# Private healthcare remittal

## Insured price analysis

**11 June 2015**

Parties wishing to comment on this paper should send their comments to Private-Healthcare@cma.gsi.gov.uk by 17 July 2015.

The Competition and Markets Authority has excluded from this published version of the working paper information which the Inquiry Group considers should be excluded having regard to the three considerations set out in section 244 of the Enterprise Act 2002 (specified information: considerations relevant to disclosure). The omissions are indicated by [✄].

**Contents**

*Page*

## Summary

1. This working paper forms part of the CMA's work on the Private Healthcare Remittal and focuses on our analysis of the prices that HCA and TLC charge to private medical insurers (PMIs) – we refer to this analysis as the Insured Price Analysis (IPA). More specifically, this paper sets out:

   *(a)* our approach to calculating any difference in the prices that HCA and TLC charge to PMIs;

   *(b)* HCA's views on the methodology that we have employed;

   *(c)* our revisions to the IPA, as a result of changes we have made (either in response to HCA's points or as a result of our own review of the IPA) and our updated results; and

   *(d)* the results of a number of robustness checks, both sensitivity tests for the IPA methodology, as well as setting out an alternative methodology.

2. We have not, in this paper, addressed parties' comments relating to how the results of the IPA relate to the overall assessment of competition - we will be giving this issue, alongside other relevant evidence, further consideration as part of our competitive assessment which will be set out in our provisional findings.

3. This paper is technical in nature - we have endeavoured to explain technical terms wherever possible and convey the intuition behind the analysis we have conducted. We have also included this summary as a high-level guide to the issues we have covered in this paper.

### *High-level approach to our analysis*

4. Our approach to calculating any difference in the prices that HCA and TLC charge to PMIs is based on the same overall methodology that we employed in our final report that we published on 2 April 2014 (the Final Report). The main aim of the IPA is to compare the prices charged by different hospital operators to PMIs, as well as the prices paid to hospital operators across different PMIs. In order to achieve this we have constructed a price index based on a common basket of treatments offered by both hospital operators to each PMI. We discuss the main steps of our IPA methodology in paragraphs 27 to 40 of this working paper.

5. We have made some revisions to our analysis; both to take into account comments made by HCA and based on our own review of the IPA. For example, HCA's economic advisers put forward detailed points on the IPA

methodology, most notably in relation to: (i) coding errors which led to incorrect R-squared and statistical significance results being reported in the Final Report; (ii) 'irrational' price predictions that were included in our calculation of the price indices; and (iii) a number of issues relating to how we have cleaned and processed the data set. We consider these issues and discuss the main revisions we have made to our analysis in paragraphs 57 to 138 of this working paper (supported further by Appendices A to D).

### *Our revised overall results*

6. Table 1 compares our revised results, which are directly comparable with those presented by HCA's economic advisors (KPMG) in the Data Room Report (DRR) and to those in our Final Report. For the 5-episode threshold, our updated results also find that HCA charges higher prices than TLC, with an overall estimated price difference of [✂]%. This is comparable to both the DRR and Final Report results. Using a 30-episode threshold produces a higher price difference of [✂]%. In addition, throughout this working paper we place reliance on the results of both the 5-episode and 30-episode thresholds.

**Table 1: Overall price differences, KPMG, FR and updated approach**

*%*

| | *Updated approach 30 episodes* | *Updated approach 5 episodes* | *KPMG DRR 5 episodes* | *Final Report 5 episodes* |
|---|---|---|---|---|
| Simple average | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis, KPMG, Final Report.

7. We have tested the statistical significance of our revised results. The results for the overall price difference between HCA and TLC across all years and all insurers show that:

   *(a)* using a 5-patient-episode threshold, HCA is [✂]% more expensive than TLC – a difference which is statistically significant at the 99% confidence level; and

   *(b)* applying a minimum threshold of 30 patient-episodes, HCA is [✂]% more expensive than TLC – a difference which is also statistically significant at the 99% confidence level.

### *Robustness checks*

8. In order to check the robustness of our analysis we have conducted a number of sensitivity checks. The sensitivity analysis shows that the price difference between HCA and TLC varies between [✂]% and [✂]% depending on the underlying assumptions. We describe the results of our sensitivity analysis in detail in paragraphs 141 to 153 of our working paper.

9. We have also used an alternative methodology to estimate the price difference between HCA and TLC. We have estimated a single regression which models each episode price as a function of whether the patient was treated at HCA or TLC, controlling for patient characteristics and for any effects that are specific to the treatment, year and insurer. This regression approach estimates a similar price difference to the price-index approach used in the IPA, showing that HCA is [✂]% more expensive than TLC, and that this is statistically significant at the 99% confidence level. We discuss the results of our regression approach in paragraphs 154 to 170 and Appendix E of this paper.

## Introduction

### *Background to the remittal*

10. On 4 April 2012, the Office of Fair Trading (OFT) made a market investigation reference to the Competition Commission (CC) regarding the supply or acquisition of privately-funded healthcare services[1] in the UK. Following an extensive market investigation, the Competition and Markets Authority (CMA), the CC's successor, published its Final Report.[2] This Final Report set out our findings based on the evidence we received and the analysis we carried out during the course of the inquiry.[3]

11. In the Final Report, we identified two structural features of the market for privately-funded healthcare services by private hospital operators,[4] which were:

    *(a)* high barriers to entry and expansion for private hospitals; and

    *(b)* weak competitive constraints exerted on private hospitals in many local markets including central London.

12. We found that these two features in combination gave rise to adverse effects on competition (AEC) in the markets for the provision of hospital services which lead to higher prices charged for inpatient and some day-case and outpatient treatments to self-pay patients at private hospitals in many local

---

[1] These are services provided to patients via private facilities/clinics including private patient units (PPUs), through the services of consultants, medical and clinical professionals who work within such facilities.
[2] On 1 April 2014 the CMA took over many of the functions and responsibilities of the CC and the Office of Fair Trading, including in relation to the private healthcare market investigation. For ease of reference, the CC, OFT and the CMA are referred to together as the CMA.
[3] The findings are set out in Section 10 of the Final Report.
[4] When referring to private hospital operators we generally mean a person who operates a private healthcare facility that has inpatient facilities including NHS PPUs. Similarly, by private hospital we generally mean a facility providing inpatient services as well as day-case and outpatient services.

markets subject to weak competitive constraints across the UK, including central London (the 'self-pay AEC decision'). These features also in combination gave rise to AECs in the markets for hospital services which lead to higher prices across the range of treatments being charged by HCA[5] to PMIs for hospital services to insured patients in central London (the 'insured AEC decision').[6]

13. In making these findings, we considered evidence from a large number of interested parties (including hospital operators, insurers and patients) and undertook a wide ranging analysis which included an assessment of:

*(a)* barriers to entry and expansion;

*(b)* local competitive constraints; and

*(c)* market outcomes, including assessing both pricing and non-pricing outcomes (ie quality and range) and the profitability of the largest UK private hospital operators.

14. As part of our assessment of market outcomes in relation to prices, we conducted amongst other things an empirical analysis of the insured prices that PMIs paid to different hospital operators (the IPA). Based on the results of this analysis we found that HCA charged higher prices to PMIs than TLC (its closest competitor in central London).

15. To address the above AECs the CMA designed a package of remedies.[7] In order to introduce greater rivalry in central London, the CMA decided to require HCA to divest either the Wellington Hospital together with the Wellington Hospital Platinum Medical Centre, or the London Bridge Hospital and the Princess Grace Hospital (the 'divestment decision'). Our assessment of the proportionality of the divestment decision was informed in part by the results of the IPA.

16. After publication of our Final Report in 2014, HCA challenged the above AEC findings and divestment decision at the Competition Appeal Tribunal (CAT) on a number of different grounds.[8] AXA also appealed, amongst other things, against the divestment decision.[9]

17. In the course of HCA's appeal the CAT ordered the CMA to disclose to HCA via a data room (the Data Room), the data and methodology used in the IPA.

---

[5] HCA International Limited and any company in the group as appropriate.
[6] The area inside the North and South Circular Roads.
[7] See section 11 of the Final Report.
[8] Further information on the HCA appeal can be found on the CAT website.
[9] Further information on the AXA appeal can be found on the CAT website.

HCA's external economic advisers, KPMG, reviewed the IPA data and methodology disclosed in the Data Room and produced a report of its findings (the 'KPMG Data Room Report – DRR'). HCA also instructed an independent economics expert, Professor Waterson, who visited the Data Room and produced a further report (the Waterson Report). The detailed points raised in the DRR also fed into arguments that HCA presented to the CAT in its Re-amended Notice of Application (RNoA).

18. As a result of that review, HCA claimed that there were substantive and significant issues regarding the robustness of the work done by the CMA for the IPA. In particular, KPMG identified a number of coding errors in the IPA. In light of these two errors, the CMA considered that the appropriate course was for the matter to remitted and for the CMA to re-consult.

