scope of activity based funding that were used by IHPA to construct the 2018/19 NEP using AR-DRG Version 9.

5.2 Purpose for private patient adjustment

Public hospitals generate revenue for private patients from three sources:

- Commonwealth sources other than ABF (such as MBS[18] and PBS)
- Health insurers (prostheses and bed day charges)
- Patients (for self-funded private patients and insured patients with excesses or fee gaps).

The National Health Reform Agreement specifically requires the ABF funding for a private patient to be adjusted to exclude these components.

Currently, there are private patient adjustments for the following:

- Admitted Acute Model:
  - Private Patient Service Adjustment (PPSA); and
  - Private Patient Accommodation Adjustment (PPAA).
- Subacute Admitted Model:
  - PPSA; and
  - PPAA.

This chapter focuses on the methods used for PPSA and PPAA calculation in the Acute Model. The findings from this were felt to be generally applicable also to the Subacute Model.

5.3 Current approach to setting the private patient adjusment

5.3.1 Private Patient Accommodation Adjustment

The PPAA calculation is the more straight forward of the two as it consists of an amount per episode equal to the private health insurance default bed day rate applied to that episode's length of stay. This review found no reason to consider changes to this method. It faithfully implements

[18] Technically, the medical officer invoices the patient for medical costs largely through direct bulk billing to the Commonwealth. There are varied arrangements in place between hospitals and individual medical officers which can vary by individual medical officers whereby hospital bill on behalf of the medical officer and an agreement is in place as how this revenue is managed.

the NHRA requirement and accurately reflects the accommodation revenue available to public hospitals for treatment of private patients.

5.3.2 Private Patient Service Adjustment

The PPSA is a more complex calculation, consistent with the less predictable nature of the private revenue relating to doctors' fees and ancillary service fees for private patient treatment. The current method:

- links Hospital Casemix Protocol (HCP), Admitted Patient Collection (APC) and National Hospital Cost Data Collection (NHCDC) data;
- uses HCP benefits to measure medical and ancillary service revenue for privately insured episodes with HCP data;
- assumes other private patient episodes on average generated the same medical and ancillary revenue per episode, taking into account state/territory, DRG, ADRG and MDC;
- inflates costs as appropriate for episodes which do not include some or all costs for ancillary services within NHCDC, using ancillary benefits as a proxy for costs; and
- derives ratios of private patient revenue to costs per episode at the DRG level.

Our literature review made the following observations and conclusions regarding the current method:

- The review did not identify an alternative data source to the Hospital Casemix Protocol (HCP) that would provide better quality data for linkage to the NHCDC, in order to derive the PPSA.
- The phenomenon of public and private patients with the same clinical presentation and treatment (in terms of ABF relevant data items) generating different amounts of hospital revenue cannot be eliminated in its entirety. Within-year variation in MBS rebates, differences in methods for setting MBS fees and NEP, and impacts of averaging within DRGs (sometimes within MDCs) mean that complete equivalence cannot be delivered via NEP modelling.
- There was scope to gain further insights into the quality of PPSA and its potential impact on non-equivalence of private and public patient revenue by comparing actual private patient benefits and modelled PPSA using historical data.

5.4 Test plan

Given the non-equivalence phenomenon cannot be eliminated, any further consideration of potential to improve this aspect of the PPSA's performance should be based on more detailed diagnostic analysis.

Research question #1: Does the PPSA produce results that systematically over or under estimate the actual private patient revenue for episodes?

This is assessed by using 2015/16 data to test the phenomenon as follows:

- Take data on private patient episodes with modelled costs and benefits (final output data sets from Stage 12 of Acute Model). Determine the PPSA ratio used for each episode.
- Take data on private patient episodes with HCP data and in-scope costs (input data sets from Stage 1 of Acute Model). Determine the ratio of medical plus ancillary benefits to in-scope costs for each episode (Benefit's Ratio), allowing for inflation of costs due to missing ancillary service charges—using the inflation method in the Acute Model.
- investigate scatter plots for the PPSA and the Benefits' Ratio, by selected episode, patient and hospital characteristics—partition, indigenous status, remoteness, paediatric specialisation, state/territory, hospital ancillary inflation factor.

- For episodes in both data sets, calculate the difference between the PPSA and the Benefits' Ratio.
- Model these differences against selected hospital and episode characteristics to identify whether specific characteristics are materially associated with large differences. Modelling was using linear regression and included the same characteristics as above.

Research question #2: If the answer to the above question is 'yes', can the Acute Model method be modified to remove or minimise such bias?

5.5 Results and discussion

5.5.1 Testing for systematic over or under estimation of private revenue

Research question #1: Does the PPSA produce results that systematically over or under estimate the actual private patient revenue for episodes?

We analysed the distribution of the difference between the PPSA and the Benefits' Ratio, with the results shown in Table 35.

Table 35: Difference between PPSA and Benefits' Ratio, key statistics

Statistic	Value
Number of episodes	517,709
Mean	-2.1%
Median	+4.4%
5th percentile	-48.2%
25th percentile	-10.3%
75th percentile	+13.0%
95th percentile	+19.0%

The results show a negatively skewed distribution for the difference measure, with both mean and median close to, but on opposite sides of, zero.

We used linear regression to test for systematic relationships between the difference measure and the following selected characteristics:

- partition;
- indigenous status;
- patient remoteness;
- hospital remoteness;
- paediatric specialisation;
- state/territory; and
- hospital ancillary inflation factor.

The results of this modelling are shown in Table 36.

Table 36: Results of regression modelling—covariates with significant coefficients

Covariate	Model coefficient
Hospital ancillary proportion	-3.9%
DRG partition	-1.0%
State – NSW	-6.1%
State – Victoria	-2.6%
State – Queensland	4.8%
State – South Australia	3.9%
State – Western Australia	5.9%
State – Tasmania	-5.9%
Territory – Northern Territory	13.8%
Patient remoteness – Major cities	4.0%
Patient remoteness – Inner regional	4.3%
Length of stay	0.1% (per day)
Sameday status	-1.3%
Indigenous status	2.0%
Specialist paediatric hospital	1.9%

These results show a range of factors that appear to be systematically associated with PPSA difference from benefits to cost ratio. Some are associated with over estimation of private revenue and others with the opposite. In most cases, the underlying mechanisms that might contribute to these associations are not apparent.

As a result, the need or otherwise to actively address any of these factors through modifications to the NEP Model and how that might be done, are unable to be ascertained.

Key findings

- The distribution of the difference between PPSA and Benefits' Ratio shows that PPSA would be likely to overestimate the medical plus ancillary revenue for private episodes more often than it under estimates that revenue.
- On average, PPSA underestimates the net value of private revenue across private episodes.
- A number of factors appear to be systematically associated with the difference between the benefits to cost ratio and the PPSA. The mechanisms driving these associations are unknown.

5.5.2 Conclusions and implications

Summary

The main conclusions for the NEP arising from this review are:

1. The PPSA is of variable effectiveness in estimating the true ratio of benefits to episode costs.
2. There are a number of episode characteristics associated with the performance of PPSA in estimating this ratio.
3. The underlying reasons for these apparent associations and their materiality or otherwise is unknown.

4. The need for remedial action to modify the Model to address this issue is consequently unknown, together with what that action might look like.

6 Price stabilisation

6.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches for price stabilisation in the following aspects of the NEP for *acute admitted care type only*. Sensitivity testing of key stabilisation parameters was undertaken by changing:

- the inlier cost parameter from +/- 20% to alternative values of 15%, 10% and 5%.
- the inlier episodes' threshold from 1,000 to alternative values of 500, 2,000, 5,000 and 1,000,000.

The testing used 2015/16 acute admitted episodes from hospitals in scope of activity based funding that were used by IHPA to construct the 2018/19 NEP using AR-DRG Version 9.

6.2 Purpose of price stabilisation

One of IHPA's objectives is to "promote funding stability and predictability for LHNs and hospital managers through satisfying two key principles within the pricing model" [19] which are:

- being sensitive to changes in activity, cost or data lags
- minimising statistical noise.

Price stabilisation within the context of the NEP refers to a set of analytical activities that IHPA undertakes and mechanisms within the NEP processes that collectively aim to ensure that changes to elements of the NEP such as price weights, price loadings, boundary points etc., reflect systemic changes in actual costs and/or clinical practice.

6.3 Current approach to price stabilisation

There are multiple elements to price stabilisation as follows (not all of which were examined in this phase of the review):

- backcasting to take into account changes to a classification, costing standards or price methodology (this aspect of the NEP is being reviewed in the third phase of this project);
- base data preparation where spurious data is identified and removed from the analysis (this was reviewed in the first phase of this project);
- loadings for unavoidable costs (which is assessed in this phase of the review but as part of a review of the actual method used to derive these loadings, see section 4); and
- placing constraints on year on year changes to price weights (the focus of this review).

This part of this phase of the review is examining the constraints that are currently placed on year on year changes to price weights. IHPA generally restricts the year-on-year changes in price weights to 20 per cent under the following conditions:

- There is no change to inlier bounds, and to the status on the same-day pricing list and bundled ICU list;
- The change in the inlier cost parameter is outside +/- 20%; and
- There are less than 1,000 inlier episodes.

Where sampling is an issue, IHPA aggregates data with previous years, although it is uncertain as to how many years are aggregated and if a sample size target drives this. Through the stabilisation

[19] National Pricing Model Stability Policy. Version 3.1 May 2018. IHPA

process, IHPA may apply further restrictions (lower than +/- 20%) to the price weights for high volume and high cost services to minimise volatility in changes across years.

The current approach appears to be driven by pragmatic decisions with no apparent overarching framework or guiding principles. Our literature review found little in the literature that represents definitive 'best practice', but observed that some other industries benefited from having a set of guiding principles/objectives that underpin their price stabilisation methods.

6.4 Test plan

Price stabilisation aims to reduce undue volatility but pass through key genuine cost trends. Our testing plan involved sensitivity analyses in order to better understand if the current processes are achieving these aims and qualitative consideration of alternative approaches.

Sensitivity of key parameters

Sensitivity testing of key stabilisation parameters was undertaken as follows:

- Changing the inlier cost parameter from +/- 20% to alternative values of 15%, 10% and 5%;
- Changing the inlier episodes' threshold from 1,000 to alternative values of 500, 2,000, 5,000 and 1,000,000.

6.5 Results and discussion

The base work of the sensitivity analysis suggests that the current price stabilisation process is relatively insensitive to the inlier episodes threshold, but can be sensitive to the inlier cost parameter. For example, if the inlier cost parameter were to change from +/- 20% to +/- 10%, the list of DRGs requiring stabilisation for 2018/19 NEP will increase from 1 to 6. Further analytical work was not undertaken as the matter is more about managing volatility rather than any theoretical or technical view of volatility in price weights. For example, there have been examples of AR-DRGs whose price weight has increased materially in one year only to decrease in a subsequent year such as haemodialysis. This 'yo-yo' phenomenon has also occurred recently with the non-admitted price weights for the Tier 2 classification. In both cases, it is likely that the underlying cost drivers have not changed. These annual movements in price weights can be problematic for health services when in DRGs that have significant volumes or significant costs and make it difficult to 'sell' to clinicians.

There are a number of alternative approaches that could be considered including:

- Reducing the threshold value from 20% say to 10%
- Normalising changes within a class of DRGs (for example to prevent some increasing and others decreasing when there is no apparent differential change in the underlying cost drivers particularly for Adjacent DRGs that have material expenditures such as the delivery DRGs)
- Mandating that DRGs that carry material health expenditures (because of volume e.g., haemodialysis or a mix of volume and cost e.g., delivery DRGs) have a relatively low threshold value (e.g., 5%) that then requires purposeful assessment before the change is passed through to the final NEP.

6.5.1 Conclusions and implications

There are multiple elements to price stabilisation, such as removal of spurious data, backcasting and loadings for unavoidable costs, The current approach to placing constraints on year on year changes to price weights as part of the overall price stabilisation approach is not sufficiently robust to ensure that changes to price weights are driven by underlying changes to actual cost drivers. This results in price weights changes in different directions from one year to the next requiring a more nuanced approach to price stabilisation. Different options have been set out in this report.

7 Indexation

7.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches for indexation in the following aspects of the NEP for acute admitted care type only:

a. Test the sensitivity of the selected number of 'cost years' used to derive the indexation rate.

b. Test how the indexation results change if the line of best fit approach is applied using more granular cost driver groupings.

7.2 Current approach to indexation

IHPA uses the most up-to-date cost data available from the National Hospital Cost Data Collection (NHCDC) for setting the NEP. However there is currently a three year gap between the year of cost data used, and the year for which the price is determined. For example, NEP14 was determined using NHCDC 2011-12 cost data and NEP15 was determined using NHCDC 2012-13 cost data.

To account for this time lag IHPA applies an indexation rate to inflate the historic costs to current values in order to set future prices. IHPA does this by estimating the expected growth rate over the three years based on historical growth in unit costs, as reported in the NHCDC.

As more up-to-date data becomes available there may be large discrepancies between the predicted growth rate and the actual growth rate. This could impact the growth in the NEP upwards or downwards between successive years.

To account for this, IHPA uses the most recently available NHCDC cost data to back-cast the previous year's NEP for the purposes of determining Commonwealth growth funding.

7.3 Alternatives to test

Our initial review did not identify issues with the method per se but more that the method was based on an unstated assumption that the underlying cost drivers behave in a consistent manner and are relatively stable over time whereas, the initial review identified a number of scenarios that suggests that this assumption is unlikely to hold true such as:

- trends in costs are likely to vary by cost component such as labour vs non-labour costs;
- systemic factors could interrupt a time series such as devaluation of the Australian dollar resulting in a step in time increase in consumable costs or an EBA in one state that has flow on effects in other states; and
- changes in the mix of intermediate product consumption that changes the cost structure for certain types of cases.

These observations are evident in other industries that have taken the decision to adopt a more nuanced approach to indexation such as the insurance industry that calculates separate index values for each of the major cost drivers (of their insurance products) and is consistent with the practice of some state treasuries that apply different indexation values to labour and non-labour costs.

The current method relies on a relatively stable year-on-year change in costliness of the mix of goods and services that make up the average cost. IHPA performs investigations from time to time to understand the underlying drivers of significant movements in price weights, cost parameters and other adjustments. However as an alternative, we propose to test the sensitivity of the current method to differential trends either through analysis of actual trend data or simulation of cost disruptors.

Examples of those scenarios under consideration include:

- different time series period (short period where there is a noticeable change in trend); and
- different cost components (note this will be constrained given that labour and non-labour costs are differentiated for some but not all cost buckets).

7.4 Results and discussion

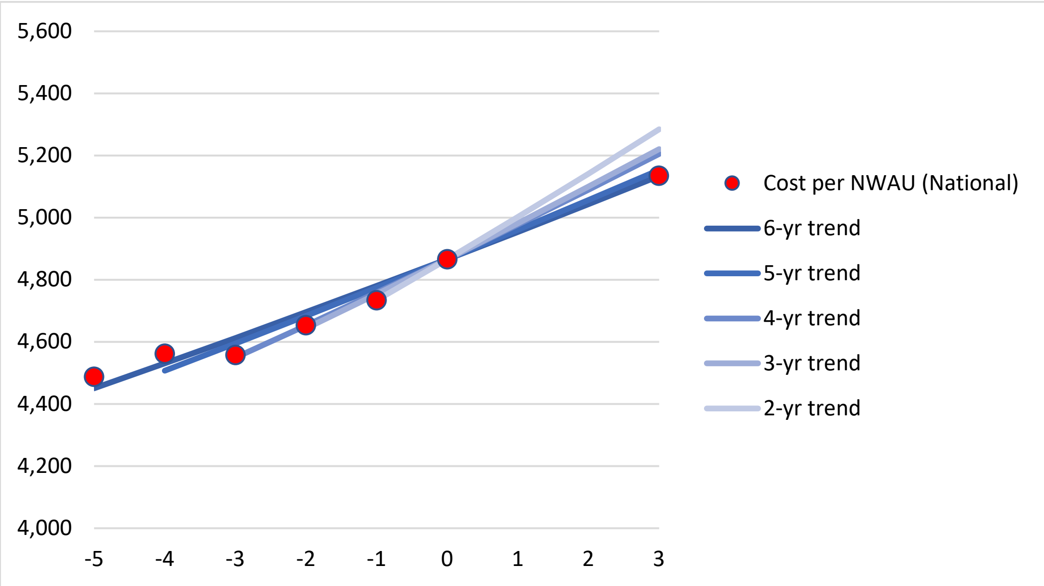
7.4.1 Period used to establish the trend

We have tested the sensitivity of the current approach in the number of periods used in establishing the indexation rate from the historical reference costs, while maintaining the current assumption of an exponential trend. The current approach fits an exponential line of best fit to the most recent 6 years' reference costs, and estimates the implied constant annual exponential growth rate. The 2019-20 indexed reference cost of $5,134 has been estimated from the 2016-17 reference cost of $4,866 indexed at an exponential growth rate of 1.8% for 3 years.

We have tested using fewer number of years to form the exponential trend line of best fit, i.e. most recent 5 years, 4 years and so on. The use of a shorter time period gives more weight to more recent inflationary trends over longer term inflationary trends.

The different scenario exponential trend lines are presented in different shades of blue in the following graph (with lighter colours represent scenarios using only the more recent data points). The red dots are the historical reference costs. The dot at time 3 is the current estimated 2019-20 indexed reference cost.

Figure 9: Time series of cost ratio and cost per weighted separation with exponential line of best fit – sensitivity analysis on number of periods used to form trend



Source: KPMG

The following table provides the actual index values using index values derived from different time periods.

Table 37: Comparison of forecasted 2019/20 cost per NWAU values using different inflationary indices

No. of years used to form trends	Indexation rate	Indexed reference cost for 2019-20	% difference with 2019-20
6	1.8%	5,134	
5	1.9%	5,154	0.4%
4	2.2%	5,205	1.4%
3	2.3%	5,221	1.7%
2	2.8%	5,284	2.9%

Source: KPMG

Key findings

- There is evidence of inflationary trend (with some degree of volatility) as observed in the historical reference costs. Using the more recent data to determine the exponential inflationary trend implies a higher indexation rate would be required to estimate the 2019-20 indexed reference cost. For example, using data from the most recent 3-4 years would have implied a 2019-20 indexed reference cost of about 1.4% to 1.7% higher than the current estimate from using 6 years of data.
- There seems to be some evidence of momentum in inflationary pressures from looking at the most recent 2 data points (i.e. an indexation rate of 2.9% between 2015-16 and 2016-17). However, it is a somewhat questionable approach to adopt this as a long term average indexation rate. Nevertheless, this highlights the sensitivity of the indexation approach to the nuance of the underlying data points and volatility over time.
- There is no theoretical basis for specifying the number of periods to use in informing a "trend". Going forward it is advisable to repeat this sensitivity analysis in order to draw out any differences in recent versus longer term trends, to enable a more nuanced decision for the indexation rate. Analysis by cost bucket groupings (i.e. labour vs consumables) will further mitigate the risk in this process.

7.4.2 Trend by cost driver components

We have examined the robustness of the current approach, which focuses on total cost trends across all cost buckets, in how it reflects any underlying trends in the component cost drivers, such as labour vs non-labour costs.

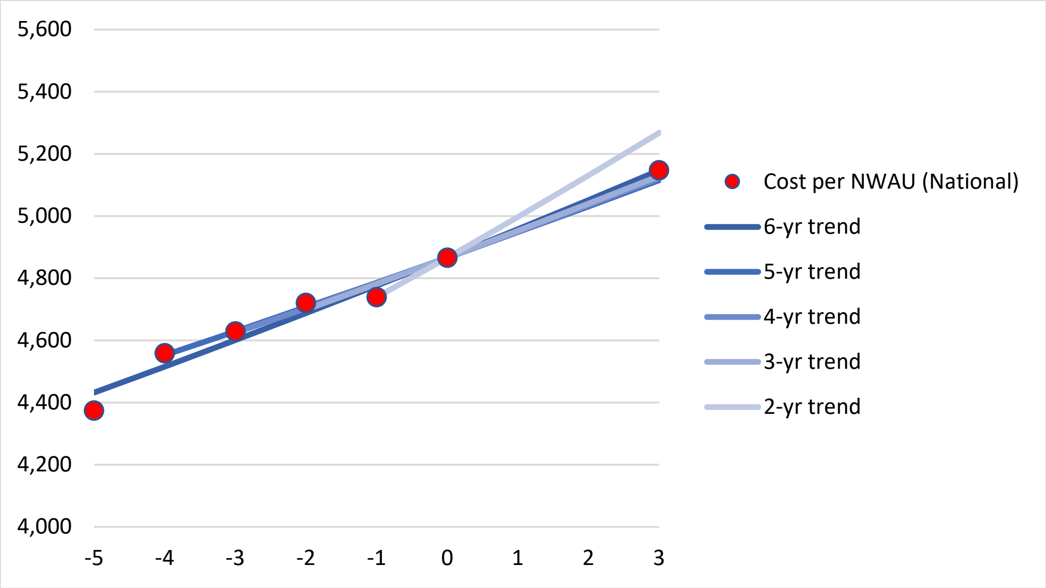
It was not possible from the existing datasets to breakdown costs further into labour and non-labour components. For the purpose of testing on a best endeavour basis, we have utilised the existing cost bucket breakdowns as a proxy, and approximately classified the cost buckets into Labour, Consumables, and Mixed as follows.

Table 38: Approximate cost bucket classifications into Labour, Consumables and Mixed groupings

Cost buckets assumed to consist largely of Labour related costs	Cost buckets assumed to consist largely of Consumables related costs	Cost buckets with Mixed labour and non-labour costs
Ward Medical	Prostheses	Pathology
Ward Nursing	Pharmacy	Imaging
Non-clinical Salaries	Ward Supplies	Critical Care
Emergency Department		Operating Rooms
On-costs		Hotel
Special Procedural Rooms		
Allied Health		

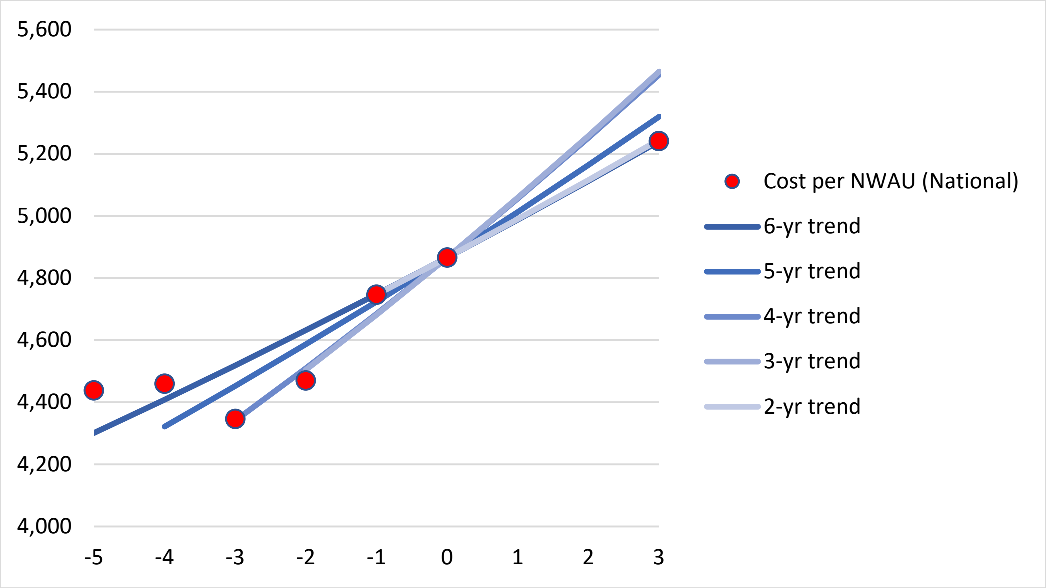
As part of this testing we have also examined the range of indexation rates from varying the number of years to form the exponential trend line of best fit by cost bucket groupings. The use of a shorter time period gives more weight to more recent inflationary trends over longer term inflationary trends. The results are shown below.

Figure 10: Time series of cost per NWAU with exponential lines of best fit – Labour cost grouping



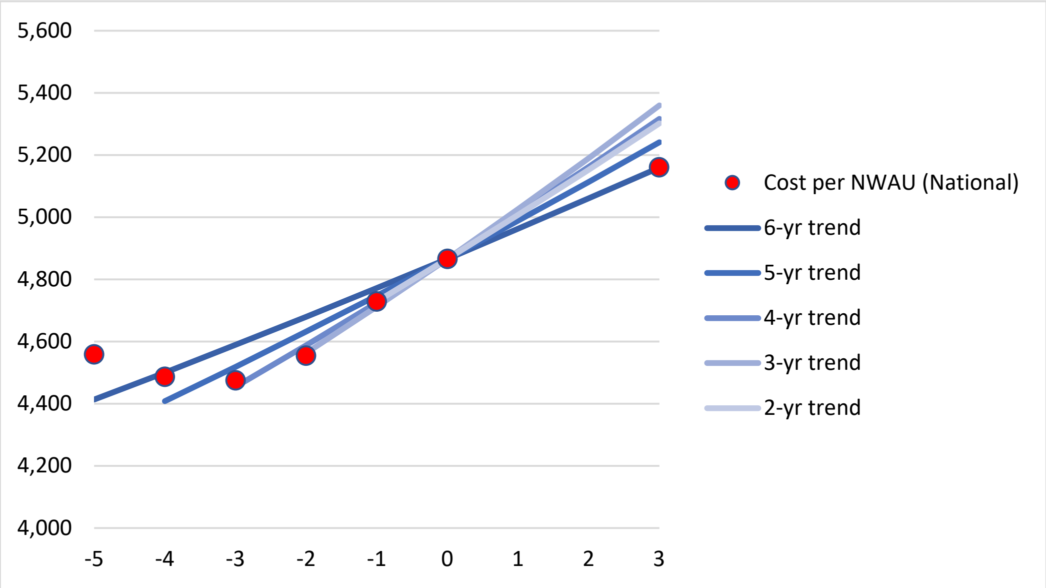
Source: KPMG

Figure 11: Time series of cost per NWAU with exponential lines of best fit – Consumables cost grouping



Source: KPMG

Figure 12: Time series of cost per NWAU with exponential lines of best fit – Mixed cost grouping



Source: KPMG

The following table compares the exponential line of best fit indexation rates implied by the different cost bucket groupings vs total cost.

Table 39: Comparison of implied indexation rates by cost bucket groupings vs total cost

No. of years used to form trends	Total cost	Labour cost bucket grouping	Consumables cost bucket grouping	Mixed cost bucket grouping
6	1.8%	1.9%	2.5%	2.0%
5	1.9%	1.7%	3.0%	2.5%
4	2.2%	1.7%	3.8%	3.0%
3	2.3%	1.8%	3.9%	3.2%
2	2.8%	2.6%	2.5%	2.9%

Source: KPMG

Key findings

- The range of indexation rates implied by Labour cost buckets seems to be lower than the range of rates for the Consumables cost buckets. This can be influenced by any EBAs in place in the recent years limiting upward pressures on labour related costs as well as differences in case mix over time.
- The apparent momentum in inflationary pressures in the total costs seems to be driven by a period of strong increases in Consumables, and subsequently by increases from the Labour cost buckets. However, we again note the sensitivity of the indexation approach to the nuance of the underlying data points and volatility over time.
- The test results show some underlying differences in the implied indexation rates by cost bucket groupings. This testing is approximate, but it highlights there is potential value in performing further investigations on the different inflationary drivers between labour and non-labour costs impacting trends in the underlying cost components.

7.4.3 Assessment

Table 40: Assessment of current indexation method

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Direct impact unlikely. However, the extent to which the current method does not adequately deal with volatility in cost changes could influence patient admission decisions if back-casting is required to adjust for changes in forecasted NEP.
Public hospital efficiency	The extent to which the current method does not reflect actual cost changes or deal with cost volatility, could lead to hospitals' taking actions to deal with perceived efficiency issues that do not exist or fail to take action if the indexation over-estimates actual cost inflation.
Funding transparency	There is a risk associated with any changes to payments that need to be made if, when updating the index value using more recent data, a different index value is determined from the same financial year.
Level 1: Feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Could arise from risks associated with patient access and public hospital efficiency referred to above.
Gaming	Meets this requirement.

Table 41: Assessment of proposed approach using a shorter time period for indexation

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Direct impact unlikely. However, any risk would be less than the current method if sensitivity analysis (as recommended) suggests that a different index value is justified because of different more recent inflationary trends.
Public hospital efficiency	Likely to mitigate some of the risk associated with the current method because it takes into account any differences in more recent trends compared to longer term trends.
Funding transparency	Risk is likely to be less than the current method as the proposed approach would better deal with volatility in trend values.
Level 1: Feasibility	
Data availability	Meets this requirement.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Could arise from risks associated with patient access and public hospital efficiency referred to above but likely to be less than the current method
Gaming	Meets this requirement.

Table 42: Assessment of proposed approach for indexation by different cost drivers

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Direct impact unlikely. However, any risk would be less than the current method if breakdown by cost drivers (as recommended) suggests that a different index value is justified because of different trajectories by inflationary drivers.
Public hospital efficiency	Likely to mitigate some of the risk associated with the current method because it takes into account any underlying signals indicating which particular parts of the system are showing cost pressures compared to just the overall trends.
Funding transparency	Risk is likely to be less than the current method as the proposed approach would better deal with volatility in trend values.
Level 1: Feasibility	
Data availability	Meets this requirement in terms of using existing cost buckets as a proxy. However a full delineation of labour and non-labour cost components may need further investigations and implemented through the NHCDC reporting requirements.
Data quality	Meets this requirement, however can be subject to changes in hospital data practice and accuracy in delineation of costs into the cost buckets (e.g. use and reliance on cost fractions), and full delineation into labour and non-labour cost components.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Could arise from risks associated with patient access and public hospital efficiency referred to above but likely to be less than the current method
Gaming	Meets this requirement.

7.4.4 Conclusions and implications

Based on our observations and cross industry experience, there is a case to be made for IHPA to consider to adopt a more nuanced approach to indexation including:

- Undertaking sensitivity analysis of inflationary trends using different time periods that, if there are material differences in trends between shorter and longer time periods, an adjusted index value would be used instead of the standard five year trend value.
- Identifying and adjusting for systemic or structural changes to the cost of services (or component cost); and
- calculating separate index values for major cost drivers where there is evidence of differential trend.

This is likely to require a change in the NHCDC reporting requirements such as full delineation of labour and non-labour costs. Consequently, an adapted approach will take time before it can implemented.

8 Back-casting

8.1 Aspects of NEP tested

This section presents the results of the testing of alternative approaches for back-casting in the following aspects of the NEP for acute admitted care type only:

a. Qualitative assessment of the merits of the current method and alternative methods for calculation of back-cast volume multipliers.

b. Quantitative assessment of the impacts of alternative methods on resulting volume multipliers.

8.2 Purpose of back-casting

The purpose of back-casting is to inform the calculation of the Commonwealth's funding, as set out in the National Health Reform Agreement (NHRA).[20] (Clause A40). The back-cast volume multipliers also serve a secondary purpose of informing the work of the Administrator of the National Health Funding Pool to prospectively determine likely payments for the coming funding year.

8.3 Current approach to back-casting

The current method involves calculating a separate volume multiplier for each jurisdiction and for each activity based funding (ABF) service category. Each such back-cast volume multiplier is calculated as:

$$\text{Volume Multiplier}_0 = \frac{\begin{array}{c}\text{NWAUs delivered by NEP}_{\text{n}}\text{model (NWAU[n])}\\ \text{applied to latest activity data}\end{array}}{\begin{array}{c}\text{NWAUs delivered by NEP}_{\text{n}-1}\text{ model (NWAU[n}-1\text{])}\\ \text{applied to latest activity data}\end{array}}$$

In practice, the volume multipliers for one year are calculated using the activity data for that year and published in the following year's NEP Determination.

As used in back-casting, the volume multiplier is analogous to the Laspeyres Index[21] and estimates the overall impact of change in NEP model structure on NWAU volumes for the target year for back-casting. The key difference from the orthodox Laspeyres Index is the volume multiplier also includes changes in the AR-DRG classification from time to time. Nonetheless, the relationship is strong enough to be used when considering alternative methods.

Note that when there is no change in AR-DRG classification between Models, this multiplier effectively estimates the weighted average change in price weights and adjustments between the years.

The Laspeyres Index is traditionally used to estimate average price inflation across a range of goods and services between periods. In that context, it tends to overestimate inflation due to a failure to consider impact of price changes on purchasers' behaviour in the subsequent year. In the NEP context, it may suffer from error due to changes in hospitals' practices and casemix stemming from changes in the NEP model and associated NEP value. This is affirmed by history with ABF, which has

[20] Clause A40 of the NHRA states "*If the IHPA makes any significant changes to the ABF classification systems or costing methodologies, the effect of such changes must be back-cast to the year prior to their implementation for the purpose of the calculations*".
[21] Australian Bureau of Statistics (2018). *6461.0 - Consumer Price Index: Concepts, Sources and Methods, 2018*. Available at https://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/6461.0Main%20Features42018.

shown that changes in the Model and the Price affect hospitals' behaviour, in terms of the amount and mix of activity undertaken, as well as how that activity is captured and coded.

8.4 Alternatives to test

The following specific issues were considered for testing:

- *Research question #1:* Is there a suitable alternative to the current back-casting method that would reduce or eliminate the potential for error due to impact on hospitals' practice?

- *Research question #2:* What is the impact of alternative back-casting methods?

8.5 Results and discussion

8.5.1 Identifying alternative methods

Research question #1: Is there a suitable alternative to the current back-casting method that would reduce or eliminate the potential for error due to impact on hospitals' practice?

Paasche Index[22] is a closely related alternative to the Laspeyres Index. It suffers from the opposite problem to Laspeyres. That is, it tends to err in the opposite direction to the Laspeyres Index. It would also require the activity data to be available for the year following the back-casting target year.

The Lowe Index[23] is similar to both the Laspeyres and Paasche Indices but uses a standard reference activity data set, which does not change from one year to the next. Such a reference data set would usually be a nominated, historical year of activity. The Australian Bureau of Statistics (ABS) adopts a practice of using as base year for Lowe Index based CPI the financial year two years prior to the present.

The Fisher Index[24] attempts to address the respective issues with Paasche and Laspeyres indices. It does so by calculating the geometric mean of those two indices.

There are other indices from economics that would not be practicable due to the unsuitability of the respective formulae to the task of back-casting ABF volumes. The key consideration for these other indices is that episodes in the same service class (DRG, AN-SNAP class, Tier 2 clinic, URG/UDG) may have different NWAU values.

Research question #2: What is the impact of alternative back-casting methods?

To address this research question, we analysed the following alternative methods for calculating back-cast volume multipliers:

- Paasche Index formula;
- Fisher Index; and
- Lowe Index.

Paasche Index method

This method replaces the target year of activity data with the activity data for the year following the one to be back-cast. As such, the formula becomes:

$$\text{Volume Multiplier}_P = \frac{\begin{array}{c}\text{NWAUs delivered by NEP}_\text{n}\text{model (NWAU[n])}\\ \text{applied to Year[n] activity data}\end{array}}{\begin{array}{c}\text{NWAUs delivered by NEP}_{\text{n}-1}\text{ model (NWAU[n}-1\text{])}\\ \text{applied to Year[n] activity data}\end{array}}$$

Key findings – Paasche Index method

[22] Ibid.
[23] Ibid.
[24] Ibid.

- This method suffers from the same issue as the current method, in that it will not take into account changes in hospital behaviour and casemix due to changes in the Model and Price.
- Errors due to this shortcoming will tend to operate in the opposite direction to those for the current method.
- This method cannot be applied until the data for the year following that to be back-cast become available. That is, approximately a year or two later than at present.

Fisher Index method

This method simply calculates the geometric mean of the current method result and the Paasche Index method result. That is, the two results are multiplied and the square root of that product gives the volume multiplier:

$$Volume\ Multiplier_F = \sqrt{Volume\ multiplier_0 \times Volume\ multiplier_P}$$

Key findings – Fisher Index method

- This method aims to reduce the potential errors due to changes in hospital behaviour and casemix, which arise in both the current method and the Paasche Index method.
- This method cannot be applied until the data for the year following that to be back-cast become available. That is, approximately one to two years later than at present.

Lowe Index method

This method replaces the target year of activity data with the activity data for the year before the one to be back-cast. As such, the formula becomes:

$$\text{Volume Multiplier}_L = \frac{\begin{array}{c}\text{NWAUs delivered by NEP}_\text{n}\text{model (NWAU[n])}\\ \text{applied to Year[n} - \text{2] activity data}\end{array}}{\begin{array}{c}\text{NWAUs delivered by NEP}_{\text{n}-1}\text{ model (NWAU[n} - \text{1])}\\ \text{applied to Year[n} - \text{2] activity data}\end{array}}$$

Key findings – Lowe Index method

- This method suffers from the same issue as the current method, in that it will not take into account changes in hospital behaviour and casemix due to changes in the Model and Price.
- This source of error will be compounded by virtue of the year of activity data being non-contemporary with both the NEP Models involved in the calculation.
- It is unclear how this compounding might affect potential (for) errors in the resulting multipliers but it is likely to be less reliable than the current method.

8.5.2 Testing alternative methods

Table 43 shows the results produced for the 2016-17 acute admitted activity multipliers using the current method and the three alternatives described in Section 8.4.

Those results show the following:

- the Paasche method consistently produces lower multipliers than the current method (Laspeyres), for all jurisdictions except the Australian Capital Territory (ACT);
- consequently, the Fisher method also produces consistently lower multipliers than the current method, with the exception of the ACT. The differences are around half of those for the Paasche method; and
- the Lowe method produces results consistent with the Paasche and Fisher results in terms of whether higher or lower than the current method, but varying in magnitude.