19. Consequently on 12 January 2015 the CAT ordered that the insured AEC decision and the divestment decision be quashed and remitted back to the CMA for reconsideration.

20. The remainder of HCA's challenge and the relevant grounds of AXA's challenge related to the divestment decision have been stayed pending our re-determination of the insured AEC decision and the divestment decision.

***Approach to the remittal***

21. The CAT provided guidance on the approach that the CMA should take to the remittal in its Ruling of 23 December 2014, explaining that:

> 'The task of the CMA will be to consult on the IPA and then re-determine the questions whether any new insured AEC decision should be made and whether any new divestment decision should be made. The CMA will have to consider what impact the new information and representations it receives in relation to the IPA has upon the existing statements of reasoning contained in the Final Report with respect to those decisions'.[10]

22. Following the CAT's ruling, on 25 February 2015 we published a notice of the launch of the remittal and we asked parties to make any written submissions on any relevant matters which they considered should be taken into account in the remittal.[11] On 15 April 2015 we published a second notice asking

[10] CAT Ruling of 23 December 2014, paragraph 56 b).
[11] See the Notice of launch of remittal and invitation to comment on the case page.

parties to make further detailed submission on specific issues (supported, as far as possible, by appropriate evidence).[12]

23. The remainder of this working paper sets out the approach we have taken to the IPA and the revisions we have made since the Final Report, responds to the comments we received from parties on the IPA and presents the results of our revised IPA. We have not addressed parties' comments relating to how the results of the IPA relate to the overall assessment of competition. We will be giving this issue, alongside other relevant evidence, further consideration as part of our competitive assessment which will be set out in our provisional findings.

### *Structure of this working paper*

24. The main body of this working paper sets out:

   *(a)* our approach to calculating any difference in the prices that HCA and TLC charge to PMIs (the IPA);

   *(b)* HCA's views on the methodology that we employed;

   *(c)* our revisions to the IPA, as a result of changes that we have made (either in response to HCA's points or as a result of our own review of the IPA) and our updated results; and

   *(d)* the results of a number of robustness checks, both sensitivity tests for the IPA methodology, as well as setting out an alternative methodology.

25. Appendices A to D set out in more detail some aspects of the IPA[13] methodology, in particular, in response to those areas where HCA has raised criticisms. Appendix E sets out the sensitivities and robustness checks that we have conducted in relation to the IPA results.

26. We also received some more general comments from other parties on the IPA, in particular on how we should interpret the results and what weight we should place on these in the context of our overall assessment of competition in central London. As these do not relate directly to the IPA methodology, they will be considered further in our provisional findings.

---

[12] See the Invitation to comment and submit further evidence on the case page.

[13] To avoid confusion, when we use 'IPA' in the remainder of this paper we refer to the price-index approach that the CMA used in the Final Report and which we have revised and present again in this paper. We will draw a distinction between this price-index-based analysis of insured prices ('IPA') and an approach that we set out later in this paper which analyses the determinants of insured prices using a regression (the 'regression approach').

## Original analysis

27. Our approach to calculating any difference in the prices that HCA and TLC charge to PMIs is based on the same overall methodology we employed in the Final Report. The IPA was presented in paragraphs 6.333 to 6.383 of the Final Report and Appendix 6.12 of the Final Report. Its findings were part of the evidence base which supported the central London insured AEC finding in the Final Report.

28. We also assessed the IPA results alongside our analysis of local substitutability and we used them to inform, in part, our assessment of the proportionality of the divestment decision (see paragraph 15, above), as the price differential was used as an estimate of the likely reduction in prices that the remedy would achieve.[14] As previously explained, this paper focuses on the IPA methodology and results. Therefore, our analysis of comparing insured prices with local substitutability, and discussions around the assessment of proportionality of any remedies (if required), will form part of our provisional findings.

29. Wer first describe the basis for, and the steps involved in, the IPA analysis. We outline the subsequent work we have undertaken in the remittal and our revised analysis from paragraph 57 onwards of this paper.

### *Constructing price indices on a like-for-like basis*

30. The aim of the IPA in relation to central London is to compare the prices charged by HCA and TLC to individual PMIs, as well as to compare the prices paid to these hospital operators across all of the PMIs across all years. This is a complex task due to the differences between hospital operators in the treatments that they offer and the mix of patients that they treat (factors for which we sought to control).

31. Our methodology aimed to construct a measure of insured prices that would be comparable between hospital operators. To do this, we constructed a ‘price index’ based on a common basket of treatments offered by both hospital operators to each PMI.

32. The index summarised prices in a single aggregated number, to reflect the process of bargaining between PMIs and hospital operators, which does not take place at the level of the individual treatment but at an overall level.

[14] Final Report, paragraphs 11.219 to 11.222.

33. We based our calculations on underlying invoice data which captured what we called the 'episode price'. An episode is defined as a single patient visit to a given hospital for a given treatment and the corresponding episode price is defined as the total amount paid for hospital services excluding consultant fees. The prices are based on data obtained from Healthcode, which is an intermediary between hospital operators and PMIs, which we further prepared for the purposes of our analysis (a process referred to as "data cleaning"). The relevant data related to inpatient and day-case episodes, which accounted for 75% of revenue in 2011. Because data relating to outpatients was not classified in a way which allowed it to be compared across operators, outpatient episodes are not included in the analysis.

34. In the Final Report, we set out different versions of the price index, two of which are relevant for the remittal. These are:[15]

*(a)* an 'insurer-specific price index' (eg, TLC's average insured price charged to Bupa in a given year). This allows for comparisons between the prices charged by different hospital operators for a given PMI; and

*(b)* an 'average price index' for each hospital operator on average across PMIs. This allows for comparisons between the prices charged by different hospital operators across PMIs.

35. We split the analysis of insured prices between central London and the rest of the UK. This allowed us to control for potential differences between hospital operators located in these different geographic areas, such as differences in labour and other input costs.[16]

36. In central London, we focused on comparing HCA and TLC. This was because they are the largest two operators in terms of size and we considered them to be the two closest competitors to each other based on their shares of admissions and capacity, overlap in terms of the range of services provided, and the views of relevant parties.[17] Because HCA and TLC are almost exclusively based in central London, we noted that as far as they were concerned, insured prices and local prices were essentially the same thing.

---

[15] The self-pay price index is not dealt with in this working paper which is focused on insured prices.
[16] See paragraph 14 (c), Appendix 6.12, Final Report.
[17] See paragraphs 6.204 to 6.218, Final Report and Appendix 6.10, Final Report.

*Insurer-specific and average price indices – step-by-step approach*

37. To calculate the insurer-specific price index for a hospital operator (eg HCA) for a given PMI in a given year (eg Bupa in 2011) we took the following steps:

*(a)* We identify the 'common basket' of treatments for the hospital operators included in the comparison (eg HCA and TLC for central London). The common basket includes treatments provided by each operator included in the price comparison for the given PMI in a given year. This step of the methodology controls for treatment mix.

*(b)* For each treatment in the common basket, we regress episode prices on patient characteristics (age, gender and length of stay) and a constant term using all episodes associated with the hospital operators for the given PMI in a given year.

*(c)* For each treatment in the common basket, we use the regression estimates from step *(b)* to estimate the price charged by the hospital operator for the given PMI in a given year for treating a 'representative patient'. The representative patient is defined separately for each treatment as a patient with median characteristics (age, gender and length of stay) across all hospital operators included in the price comparison. In combination with step *(b)*, this step of the methodology controls for patient mix.

*(d)* We then calculate the insurer-specific price index as a weighted average of the estimated prices for each treatment obtained in step *(c)*. Each treatment receives a weight equal to the number of admissions for that treatment across all operators included in the price comparison (eg HCA and TLC in central London).

38. Repeating the above steps for each hospital operator in the price comparison produces insurer-specific price index results for a PMI and year pair (eg HCA and TLC for Bupa 2011). We then repeat this process for all PMIs and all years to produce the full set of results.

39. To calculate the average price index, we use the weighted average of the insurer-specific price index results described above. We weight each insurer specific price index by the size of the common basket of treatments according to the number of admissions.