Based on the discussion of Laspeyres, Paasche and Fisher in Section 8.4, these results imply that the current method produced back-cast multipliers that tend to overestimate the effect on calculated NWAU due to the implementation of the NEP17 Acute Model. At the national level, the degree of over estimation is 0.03 per cent. At the State and Territory level, it varies from 0.11 per cent understated (ACT) to 0.32 per cent overstated (Northern Territory (NT)).

Table 43: Acute admitted activity back-casting multipliers for 2016-17, using different calculation methods

Jurisdiction	Laspeyres method	Paasche method	Fisher method	Lowe method
NSW	0.9938	0.9924	0.9931	0.9937
Victoria	0.9982	0.9978	0.9980	0.9978
Queensland	0.9981	0.9978	0.9979	0.9967
SA	0.9956	0.9957	0.9956	0.9965
WA	1.0029	1.0026	1.0027	1.0020
Tasmania	0.9955	0.9952	0.9953	0.9949
NT	0.9907	0.9875	0.9891	0.9890
ACT	0.9922	0.9933	0.9927	0.9936
Australia	*0.9967*	0.9960	*0.9963*	*0.9963*

8.6 Source: KPMG analysis

In calculating growth funding from year to year, these results can be used to estimate the impact on growth factors of applying each of the tested methods. In short, the impact will be of the same magnitude as the change in the value of the relevant multiplier but in the opposite direction. These impacts are summarised in Table 44.

To interpret this table, consider a hypothetical growth for NSW of 6.0 per cent, calculated using the current method. The impact of using the Paasche method would be to increase that growth to 6.009 per cent.

Table 44: Impact on growth funding calculations, using different back-casting multiplier methods

Jurisdiction	Laspeyres method	Paasche method	Fisher method	Lowe method
NSW	-	+0.15%	+0.07%	+0.01%
Victoria	-	+0.04%	+0.02%	+0.05%
Queensland	-	+0.03%	+0.01%	+0.14%
SA	-	-0.00%	-0.00%	-0.09%
WA	-	+0.03%	+0.02%	+0.09%
Tasmania	-	+0.03%	+0.01%	+0.06%
NT	-	+0.32%	+0.16%	+0.17%
ACT	-	-0.11%	-0.06%	-0.15%
Australia	-	*+0.07%*	*+0.03%*	*+0.03%*

8.7 Source: KPMG analysis

8.7.1 Assessment

Table 45: Assessment of current back-casting approach

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Meets this requirement. However, these results suggest changes in the NEP Model and Price between years operate in a way to produce volume multipliers that overestimate the true values.

Level/Criterion	Assessment
	Over estimation works to reduce the amount of funding available to hospitals over time, leading to reduced access to services.
Public hospital efficiency	Meets this requirement noting that this current method risks underfunding public hospitals over time.
Funding transparency	Does not meet this requirement. The volume multipliers for a given year are not available until the NEP Determination for the following year is released. This gives little visibility to States and Territories and to LHNs of likely growth funding impacts until almost one year into the following funding year.
Level 1: Feasibility	
Data availability	Meets this requirement. Activity data available in the year after the target year is ended.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Meets this requirement.
Gaming	Meets this requirement.

Table 46: Assessment of the Paasche Index method

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Meets this requirement. However, these results suggest changes in the NEP model and price between years operate in a way to produce Paasche volume multipliers that underestimate the true values. Under estimation works to increase the amount of funding available to hospitals over time, leading to risk of reduced efficiency and consequential impacts on access to services.
Public hospital efficiency	Meets this requirement noting that this method risks overfunding public hospitals over time.
Funding transparency	Does not meet this requirement. This method would exacerbate the current delays in availability of back-cast volume multipliers by an additional year.
Level 1: Feasibility	
Data availability	Meets this requirement. Activity data available a year later than current method.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Meets this requirement.
Gaming	Meets this requirement.

Table 47: Assessment of the Fisher Index method

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Meets this requirement. Reduced risk of over or under estimation of true volume multiplier values, reducing risk of negatively affecting patient access to hospital services.
Public hospital efficiency	Meets this requirement. Reduced risk of overfunding or underfunding public hospitals over time, relative to both the current method and the Paasche Index method.
Funding transparency	Does not meet this requirement. The volume multipliers for a given year are not available until the NEP Determination for the following year is released. This gives little visibility to States and Territories and to LHNs of likely growth funding impacts until almost a year into the following funding year.
Level 1: Feasibility	
Data availability	Meets this requirement. Activity data available a year later than current method.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Meets this requirement.
Gaming	Meets this requirement.

Table 48: Assessment of the Lowe Index method

Level/Criterion	Assessment
Level 1: Effectiveness	
Patient access to services	Meets this requirement. The results of testing suggest increased risk of both over and under estimation of true volume multiplier values, increasing the risk of negatively affecting patient access to hospital services over time.
Public hospital efficiency	Meets this requirement noting that it risks underfunding and overfunding public hospitals over time. It may do so to a larger extent than the current and Paasche methods.
Funding transparency	Does not meet this requirement. The volume multipliers for a given year are not available until the NEP Determination for the following year is released. This gives little visibility to States and Territories and to LHNs of likely growth funding impacts until almost one year into the following funding year.
Level 1: Feasibility	
Data availability	Meets this requirement. Activity data available one year earlier than for other methods.
Data quality	Meets this requirement.
Data volume	Meets this requirement.
Level 1: Risks	
Perverse incentives	Meets this requirement.
Gaming	Meets this requirement.

8.7.2 Conclusions and implications

KPMG's qualitative assessment is that the preferred method is the current method. While it suffers from the identified risk with regard to not accounting for NEP Model and Price induced changes in hospitals' behaviour and casemix, there is no basis on which to conclude this risk is larger than for the Paasche Index and Lowe Index methods. While the Fisher Index method would reduce this risk, the additional year of delay inherent in the Fisher Index method is not commensurate for that reduced risk.

The results of testing suggest the current method does in fact overestimate the 'correct' back-casting multipliers for the Acute Model. The order of magnitude of the impact on growth funding is relatively small – less than $800 per $1m of growth funding for jurisdictions below the growth funding cap. This is based on the impact of applying the Fisher method (see Table 44). Even after five years, the cumulative impact of such funding differences on a jurisdiction's base funding would still be small.

Given this small impact from the current method's imperfections and the transparency costs of moving to the Fisher method, KPMG's position remains that the current method is the preferred method for calculating the back-casting multipliers.

9 Calculation of the Reference Cost

This section presents KPMG's review of the calculation of the Reference Cost. The part of the Review did not involve data analysis nor testing of alternative methods.

9.1 Purpose of calculating the Reference Cost

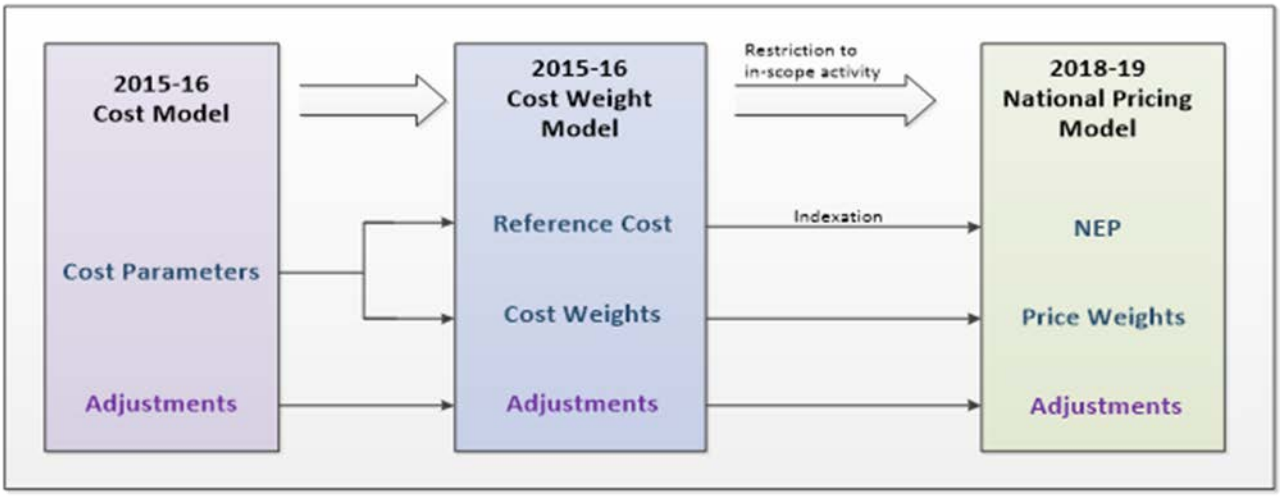
The Reference Cost is used in the transformation of the NEP cost models into pricing models (Figure 13). It is used firstly to convert the cost parameters in the various NEP cost models into cost weights. This is done by:

- for each DRG, dividing modelled average inlier cost, average same-day cost, average short stay outlier base cost, average short stay outlier per diem cost and average long stay outlier per diem cost by the Reference Cost (Acute Admitted Model);
- for each AN-SNAP class, dividing modelled average inlier cost, average same-day cost, average short stay outlier per diem cost and average long stay outlier per diem cost by the Reference Cost (Sub Acute Admitted Model);
- for paediatric palliative care, dividing modelled average per diem cost by the Reference Cost (Sub Acute Admitted Model);
- for each Tier 2 Clinic, dividing modelled average service event cost by the Reference Cost (Non-Admitted Model); and
- for each URG and UDG, dividing modelled average service event cost by the Reference Cost (Emergency Services Model).

Those cost weights become the price weights for the pricing model.

The Reference Cost is then indexed forwards to produce the National Efficient Price itself.

Figure 13: Transforming the 2015-16 Cost Model to the 2018-19 National Pricing Model.



Source: IHPA, National Efficient Price Determination 2018-19.

9.2 Current approach to calculating the Reference Cost

The Reference Cost is calculated using acute admitted activity and costs available from the NHCDC. It is calculated in such a way as to control for changes in the casemix profile of admitted activity reported to the NHCDC over time. This includes the range of included episodes, hospitals and costs.

The Reference Cost is calculated as follows:

$$\text{Reference Cost}_n = \text{Reference Cost}_{n-1} \times \frac{\begin{array}{c}\text{Average episode cost delivered by NEP}_\text{n}\text{cost model}\\ \text{applied to latest activty data for year } n\end{array}}{\begin{array}{c}\text{Average cost delivered by NEP}_{\text{n}-1}\text{ cost model}\\ \text{applied to activity data for year } n\end{array}}$$

This formula is mathematically equivalent to the following:

$$\text{Reference Cost}_n = \frac{\begin{array}{c}\text{Total cost delivered by NEP}_\text{n}\text{cost model}\\ \text{applied to activity data for year } n\end{array}}{\begin{array}{c}\text{GWAU}_{\text{n}-1}\text{ model}\\ \text{applied to activity data for year } n\end{array}}$$

where GWAU_{n-1} is the number of gross weighted activity units produced by applying the price weight model for year *n-1*, without private patient adjustments.

That is, in effect the Reference Cost for the current year is the average cost per GWAU for the current year's activity data, with GWAU calculated using the previous year's model. The activity data is limited to episodes in scope for ABF funding.

As discussed in relation to back-casting (see Section 8), the Reference cost is analogous to calculating a Paasche Index and applying that Index to the previous year's Reference Cost.

Furthermore, this method results in the Reference Cost being equal to the first ever Reference Cost multiplied by a chained product of Paasche Indices for each subsequent year. That first ever Reference Cost is the modelled average cost per episode (excluding private patient adjustments) from the 2009-10 activity data, as produced for the 2012-13 NEP.[25] That original Reference Cost was $4,260.

As a result of this chaining, the Reference Cost for the 2018-19 NEP Model estimates the change in unit cost of a GWAU between the 2009-10 and 2015-16 cost models. More precisely, it estimates the change in modelled average acute admitted episode cost over that period, excluding effects due to casemix changes.

9.3 Test plan

In KPMG's view, the principles and objectives of the current method – to produce a Reference Cost that maintains equivalency of NWAU values over time, independently of changes in the casemix present in the activity data– are reasonable. Given that, the current chained index approach is sound.

Nonetheless, the method as currently applied has a number of weaknesses that should be addressed. We briefly summarise these weaknesses and propose remedies, below.

9.4 Results and discussion

The main weaknesses of the current method identified by our review are:

1. Bias inherent in the choice of the chained Paasche Index;
2. Risk of bias due to substantial changes in the AR-DRG classification, activity counting rules and costing standards since the 2009-10 activity data were collected; and
3. Exclusion of subacute, non-admitted and emergency services' costs when calculating the Paasche Indices.

[25] Trent Yeend (16 May 2013). *The National Pricing Model Explained*. Presentation to the 2013 ABF Conference.

9.4.1 Choice of Paasche Index

As stated above, there are two purposes for calculating the Reference Cost. The first to convert the NEP_n cost model into a price model and the second is to set the NEP to be applied for funding purposes.

Price model conversion considerations

The choice of method for calculating the Reference Cost was originally made with a focus on ensuring stability of NWAU values between successive NEP models. The current method aims to do this by standardising the value of a GWAU relative to a selected year of activity – the current year's activity data – and setting the Reference Cost accordingly.

However, this method works only for the selected year of activity. For example, standardising relative to the previous year's activity data would produce a different result for the Reference Cost.

The reason for this outcome relates to the nature of the Paasche Index as a measure of price change. The Paasche Index is traditionally used to estimate average price inflation across a range of goods and services between periods. In that context, it tends to underestimate inflation due to a failure to consider impact of price changes on purchasers' behaviour between the years.

In the NEP context, it would suffer from bias due to changes in hospitals' practices and casemix stemming from changes in the NEP model and associated NEP value used for funding activity in the relevant years. This is confirmed by history with ABF, which has shown that changes in the Model and the NEP affect hospitals' behaviour, in terms of the amount and mix of activity undertaken, as well as how that activity is captured and coded.

An alternative to the Paasche Index is the Laspeyres Index.[26] In the context of setting the Reference Cost, a method that applied the Laspeyres Index approach would standardise the value of a GWAU relative to the previous year's activity data, instead of the current year. That is, it would calculate Reference Cost as the average cost per GWAU, as follows:

$$\text{Reference Cost}_n = \frac{\begin{array}{c}\text{Total cost delivered by } \mathrm{NEP_n} \text{cost model} \\ \text{applied to activity data for year } (n-1)\end{array}}{\begin{array}{c}\mathrm{GWAU_{n-1}} \text{ model} \\ \text{applied to activity data for year } (n-1)\end{array}}$$

In economic terms, the Laspeyres Index suffers from the opposite problem to the Paasche Index. That is, it tends to overestimate inflation.

In the NEP context, calculating the Reference Cost using this method would be likely to produce a value lower than the current method. In reality, the perfect Reference cost would lie somewhere between the values produced by these two methods.

A third index – the Fisher Index[27] – attempts to compensate for the biases inherent in the Paasche and Laspeyres indices by calculating the geometric mean of both indices. This method is accepted by economists as producing better estimates of true inflation than either of the Paasche or Laspeyres methods. Furthermore, testing of the Laspeyres and Fisher indices for back-casting (see Section 8) confirmed the Fisher Index is technically preferable to either the Paasche or the Laspeyres Index alone.

In terms of delivering stability in the value of a GWAU, the current Paasche method ensures that the total GWAU for the **current** year's activity data is the same using the current year's NEP Model and the previous year's Model. The Laspeyres method would ensure that the total GWAU for the **previous** year's activity data is the same using both models. Under the Laspeyres method, the total GWAU would be expected to be higher than that generated with the Paasche method, without necessarily being so.

The Fisher Index method would produce different GWAU values for both years. The extent to which those values vary from each other would depend on the size of the relative change in average

[26] Ibid.

[27] Ibid.

episode cost between the two years. The larger the change in average cost per episode, the larger the difference in the GWAU values produced.

Setting the NEP

As described above, the Reference Cost is intended to represent the unit cost of a GWAU, taking into account changes in hospital production costs since the original Reference Cost was set using 2009-10 activity data.

As discussed in the preceding section, the current method would be expected to produce a biased result that is slightly higher than the true cost per GWAU, inflated for higher production costs. Using the same rationale as presented in that discussion, the Fisher Index method would produce a better estimate of the true production cost per GWAU, after inflation, than the current method.

9.4.2 Changes in classifications, counting rules and costing standards

In the 2009-10 year, admitted data were collected using AR-DRG v6.0. The counting rules and costing standards in place at that time preceded the implementation of national ABF, in 2013-14, by four years. The NHCDC for that year also did not separately capture costs for non-admitted activity nor for emergency services activity.

Subsequently and leading up to the implementation of national ABF in 2013-14 there were major changes to counting rules and costing standards. Those changes significantly affected the DRG profile of acute admitted activity, the amount of that activity and the costs associated with that activity. Examples include introduction of AR-DRG v7.0, bundling of emergency department (ED) costs with admitted episode costs for patients admitted through ED and treatment of ED-only episodes as non-admitted.

Further changes have happened following implementation of ABF, such as the changes to AR-DRG classifications in subsequent years. An example is the introduction of AR-DRG v8.0 in 2016-17 and its impact on coded complexity of episodes.

In economic terms relevant to the Paasche Index used to calculate Reference Cost, these changes constitute changes to the basket of goods and services used to calculate each year's index. Traditionally, economic indices are rebased when the underlying basket of goods and services is changed. In the context of the NEP's Reference Cost, this would comprise recalculating the original Reference Cost for a new base activity year.

Future years' Reference Costs would then be calculated using chained indices starting from that rebased Reference Cost.

9.4.3 Partial reporting of indexed costs

The 2009-10 NHCDC included costs for non-admitted activity for the first time but only for a subset of hospitals and Tier 2 clinics with the inevitable impact on the reliability of these estimates[28]. There were similar reliability issues for ED costs. However, the NEP has included these activity streams since NEP15.

Given these activity streams are now fully included in the basket of goods and services for which NWAU values are calculated, based on the Reference Cost, the calculation method should include (modelled) production costs for these activity streams.

Excluding these activity types introduces a source of bias in the resulting value of a GWAU. For example, a systematic movement of activity from an admitted setting to a non-admitted setting could be accompanied by a rise in average admitted episode cost (as lower complexity episodes are

[28] "The majority of jurisdictions have attempted to provide patient level cost data for the subacute classification 'AN-SNAP' and the outpatients classification 'NHCDC Tier 2 Clinics' for Round 14 and 15 respectively. However regardless of best efforts the data coverage was lower than ideal and exhibited wide variations". Pp 6 of the Round 15 NHCDC Cost Report. https://www.ihpa.gov.au/sites/default/files/publications/rd15-nhcdc-cost-report-2010-11.pdf.

removed) while non-admitted costs remain unchanged or increase less rapidly. The current method, limited to acute activity data, would over value the cost of a GWAU in this scenario.