*Sensitivity analysis and statistical significance testing*

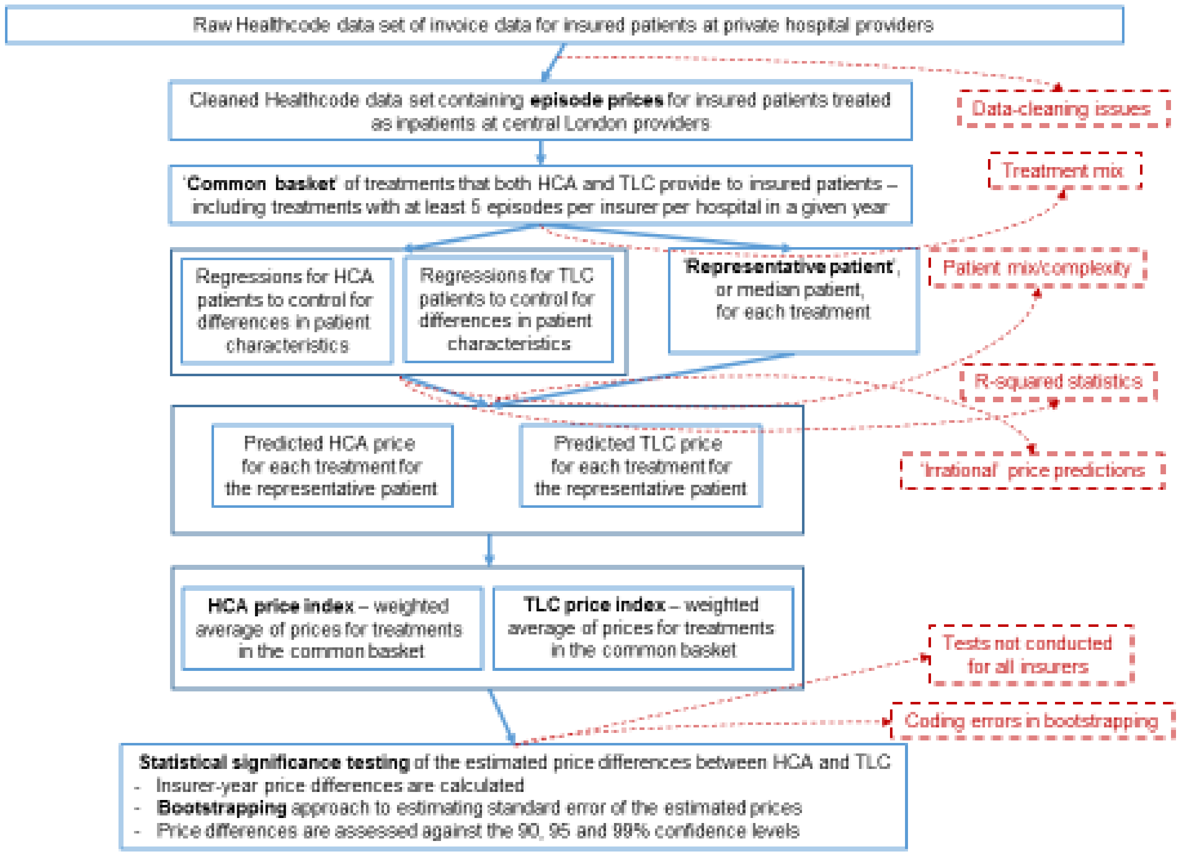
40. To assess the robustness of our results we carried out a number of sensitivity tests which tested the impact of modifications to both the data used and the

methodology employed. We also assessed whether its results were statistically significant.[18] In the Final Report, we reported that the price differences were, 'in the large majority of cases, statistically significant',[19] and that in general the price index estimates were 'very precise' (in statistical terms),[20] and that the larger price differences estimated were robust to the sensitivity analysis that we undertook.

## HCA's views on the IPA

41. Figure 1 below sets out a simplified version of the overall IPA analysis for central London, highlighting the seven main areas where HCA and its economic advisers have raised issues.

**Figure 1: Simplified version of the IPA for central London**




Source: CMA analysis.
Note: Solid boxes set out the various steps in the analysis, while the boxes with dashed borders indicate issues that HCA's advisers have raised.

[18] That is, whether the estimated price differences were likely to reflects genuine price difference or whether they may be the result of random variation or statistical 'noise' in the data.
[19] Paragraph 6.341, FR.
[20] Paragraph 28, Appendix 6.12, Final Report.

42. We treat HCA's various points in two ways:

*(a)* The points that relate directly to the methodology and results of the IPA are dealt with in this working paper and, in more detail, in Appendices A to D.

*(b)* Other points that HCA has made that do not relate directly to the methodology of the IPA, for example, issues that relate to how we interpret the results of the IPA and how these relate to any overall assessment of competition, are not covered in this paper. These, together with views on similar issues from other parties, will be considered in our provisional findings, rather than in this narrower, more technical, working paper which focuses on the IPA itself.[21]

43. There are six main areas where HCA have raised issues, as set out in the following paragraphs.

44. First, HCA raised a number of data-related issues, which fall into three categories:

*(a)* HCA put forward the view that, contrary to an assertion made by the CMA at footnote 8, Appendix 6.12, our analysis of episode prices could have included additional variables contained within the raw pricing data, for example, the diagnosis code assigned to the patient and the medical specialty through which the patient was admitted and that this 'could have improved the robustness of the CMA's analysis…'.[22]

*(b)* The IPA analysis in the Final Report only included those episodes where the patient received a single treatment (or single-CCSD[23] episodes). HCA put forward the view that the IPA results are sensitive to the inclusion of those episodes where the patient received more than one treatment (multiple-CCSD episodes). KPMG, in its DRR, carried out the IPA analysis including those episodes with multiple CCSDs and reported different results to the Final Report, concluding that 'these results highlight the fragility of the CMA's results with respect to the composition of the

---

[21] HCA has put forward the view that it provides higher quality care and treats more complex patients than TLC, which are issues that could potentially affect HCA's costs. Failing to appropriately take account of these differences could potentially mean that the price difference we calculate is an overestimate. The approach we have taken in this paper controls for differences in patient characteristics between HCA and TLC given the constraints of the data set. As HCA has argued, it is, however, possible that there are other, unobservable, factors that may affect hospital providers' costs and, hence, the price difference between HCA and TLC. We will deal with the issue of whether there are any significant differences between HCA and TLC, for example any quality differences, when we assess this and other issues that are relevant to our overall assessment of competition in our provisional findings.
[22] Re-amended Notice of Application, paragraph 100.
[23] CCSD, or Clinical Coding and Scheduling Development, is a system of classifying treatments and diagnostic procedures. See the CCSD website.

common basket, casting serious doubts on the robustness of the CMA's analysis'.[24]

(c) Various other data-related issues, such as whether the CMA had stripped out all of the instances of consultant fees that were included in the data set, as it had intended to do, were also raised, and are dealt with in Appendix A.

45. Second, HCA pointed out that a coding error in how the CMA had calculated the R-squared statistics resulted in the Final Report overstating these figures. In other words, the Final Report overstated how much of the variation in prices that insurers paid to hospitals operators were explained by the regression analysis that we had conducted. The implication is that, according to HCA, the variables included in the regression analysis explain a lower share of the variation in insured prices than we had reported. HCA's views are that "the CMA's R-squared calculation is incorrect",[25] that "[i]n consequence of this error, the R-squared values reported in the Final Report are incorrect and overstated"[26] and that "[t]he CMA's conclusion that the factors it has controlled for account for the large majority of observed price variation is therefore incorrect".[27]

46. Third, KPMG's DRR pointed out a number of treatments where the predicted prices in the IPA methodology were 'irrational'. These 'irrational' price predictions fell into three categories:

(a) Zero price predictions: a coding error led to the predicted prices for four treatments being zero.

(b) Negative price predictions: for four treatments the methodology used in the IPA resulted in a negative price being predicted.

(c) Out-of-sample price predictions: for two treatments the IPA used a representative patient to predict prices where one of the hospital operators (HCA or TLC) did not treat patients with these characteristics.[28]

47. Assessing the impact of these, KPMG states that: 'fixing these errors [irrational price predications] leads to a decrease in the estimate of pricing differences between TLC and HCA in all years except [✂] and [✂]. The

[24] DRR, Annex 5, paragraph 5.
[25] Re-amended Notice of Application, paragraph 95.
[26] Re-amended Notice of Application, paragraph 96.
[27] Re-amended Notice of Application, paragraph 97.
[28] On one occasion the representative patient is female, while TLC did not treat female patients. On the second occasion, the average length of stay of the representative patient is positive, while TLC did not treat any inpatients.

average difference across the five years also decreases'.[29] The DRR goes on to report that across all years the price difference between TLC and HCA decreases by [✂] percentage points if the treatments with irrational price predictions are dropped. The price difference falls from [✂]% to [✂]% overall.[30]

48. Fourth, in relation to statistical significance testing, HCA made two main points:[31]

*(a)* HCA pointed out that in testing the statistical significance of price differences between HCA and TLC for each insurer, the CMA had not tested all of the price differences and so was inaccurate in its presentation of these results in the Final Report.

*(b)* The DRR identified errors in the code that the CMA had written for use in the statistical program, Stata, to test the statistical significance of the price differences between HCA and TLC, resulting in incorrect results being reported in the Final Report.

49. The DRR also presented KPMG's 'corrected' results for the statistical significance tests. These results indicated that price differences for [✂] out of [✂] insurer-year pairs were not statistically significant. The report states that these represent [✂]% of HCA's revenues and [✂]% of its admissions.[32]

50. An additional point raised in the DRR was an analysis conducted in the data room that indicated that the inclusion of the King Edward VII Hospital (KEVII), another central London private hospital, changed the relative price differences between HCA and TLC and gave results that showed that TLC was more expensive than HCA in three of the five years in our data set.[33] Furthermore, averaging over all years, the DRR showed that TLC is more expensive than HCA, while KEVII is the least expensive hospital.

51. HCA also made a number of additional points in relation to the IPA methodology in its most recent submissions as part of the Remittal, although some of these are related to points that were made in the DRR.

52. HCA's economic advisers, KPMG, stated that the price comparisons in the IPA were 'not conducted on a like-for-like basis. Differences in treatment mix were not properly taken into account. Furthermore, systematic differences in

---

[29] DRR, paragraph 79.
[30] The price difference reported without taking into account the irrational price predictions is [✂]%, while the reported price difference excluding those treatments where irrational price predictions arise is [✂]%.
[31] Paragraph 97, DRR.
[32] DRR, paragraph 114 and Table 10.
[33] DRR, Table 12.

patient complexity were not appropriately controlled for. Both are likely to influence the reliability of price indices. This means that observed differences in price are – at least partly – explained by higher costs of providing higher quality procedures, a greater scope of services overall, and lastly of treating patients with more complex medical needs. Observed differences in prices would then be due to legitimate differences in the nature and/or quality of services provided, and not necessarily to discretionary price setting behaviour'.[34]

53. It pointed to a number of examples of academic research and empirical work by competition authorities in the US where differences in treatment mix and patient complexity – or 'case mix' – need to be taken into account when measuring prices. It also stated that the academic literature 'shows that age gender and length of stay are insufficient to account for case mix and complexity'.[35] These are the patient characteristics that we have used in constructing like-for-like comparisons between the prices charged for episodes for each treatment by HCA and TLC.

54. KPMG went on to outline why the approach to controlling for treatment mix in the IPA was inadequate:

*(a)* It stated that the common basket approach was not representative of either HCA's or TLC's businesses. It stated that:

> 'analysis conducted in the Data Room showed that the common basket, from a revenue perspective, is not representative of HCA's or TLC's businesses. For [✂]% of HCA's PMI-year pairs, the proportion of in-patient and day-case patient revenue associated with the common basket was less than [✂]%. Similarly, for [✂]% of TLC's PMI-year pairs, the proportion of in-patient and day-case patient revenue associated with the common basket was less than [✂]%'.[36]

*(b)* According to KPMG, the selection of treatments in the common basket did not reflect the treatments that HCA performs, with HCA performing a 'far larger proportion of high complexity treatments than TLC'.[37]

*(c)* It also stated that the 'extrapolation of estimated price differences based on a common basket approach was flawed',[38] as it extrapolates the price differences found in the common basket to those treatments outside it.

---

[34] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 8.
[35] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 17.
[36] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 25.
[37] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 28.
[38] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, section 2.1.2.

Table 4 of its submission compares the price differences in the IPA for treatments of different levels of complexity and shows that for more complex treatments, where HCA tends to concentrate its activity, HCA is less likely to be more expensive than TLC, whereas for less complex treatments, HCA is more likely to be charge higher prices. It concluded that:

> '…to the extent that the CMA considers it possible to extrapolate from the common basket to treatments outside it, it should take into account the possibility that treatments outside the common basket, being mostly high complexity treatments, might also be not significantly different from what the CMA considers a competitive price benchmark'.[39]

55. KPMG then set out a number of ways in which the IPA did not appropriately control for differences in patient characteristics:

*(a)* It pointed out that the academic literature and previous empirical work by competition authorities recognises the importance of controlling for differences in clinical need across patients, for example, comorbidities that that patient may suffer from, the number of conditions, race of the patient, the final outcome, etc. It concluded that: 'without appropriate control variables in place to risk-adjust patients across competitors, the validity of the estimated price differences in the IPA cannot be assured'.[40]

*(b)* KPMG then repeated its earlier point about the level of the R-squared statistics and stated that:

> 'given the large unexplained price variation and the lack of adequate control variables that adjust for patient complexity within a treatment, without considerable changes the IPA is not conducting a meaningful like-for-like comparison of treatment charges and thus of overall prices. Unless the CMA's new model compares prices on a like-for-like basis, the comparison will be meaningless and of no probative value'.[41]

*(c)* KPMG also questioned whether the IPA takes account of all relevant information and is free of empirical errors.[42] In particular, it pointed out that:

[39] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 34.
[40] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 42.
[41] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 46.
[42] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, see paragraphs 47 to 50.

(i) some PMIs receive rebates from HCA, so the invoiced amounts may not be reflective of HCA's revenues;

(ii) some PMIs 'shortfall' their patients, that is, they pay only part of the invoiced amount and the hospital operator may not receive the full amount invoiced; and

(iii) there were some negative and zero price predictions, as well as out-of-sample price predictions, in the IPA methodology, as set out above.

56. KPMG's submission on the IPA concluded that it 'suffered from significant flaws which render it unfit to assess either the existence of a price differential between HCA or TLC or the relationship between any such difference and local market concentration'[43] and that the DRR had already demonstrated that the results of the IPA 'did not show that there was a significant price difference between HCA and TLC'.[44]

## The updated IPA

57. In the below paragraphs we discuss the following:

*(a)* First, we deal with a methodological issue which, although not raised by the parties, appeared to us to require further consideration. This is the issue of whether we should continue to base our results solely on the analysis of treatments with a minimum of 5 episodes for any one insurer in a given year at a given hospital operator.

*(b)* Second we deal with issues that might directly impact the calculation of the price difference between HCA and TLC. In particular, we focus on the following steps:

(i) We deal with a number of data-related issues. We assess whether additional variables (diagnosis code and the medical specialty of the treating consultant) should be included in the regressions in the IPA in response to HCA's view that these may provide additional information on the complexity of the patient's condition. We also assess whether those episodes where multiple treatments ('multiple-CCSD episodes') were delivered should be included in the IPA.

(ii) We respond to three issues that HCA has raised in relation to 'irrational' price predictions, where a small number of regression

[43] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 89.
[44] KPMG, 'A Submission on the Analysis of Insured Prices', dated 1 May 2015, paragraph 53.

results yielded predicted prices that were zero (implying that these treatments were provided for free), negative (implying that insurers would be paid by HCA or TLC rather than the other way around) and out of sample (as explained below).

(iii) We present our revised results for the price difference between HCA and TLC following our revisions relating to data issues and irrational price predictions.

*(c)* Third, we outline the coding error identified in the DRR that led to erroneous R-squared statistics for the regressions in the IPA and we report our corrected R-squared statistics.

*(d)* Fourth, we outline the coding error identified in the DRR that led to erroneous results being reported for the statistical significance of the price differences between HCA and TLC for each insurer in each year. We correct this error, make some additional changes to our statistical significance tests and report the corrected results.

*(e)* Finally, we set out the results of a number of robustness checks, which we considered were particularly relevant checks on our results. We begin by checking the sensitivities in our IPA approach. We also calculate the price difference between HCA and TLC using a single regression[45] to control for other factors that are likely to determine episode prices.

## Minimum episode threshold for treatments included in IPA

58. We start by drawing attention to a methodological issue to which we have given further consideration and which is relevant to several aspects of our analysis.

59. We have used the raw data from Healthcode to form a data set which consists of data at the patient episode level across different treatments and insurers for each year between 2007 and 2011. In the Final Report, we reported results – both for the estimated price differences between HCA and TLC and for the statistical significance testing of these price differences - based on a minimum of 5 episodes per treatment per insurer per year per hospital operator. As set out in the Final Report, we checked the sensitivity of these results using a 30-episode threshold.

[45] Although the IPA methodology also involves the use of regressions to control for differences in patient characteristics between HCA and TLC, we refer here to our 'regression approach' as it involves estimating a single regression, rather than constructing multiple price indices and then testing the statistical significance of any differences in their levels, as we do in the IPA.

60. Looking at the price difference between HCA and TLC, the results of the 30-episode approach reported in the Final Report showed lower average prices for both providers compared to using a 5-episode threshold, while the price differences were broadly similar.[46]

61. The Final Report set out briefly the relative merits of using a 5-episode threshold and of using a 30-episode threshold:

*(a)* In relation to the 5-episode threshold, the Final Report stated that:

> '…because negotiations between a PMI and a hospital operator focus on all of a PMI's expenditure, we thought it was more appropriate to compare prices over as wide a range of treatments as possible.[47] For the same reasons, we did not separately examine inpatient and day-case treatments. Note that as part of our sensitivity analysis, one analysis considered only those treatments with more than 30 patients per operator for a given PMI in a given year, and these results are therefore relevant to the more common treatments'.[48]

*(b)* On the 30-episode sensitivity, the Final Report stated that:

> '…a higher threshold of 30 patient episodes … allows for a higher number of observations per regression and as a result may mitigate the impact of any outlying or extreme price observations and produce more precise price predictions…'.[49]

62. Bearing these points in mind, we have further considered whether we should continue to treat the results based on the 5-episode threshold as our main analysis, while using the 30-episode threshold as a sensitivity check, when coming to a view on any price difference between HCA and TLC. We came to the view that we should place reliance on both sets of results in our analysis of the price difference (and not rely on the 30-episode threshold as a sensitivity as we did in the Final Report). We set out below two main reasons for this decision.