By way of analogy, consider applying the current method (Paasche Index) including only NHCDC admitted acute activity for medical episodes. It is unlikely that production costs for medical episodes and surgical episodes are changing at the same rates. Applying a medical episodes only Reference Cost to surgical NWAU calculations would not be appropriate.

An argument against including the non-admitted costs' data in the Reference Cost calculation is that the cost data for these data streams is of variable quality. However, the calculation of Reference cost does not rely on the cost data per se, it relies only on the relevant cost models that will be converted into price weight models and the base year activity data. Given the Non-Admitted Cost Model is deemed adequate for generating the Non-Admitted Price Weight Model, implicitly it is adequate for use in calculating the Reference Cost. Similarly, as the activity data is deemed adequate for ABF funding purposes it is implicitly adequate for calculating the Reference Cost used for ABF funding.

9.5 Conclusions and recommendations

Based on the issues identified with the current method, this review makes the following suggestions:

1. That the index used for the calculation of the Reference Cost be changed to the (chained) Fisher Index.
2. That each year's index be calculated using activity data for all activity streams funded through the NEP.
3. That the index be rebased to a new base activity year. The specific year chosen should take into account the magnitude of changes in classifications, counting rules and costing standards since national ABF was introduced and changes planned for the near future.
4. That the rebased Reference Cost be set to the modelled average acute admitted episode cost for the chosen base year and cost weights be calculated using that new base Reference Cost, in that year.
5. That IHPA include the question of rebasing the Reference Cost in annual review of the NEP process and scope. This question should focus on whether there has been significant change to the activity classifications used for any of the activity streams, significant change to counting rules or related policies (such as admission criteria), and whether there has been significant change that would affect comparability of costing data over time.
6. That whenever IHPA rebases the Reference Cost, it publishes a time series of rebased historical Reference Costs and NEPs, to allow conversion of cost and price weights between NEP Determinations produced with different Reference Cost base years.

Appendix 1: Literature review

Executive Summary

Background and scope

The Independent Hospital Pricing Authority (IHPA) engaged KPMG to review current processes and statistical techniques used in the development of the NEP, drawing upon the literature and internal IHPA documentation. Findings from this process will be assessed against a performance framework that is to be developed as part of the review.

The objectives of the review are to:

- Conduct a literature review of modern data analysis and statistical modelling techniques applicable to Activity Based Funding (ABF) of hospital services;
- Review all current processes and statistical techniques used in the development of the NEP; and
- Produce a list of recommendations of improvements to the current processes and statistical techniques used in the development of the NEP, quantified against an appropriate metric.

The focus of the literature review is on:

1. Methods used to derive the NEP – including the underpinning views/philosophy of the NEP. Currently, the NEP reflects the mean cost of services for in-scope costs and for in-scope services;
2. Methods used to derive price weights – this includes methods used to determine whether multiple price weights are required and the methods for deriving the actual price weights;
3. Methods used to 'clean' the data – this relates to the removal of poor quality data;
4. Methods used to trim data – this relates to establishing the inlier population but associated with this is point one above and also the underlying reason for basing the price on an inlier population;
5. Methods for price stabilisation;
6. Methods to transform costs to price; and
7. Criteria for assessing the performance of a pricing model.

Areas of focus

Based on the scope and objectives described above, the literature review focused on the following areas:

- Review of all data preparation methodology;
- Review the fundamental calculation of all base price weights for each of the models;
- Review adjustments for legitimate and unavoidable variation in costs;
- Review of IHPA's stabilisation policies;
- Review the transformation of the cost models to pricing models; and
- Review the methodology for back-casting the NEP.

Purpose of this report

This report contains KPMG's review of the literature focusing on issues that have arisen in assessing different aspects of the National Efficient Price (NEP). In this way, the literature review is focused and will inform aspects of the NEP where there is the potential to benefit from the literature. Hence, KPMG's review and the use of the literature is developmental in nature and will continue to be undertaken as new issues are identified throughout the project.

Key findings

Many of the NEP processes and methods are targeted at dealing with a specific aspect of the NEP which are not necessarily found in or relevant to other industries. As a result, findings from the literature have limited applicability in some instances of the Review focus areas. However, there are aspects of the NEP process that mirror the challenges and context of other industries, such as the insurance and utility industries, where the literature is more informative. While the literature does not refer explicitly to 'best practice', it is instructive in considering alternative methodological approaches which are fit for purpose depending on context and policy drivers. Thus, what is common practice in one setting is not automatically transferable to other settings such as national Activity Based Funding (ABF). However, the literature is relevant insofar that it raises options and discusses their applicability.

To date, this review has found that the NEP process and methods are broadly consistent with what could be deemed from the literature to be 'common practice'. The review has identified a number of areas where there is the potential to consider an alternative approach. The findings to date, recognising that KPMG's literature work is designed to be developmental, are summarised in the following table.

Table 49: Findings by NEP review domain

NEP component	Review finding
Use of threshold values in data preparation	The current approach is to establish a standard value (e.g. to identify cost outliers). There are a range of methods referred to in the literature whereby these can be set dynamically.
Data matching	The current approach is deterministic which is generally preferred over the alternative probabilistic approach. However, it is important that analysis be undertaken of unmatched data to determine the potential to modify the current approach.
Inlier boundary points	The current approach is based on historical practice that itself was the subject of a formal review in the 1990s. This approach is based on an assumption which may not be robust today given changes in length of stay patterns. There have also been important developments internationally which suggest that alternatives need to be tested.
Determining adjustment factors in NWAU model	The current approach largely overlooks the potential for the existence of interactions between factors that have a loading. There are a number of approaches evident in the literature to address this issue.
Price stabilisation	The current approach appears to be driven by pragmatic decisions with no apparent overarching framework. While the literature itself is not overly informative in this area, there is the opportunity to ensure that the business rules are developed with a framework that is driven by agreed principles.
Transformation of the cost model	There are a number of specific issues identified that relate to a mix of process and

NEP component	Review finding
	methodological decisions. The literature does point to some alternative methods for calculating index values which are used are various steps in the process.
Back casting	The current approach largely ignores the potential for material changes in the mix of cases. The literature provides guidance on different approaches where this risk is considered material.

Source: KPMG

KPMG will continue to finalise its preliminary review of all aspects of the NEP. As new issues are identified and where there is relevant literature, the literature review will be progressively updated and discussed with IHPA.

Introduction

This section provides a background and purpose of the literature review. It describes the current National Efficient Price (NEP) process, the statistical methods applied and provides an outline of alternative models based on an assessment by the KPMG review team.

Scope

The Independent Hospital Pricing Authority (IHPA) engaged KPMG to review current processes and statistical techniques used in the development of the NEP, drawing upon the literature and internal IHPA documentation. Findings from this process will be assessed against a performance framework that is to be developed as part of the review.

The objectives of the review are to:

- Conduct a literature review of modern data analysis and statistical modelling techniques applicable to Activity Based Funding (ABF) of hospital services;
- Review all current processes and statistical techniques used in the development of the NEP; and
- Produce a list of recommendations of improvements to the current processes and statistical techniques used in the development of the NEP, quantified against an appropriate metric.

The focus of the literature review is on:

1. Methods used to derive the NEP – including the underpinning views/philosophy of the NEP. Currently, the NEP reflects the mean cost of services for in-scope costs and for in-scope services;
2. Methods used to derive price weights – this includes methods used to determine whether multiple price weights are required and the methods for deriving the actual price weights;
3. Methods used to 'clean' the data – this relates to the removal of poor quality data;
4. Methods used to trim data – this relates to establishing the inlier population but associated with this is point one above and also the underlying reason for basing the price on an inlier population;
5. Methods for price stabilisation;
6. Methods to transform costs to price; and
7. Criteria for assessing the performance of a pricing model.

Areas of focus

Based on the scope and objectives described above, the literature review focused on the following areas:

- Review of all data preparation methodology;
- Review the fundamental calculation of all base price weights for each of the models;
- Review adjustments for legitimate and unavoidable variation in costs;
- Review of IHPA's stabilisation policies;
- Review the transformation of the cost models to pricing models; and
- Review the methodology for back-casting the NEP.

Conceptual framework

KPMG's starting point for the literature review was to gain an understanding of the current process in order to identify areas where the literature could support further investigation of specific statistical techniques currently used in the development of the NEP.

A conceptual framework was developed to understand the data analysis and statistical modelling techniques and their applicability to ABF of hospital services in Australia. It incorporates the following elements:

- **Technical specifications:** including the process applied to achieve the desired output (i.e. the overall process, from data collection, data preparation, adjustments and final output, such as price);
- **Methodological elements:** including the specific methods used in preparing and transforming data to the final output; and
- **Provider and setting characteristics:** including the specific providers, settings and streams for which the method is applied (not limited to health services).

The conceptual framework provides a foundation for considering areas for further investigation and in undertaking modelling of alternative data analysis and statistical modelling techniques.

Approach

The focus of the literature review has been on methodological elements associated with the current NEP processes and where there is a prima facie case for potential improvement to an aspect of the NEP. In this way, the literature review is focused and will inform aspects of the NEP where there is the potential to benefit. Hence, KPMG's review and the use of the literature is developmental and grounded in the issues as they arise.

To gain an understanding of the existing methods applied in the NEP process, initial consultations and review of internal IHPA documentation was undertaken as part of the planning phase of the project. The literature review builds on this and provides a focus for the review.

A review of the literature was designed to address the areas of focus and was undertaken according to the following:

- Identification of publications (grey and peer-reviewed literature) according to relevant search terms through an academic database;
- Publications were selected based on the relevance to the principal review domains;
- The data and findings in the relevant literature were analysed according to the review objectives and domains; and
- Findings were synthesised and incorporated into a literature review report (this document).

The literature review will inform the further areas of investigation and modelling to be undertaken as part of the fundamental review.

Findings from the literature

This section provides an overview of the findings of the literature review as they relate to issues that have emerged from KPMG's preliminary assessment of the current NEP process and associated methods. Discussion of the literature and the issues to which is relates is provided in relation to the following review domains:

- Review of all data preparation methodology;
- Review the fundamental calculation of all base price weights for each of the models;
- Review adjustments for legitimate and unavoidable variation in costs;
- Review of IHPA's stabilisation policies;
- Review the transformation of the cost models to pricing models; and
- Review the methodology for back-casting the NEP.

Data preparation

Context and discussion

The National Pricing Model technical specifications[29] describes the data preparation phase of the NEP. This process comprises a large number of steps for which a rules based approach has been adopted by IHPA to make adjustments to data and to undertake various calculations. Many of these rules are bespoke and, as such, the literature is not overly informative in many cases.

One such example is where episodes are tripped in a particular order. A literature review is not entirely informative in this instance because the first step involves episodes being trimmed based on advice from jurisdictions. In 2017-18 there were seven establishments and 13,114 episodes trimmed and in 2018-19 there was one establishment and 125 episodes trimmed in the first step. Removing episodes based on advice that the data should not be used is self-evident and as such would not be considered in any literature review. The only comment that could be made here is that there was a substantial difference in the number of episodes trimmed between 2017-18 and 2018-19 and that this was presumably due to the relevant hospitals' data being remediated between these two reporting periods.

Some of the following steps in this data preparation trimming process can be informed by literature though. The following steps in this process involve:

- Removing episodes from hospital-Diagnostic Resource Group (DRG) combinations with extremely high or low cost-to-funding ratios;
- Removal of records with total in-scope costs ≤ $23;
- Observations with extreme outlier costs; and
- Extremely high or low cost ratios removed after deriving the preliminary regression model.

Essentially these four steps are approaches to removing one form of outlier; namely, episodes with a cost value that is deemed to be due to error because of its extreme value relevant to a reference point. Specific issues such as this example are considered in this review and illustrated in section 3.1.3 of this report**.**

Review of literature

The principal issue regarding the data preparation element of the review relates primarily to the range of targeted rules, some of which may have relevant literature for those issues that warrant further examination.

[29] Independent Hospital Pricing Authority (IHPA), 2018. Technical Specifications National Pricing Model 2018-19. IHPA

Assessment

The following table provides a summary of our assessment for each of the key elements of the data preparation followed by a more detailed discussion.

Table 50: Examples of data preparation rules and potential issues for consideration

Data preparation step	Approach	Issue for consideration
Data preparation	• Where reported, removing blood costs and/or any identified amounts related to Commonwealth pharmaceutical payments (PBS costs). Currently, a staged, deterministic matching approach using Medicare PIN, Hospital, State, Gender and Date of Birth combinations is used to remove the PBS costs.	• Literature[30] discusses two approaches to linking (deterministic or probabilistic). The IHPA rules based (deterministic) linkage is most applicable as the records from different sources consistently report sufficient information to efficiently identify links. However, there remain some records unmatched. • In probabilistic data linkage, records from two datasets are compared and brought together using several variables common to each dataset[31] Implementing probabilistic methods of linkage can be computationally intensive[32] due to the need to calculate all possible matches (a Cartesian product).
Data preparation	• Removing episodes from hospital-DRG combinations with extremely high or low cost-to-funding ratios. • Removal of records with total in-scope costs ≤ $23. • Observations with extreme outlier costs.	• Many alternative approaches to outlier identification are available (Z-score, IQR detection, Grubbs' test, K-Means, Nearest neighbours, Isolation forests[33,34]). Some are used in organisations similar to IHPA (e.g. Grubbs' test, Ireland Health Service Executive (HSE) uses a two stage IQR method[35]).

[30] Australian Bureau of Statistics (ABS) 2018. Linking Death registrations to the 2016 Census, 2016-17. Cat no. 3302.0.55.004. ABS, Canberra.
[31] Fellegi, Ivan; Sunter, Alan (December 1969). "A Theory for Record Linkage" (PDF). Journal of the American Statistical Association. 64 (328): pp. 1183-1210.
[32] Linking Data for Health Services Research: A Framework and Instructional Guide. https://www.ncbi.nlm.nih.gov/books/NBK253312/ accessed December 2018.
33 Seo, S (2006). A Review and comparison of methods for detecting outliers in univariate data sets. University of Pittsburgh.
34 Santoyo, S (2017) A Brief Overview of Outlier Detection Techniques. Available at https://towardsdatascience.com/a-brief-overview-of-outlier-detection-techniques-1e0b2c19e561, accessed December 2018.
[35] Bane, F (2015) Introduction to the Price Setting Process for Admitted Patients V 1.0. Health Service Executive.

	• Extremely high or low cost ratios removed after deriving the preliminary regression model.	Others are more complex and technically difficult to implement (e.g. Isolation Forests) and no literature was found of IHPA equivalents using such approaches. • In scope cost cut-off (currently set at $23) could be set dynamically for each DRG depending on the need to balance the need for simple and transparent methods with methodological robustness.[36] • Regression model DFFITS used (this statistic is very similar to Cook's D) and values above |0.3| are excluded. Belsley et al suggest an alternative approach to determining cut off values (i.e. 2p/sqrt(n)).[37]
Private patient adjustments (Calculation of percentage benefits, which relates to private patient adjustment – discussed further below)	30 or more separations used as a cut-off for DRG/ARDG calculation of medical and ancillary costs[38]	While the value 30 is a widely used value for minimum sampling, this value could be set dynamically by applying different methods using the characteristics of the underlying data.
Private patient adjustments (Calculation of percentage benefits, which relates to private patient adjustment – discussed further below)	The variable *AncillaryProportion* is set up to reflect this: taking a value of either 1, 0.5 or 0 for inflation.	The value could be set dynamically by using the characteristics of the underlying data.
Determine ICU adjustment level and deduct associated costs ICU bundled	• Patient-level cost data for episodes in hospitals with an eligible Intensive Care Unit (ICU) or Paediatric ICU (PICU) with ICU hours reported are analysed to estimate an average cost per ICU hour. Similar to 2.2.1 Data preparation episodes are excluded through several steps.	• ICU adjustments could be investigated further, particularly for those DRGs where ICU stay is core component of the DRG. • The cut-off for inclusion (24,000 hours, 20% mechanical ventilation) be driven by the data i.e. how much ICU activity and which particular hospitals are excluded?

[36] http://www.statisticaloutsourcingservices.com/Outlier2.pdf.
[37] Belsley, P.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. John Wiley, New York.
[38] Independent Hospital Pricing Authority (IHPA), 2015. IHPA Cost and Pricing Models Expert Guide. IHPA.

	• The eligible ICUs and PICUs are those belonging to hospitals that report more than 24,000 ICU hours and have more than 20 percent of those hours reported with the use of mechanical ventilation. • Linear regression by State/Territory was used to derive State/Territory hourly ICU costs. DFFITS statistics are used to exclude overly influential observations.	• Regression model DFFITS used (this statistic is very similar to Cook's D). Belsley et al suggest an alternative approach to determining cut off values (i.e. 2p/sqrt(n)).[39]

Data matching

The current deterministic approach to data matching using the available identifiers is logical and justifiable. However, there may be an opportunity to further investigate the records that remain unmatched and whether there are hospitals or specific circumstances that have better or worse match rates. After this exploratory analysis to understand the match rates, consideration should be given to whether alternative techniques to matching may be used on the records that remain unmatched.

Trimming outliers

With respect to the current trimming approach in identifying outliers, there are a number of statistical methods that can be used to identify and remove outliers from a data set. Among these is the "maximum normed residual test". This approach is well documented in the literature.[40] The maximum normed residual test is based on the largest absolute deviations from the average of each case mix class which is the approach used in the UK. This test is potentially less sensitive to the assumption that unit cost distributions for DRGs are statistically normal, than the current approach to outlier detection (although the test does assume that cost distributions are approximately normally distributed).[41] The NHS currently applies this method, along with other data cleaning rules, based on an independent review of its methodology.[42] Note that the method adopted by the NHS based on the review, referred to as the Grubbs' test, assumes an underlying normal distribution which is often not the case with costing data.

When determining cost outliers the Ireland HSE analysis is carried out on the average cost per inlier equivalent case per DRG and hospital combination rather than on the cost per case.[43] The first "pruning stage" removes records outside the range Q1-0.3*IQR and Q3+1.5*IQR. In the second stage of the process, a t-test is carried out of the average cost per inlier equivalent for each DRG by hospital versus the results from the first stage.

The five step trimming approach used for the NEP removes just under 0.5 per cent of the initial activity-level cost sample of admitted acute records (or about 1 in 207 records). The first step, which removes data based on advice from jurisdictions, should remain. However, the remaining four steps to remove records should be reviewed to:

- understand whether the four steps are mutually exclusive (i.e., do some records with low cost to funding ratios also have a total in scope cost of < $23);

[39] Belsley, P.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. John Wiley, New York.
[40] Prescott, P (1979), Critical values for sequential test for many outliers. Journal of applied statistics, Vol 218(1), pp36-39.
[41] NHS (2016*), 2017/18 and 2018/19 national tariff payment system.*
[42] Deloitte (2014), *Reference cost data quality.*
[43] Bane, F (2015) Introduction to the Price Setting Process for Admitted Patients V 1.0. Health Service Executive.

- understand whether the records that are trimmed are distributed evenly amongst hospitals and DRGs; and
- once the results of these steps have been clearly understood, consider alternative approaches to trimming data that achieve similar overall results (i.e., only removing under 0.5 per cent of records). An example of alternative approach may be to set a low cost and a high cost trim point overall or per DRG or use a different trimming approach as per Ireland HSE.

ICU adjustment

The current approach to ICU adjustment uses cut-offs that exclude hospitals with insufficient ICU activity and the type of ICU activity. Prior to implementing an alternative approach, analysis should be undertaken to understand the hospitals that are excluded based on the 24,000 hour cut-off (approximately 3 to 4 ICU beds at optimal occupancy[44]) and 20 per cent mechanical ventilation criteria and how including these hospitals affects the national ICU hourly rate of $210.

Private patient adjustment

Context and discussion

KPMG understands that the private patient adjustment aims to mirror actual costs. Thus, the total funding from all funding sources for public patients (Commonwealth and State/Territory funding through ABF) should be similar to the total value of funding for private patients from all sources Commonwealth and State/Territory funding through ABF, MBS and Private Health Insurance (PHI) rebates to patients). An example provided by IHPA suggests that this is not always the case. The specific example compared funding from all sources for a high volume DRG, namely O60C: child birth (vaginal delivery, minor complexity) for a specific jurisdiction. The following table summarises the illustrative example.

Table 51: Private patient adjustment and funding sources example: O60C (2 days stay, no loadings)

Funding Source	Public Patient	Private Patient
Commonwealth ABF	$1843	$1055
State ABF	$2231	$1772
MBS rebate	n/a	$520
PHI (medical fees)	n/a	$174
PHI Accommodation - minimum default benefit	n/a	$788
Total	$4074	$4209

Assessment

The NEP determination adds MBS fees from 2015/16 Hospital Casemix Protocol data to the 2015/16 National Hospital Cost Data Collection (NHCDC) costs, calculates the average cost, then works out the NEP model parameters, including private patient adjustment. The average cost per National Weighted Activity Unit (NWAU) is then indexed to 2018/19 to calculate the NEP.

Initial analysis identifies a number of reasons why it would be unlikely for the total private patient funding to match the total public patient funding under 2018/19 NEP:

- MBS schedule fees are unlikely to have grown at the same rate as the indexation rate used to inflate the 2015/16 average cost per NWAU to 2018/19 NEP;

[44] https://www.sciencedirect.com/science/article/abs/pii/S1036731413002622, accessed December 2018.

- The MBS fees added into the 2015/16 NHCDC data at the beginning of the NEP determination are based on 2015 and 2016 MBSs. There would have been at least two updates to the MBS that year (November 2015 and May 2016). This means there could have been benefits added in at three different schedule fees for the same procedure during that year (May 2015 MBS fees for July-October; November 2015 MBS fees for November-April; and/or May 2016 MBS fees for May-June);
- Hospital Casemix Protocol (HCP) data are scaled up to NHCDC data to make private patient episode numbers match. Scaling does not take into account any seasonality in private patient activity that may interact with the different MBS fee changes through the year; and
- HCP scaling also includes caps on the degree of scaling for selected hospitals.

There is also the introduction of sources of error through the total funding calculation, unrelated to the NEP (for example, the PHI contributions to accommodation fees and service fees). This is based on the assumption that the ratio of PHI relates to claims for the most recent year of available data. This is probably a different year to 2015/16 and so it may be a different ratio than what would be calculated for 2015/16. This is then applied to 2018/19 accommodation rates.

Given these considerations, getting within 5 per cent (example provided by IHPA) of total funding is arguable a good result. While the literature is not informative about these issues, it does warrant further investigation in the analysis phase of the review.

Base price weights

Context and discussion

The calculation of base price weights conceptually comprises four steps:

1. Isolate the component of care for which weights are to be set;
2. Isolate the in-scope costs associated with that component of care;
3. Calculate the average cost per instance of care (admitted episode, Emergency Department (ED) presentation or non-admitted service event), for each class within the service type classification; and
4. Divide the average by the reference cost.

There are a number of related steps (such as calibration, calculating adjustments, stabilisation and calculating reference cost), which are dealt with elsewhere within this document.

The complexity of the weights' calculations themselves principally arises in the first step listed above. The second step depends to some extent on the outcomes of the first (i.e., ensuring costs from other components of care are excluded) while the third and fourth steps are (relatively) straight forward. This complexity arises in the context of admitted care and so relates only to the acute and sub-acute models. The remainder of this section discusses potential issues relating to the acute model.

The complexity of isolating the in-scope component of care relates to the use of sameday payments, one day episode payments and both short and long stay outlier per diems in the NEP model. This requires determination of DRGs for which the various payment types will apply as well as calculation of suitable price weights. The process of determining sameday and one day payment DRGs has been assumed to be out of scope as it is not part of the model process.

The process of determining long and short stay outliers hinges on the determination of upper and lower length of stay bounds for DRGs. Episodes with lengths of stay below the lower bound (but neither sameday nor one day episodes) are then treated as short stay outliers in subsequent steps, while episodes with lengths of stay above the upper bound are treated as long stay outliers.

The acute model uses one of two methods to set the upper and lower bounds for a DRG:

- The L1.5H1.5 method – applied to DRGs in Major Diagnostic Category (MDCs) 18 and 19 (mental health care). The lower bound is set as the DRG average length of stay divided by 1.5, rounded down to a whole number. The upper bound is set at 1.5 times the DRG average length of stay, rounded down; or

- The L3H3 method – applied to all other DRGs. The method is analogous to L1.5H1.5 but uses a scale factor of 3 instead of 1.5.

Given the similarity of the two methods, only the L3H3 method is referred to in the remainder of this section. Observations made apply equally to the L1.5H1.5 method.

Review of literature

A review of ABF in five European countries identified average costs as the basis for determining the payment for each class of healthcare[45]. However, within these countries, there is some variation in how those costs are applied. For example, the UK uses them to determine a tariff of prices for individual Health Resource Groups (HRGs) and to benchmark hospital efficiency in terms of casemix weighted average price. Ireland uses its average prices to calculate price weights that are used, similarly to Australia, to calculate Weighted Units (WUs) of activity for funding purposes.

These countries also use length of stay trim points of one kind or another, similar to the acute model. The reasons for their use are similar to Australia, to accommodate for the inherent limitations of casemix classifications and of a having a set casemix price in handling extreme outlier cases.

The UK system differs materially to Australia in this regard, in not setting short stay outlier trim points and only using long stay trim points[46]. The long stay trim points are set at the upper quartile length of stay plus 1.5 times the interquartile range[47]. This method is derived from the Tukey method for boxplots[48]. In setting these upper bounds for long stay outlier identification, the UK sets a floor of 5 days as the minimum value for any such bound.

Ireland uses both short stay and long stay trim points, similar to Australia, but sets them in a materially different way[49]. To accommodate the typically skewed nature of DRG length of stay distributions, the data are first log transformed. The mean and standard deviation of the transformed data are calculated, then used to calculate lower and upper bounds at 2 times the standard deviation below and above the mean, respectively. The resulting bounds are then exponentiated to produce the lower and upper trim points for short and long stay outlier identification. The resulting trim points are also restricted to a maximum distance of 17 days from the average length of stay.

The L3H3 method was originally developed in the 1990s as part of the casemix development activities taking place at the time. The rationale for the method was based on the "the objective of reducing exceptional case risk"[50]. It was proposed as an alternative to traditional confidence interval methods based on multiples of standard deviations from the mean. The H3 method was put forward on the basis that, when length of stay distribution is sufficiently dispersed (variable), then the standard deviation based method places the high trim point further from the average, allowing greater risk to be borne by the provider.

[45] O'Reilly et al. (2012). *Paying for hospital care: the experience with implementing activity-based funding in five European countries*. Health Economics, Policy and Law (2012), 7: 73–101.
[46] NHS (March 2017). *2017/18 and 2018/19 National Tariff: currencies and prices*. Available at https://improvement.nhs.uk/documents/597/Copy_of_Annex_A_-_National_tariff_workbook.xlsx.
[47] NHS England and NHS Improvement (2017). *2017/18 and 2018/19 National Tariff Payment System*. Available at https://improvement.nhs.uk/documents/1044/2017-18_and_2018-19_National_Tariff_Payment_System.pdf.
[48] Tukey, JW (1977). *Exploratory data analysis*. Addison-Wesley.
[49] Fiachra Bane (20165). *Introduction to the Price Setting Process for Admitted Patients V1.0 26 May 2015*. Healthcare Pricing Office, HSE.
[50] McGuire, T. and Bender, J. (1995). *Two Sd or not two Sd: high exception boundary options*. Australian Casemix Bulletin. Volume 7, Number 1 February 1995; pp42–45.

The L3H3 method also was not referred to in the statistical outlier methodological literature located for this review. However, it should be noted that literature generally is focused on outlier methods for statistical modelling purposes rather than for payment system design.

This variability in practice is evident in other contexts. A review of outliers in the Portuguese health system included a comprehensive review of outlier literature. The authors concluded "there is no universal technique for the detection of outliers; … both in statistics and machine learning it is possible to find many different methodologies"[51]. The authors, in the context of their work, defined a high stay outlier as being 2 standard deviations above the geometric mean to take into account the skewness of the distribution of length of stay for the majority of DRGs. They also were influenced by another review undertaken by Cots et. al. which compared a number of methods to identify high length of stay outliers.[52] Importantly, that review noted that this method would not be appropriate for identifying low stay outliers which raises the prospects of having two methods: one for high stay outliers; and one for low stay outliers.

Assessment

The use of average costs in the weight setting process is consistent with practice in other jurisdictions. The use of inliers, outliers and associated per diem weights or payments is likewise largely consistent with international practice. The use of the L3H3 method is unique in the ABF systems reviewed.

The L3H3 method was developed in part to recognise the fact that DRG length of stay distributions are typically not Gaussian (normal). This is consistent with the Irish and Portuguese health systems for setting outlier trim points. The similarity extends further when one realises that both methods generate inlier ranges that are skewed relative to the average length of stay and which involve scaling the mean by a single factor when generating the trim points. The L3H3 method scales the mean by a factor of 3 (either dividing or multiplying by this factor). In contrast, the Irish method scales the mean by a factor equal to the power of 2 times the standard deviation of the log transformed distribution of the length of stay and the Portuguese method scales the geometric mean by a factor of 2. The Irish method introduces an additional multiplicative factor related to the standard deviation of the log transformed data, which is a consequence of the log transformation.

A key assumption in the adoption of the L3H3 method was that the coefficient of variation (CV) for DRG lengths of stay would usually lie between 0.25 and 1.5. This assumption was important to the mathematics of comparing the H3 upper trim point with the use of a mean plus 3 times standard deviation upper trim point. The analysis of length of stay data at the time validated that assumption. Importantly, that analysis used data classified to an early version of the Australian DRG classification and not to AR-DRGs. It also was carried out in an era with longer average lengths of stay (for overnight stay patients) and no exclusion of one day of stay patients.

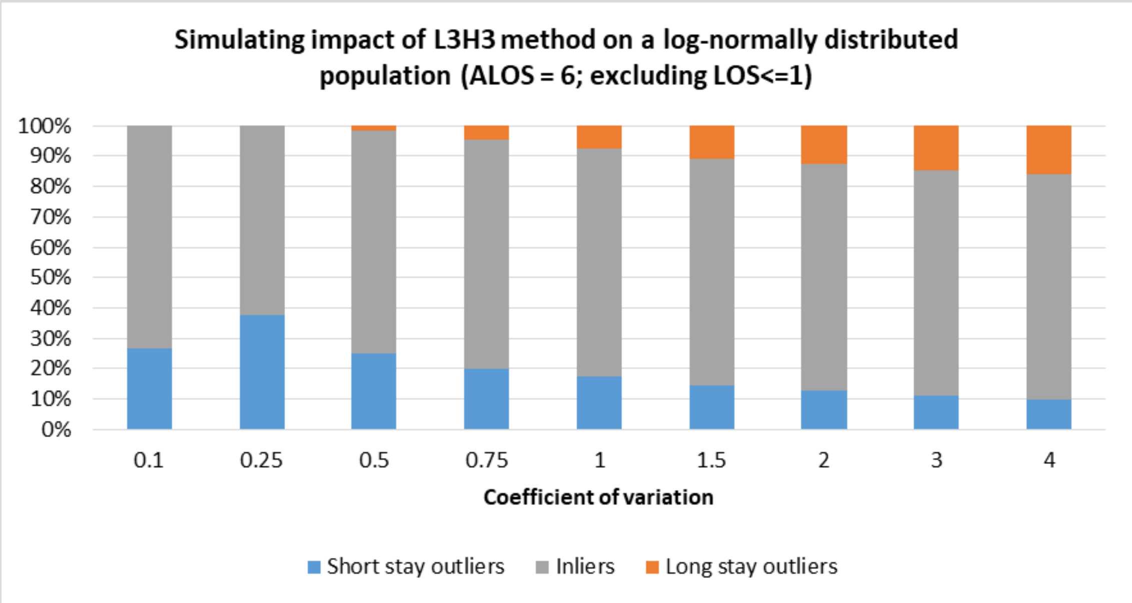
Figure 1 illustrates the impact of changes in CV on outlier numbers, for log normally distributed length of stay data. This was generated using simulated data, for the scenario of average length of stay equal to 6 days. For this scenario, the lower trim point is 2 days and the upper trim point is 18 days. Simulated episodes with length of stay less than or equal to 1 day were excluded.

From the figure it is clear there is significant potential for different DRGs to generate materially different numbers of outliers. The general trend is for the share of long stay outliers to increase as CV grows and for short stay outlier share to decrease as CV grows.

[51] Alberto et. Al. (2012). *Factors influencing hospital high length of stay outliers.* British medical Journal Health Services Research (2012), 12:265.

[52] Cots et. Al (2003). *Relevance of outlier cases in casemix systems and evaluation of trimming methods.* Health Care management Service 2003, 6:27-35.

Figure 14: Comparing proportions of outliers and inliers for different coefficients of variation, assuming log normal distribution of length of stay



Source: KPMG.

The principal reason for these observed trends is the fact that L3H3 is based solely on the average length of stay and does not include any adjustment for inherent variability of episode length of stay within a DRG.

Given this assessment, it is worth revisiting the L3H3 method and considering alternatives. The Irish and Portuguese methods present useful alternatives to test. However, the Portuguese method raises the prospect of having a different method for high and low stay outlier boundary points. The log transformation of skewed data such as episode lengths of stay is a commonly used statistical method. It is generally used to make the resulting distribution more symmetrical and closer to a Gaussian distribution. A risk is that the inclusion of a measure of variability in the length of stay distribution, combined with the exponentiation to reverse the log transformation, may introduce greater volatility in resulting trim points.

Given this assessment, there is benefit in:

- analysing length of stay distributions to test the extent to which CVs lie within the 0.25 to 1.5 range;
- considering the use of different methods to identify high and low boundary points; and
- testing alternative methods and, most notably, adaptations of the Irish log transformation method for setting trim points and/or the Portuguese method for setting trim points.

Unavoidable costs

Context and discussion

There are a range of adjustments that relate to unavoidable costs that also differ across care types. For example, in relation to acute care, currently the Paediatric Adjustment is the first of the unavoidable costs' adjustments calculated. It is calculated using a multivariate regression model. The Specialist Psychiatric Age Adjustment is applied next, again using an analysis of costs for eligible episodes.

The remaining adjustments are then calculated in subsequent steps, using further regression model(s) and other analyses. The model(s) and analyses are applied to cost data incorporating the Paediatric Adjustment.

This sequential approach means that any interactions (e.g., between paediatric status such as eligibility for the Paediatric Adjustment) and the other adjustment parameters (Indigenous status, residential remoteness, treatment remoteness, radiotherapy, dialysis) are not dealt with. For example, if there is a higher proportion of Indigenous status episodes among paediatric episodes than non-paediatric episodes within an AR-DRG, then the relevant Paediatric Adjustment may be inflated and the Indigenous Adjustment may be deflated.

Private patient adjustments

The Private Patient Service Adjustment (PPSA) and the Private Patient Accommodation Adjustment (PPAA) are also included in this section, although they do not relate to unavoidable costs but rather to the NHRA principles of:

- the Commonwealth should not pay for services twice[53]; and
- the Commonwealth should pay at a lower rate to offset revenue from other sources[54].

The PPAA method is straight forward, consisting of indexing the most recently available default insurer per diem amounts and indexing them for inflation at the rate specified by legislation. These are then divided by the NEP to produce fixed sameday and per diem PPAA weights by State and Territory.

The PPSA is also relatively straight forward, once the private patient benefits' data have been linked to the NHCDC costs' data, in the data preparation stage. In essence, the PPSA is calculated as the average percentage ratio of benefits to costs for private patients, by DRG or by adjacent DRG (ADRG) where DRG numbers are small (less than 30 private patient episodes).

Review of literature

The UK adopts the Market Forces Factor (MFF) which provides an estimate of unavoidable cost differences between healthcare providers, based on geographical location. It is used to adjust resource allocations in the NHS for local variation in input prices, so that patients are neither advantaged nor disadvantaged by such variation across the country.

Each NHS organisation has an individual MFF value, expressed in two indices:

- **Underlying index:** used to adjust funding flows: in higher cost areas, commissioners receive higher levels of funding through the allocation formula so that they are able to meet the higher costs of providers for the same level of healthcare; and
- **Payment index:** used in the national tariff to adjust prices at the local level for each provider. To create an overall MFF value for an organisation, the index value for each element of the MFF is multiplied by its proportion of total running costs. This approach applies a weight to each element of the MFF equal to its weight within total costs. The weighted index values are then added together to give an overall figure for the organisation[55].

---

[53] National Health Reform Agreement, clause A6: "*…the Commonwealth will not fund patient services through this Agreement if the same service, or any part of the same service, is funded through any of these benefit programs* [MBS, PBS, Private Health Insurance Rebate] *or any other Commonwealth program*".

[54] National Health Reform Agreement, clause A65: "*…a lower payment is appropriate, having regard to the actual cost of service delivery and the Local Hospital Network's capacity to generate revenue from other sources*".

[55] NHS (2016), *Guidance on the market forces factor: a supporting document for the 2017 to 2019 national tariff payment system.*

For the current approach adopted by IHPA in calculating the NEP, the principal issue currently under consideration is that the current approach does not explicitly consider the interactive effects of the various factors. Various methods could be used to assess these effects. Prior to considering the literature on methodological approaches, consideration first needs to be given to the likely benefit of doing so.

Private patient adjustments

In France, the tariff applied to public and private not-for-profit hospitals includes all costs associated with a hospital stay, while the equivalent for private-for-profit hospitals excludes doctors' fees and costs of pathology and imaging[56,57]. This is somewhat analogous to the Australian situation, but only tenuously, given the public sector deals only with public patients. As a result, no private patient adjustment is required and there are simply separate prices set for public and private hospitals, for each homogenous clinical group (GHS).