63. First, the 5-episode threshold includes treatments with very low patient volumes, which has the advantage of increasing the number of different

[46] The results of the 30-episode sensitivity are presented in Figure 2 of Annex B to Appendix 6.12 of the Final Report, and are compared to the 5-episode results presented in Figure 1 of Appendix 6.12.
[47] Looking at those treatments where there are at least 5-episodes per treatment per year per hospital operator provides us with a larger data set than when we restrict this to only those treatments that meet a 30-episode threshold.
[48] Final Report, Appendix 6.12, footnote 16.
[49] Final Report, Appendix 6.12, paragraph 25 (c).

treatments included in the common basket. However, this approach has the disadvantage of not allowing us to be as confident as we could that the treatment-level regressions in the IPA precisely identify the relationship between the patients' characteristics and the episode prices.[50] Increasing the minimum number of episodes per treatment increases our confidence that we are getting more precise estimates of the relationship between patient characteristics and prices.

64. Second, increasing the minimum number of episodes per treatment to 30 increases the reliability of our statistical significance testing of any estimated price differences. The more observations are available for a given treatment, the more information is available about the underlying true distribution of the episode prices for that treatment. This means that we are able to estimate the standard errors of our estimated price differences through the bootstrapping procedure (as set out in Appendix D) with a higher degree of accuracy.[51] Thus the precision and reliability of our statistical significance testing improves when we apply a higher minimum threshold for the number of episodes per treatment.

65. We also recognise that while increasing the threshold to a minimum of 30 episodes increases the precision and reliability of our methodology, it reduces the size of the common baskets considered and the number of insurers we are able to consider. In particular, when increasing the threshold to a minimum of 30 episodes, we are able to conduct the statistical significance test for only 23 out of 36 insurer-year observations. We cannot do so in relation to the price differences for the remaining 13 insurer-year pairs because of insufficient number of observations due to low patient volumes for these smaller insurers.

66. In terms of coverage, we set out below four measures (all in terms of nominal revenue):

*(a)* The proportion of the hospital operators' insured revenue that is covered in the Healthcode raw data;

*(b)* The proportion of the Healthcode data that is included in the final cleaned data set that we use in our analysis;

---

[50] Having larger sample sizes – in our case, analysing treatments with higher numbers of patients being treated – leads to better estimates. In technical terms, larger sample sizes improve the consistency of our estimates meaning that the larger the sample, the less risk that the estimates that are produced will be biased.
[51] The principle of the bootstrap assumes that the observed distribution of the data in our sample is the best approximation for the true underlying distribution in the population. This may be a questionable assumption for treatments in which we observe very small numbers of patient episodes. For the bootstrap, it is, therefore, preferable to use those treatments with a higher number of observations, and thus have a higher threshold, in order to obtain more robust results.

(c) The proportion of this final cleaned data set that we use in our IPA – both the 5- and 30-episode analyses; and

(d) The proportion of the spend of the two major insurers in the final data set that is included in the IPA – again both 5- and 30-episode analyses.

67. These figures are set out in Tables 2 and 3, below.

68. Looking at 2011, we note that the Healthcode data set accounts for approximately [✂]% of HCA's insured revenue, while for TLC the equivalent figure is [✂]%. The final cleaned version of the data set that we use in our IPA includes invoices accounting for [✂]% and [✂]%, respectively, of the hospital operators' insured revenue covered by the Healthcode data.

69. Our IPA analysis is based on a smaller subset of this data set for two reasons. First, we are comparing the price of HCA and TLC and so can only conduct our analysis on treatments that both HCA and TLC provide – the 'common basket'. Second, our IPA analysis only covers those treatments where at least 5 episodes are observed per treatment per insurer per year per hospital operator, which reduces the coverage of the sample further.[52]

70. The IPA based on the 5-episode threshold covers episodes accounting for [✂]% of HCA's revenue in the final data set, while for TLC it accounts for [✂]%. Looking at the IPA conducted using the 30-episode threshold, the data set is further reduced, as treatments with lower patient volumes are no longer included. Overall this reduces the activity covered by the common basket, that is, those treatments that both HCA and TLC provide to the same insurer in the same year. The data set used in the 30-episode analysis accounts for [✂]% of HCA's revenue in the final cleaned data set and [✂]% of TLC's.

**Table 2: Share of insured revenue included in the IPA**

| | | % |
|---|---|---|
| | *HCA* | *TLC* |
| Healthcode data as a share of insured revenue, 2011* | [✂] | [✂] |
| Final data set that we use as a share of Healthcode data, 2011 | [✂] | [✂] |
| Data used in the IPA (5-episode) as a share of final data set, all years | [✂] | [✂] |
| Data used in the IPA (30-episode) as a share of final data set, all years | [✂] | [✂] |

Source: CMA analysis.
*HCA's and TLC's insured revenues for 2011 are presented in Tables 3.3 and 3.6, respectively, Final Report.

52 Our regression approach (see paragraph 154 onwards and Appendix E) covers a larger proportion of the Healthcode data, as it includes all treatments with at least two episodes.

71. Table 3 below sets out the share of insurer spend at HCA and TLC in our final cleaned data set that is used in our IPA analysis. Looking at the two main insurers, our IPA (based on 5episodes) uses data representing, for example, [✂]% of Bupa's spend with HCA that is contained in the final cleaned data set, while for AXA PPP the equivalent share is [✂]%. Applying the 30-episode threshold, the coverage of the IPA falls, as lower volume treatments are no longer included in the analysis.

**Table 3: Share of insured revenue included in the IPA, Bupa and AXA PPP**

| | | | | *%* |
|---|---|---|---|---|
| | *HCA* | | *TLC* | |
| | *Bupa* | *AXA PPP* | *Bupa* | *AXA PPP* |
| 5-episode IPA as a share of final data set, all years | [✂] | [✂] | [✂] | [✂] |
| 30-episode IPA as a share of final data set, all years | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.

72. Based on these figures, the IPA covers less than [✂]% of the revenue accounted for by the Healthcode data for both TLC and HCA. We do not consider that this invalidates our analysis, for a number of reasons.

73. First, looking again at the revenue coverage of our IPA, we consider that the most relevant measure of its coverage is to focus on the 'overlap' treatments which both HCA and TLC provide to insured patients. For HCA, the IPA (5-episode version) accounts for [✂]% of the revenue generated by overlapping treatments in the final cleaned data set, while for TLC the equivalent figure is [✂]%. As such, our analysis does cover a substantial proportion of those treatments for which a price comparison between HCA and TLC is meaningful.

74. Second, and more importantly, in order to make a meaningful comparison between HCA and TLC prices we only compare those treatments that are provided by both operators. Given that the range of services that HCA and TLC provide is not identical, there are treatments which HCA provides that TLC does not and vice versa. Therefore, there are many treatments that HCA and TLC provide, and which generate insured revenue for them, which are not relevant to our analysis.

75. We are comparing prices in those treatments where HCA and TLC overlap and, hence, actually or potentially compete for insured patient business. Comparing price differences for those treatments where HCA and TLC do not overlap would be impossible. Furthermore, to the extent that we consider TLC to be HCA's closest competitor, we would expect a comparison of prices in

those treatments where they overlap to be representative of HCA's pricing more generally and to be a reasonable proxy for HCA's relative market power.

### *Conclusion on minimum episode thresholds*

76. As presented in Table 2 above, there is a reduction in the coverage of the common baskets used in the IPA when we use the 30-episode minimum threshold. However, this higher threshold has clear advantages in terms of statistical robustness, as set out above. Therefore, we have come to the view that we should place reliance on both sets of results in our analysis of the price difference (and not rely on the 30 episode threshold as a sensitivity as we did in the Final Report).

## Issues that might directly impact the calculation of the price difference

### *Data-related issues*

77. The data set we have used for the IPA is based on invoice data received from Healthcode, an intermediary between hospital operators and PMIs.[53] The Healthcode data provides information on the hospital visits of insured patients. It includes details of the hospital visited, the treating consultant, the treatment received, and the actual prices paid by PMIs.[54] We have cleaned and consolidated the data to produce a cleaned data set for our analysis that covers the period 2007 to 2011. Each row in this data set is an 'episode', which we have defined as a single visit to a hospital by a patient.

78. The specific areas that are dealt with in this section are:

   *(a)* Whether the regression analysis should include variables that relate to diagnosis codes and the medical specialty of the treating consultant among the explanatory variables in the price regressions; and

   *(b)* Whether the regression analysis should include episodes where more than one treatment (that is, multiple CCSD codes) has been recorded for a single episode.