Assessment

It is feasible to undertake profiling of episodic cost data (such as in data sets used in the relevant steps of the acute admitted model process) to see if there are significant interactions such as between paediatric status and status for other adjustment factors.

If the profiling suggests that there is the potential for an interaction to be material, it would be worthwhile to model the associated impacts (for example, running a regression model including paediatric status with the other adjustment factor variables). The literature can be used to inform the appropriate method to test and measure the interactive effects. The results of that modelling would indicate the extent to which there is a material problem or not, and where it might sit (in terms of AR-DRGs and adjustment factors affected).

There are a range of approaches such as:

- Running a single multivariate linear regression model still using cost ratio as the dependent variable, but introducing most or all of the adjustment factor flag variables as covariates; or
- Running a single model with episode cost as the dependent variable.

An advantage of using a single model is that it directly addresses the risk of interactions between adjustment factors and reduces the number of steps and adjustments to predicted costs along the way. It is also simpler and more easily understood.

The use of generalised additive modelling may be an option to preserve some of the structural features of the current NEP model (for example, the Paediatric Adjustment is multiplicative while the ICU adjustment is additive).[58]

This issue (i.e., potential for interactive effects), is also relevant for the hospital acquired complications (HAC) adjustment factor (KPMG notes that this is out of scope of this review; however, the issue of interactive effects is a broader issue). In developing the risk adjustment method that underpins the HAC adjustment, patient age and ICU status were both identified as risk factors for presence of a HAC. These are also known risk factors for increased episode cost, for which adjustments are made in the acute admitted model. IHPA elected not to include these or other avoidable cost drivers as covariates in its linear regression modelling to "retain simplicity". This creates a risk that the cost differentials between episodes with and without HACs are biased by interactive effects between these risk factors. This may be particularly a risk for age and ICU status, which have been identified as being correlated with both risk of a HAC and episode cost.

---

[56] Or, Z. (2009), *Activity based payment in France.* Euro Observer, 11(4): 5–6.
[57] The Health Systems and Policy Monitor (2018). *Health Systems in Transition (HiT) profile of France*. Available at https://www.hspm.org/countries/france25062012/livinghit.aspx?Section=3.7%20Payment%20mechanisms&Type=Section.
[58] Barrio et al. (2013), *Use of generalised additive models to categorise continuous variables in clinical prediction*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3716996/.

Private patient adjustments

This review has not identified an alternative data source to the Hospital Casemix Protocol (HCP) that would provide better quality data for linkage to the NHCDC, in order to derive the PPSA. During this review, IHPA raised a concern with the quality of the PPSA, in relation to the fact that its application can lead to different total hospital revenue for a public patient and a private patient with identical characteristics (in terms of factors included in the NEP) except for private patient status.

It is not feasible to eliminate this phenomenon entirely. There are a number of reasons for this, including:

- MBS fees are set using a different process to the NEP, which take into account economic value of the service as well costs of service delivery (often in a community medical practice) rather than simply medical costs in a hospital setting;
- MBS fees are unlikely to have grown at the same rate since 2015/16 as the indexation rate used to index 2015/16 average cost per NWAU is not the same as the 2018/19 NEP;
- MBS fees added into the 2015/16 NHCDC data at the beginning of the NEP determination process are based on 2015 and 2016 Schedules. There were 7 MBS updates issued that cover the period from 1 July 2015 to 30 June 2016, meaning benefits may have varied for equivalent services provided to private patients at different times in that year;
- Similarly, there will be multiple MBS updates issued during any given NEP funding year, potentially changing the MBS fee payable while the PPSA remains unchanged for a given patient scenario; and
- While two patient scenarios may appear the same for NEP purposes, they may attract different MBS fees when one generates two MBS items and the other generates only one MBS item. As a result, the MBS benefits' average will not match the exact benefit for either case.

Given the phenomenon cannot be eliminated, any further consideration of potential to improve this aspect of the PPSA's performance should be based on more detailed diagnostic analysis. This could be done using 2015/16 data to test the phenomenon:

- Take a private patient episode with modelled costs and benefits (final output data sets from Stage 12 on=f acute model);
- Calculate the difference between the benefit amounts and the PPSA proportion applied to the total costs;
- Investigate the distribution of these differences to identify whether specific DRGs or patient characteristics are more strongly associated with large differences;
- Interpret the results to decide what the underlying problem cause/s is/are;
- Consider remedial action directly address the hypothesised cause/s' and/or
- Implement remedy(ies) and test results.

Price stabilisation policies

Context and discussion

IHPA generally restricts the year-on-year changes in price weights to 20 per cent under the following conditions:

- There is no change to inlier bounds, and to the status on the same-day pricing list and bundled ICU list;
- The change in the inlier cost parameter is outside +/- 20 per cent; and
- There are less than 1,000 inlier episodes.

Through the stabilisation process, IHPA may apply further restrictions (lower than +/- 20 per cent) to the price weights for high volume and high cost services to minimise volatility in changes across years. There is no clear rationale for selection of +/- 20 per cent.

Where sampling is an issue, IHPA aggregates data with previous years, although it is uncertain as to how many years are aggregated and if a sample size target drives this.

Review of literature

KPMG searched the international literature on ABF of health services and the determination of the efficient prices and costs in other countries. Our search terms included: "activity based funding"; "price stabilisation"; "adjusting variation prices"; and variations thereof.

We reviewed a number of international literature on ABF funding of health services such as the studies undertaken by Centre for Health Economics - University of York[59], and Canadian Institute for Health Information[60] to find out how other countries stabilise the movement in price/cost weights across years prior to determining the efficient prices and costs. International studies primarily discussed patients' classification, as well as how health care prices in those countries are set.

While there are a number of studies that explain why price/cost weights vary over time in other countries, no international studies were identified which address how those jurisdictions stabilise year-to-year differences in price/cost weights.

Similarly, the Australian grey literature across other utilities provide a short, sometimes one line, description of price stability. However, there is an emphasis on setting principles to underpin price stabilisation rules.[61]

Assessment

The current approach appears to be driven by pragmatic decisions with no apparent overarching framework.

KPMG has found little in the literature that represents definitive 'best practice'. Other industries also tend to take a bespoke approach to price stabilisation (e.g. utilities such as electricity and water) where a glide-path or transition adjustment are typically employed with some having a set of guiding principles/objectives that underpin their price stabilisation methods.[62]

Given this, there may be benefit in:

1. Framing the problem of (and the methodological response to) price stability more clearly (e.g., variation over time exists due to the following three potential reasons):
   - Actual variation – Australia's healthcare system is continually changing, and there is inherent variability in the health care data that IHPA uses in ABF models;
   - Methodological differences – Significant changes in the classification system, costing and pricing methodology, can cause variability in the data set. In this case, IHPA uses a back-casting approach (back-casting these changes to the year prior to implementation) to adjust significant data variation and assess year-on-year impact; and/or
   - Sampling differences – The variation in the data set can be due to changes in sample size, and/or coding practices. In the case of existing sampling differences in the data collected over years, IHPA generally reviews the data preparation and adjusts the changes in the input data before the methodology being presented to the Technical Advisory Committee.
2. Designing an empirically based method to determine allowable volatility based on historical data.

---

[59] Centre for Health economics (2007), Introducing activity-based financing: a review of experience in Australia, Denmark, Norway and Sweden. https://ideas.repec.org/p/chy/respap/30cherp.html.

[60] Canadian Institute for Health Information (2013), *The Why, the What and the How of Activity-Based Funding in Canada: A Resource for Health System Funders and Hospital Managers*. Retrieved from: https://www.cihi.ca/en/activity-based-funding.

[61] Independent Competition and Regulatory Commission (2018), *Regulated water and sewerage services prices 2018-23*. Retrieved from: http://www.icrc.act.gov.au/wp-content/uploads/2016/03/Report-1-of-2018-Final-Report-Water-Sewerage-Services-2018-23.pdf.

[62] Unpublished documentation sourced from work undertaken by KPMG for various utility service providers.

Review of the transformation of the cost model to pricing models

KPMG identified a number of specific issues in relation to the process and methods used to calculate the price index value. These issues relate to:

- Setting of the reference cost;
- Casemix changes between years;
- Number of cost years used to derive the indexation rate;
- Line of best fit against the time series; and
- Monitoring process for indexation rate.

Each of these is discussed individually in the remaining of this section.

Setting of the reference cost

Context and discussion

Section 7.3 of IHPA's Technical Specifications states that "The second step in the transformation of cost model to pricing model is the derivation of a reference cost (or a mean standardised to ensure the measure of an NWAU remains constant over time) that is used to convert the cost model into a cost weight model. Put simply, the parameters of the cost model are divided by this reference cost, converting the parameters to cost weights."

It also states that "There are two intended consequences of the selection of the reference costs:

- The change in reference costs represents change in unit costs excluding the effect of any changes in casemix; and
- The 2014-15 and 2015-16 cost weight models give the same total weighted volume when applied to the 2015-16 activity data on which the standardised growth rate is derived."

Review of literature

Given the "reference cost" is a specific concept within the NEP framework, there is no formal literature on this topic. Therefore, we have formed our assessment largely based on grey literature and experience from an alternative industry.

Assessment

Although there is no formal literature on the topic of "reference cost", based on our experience from the insurance industry –where the rationale of actuarial pricing is well established with a matured range of generally acceptable practices –we have drawn out similarities in approach and provide comments, noting differences in contexts.

In actuarial pricing for insurance risk, there is a concept of "base premium" that is analogous to the concept of "reference cost", which equates to the actuarial price of the expected losses arising from the baseline risk profile. Analytical techniques (such as generalised linear modelling) are used to determine the relative risk among different policies with different risk factors and levels. The relative risk is quantified and expressed as a risk relativity loading to the "base premium", which in concept is analogous to "cost weights" in the current approach.

Our preliminary assessment is that the current conceptual approach to "reference cost" appears reasonable in light of acceptable practices seen in an alternative industry, though we note there are potential alternatives in the details (as discussed in the following sections).

The method used to calculate the reference cost is susceptible when there are structural differences between the two years of activity data. For example, if the prior year has no cost data for a casemix class (DRG), but the subsequent year does, then this distorts the reference cost unless data for the prior year has deemed values.

Casemix changes between years

Context and discussion

Section 7.3 of IHPA's Technical Specifications states the following "From 2010-11 onward, the reference cost is defined so that change in the reference cost over time reflects change in unit costs, excluding any influence of underlying changes in activity profiles between years (i.e., casemix change)."

The reference cost and indexation approach used in IHPA's current framework refers to the latest year casemix as the "base", which is an approach analogous to the Paasche Index. We observe that a consistent approach is also adopted for back-casting (see section 3.5).

Review of literature

KPMG's literature review included:

- IHPA document, in particular Section 7 of the Technical Specifications 2018-19 National Pricing Model March 2018; and
- Grey literature around different conceptual designs of a consumer price index to provide context on pros and cons of potential alternative approaches where relevant.

Assessment

Paasche Index has relative pros and cons compared to other approaches, which are discussed in more detail in relation to back-casting. It is likely to be desirable to maintain internal consistency of concepts and approach throughout the IHPA pricing framework, so we recommend considerations of alternatives to also include any flow-on implications.

The current approach determines the underlying growth in reference cost by excluding the casemix related change between two consecutive years. The use of "curve fitting" techniques to extrapolate future growth rates assumes the casemix to be stable across all years to provide a reasonable estimate of the indexation rate and future cost. We recommend that IHPA assesses and tests stability of casemix over time.

Number of cost years used to derive the indexation rate

Context and discussion

Section 7.4 IHPA's Technical Specifications states the following "To derive this rate, the 2015-16 cost model is applied retrospectively to the five years of patient costed admitted acute activity data prior to 2015-16…"

The Specifications do not discuss the rationales for deciding on using a rolling most recent five years' experience for indexation, or how well this approach been historically in predicting cost growths. The document included a graph showing the adopted extrapolation using past five years of history, but did not include scenarios or discussions of how sensitive the overall results are to the number of retrospective years used.

Review of literature

KPMG's literature review included:

- IHPA document, in particular Section 7 of the Technical Specifications 2018-19 National Pricing Model March 2018; and
- Grey literature (e.g., actuarial advice and reports) and experience from the insurance industry around the rationale of selecting the number of years' data for modelling. For actuarial modelling in insurance, all else being equal, this decision generally depends on the duration of the insurance liabilities.
  - "Long tailed" duration products such as public liabilities and compulsory third party (CTP) bodily injury claims can take a long time from the date of the accident to when the claims are reported, assessed and settled. Therefore the actuarial models are often built to examine data over a long time (e.g., more than 10 years).

- "Short tailed" duration products such as domestic motor or home contents insurance claim amounts are often known with greater degree of certainty and paid out quickly, and therefore the models are built on more recent trends over a fewer number of years' data (e.g. 3-5 years).

Assessment

National ABF deals with episodes of a relatively fixed and short term duration which is equivalent to what is referred to as "short tailed" products in an insurance context. To this end, the use of five years' data is consistent with standard practice in other industries such as the insurance industry. We suggest exploring the value in considering an approach that examines historical inflation trends corresponding to the duration and structure of the underlying cost drivers if it is reasonable and practical. There is an implicit assumption, however, that there are no structural changes in underlying cost drivers across the five year period compared to future years as, if this does occur, the current approach may not produce as a reliable result. Thus, regardless of the time period selected, an assessment in differential trends across the various cost drivers (for example, labour costs and non-labour costs) should be considered.

Line of best fit against the time series

Context and discussion

Section 7.4 of IHPA's Technical Specifications states the following "The next step in the process of deriving an annual indexation rate is to model a line of best fit against the time series of cost ratios (or equivalently, against the time series of costs per weighted separation)".

The current indexation approach used to derive the NEP is primarily based on an extrapolation of past cost ratio (or costs per weighted separation) trends. This method assumes that historical cost drivers, including efficiencies gained, continue in the future in the same manner. The extent to which the future behaviour of the principal cost drivers differs from historical trends introduces a potential error in the predicted values.

Review of literature

KPMG's literature review included:

- IHPA document, in particular Section 7 of the Technical Specifications 2018-19 National Pricing Model March 2018; and
- James, Witten, Hastie & Tibshirani (2017), An Introduction to Statistical Learning, Springer.

The current "curve fitting" approach is primarily based on a "prediction" philosophy to modelling, whereas another philosophy is "inference" which aims to identify the underlying drivers of the phenomenon being modelled (e.g. regression analysis). There is literature that directly or indirectly discusses the difference between "prediction" and "inference" as a fundamental design consideration in modelling. The literature itself generally does not tend to prefer one method over the other as much depends on context and complexity/feasibility of using one method over the other. A discussion of this issue is found in the following literature:

- Hart, Buchanan & Howe (2007) Actuarial Practice of General Insurance. The Institute of Actuaries of Australia;
- Evans & Miller (2017) Challenge, Inflation, Opportunity. Retrieved from: https://www.actuaries.asn.au/Library/Events/%20InjuryDisabilitySchemesSeminar/2017/ChallengeInflationPaper.pdf;
- De Ravin & Fowlds (2010) Inflation Risk in General Insurance. Retrieved from: https://www.actuaries.asn.au/Library/Events/GIS2010/GIS10_Paper_Deravin.pdf;
- Miller (2010) Towards a Better Inflation Forecast. Retrieved from: https://www.actuaries.asn.au/Library/Events/GIS2010/GIS10_Paper_Miller_Inflation.pdf; and

- Pearson & Beynon (2007) Superimposed Inflation – Australian Accident Compensation Landscape in 2007. Retrieved from: https://actuaries.asn.au/Library/p.2_ACS07_paper_Pearson_Superimposed%20inflation_paper.pdf .

Assessment

In the insurance industry, as an example, there are different drivers of the year-on-year increase in claims costs:

- Building and medical equipment costs are more closely linked to CPI;
- Labour and expert costs are more closely linked to wages and salaries (such as AWE); and
- Residual increase in addition to normal CPI/AWE inflation may be referred to as "superimposed inflation" in the insurance industry. Examples are inflation in court awards or new and unexpected heads of damage introduced by the legal and judicial processes.

The current approach uses a single overall index, but there may be scope to model indexation at a more granular level where there are significant differences in inflation levels and/or cost drivers. In the insurance industry, it is common practice to examine and arrive at different inflation assumptions by different product or claim types.

The current indexation approach, which uses a curve fitting technique applied to mix-adjusted historical average costs, appears to be in line with acceptable practices seen in the insurance industry. Alternative approaches (as a variation of the same theme) may be adopting a different parametric curve instead of the exponential, and changing how far back in history data sets are used as the basis for extrapolation. These are also discussed in our other observations.

There is a more fundamental question of design on "prediction" versus. "inference" (or a combination of the two). The design basis for modelling depends on the stability of the underlying cost drivers, and the importance (and level of interest) in understanding the way that inflation is affected by these drivers. Modelling design with emphasis on "prediction" tends to place a higher reliance on data quality, stability and comparability over time. In some situations, we wish to understand the relationship between future inflation and its drivers (i.e., inference), and our goal is not necessarily limited to making blind predictions.

Monitoring process for indexation rate

Context and discussion

Section 7.4 does not refer to an ongoing monitoring process in place to compare the modelled indexation against actual inflation over time. We understand this is performed on an ad-hoc basis as a separate process.

Review of literature

KPMG's literature review included:

- IHPA document, in particular Section 7 of the Technical Specifications 2018-19 National Pricing Model March 2018; and
- Bellis, C., & Shepherd, J. (2003). Applying the Actuarial Control Cycle. In C. Bellis, J. Shepherd, & R. Lyon (Eds.), Understanding Actuarial Management: the actuarial control cycle (pp. 449-456). Sydney Australia: The Institute of Actuaries of Australia.
  - This literature outlines the actuarial control cycle as a fundamental conceptual framework for actuarial approaches.
- Grey literature and industry experience around accepted actuarial practice in evaluating model performance and monitoring emerging actual experience against expectations.

Assessment

Monitoring of actual experience against expectations is standard practice in the insurance industry. This is generally referred to as the "actual versus expected" analysis and "analysis of change". These approaches are effectively an attribution analysis performed to provide an understanding of the key drivers for actual experience to be different to expectations implied by the actuarial assumptions. This monitoring process can also inform the ongoing performance of the selected indexation approach compared to alternative approaches.

Back-casting the NEP

Context and discussion

As used in back-casting, the volume multiplier is a type of Paasche Index that attempts to estimate the overall impact of change in the NEP structure on NWAU volumes between the previous year and the current year. The key difference from the orthodox Paasche Index is the volume multiplier also includes change in AR-DRG classification from time to time. Nonetheless, the relationship is strong enough to be used to inform consideration of alternatives. Note that when there is no change in AR-DRG classification, the index effectively estimates average change in price weights between the years.