---

[53] See paragraphs 9 to 13 and Annex A of Appendix 6.12 of the Final Report.
[54] We considered this data, on actual prices paid, to be a better basis for our analysis than the (paper or electronic) contractual agreements between hospital operators and PMIs. The latter were not easily available in a format that was comparable between hospital operators or PMIs, and are typically based on a detailed contract which may span several documents. We noted that PMIs also use the actual prices paid, rather than their contractual agreements, to compare the prices of hospital operators.

79. We deal with other data-cleaning issues in Appendix A, where we describe any changes that we have made to our data set as a result. The overall impact of these changes on our results is minor.

*Diagnosis code and the medical speciality of the treating consultant*

80. The Healthcode data set includes variables on the diagnosis code for each patient episode, as well as the medical specialty of the treating consultant. In theory, including these variables in the regressions explaining treatment-prices could add explanatory power (in terms of how well our model explains the variation in prices), because, as HCA has argued, these variables could play a role in explaining the costs that providers face in treating these patients. HCA has argued that:

> ‘large price variations within treatments are driven by patient medical need as related to complications and comorbidities. Some patients, for example, may require more or more costly diagnostic procedures, drugs or nursing care at different levels of intensity, and each of these factors would result in a higher episode charge’.[55]

81. We have given some consideration to using the diagnosis code variable and we have reconsidered its use in response to the points made by HCA’s economic advisers, KPMG, in the DRR.

82. The data provider, Healthcode, stated that: ‘the quality of diagnosis coding in the sector is very poor’.[56] In the course of further discussions with Healthcode on the possible use of the diagnosis code variable in the regression approach, Healthcode stated that the diagnosis codes ‘cannot be used as an accurate barometer of patient’s condition’.[57] Healthcode also stated that ‘that data [on diagnosis code] is unreliable in this data set’. We therefore consider the diagnosis code variable not to be a reliable source of information in the econometric analysis conducted as part of the IPA.[58]

83. In relation to the consultant’s medical specialty, the fact that our regressions in the IPA are already estimated for each treatment separately means that including this variable in our analysis separately would add little explanatory power. For example, when explaining the price of hip replacements adding a variable that indicates when an orthopaedic surgeon has been used is unlikely

[55] KPMG, ‘A Submission on the Analysis of Insured Prices’, dated 1 May 2015, see paragraph 37.
[56] Document provided by Healthcode by email on 11 January 2012.
[57] Email from Healthcode, 5 February 2015.
[58] From an econometric perspective, measurement error leads to a bias in the estimated coefficient and would therefore lead to a bias in the average price indices.

to add to the accuracy of our analysis if all such procedures are delivered by an orthopaedic surgeon. In addition, adding a variable for the consultant's specialty would complicate the construction of the representative patient; it is not clear what the median medical speciality would be if more than one is relevant for a given treatment. Alternatively, we would have to calculate separate predicted prices for the same treatment when delivered by consultants with different specialties: this would reduce the size of our sample for each treatment further and decrease the accuracy of our results.

84. Healthcode's view is that the treating consultant's medical specialty 'does not provide information on the patient's medical condition'.[59] This means that, for each treatment, there is unlikely to be any meaningful variation in the medical specialty variable that would add useful information in the regressions.

85. For the reasons stated above, we have not used the information on diagnosis code or the consultant's speciality in our regression analysis as part of the IPA.

*Multiple-CCSD episodes*

86. In cleaning the data for the insured price analysis we defined a treatment by its CCSD code. For example, the CCSD code for a common cataract procedure is C7122.[60] However, the data set also includes episodes that are associated with multiple CCSD codes, where more than one treatment has been recorded for the same patient within the same episode. Our initial analysis did not include episodes with multiple CCSD codes.

87. On its webpage, Clinical Coding and Scheduling Development (the organisation that developed and maintains the CCSD system of classification) clarifies that when recording CCSD codes for a clinical procedure 'users should use a single CCSD code to describe the majority of common clinical interventions. This single code will usually fully describe the procedure from start to finish'.[61]

88. We queried with Healthcode whether multiple CCSD codes are comparable across hospital providers. Healthcode stated that multiple CCSD codes are sometimes used by hospital providers, in particular if a single CCSD code does not cover the whole procedure.[62] This is, to a limited degree, accepted

[59] Email from Healthcode, 5 February 2015.
[60] C7122 relates to 'Phakoemulsification of cataract, with lens implant - unilateral (including topical or local anaesthetic)'.
[61] See CCSD webpage Single Codes.
[62] Email from Healthcode, 5 February 2015

practice by PMIs.[63] However, the relationship between the number of CCSDs recorded for an episode and the price charged may not be straightforward, as the extent to which hospital operators are reimbursed fully for each individual CCSD recorded can vary depending on the specific contracts in place between the hospital operator and the insurer. For example, an insurer may pay in full for the main (most expensive) CCSD, but only partially cover the costs of the additional CCSDs.

89. Given the above discussions, especially the risk that episodes with multiple CCSD codes might not be comparable between hospital providers, we have excluded episodes with multiple CCSD codes from our analysis.

90. Nevertheless, we have checked the sensitivity of our results to the inclusion of those episodes with multiple CCSD codes (see Table 4).

91. Column A in Table 4 reports the price differences for single-CCSD episodes, while column B reports the price differences for multiple-CCSD episodes – in both cases based on a minimum of 5episodes. Comparing the price differences we observe some changes in the insurer-year price indices. For example, for AXA in 2008 the price difference turns from [✄]% to [✄]%. This substantial change in the price difference appears to be out of line with the remaining price differences for AXA. Less dramatic, though also notable, changes occur for some other insurers in some years, for example, for WPA the price difference increases by [✄] percentage points in 2011, while it falls by [✄]percentage points in 2010. The overall estimated price difference decreases from [✄]% to [✄]% when we include multiple-CCSD episodes.

92. In Table 4, columns C and D, we have also considered the price difference for a minimum of 30 episodes. Here, we observe fewer changes (three out of 23 insurer-year pairs) and much smaller changes in the price differences. For example, the largest change in the price difference is [✄] percentage points for Bupa in 2011. The reason for the smaller price changes is that treatments with multiple CCSDs tend to have few patients and so are unlikely to be in the common basket when the 30-episode threshold is applied. Therefore, they are not part of the calculation of the price index when the 30-episode threshold is applied. The overall price difference changes by just over one percentage point when moving from the single- to multiple-CCSD analysis.

93. We conclude that our overall estimated price differences between HCA and TLC are robust to the inclusion of multiple-CCSD episodes.

[63] Healthcode stated that there are limits to using multiple CCSDs, for example that an insurer might only pay 50% of a second CCSD on the invoice.

**Table 4: Price differences between HCA and TLC for single- and multiple-CCSD episodes**

| | | Price difference for single-CCSD episodes (5episodes) | Price difference for multiple-CCSD episodes (5-episodes) | Price difference for single-CCSD episodes (30-episodes) | Price difference for multiple-CCSD episodes (30-episodes) % |
|---|---|---|---|---|---|
| | | *A* | *B* | *C* | *D* |
| 2011 | Aviva | [✂] | [✂] | [✂] | [✂] |
| 2007 | AXA PPP | [✂] | [✂] | [✂] | [✂] |
| 2008 | AXA PPP | [✂] | [✂] | [✂] | [✂] |
| 2009 | AXA PPP | [✂] | [✂] | [✂] | [✂] |
| 2010 | AXA PPP | [✂] | [✂] | [✂] | [✂] |
| 2011 | AXA PPP | [✂] | [✂] | [✂] | [✂] |
| 2007 | Bupa | [✂] | [✂] | [✂] | [✂] |
| 2008 | Bupa | [✂] | [✂] | [✂] | [✂] |
| 2009 | Bupa | [✂] | [✂] | [✂] | [✂] |
| 2010 | Bupa | [✂] | [✂] | [✂] | [✂] |
| 2011 | Bupa | [✂] | [✂] | [✂] | [✂] |
| 2007 | Bupa int'l | [✂] | [✂] | [✂] | [✂] |
| 2008 | Bupa int'l | [✂] | [✂] | [✂] | [✂] |
| 2009 | Bupa int'l | [✂] | [✂] | [✂] | [✂] |
| 2010 | Bupa int'l | [✂] | [✂] | [✂] | [✂] |
| 2011 | Bupa int'l | [✂] | [✂] | [✂] | [✂] |
| 2007 | Cigna | [✂] | [✂] | [✂] | [✂] |
| 2008 | Cigna | [✂] | [✂] | [✂] | [✂] |
| 2009 | Cigna | [✂] | [✂] | [✂] | [✂] |
| 2010 | Cigna | [✂] | [✂] | [✂] | [✂] |
| 2011 | Cigna | [✂] | [✂] | [✂] | [✂] |
| 2010 | Exeter | [✂] | [✂] | [✂] | [✂] |
| 2008 | Pruhealth | [✂] | [✂] | [✂] | [✂] |
| 2009 | Pruhealth | [✂] | [✂] | [✂] | [✂] |
| 2010 | Pruhealth | [✂] | [✂] | [✂] | [✂] |
| 2011 | Pruhealth | [✂] | [✂] | [✂] | [✂] |
| 2009 | Simplyhealth | [✂] | [✂] | [✂] | [✂] |
| 2010 | Simplyhealth | [✂] | [✂] | [✂] | [✂] |
| 2011 | Simplyhealth | [✂] | [✂] | [✂] | [✂] |
| 2007 | SLH | [✂] | [✂] | [✂] | [✂] |
| 2008 | SLH | [✂] | [✂] | [✂] | [✂] |
| 2009 | SLH | [✂] | [✂] | [✂] | [✂] |
| 2010 | SLH | [✂] | [✂] | [✂] | [✂] |
| 2011 | SLH | [✂] | [✂] | [✂] | [✂] |
| 2010 | WPA | [✂] | [✂] | [✂] | [✂] |
| 2011 | WPA | [✂] | [✂] | [✂] | [✂] |
| | *Overall* | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.
Note: Price differences based on a 30-episode threshold cannot be calculated for some insurers in some years due to low patient volumes per treatment for some smaller insurers.

## *'Irrational' price predictions*

94. As part of the DRR, HCA's economic advisors (KPMG) identified a number of issues in relation to the 'irrational' price predictions produced by the treatment-level regressions that are used in constructing the price indices in the IPA. These issues are:

    *(a)* zero price predictions occurred in four out of 694 treatment-insurer-year prices. This resulted from a coding error, which we have now rectified;

(b) negative price predictions occurred in four out of 694 treatment-insurer-year prices. This was not an error, but rather a result of the regression methodology; and

(c) out-of-sample price predictions occurred in two out of 694 treatment-insurer-year prices, where our representative patient for those particular treatments was not representative of both operators' patients' characteristics.

95. We agree with KPMG that the issue of 'irrational' price predictions should be addressed, however we disagree with its approach. KPMG addresses this issue by excluding treatments with 'irrational' prices from the analysis on the basis that they produce odd results for certain years. We do not agree that we should simply exclude treatments on the basis that they produce odd results. As we detail below, we do not encounter this issue when we increase the threshold of minimum number of patient episodes. This approach has the additional desirable effect that it increases the precision of the estimates in our regressions.

96. In the remainder of this section we deal with each of the 'irrational' price predictions in turn below.

*Negative price predictions*

97. In the DRR KPMG identified four treatments for which negative prices are observed.[64] Negative price predictions occur as a result of the methodology we have used. In particular, the estimator we base our regression analysis on does not restrict the dependent variable, eg in this context the predicted price of a patient episode, to be positive.[65] However, observed prices in this context, as opposed to the *predicted* prices that our methodology produces, are by their nature only ever positive. The fact that we are not restricting prices to be positive in our methodology is not a problem as such, in particular when prices are predicted using explanatory variables that are within the sample.[66] We therefore do not agree with KPMG that negative price predictions are an error.

98. In addition, when we calculate price differences on the basis of a 30-episode threshold, we do not observe negative prices. The reason is that with a larger number of patients available for the estimation, the estimates are more robust

[64] DRR, Table 2 on page 17.
[65] We use an OLS estimator.
[66] In our case explanatory variables are the patient's age, gender and length of stay.

towards outlier observations.[67] For example, an outlier may include a patient for a specific treatment that stayed more than 50 days in one of the hospitals, whereas the average length of stay for that treatment is around 5 days. The effect of the outlier is to bias the estimates, and, therefore, to misrepresent the effect of length of stay on prices.

99. Notwithstanding our arguments that negative price predictions are a by-product of our methodology and the fact that they do not arise when we calculate prices based on a 30-episode threshold, we also present our results based on KPMG's methodology by excluding the treatments which give rise to negative predicted prices from our data. These results are presented in column A of Table 5, below, and illustrate the impact that those treatments with negative prices have on the estimated price differences, based on a 5-episode threshold.

*Zero price predictions*

100. In the DRR, KPMG showed that for four out of 694 treatment-insurer-year prices the analysis predicted that one of the hospital operators had charged a price of zero. We agree with HCA that zero price predictions should be avoided. We therefore implemented an approach in the computer code for the price index calculation that resolves this issue.[68]

101. Zero price predictions occur where Stata deals with collinear variables[69] in the data set by randomly dropping one of them. For example, in our dataset we observe that for a small number of treatments a hospital is treating female patients only. If this corresponds with a treatment where the hospital operator is only treating day-case patients, ie patients that have a length of stay equal to zero nights, then both patient characteristics have the same values for all patients and so are perfectly correlated or 'collinear'. Therefore, the relationships between these variables and the episode price cannot be estimated. This can lead to problems when we use our regression results to predict the episode price.[70]

102. In order to calculate the average price per treatment, we use a representative patient, ie a patient with the median characteristics for patients of that specific treatment. The representative patient gives us specific values for the patient's

[67] An outlier is classified as a patient with a, or several, characteristics that are visibly not in line with the other observations.
[68] We use the Stata command '_rmcollright', which excludes collinear variables in a specified.
[69] In statistics and econometrics, two variable are described as being 'collinear' when they are highly or perfectly correlated.
[70] If the variable that is dropped is the variable for the average net price charged by a hospital operator, then we do not have this 'constant' available in the data.

characteristics, that is, age, gender and length of stay. Our regression results generate a 'constant' – in other words, it tells us what the price would be before adding the effects of patient characteristics.[71] However, for some regressions this 'constant' element is incorrectly removed from the regression, while another term assumes the role of the constant.[72] Our methodology then multiplies the patient's characteristic with this term, which is not missing, and so we get a zero price prediction. For example, if patient gender is constant within a treatment as in the above example, we would multiply the median gender – that is female – by the constant term, which results in a zero.

103. After implementing the solution to our computer code we do not observe zero price predictions.

*Out-of-sample price predictions*

104. In the DRR, KPMG showed that for two out of 694 treatment-insurer-year prices the IPA used a representative patient to predict prices where one of the hospitals did not treat any patients with these characteristics.[73] We do not agree with KPMG that the out-of-sample price prediction is a mistake, but agree that it potentially raises an issue in relation to the precision of our predicted price estimates. Specifically, the out-of-sample predictions are part of our methodology and are a result of the 'representative patient' assumption. To understand this, one could consider the following stylised example. If for a specific treatment, patients at one hospital tend to stay for one night and patients at the other hospital are always treated as day cases, then the median patient may be one that stays for zero or for one night. Depending on which hospital operator treats more patients for that particular treatment, using the median patient to calculate the price for the treatment would lead to an 'out-of-sample' prediction for one or other of the hospital operators.

105. We have tested the sensitivity of our results to out-of-sample price predictions by using alternative definitions of the representative patient (see paragraphs 145 to 147). Taking again the example from the previous paragraph, using a representative patient based on the mean, rather than the median, characteristics of the relevant patients would give a mean length of stay somewhere between zero and one night. While not fully correcting for the difference between day-case and longer hospital stays, this alternative definition would mitigate the problem. Furthermore, based on a minimum

[71] For example, the price of a hip replacement might be made up of a 'constant' of £500 plus £10 for every extra year of patient age, plus £1,000 for every additional night spent in hospital, plus £100 more if the patient is male rather than female.
[72] The reason for the omission of the constant term is collinearity.
[73] On one occasion the representative patient is female, while TLC did not treat female patients. On the second occasion, the average length of stay of the representative patient is positive, while TLC did not treat inpatients.

threshold of 30 episodes per treatment, it does not appear that the issue of out-of-sample predictions arises. We present the results of this analysis - using alternative definitions of the representative patient – in the sensitivities section of this working paper, below.

### *Conclusions and updated results*

106. We present our revised results of the percentage price difference between HCA and TLC in Tables 5 and 6 below.

107. As discussed in detail above, these results are produced on the basis that we have:

*(a)* excluded episodes with multiple CCSD codes from our analysis; and

*(b)* rectified the coding issue which gave rise to zero price predictions;

*(c)* not used the information on diagnosis code or the consultant's speciality in our regression analysis; and

*(d)* not removed those treatments that lead to negative or out-of-sample price predictions.

108. First, we note that using a 30-episode threshold, 13 out of the 36 insurer-year price differences cannot be calculated. This is because, for these smaller insurers, the patient volumes are too low and the relevant treatments do not meet the 30-episode threshold in any year. Second, the results show that the individual insurer-year price differences are somewhat higher under a 30-episode threshold for some insurers, while for others the 30-epsiode results show lower price differences. The overall price difference is [✄]% for the 30-episode threshold, compared to [✄]% for the 5-episode threshold.