Review of literature

Indices such as the Paasche Index and its alternatives are used in various contexts so assessing literature conclusions about the best method needs to consider the context of application. For example, the Australian Bureau of Statistics (ABS) publishes extensively about its decision on which method to use for the purposes of calculating the CPI.[63] The ABS considers both the technical merit of different methods and the importance for the CPI to reflect changes in consumer patterns noting that price changes can affect those patterns. The ABS has also published its assessment of these alternative indices in order to assess any bias of the estimates using various approaches to mathematically assess alternative indices.[64] Given that IHPA uses what is in essence a form of the Paasche Index for the purposes of back-casting, some of the alternatives could not be practically applied by IHPA (further discussed below).

Paasche Index is often used to estimate average price inflation across a range of goods and services between periods. In that context, it tends to underestimate inflation due to the failure to consider impact of price changes on providers' behaviour. In the NEP context, it is likely to overestimate average impact of NEP model changes, as it does not consider changes in provider behaviour. Changes in price weights could impact on hospitals' behaviour, in terms of the amount and mix of activity undertaken and on coding practices. This is evidenced also by ABF's past and continued use as a tool by funders and purchasers to change hospitals' behaviour.

Laspeyres is a closely related alternative. It suffers from the opposite problem (i.e., that is, it would tend to under estimate average NEP model impact). The limitations of both indices are widely acknowledged in the literature.[65]

Lowe Index could be implemented using a standard casemix (different to previous and current year) to measure relative impact of changes in price weights. This would be hard to use for back-casting purposes when there are changes in coding patterns. The base year would need to be regularly changed. However, it is difficult to adjust for changes to historical data where there are changes in coding behaviours/pattern in subsequent years. It would similarly suffer from changes in provider behaviour, and potentially more so, due to compounding changes for multiple years between the chosen base year and the current year.

---

[63] ABS (2017). Consumer Price Index. *Concepts, sources and methods.* Retrieved from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/6461.0Main%20Features112017
[64] Yehezkael, R (2009), *Computational tests as a means of assessing the inaccuracies of index numbers*. Retrieved from: http://homedir.jct.ac.il/~rafi/inaccura.pdf
[65] Policonomics (2017), *Laspeyres and Paasche indices.* Retrieved from: https://policonomics.com/laspeyres-paasche/

Marshall-Edgeworth attempts to address the above issues identified for Paasche and Laspeyres Indices. It does this by averaging the quantities of goods and services between the two years before applying the respective prices.[66] As such, it relies on the classifications of goods and services to be the same for the two years. This would not be practicable for the NEP as AR-DRG versions regularly change between years.

Fisher Index likewise attempts to address the respective issues with Paasche and Laspeyres indices. It does so by calculating the geometric mean of the two indices.[67]

Assessment

The current method is known from the literature to underestimate the 'true value', particularly given the likely impact of changes in supplier behaviour on the current volume multipliers. Therefore, it would be worth testing the impact of adopting a Laspeyres version of the multiplier and a Fisher version of the multiplier which attempt to address this shortcoming, although both methods have their own potential bias. The methods can be compared for bias using mathematical constructs used by the ABS in its assessment of alternative indices.

[66] Hansen, C (2006), Price-updating of weights in the CPI. Retrieved from: http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+9/$file/2006%209th%20Meeting%20-%20Carsten%20Hansen%20-%20Price%20updating%20of%20weights%20in%20the%20CPI.pdf

[67] Ibid.

Conclusion and Next Steps

Many of the NEP processes and methods are targeted at dealing with specific aspects of the NEP which are not necessarily found in, or relevant to, other industries. As a result, findings from the literature have limited applicability in some instances of the review focus areas. However, there are aspects of the NEP process that mirror the challenges and context of other industries, such as the insurance and utility industries, where the literature is more informative. While the literature does not refer explicitly to 'best practice', it is instructive in considering alternative methodological approaches which are fit for purpose depending on context and policy drivers. What is common practice in one setting is not automatically transferable to other settings such as national ABF. However, the literature is relevant insofar that it raises options and discusses their applicability.

To date, this review has found that the NEP process and methods are broadly consistent with what could be deemed from the literature to be 'common practice'. The review has identified a number of areas where there is the potential to consider an alternative approach. The findings to date, recognising that KPMG's literature work is designed to be developmental, are summarised as follows.

Use of threshold values in data preparation

The current approach is to establish a standard value (e.g., to identify cost outliers). There are a range of methods referred to in the literature whereby these can be set dynamically.

Setting inlier boundary points

The current approach is one that was first adopted in Australia in the 1990s based on an assumption regarding values of coefficient of variation that may not be applicable today given changes in length of stay patterns. Further, there has been considerable developments in this area internationally that warrants an assessment of alternative methods two of which have been noted in this report.

Determining adjustment factors in NWAU model

The current approach largely overlooks the potential for the existence of interactions between factors that have a loading. There are a number of approaches evident in the literature to address this issue.

Price stabilisation

The current approach appears to be driven by pragmatic decisions with no apparent overarching framework. While the literature itself is not overly informative in this area, there is the opportunity to ensure that the business rules are developed with a framework that is driven by agreed principles.

Transformation of the cost model

There are a number of specific issues identified that relate to a mix of process and methodological decisions. The literature does point to some alternative methods for calculating index values which are used are various steps in the process.

Back casting

The current approach largely ignores the potential for material changes in the mix of cases. The literature provides guidance on different approaches where this risk is considered material.

Areas for further investigation

KPMG will continue to finalise its preliminary review of aspects of the NEP.

As new issues are identified, and where there is relevant literature, the literature review will be progressively updated and discussed with IHPA.

Appendix 2: Assessment metrics

The metrics used to assess the alternative methods are described in detail in this Appendix.

Level 1: funding model assessment criteria

These criteria relate how the proposed change impacts on the funding model in terms of:

1) Its effectiveness in achieving ABF objectives
2) Feasibility of application of the finding model
3) Risks associated with the funding model

Effectiveness

The objectives of national ABF as stated in the National Health Reform Agreement[68] that relate specifically to ABF are:

- improve patient access to services and public hospital efficiency through the use of activity based funding (ABF) based on a national efficient price (Schedule A);
- ensure the sustainability of funding for public hospitals by increasing the Commonwealth's share of public hospital funding through an increased contribution to the costs of growth (Schedule A);
- improve the transparency of public hospital funding through a National Health Funding Pool and a nationally consistent approach to ABF (Schedule A and B);

The first and third of the objectives above relate specifically to the functional form of ABF (i.e. methods used to derive the NEP and the NWAU model). The second of the objectives above is not dependent on the functional form of the model and this is not used to assess potential changes to the NEP methodology. The first of these objectives has two distinct components namely, patient access and hospital efficiency. Therefore, three distinct criteria have been formed that relate to the effectiveness of the funding model in terms of meeting the national ABF objectives:

- Patient access to services: does the change impact on the capacity of hospitals to enhance patient access to services?
- Public Hospital efficiency: does the change better reflect the efficient cost of service?
- Transparency of public hospital funding: does the change lead to a more transparent link between funding and cost of services?

Feasibility

Feasibility relates to the extent to which the proposed change is able to be operationalised and sustained with the practical constraints of data availability. The following specific criteria will be applied:

- Does the change require data that is not yet available/collected and would be difficult for data providers to maintain the data collection required?
- Does the change require data whose quality is difficult to sustain at a level essential to ensure reliability of the outcome of applying the change?
- Does the change require a volume of data that is difficult to sustain (that is, the data is available, but the change requires of volume of data that is not able to be sustained in normal operations of the NEP)?

---

[68] The NHRA has nine objectives of which three relate specifically to ABF. http://www.federalfinancialrelations.gov.au/content/npa/health/_archive/national-agreement.pdf.

Risks

Two risk criteria will be applied as follows:

- Perverse incentives: does the change create perverse incentives that cannot be mitigated?
- Gaming: can the proposed change lead to gaming that cannot be mitigated?

Level 2: NEP attribute specific metrics

When applicable additional metrics may be used (e.g., there is a specific nuance associated with the aspect of the NEP being assessed where the standard metrics used in the assessment may not be sufficiently sensitive to the issue under consideration).

Level 3: macro impacts

The method improves the analytical performance of the NEP in terms of:

- Explanation of costs (measured by R^2)
- Cost ratio (ratio of modelled total cost to actual total cost for a LHN)
- Cost model costs' accuracy (measured by SMAPE)

Note that Symmetric mean absolute percentage error (SMAPE)) is an accuracy measure based on percentage (or relative) errors.[69]

[69] Tofallis, C (2015) "A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation", Journal of the Operational Research Society, 66(8), 1352-1362.

Appendix 3: AR-DRGs with no variation in length of stay

A number of DRGs with zero standard deviation in length of stay were identified in the analysis assessing the sensitivity of boundary points to CV values (see section 3.5.2). These are listed in the following table.

Table 52: DRGs with a LOS standard deviation of zero

DRG	Description
B40Z	Plasmapheresis W Neurological Disease, Sameday
I80Z	Femoral Fractures, Transferred to Acute Facility <2 Days
L41Z	Cystourethroscopy for Urinary Disorder, Sameday
L68Z	Peritoneal Dialysis
M40Z	Cystourethroscopy for Male Reproductive System Disorder, Sameday
U40A	Mental Health Treatment W ECT, Sameday, Major Complexity
U40B	Mental Health Treatment W ECT, Sameday, Minor Complexity
U60Z	Mental Health Treatment W/O ECT, Sameday

Appendix 4: Technical description of the dynamic inlier method

Overview

A dynamic method to set boundary points was developed to deal with the variation the underlying length of stay distributions. Four specific types of statistical distributions and one non-specific distribution were identified as typifying many of the DRG distributions namely:

- Normal distributions
- Log normal distributions
- Dispersed distributions
- Multi-modal distributions
- Other (non-described) distributions

A boundary point method was then chosen for each distribution as follows:

Method description

The following steps would replace most of the procedure currently carried out in *Stage 04.1 Inlier Bounds – CI* module. Specifically it would replace the steps between the *import and filter apc* step and the *add bounds to masterlist* step.

Bounds would be set as follows:

1. Test each DRG for Normality using PROC UNIVARIATE. For DRGs returning a positive result (i.e., effectively Normally distributed), exclude influential extreme values as follows:

- Count the number of episodes present (N).
- Calculate the mean length of stay (ALOS) and the standard deviation.
- Calculate an influence threshold (INFL) as follows:
  INFL = ALOS + [(N – 1)*0.05 + 1]*ALOS
  This is designed to detect episodes that would decrease the ALOS value by at least 5% (or 1 day) if that episode was not present. The smallest allowable value for this influence threshold is one day of stay above.
- Calculate a threshold equal to 3 standard deviations above the mean (UB):
  UB = ALOS + 3*SD
- Check each episode length of stay above UB to see if its length of stay is greater than the influence threshold:
  LOS > INFL?
  If this is the case, flag the episode for exclusion.
- Delete all of the episodes flagged for exclusion.

  Then, for the reduced data set of episodes, calculate a 95% confidence interval and use its bounds as the inlier boundary points.

2. For DRGs returning a negative result for the Normality test in #1, apply PROC UNIVARIATE to test for Log Normality. If the test returns a positive result, exclude influential extreme values as follows:

- Count the number of episodes present (N).
- Calculate the mean length of stay (ALOS).

- Calculate an influence threshold (INFL) as follows:
  INFL = ALOS + [(N – 1)*0.05 + 1]*ALOS

- Log transform and calculate a Mean+3*SD value for the transformed data. Exponentiate this value to generate a threshold (UB).

- Check each episode length of stay above UB to see if its length of stay is greater than the influence threshold:
  LOS > INFL?
  If this is the case, flag the episode for exclusion.

- Delete all of the episodes flagged for exclusion.

Then, for the reduced data set of episodes, log transform and generate a 95% confidence interval and exponentiation to produce high and low bounds. Set a minimum for the low bound of L3 (one third of the ALOS).

3. For DRGs returning a negative result to the Log Normality test in #2, apply PROC UNIVARIATE to 1/LOS (inverse of length of stay) for Log Normality. If the test result is positive then do analogously as for #2. This time, set a maximum upper bound value equal to H3 (i.e., 3 times the ALOS).
4. For DRGs returning a negative result to the inverse Log Normality test in #3, test for multimodality using the Silverman test[70]. The Silverman Test has been chosen as it is one of the few multimodality tests that can be executed from SAS. We propose to use the relevant macros designed by Neville and Brownstein[71]. Note that these macros use PROC IML to call the R language. This will require R to be installed on the IHPA Portal SAS server. This appears to be the case.

If the Silverman Test returns a positive result (2 or more modes), exclude influential extreme values as follows:

- Count the number of episodes present (N).

- Calculate the mean length of stay (ALOS).

- Calculate an influence threshold (INFL) as follows:
  INFL = ALOS + [(N – 1)*0.05 + 1]*ALOS

- Set a threshold length of stay equal to the 98th percentile of the length of stay distribution (UB).

- Check each episode length of stay above UB to see if its length of stay is greater than the influence threshold:
  LOS > INFL?
  If this is the case, flag the episode for exclusion.

- Delete all the episodes flagged for exclusion.

Then, for the reduced data set of episodes apply the current inlier bounds' calculation method (L3H3 or L1.5H1.5).

[70] Silverman B. W. (1981). *Using Kernel Density Estimates to Investigate Multimodality*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 43, No. 1, pp. 97-99. https://www.jstor.org/stable/2985156. Accessed: 19-02-2019 09:23 UTC
[71] Neville Z. and Brownstein N. (2018). *Macros to Conduct Tests of Multimodality in SAS*.

5. For DRGs returning a negative result to the Silverman test in #4, apply the current method applicable to that DRG (i.e., either L3H3 or L1.5H1.5), follow the procedure described for multimodal DRG distributions in #4.

Appendix 5: Portuguese inlier boundary results

The Portuguese inlier boundary point method was included in the initial exploratory analysis but not included in the full analysis as this method only considers high boundary points. This Appendix shows the results of the exploratory analyses.

Table 53: Comparison of boundary limits using current NEP trimming method

	Current	Portuguese
Lower		
Mean (days)	3	N/A
Median (days)	1	N/A
Min (days)	1	N/A
Max (days)	45	N/A
Relative to the current method, share of DRGs where boundary:		
Decreased	[redacted]	N/A
Stayed the same	[redacted]	N/A
Increased	[redacted]	N/A
Upper		
Mean (days)	22	22
Median (days)	13	14
Min (days)	2	2
Max (days)	363	281
Relative to the current method, share of DRGs where boundary:		
Decreased	[redacted]	60%
Stayed the same	[redacted]	13%
Increased	[redacted]	27%

The underlying nature of the length of stay distributions influences the impact of these methods on whether the high boundary point increases or decreases.

Table 54: mean CV values by direction of change in high boundary point

Direction of change in boundary value	Mean CV value		
	Irish	Portuguese	IQR
Decreased	0.93	0.84	1.09
Increased	1.20	1.62	1.13

Appendix 6: Impact of Irish trimming method on the Irish inlier boundary method

As noted in the main part of this report, the Irish trimming method is an integral part of its inlier boundary method as the trimming method remove episodes with extreme (valid) length of stay values from setting boundary points as otherwise, these extreme values would adversely impact on the boundary points.

This appendix, compares the impact on boundary points set by the current NEP method.

The following table illustrates the impact of this interdependency.

Table 55: comparison of Irish method using NEP trimming method with using the Irish trimming method

Statistical parameter	**Current NEP trimming plus Irish boundary methods**		**Irish trimming plus Irish boundary methods**	
	Number	**%**	**Number**	**%**
WIP/No cost	270,006	4.8%	270,006	4.8%
Trimmed	25,723	0.5%	371,474	6.6%
Same day	412,059	7.3%	412,059	7.3%
Short stay outliers	68,519	1.2%	110,484	2.0%
Inliers	4,821,011	85.9%	4,403,127	78.5%
Long stay outliers	15,048	0.3%	45,216	0.8%
Total acute episodes	5,612,366	100.00%	5,612,366	100.0%
Relative to the current NEP trimming plus Irish boundary method, share of DRGs where the lower boundary point:				
Decreased	[redacted]		-	-
Stayed the same	[redacted]		591	74%
Increased	[redacted]		207	26%
Relative to the current NEP trimming plus Irish boundary method, share of DRGs where the high boundary point:				
Decreased	[redacted]		750	94%
Stayed the same	[redacted]		48	6%
Increased	[redacted]		-	-

The above table illustrates the impact of using trimming for the purposes of desensitising boundary point method to extreme values rather than using trimming just to eliminate spurious data. For example, using trimming to remove extreme values from boundary point calculations results in a significant increase in the number of episodes trimmed (25,723 to 371,474), and tends to increase the lower boundary point, decrease the high boundary point, and reduce the range for inliers. This then has the effect as expected of increasing the low boundary point value and reducing the high boundary point value for many DRGs as illustrated in the following table:

Table 56: Comparison of boundary limits using current trimming + Irish method and Irish trimming + Irish method.

	Irish method + NEP trimming	Irish method + Irish trimming
	Lower	
Mean (days)	2	2
Median (days)	1	1
Min (days)	1	1
Max (days)	19	22
Share of DRGs where boundary:		
Decreased		-
Stayed the same		74%
Increased		26%
Upper		
	Upper	
Mean (days)	30	21
Median (days)	17	13
Min (days)	2	2
Max (days)	450	289
Share of DRGs where boundary:		
Decreased		94%
Stayed the same		6%
Increased		-

Appendix 7: Sensitivity of boundary points to CV values

This Appendix shows the results of the analysis undertaken in section 3.5.2 of the sensitivity of boundary points to CV values for other boundary points methods considered during the exploratory phase of the work.

Figure 15: Outliers versus coefficient of variation – Irish method using NEP trimming method

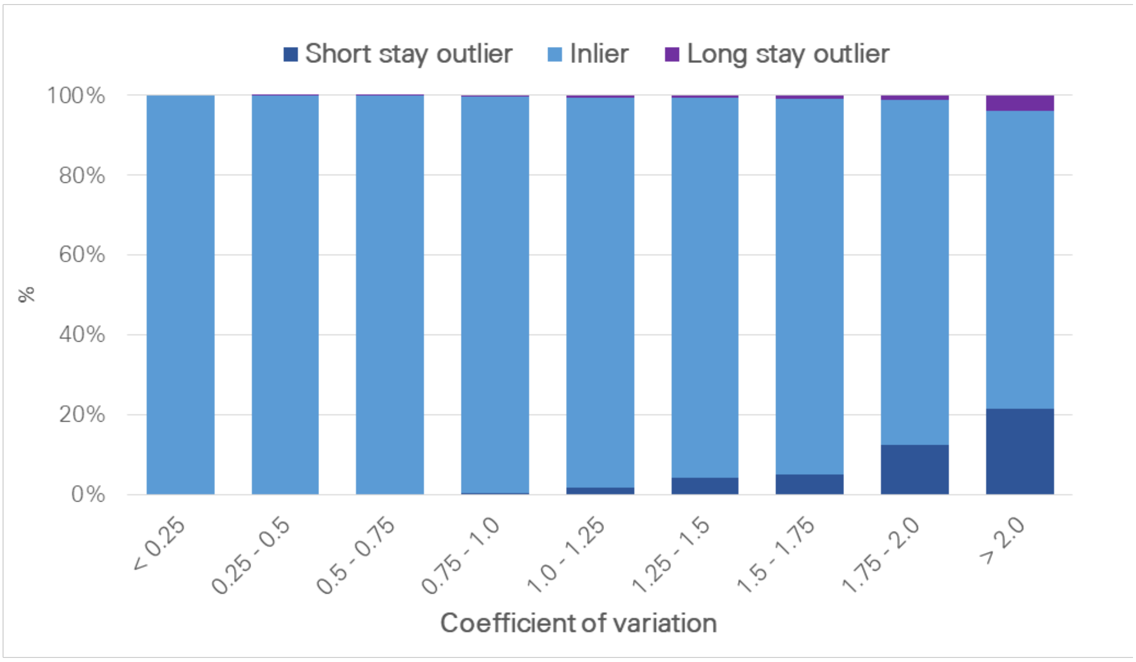


Figure 16: Outliers versus coefficient of variation –Irish inlier and Irish trimming method

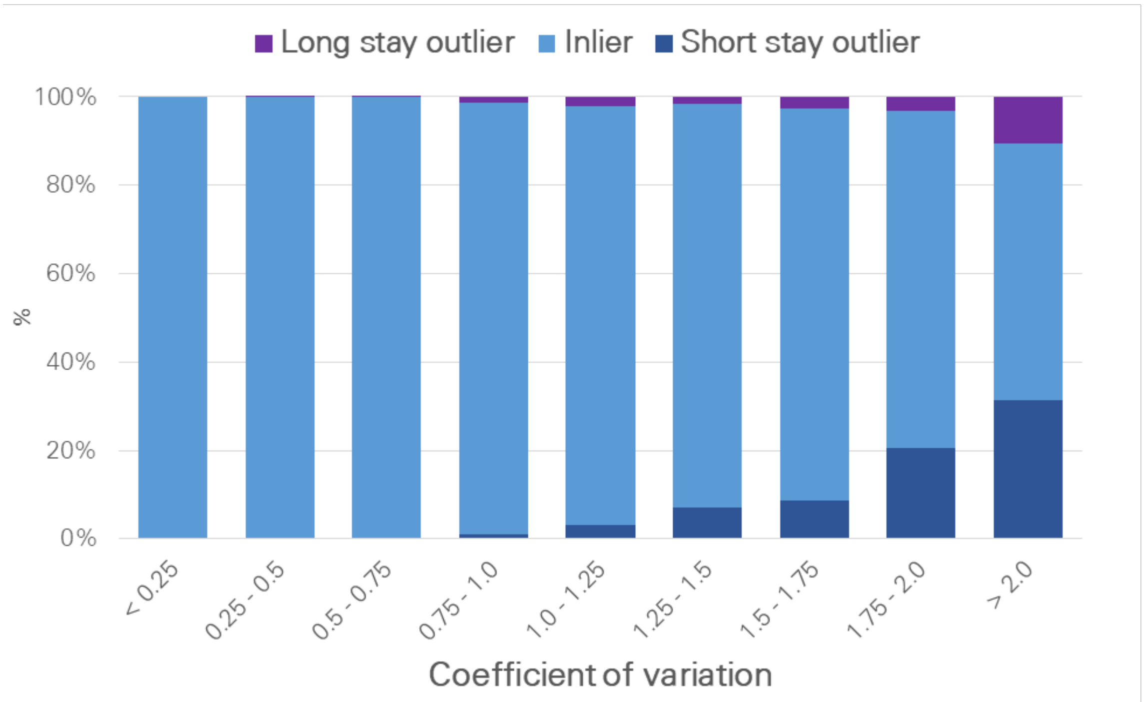


Figure 17: Outliers versus coefficient of variation – Portuguese method using NEP trimming method

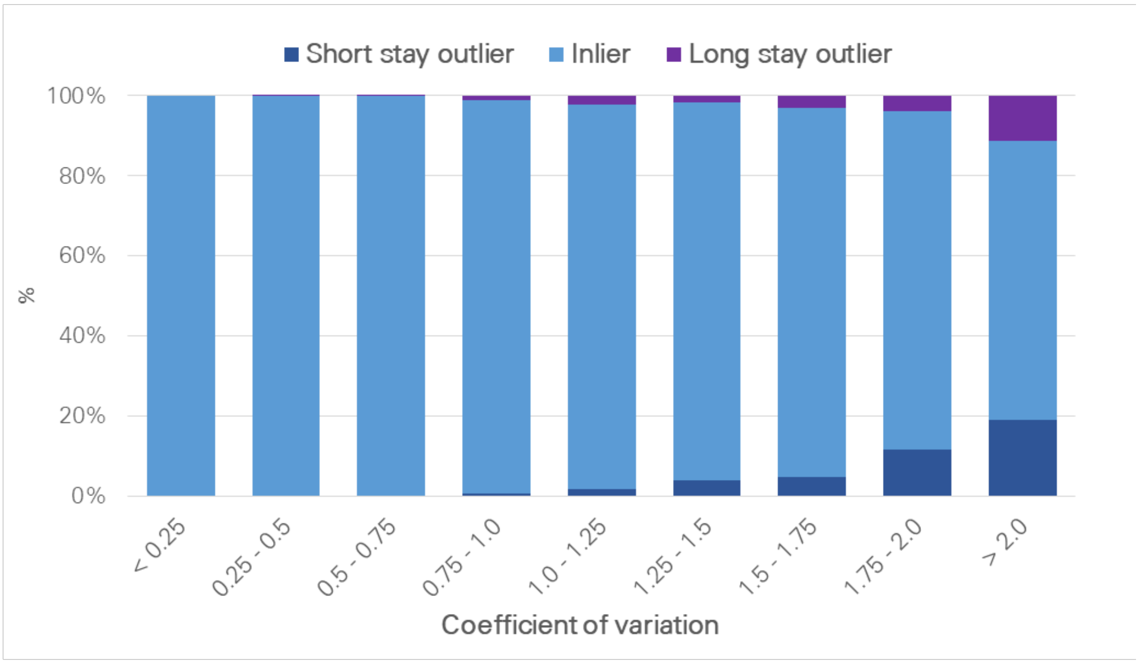
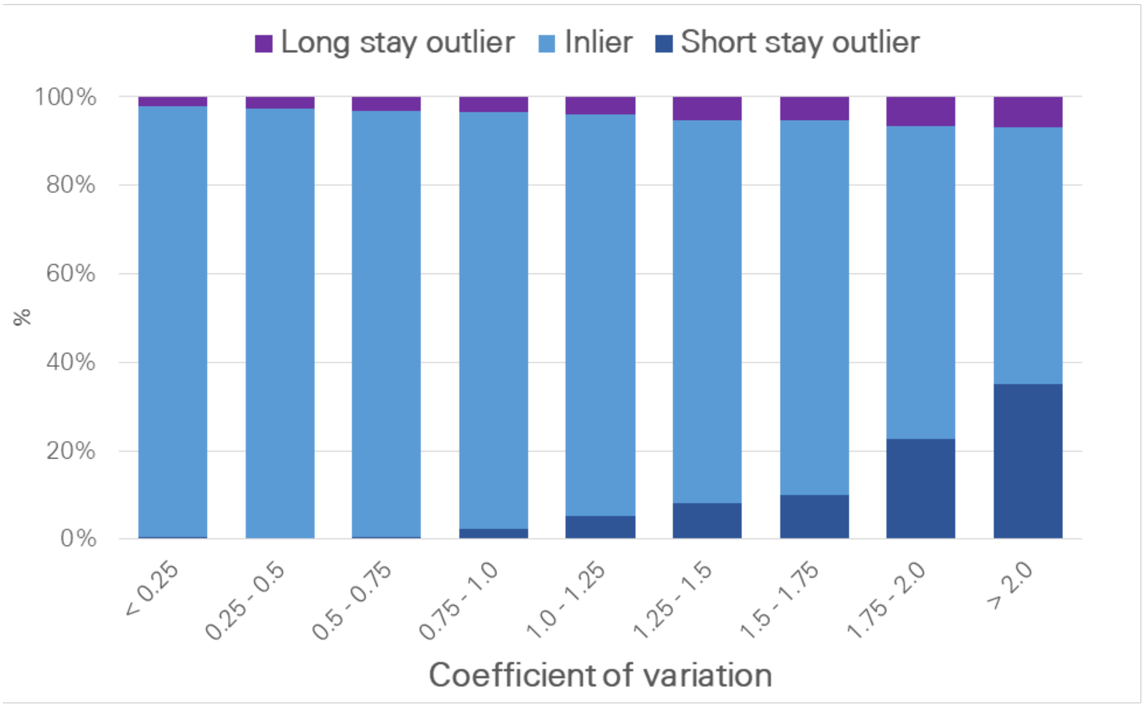


Figure 18: Outliers versus coefficient of variation –IQR method using NEP trimming method



Appendix 8: Results from analysis of hospital level changes in cost ratios

These results relate to the analysis discussed in the price loadings section 4.5.2.

Figure 19: Box plots of change in cost ratio by State or Territory of hospital

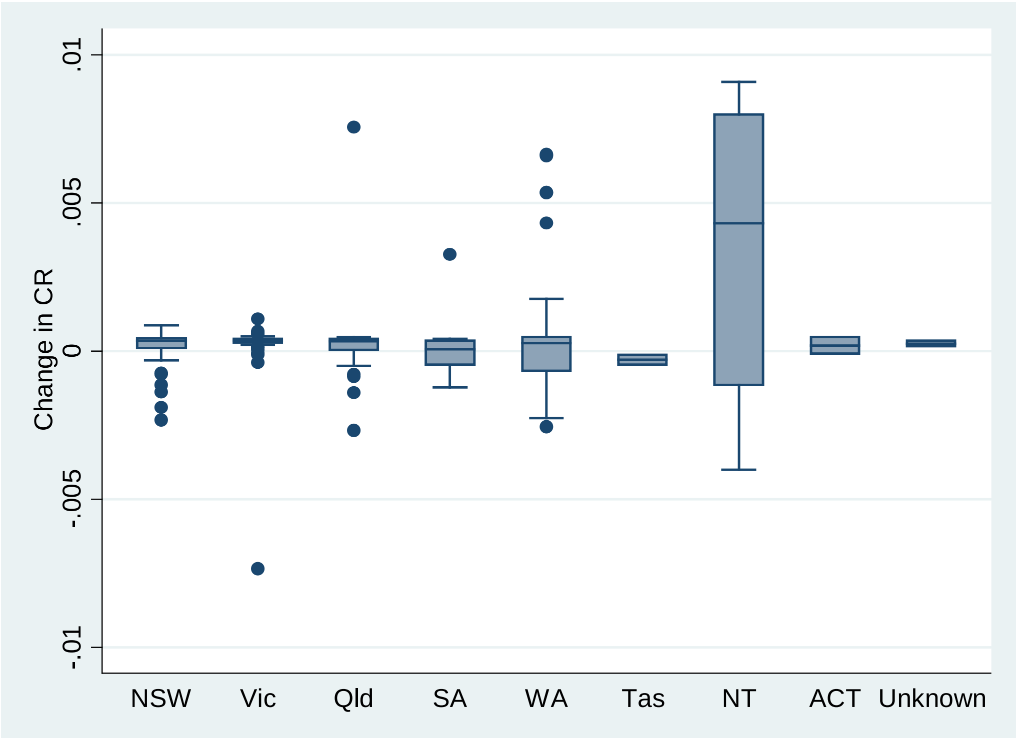


Figure 20: Box plots of changes in cost ratio by remoteness of hospital

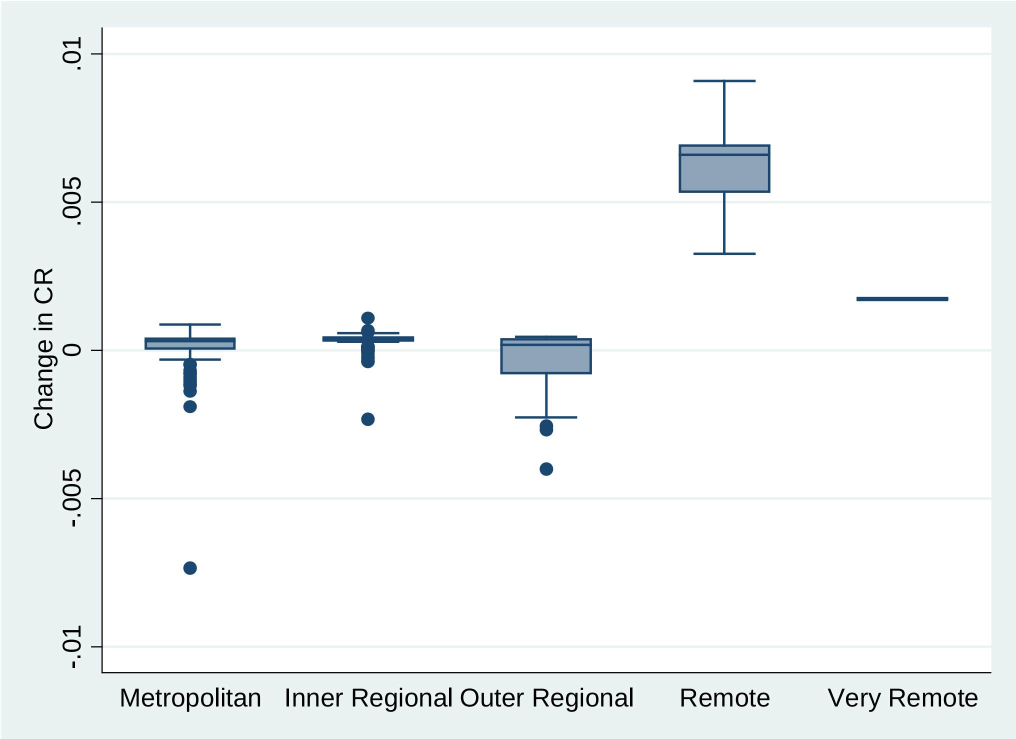


Figure 21: Box plots of changes in cost ratio by peer groups

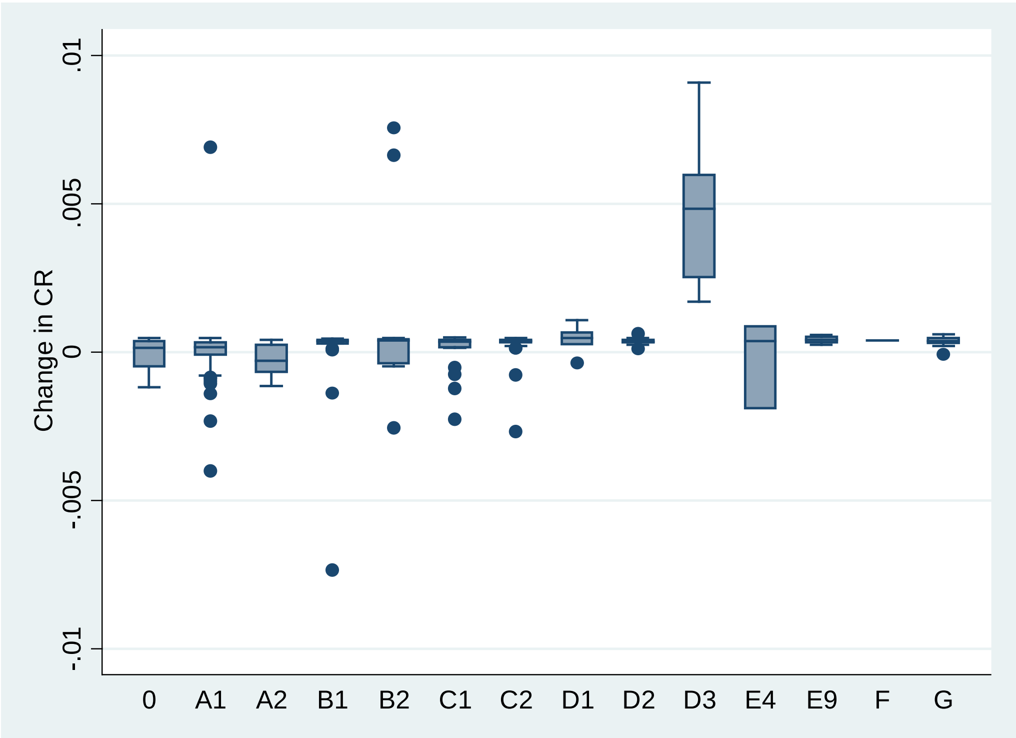


Figure 22: Box plots of change in cost ratios by Specialist Paediatric status of hospital

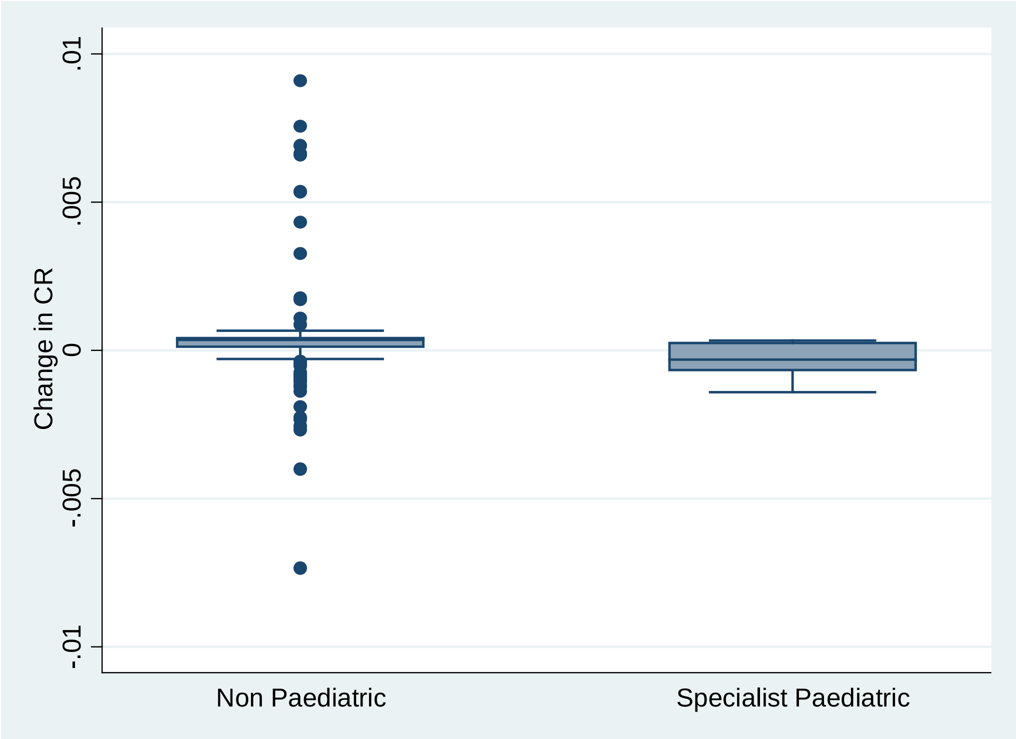


Figure 23: Box plots of change in cost ratio by ICU status of hospital