**Table 5: Revised average percentage price difference between HCA and TLC by insurer, 2007 to 2011**

| | | | % |
|---|---|---|---|
| | | *Median patient (30-episodes)* | *Median patient (5-episodes)* |
| | | *A* | *B* |
| 2011 | Aviva | [✂] | [✂] |
| 2007 | AXA PPP | [✂] | [✂] |
| 2008 | AXA PPP | [✂] | [✂] |
| 2009 | AXA PPP | [✂] | [✂] |
| 2010 | AXA PPP | [✂] | [✂] |
| 2011 | AXA PPP | [✂] | [✂] |
| 2007 | Bupa | [✂] | [✂] |
| 2008 | Bupa | [✂] | [✂] |
| 2009 | Bupa | [✂] | [✂] |
| 2010 | Bupa | [✂] | [✂] |
| 2011 | Bupa | [✂] | [✂] |
| 2007 | Bupa int'l | [✂] | [✂] |
| 2008 | Bupa int'l | [✂] | [✂] |
| 2009 | Bupa int'l | [✂] | [✂] |
| 2010 | Bupa int'l | [✂] | [✂] |
| 2011 | Bupa int'l | [✂] | [✂] |
| 2007 | Cigna | [✂] | [✂] |
| 2008 | Cigna | [✂] | [✂] |
| 2009 | Cigna | [✂] | [✂] |
| 2010 | Cigna | [✂] | [✂] |
| 2011 | Cigna | [✂] | [✂] |
| 2010 | Exeter | [✂] | [✂] |
| 2008 | Pruhealth | [✂] | [✂] |
| 2009 | Pruhealth | [✂] | [✂] |
| 2010 | Pruhealth | [✂] | [✂] |
| 2011 | Pruhealth | [✂] | [✂] |
| 2009 | Simplyhealth | [✂] | [✂] |
| 2010 | Simplyhealth | [✂] | [✂] |
| 2011 | Simplyhealth | [✂] | [✂] |
| 2007 | SLH | [✂] | [✂] |
| 2008 | SLH | [✂] | [✂] |
| 2009 | SLH | [✂] | [✂] |
| 2010 | SLH | [✂] | [✂] |
| 2011 | SLH | [✂] | [✂] |
| 2010 | WPA | [✂] | [✂] |
| 2011 | WPA | [✂] | [✂] |
| | *Overall* | [✂] | [✂] |

Source: CMA analysis.

109. Looking at Table 6 below where we average across insurers in each year, we see that, on average, increasing the threshold leads to increases in the price difference for later years but makes little or no difference in 2007 and 2008.

**Table 6: Updated overall average price differences between HCA and TLC, 2007 to 2011**

| | | | % |
|---|---|---|---|
| | *Updated results* | | *Original results in FR* |
| | *Median patient 30-episodes* | *Median patient 5-episodes* | *Median patient 5-episodes* |
| | *A* | *B* | *C* |
| 2007 | [✂] | [✂] | [✂] |
| 2008 | [✂] | [✂] | [✂] |
| 2009 | [✂] | [✂] | [✂] |
| 2010 | [✂] | [✂] | [✂] |
| 2011 | [✂] | [✂] | [✂] |
| *Overall* | [✂] | [✂] | [✂] |

Source: CMA analysis.
Notes:
1. This table presents the percentage price differences between HCA and TLC, averaged over all insurers in a given year.
2. A positive number means that HCA is more expensive than TLC.

110. Overall, we consider that both the price differences calculated using the 5- and 30-episode thresholds indicate that HCA charges higher prices than TLC in the region of [✂]% to [✂]% averaged across all five years of our data set.

## R-squared statistics

111. As part of the IPA, we have estimated a number of regressions that sought to explain the prices that PMIs paid to hospitals operators for each treatment in terms of patient characteristics. We have used the result of these regressions to construct the price indices for HCA and TLC of insured patients in central London. To explain the variation in prices within a treatment we have used patient characteristics – age, gender and length of stay – to ensure a like-for-like comparison. The R-squared figure of a regression is a measure of how much of the variation in prices is explained by the explanatory variables in the regression model.[74]

112. In the DRR HCA's economic advisers, KPMG, identified an error in the computer code we had used to calculate the R-squared figures (set out in detail in Appendix B). This coding error resulted in an overstatement of the R-squared figures that were reported in the Final Report.[75] KPMG corrected this coding error in the data room and reported their R-squared figures.

113. The implication of this error is that the variables included in the regression analysis explain a lower share of the variation in insured prices than we had reported in the Final Report. If the correct R-squared figures were so low that it appeared as if the regression model did nothing to explain the variation in prices then this could call into question these regressions and the 'representative patient' approach that we use to calculate the price indices and the resulting price differences between HCA and TLC.

114. We have now corrected this error and we present our own corrected R-squared statistics in column B of Table 7, below, alongside the R-squared figures not including our correction in the coding, which are those reported in the Final Report (column A), as well as those reported by KPMG in the DRR.[76] Table 7 below, presents R-squared statistics in terms of the proportion of regressions for which the R-squared is above the threshold specified in the first column. An alternative way to summarise the underlying results for our updated results would be to calculate an average R-squared statistic, weighing each regression by its relative importance in the basket[77] We

[74] This is explained in more detail in Appendix B of this paper.
[75] In the Final Report, the CMA stated that 'the adjusted R-squared varied… between 60 and 99% … the large majority of regressions have an adjusted R-squared that is above 80%'. See paragraph 17 (b) and footnote 19 in Appendix 6.12, Final Report.
[76] Note that we report the adjusted R-squared figures. The adjusted R-squared takes a similar approach to the unadjusted R-squared but takes account of the number of explanatory variables in the model, so that adding extra explanatory variables does not automatically increase the adjusted R-squared. The adjusted $R^2$ is generally lower (or, at least, equal to) the unadjusted R-squared.
[77] The R-squared statistics summarised in Table 7 should be interpreted bearing in mind that each of the regressions carries a different weight depending on the number of episodes per regression in the insurer-specific price index.

calculated this the weighted average adjusted R-squared statistic for all regressions for each year in the period from 2007 to 2011 and present these in Table 8, below. Looking at both of these tables, our corrected R-squared statistics show that:

*(a)* the large majority (68%) of treatment-level regressions have an adjusted R-squared statistic of over 50%;

*(b)* 45% of regressions have an adjusted R-squared that is 80% or higher; and

*(c)* the weighted average adjusted R-squared statistic (that is weighted by the number of patient episodes per treatment) ranges from 64% to 71%, depending on the year.

**Table 7: Distribution of R-squared statistics for treatment-level regressions for the HCA and TLC price comparison**

| | | | % |
|---|---|---|---|
| *R-squared* | *CMA adjusted $R^2$ referred to in the FR* | *CMA adjusted $R^2$ following corrections** | *$R^2$ (unadjusted) calculated by KPMG†* |
| | *A* | *B* | *C* |
| 90% or above | 89 | 27 | 32 |
| 80% or above | 99 | 45 | 49 |
| 70% or above | 100 | 54 | 58 |
| 60% or above | 100 | 62 | 67 |
| 50% or above | 100 | 68 | 75 |
| 40% or above | 100 | 65 | 81 |
| 30% or above | 100 | 81 | 88 |
| 20% or above | 100 | 86 | 93 |
| 10% or above | 100 | 92 | 98 |

Source: CMA analysis, KPMG Data Room Report (Table 9).
* R-squared results presented in this column incorporate the correction of the error in the calculation of adjusted R-squared, and corrections in data cleaning. Because of differences in data error corrections between KPMG and the CMA, as well as due to differences in adjusted and unadjusted R-squared statistics, our corrected results differ from KPMG's corrected results.
†DRR, Table 9.
Notes:
1. Each row in the table shows the proportion of regressions for which the R-squared was at or above the threshold specified in the first column.
2. The adjusted R-squared takes the value of the unadjusted R-squared and adjusts it for the number of explanatory variables (and observations). Adjusted R-squared is generally lower (or equal to) R-squared.

**Table 8: Weighted average of the (adjusted) R-squared statistics**

| | % |
|---|---|
| *Year* | *CMA adjusted $R^2$ following corrections (weighted average)* |
| 2007 | 64 |
| 2008 | 67 |
| 2009 | 71 |
| 2010 | 68 |
| 2011 | 69 |
| *2007 – 2011* | 68 |

Source: CMA analysis.

115. We accept that the R-squared figures reported in the Final Report were overstated, however both KPMG's corrected (unadjusted) R-squared figures

(75% of regressions with an R-squared above 50%, see column C, Table 7) and our updated figures show that our explanatory variables explain the majority of the variation that we observe in episode prices (68% of regressions with an R-squared above 50%, see column B, Table 7). We note that there is no absolute benchmark value for the R-squared statistic that we can measure any of the above numbers against. However, the majority of our corrected R-squared values are comparable with, or higher than, those R-squared values typically considered, for example in econometric textbooks (for similar types of regression models to those that we have used),[78] or observed in relevant peer-reviewed academic publications.[79] Thus, while there was an error that resulted in overstating the R-squared statistics in the Final Report, our corrected R-squared statistics still support the view that the patient characteristics included in the treatment-level regressions in the IPA explain the majority of the variation in episode prices.

***Coefficient estimates in the treatment-level regressions***

116. The preceding analysis presents our recalculated R-squared statistics for the treatment-level regressions in our IPA methodology and shows that, generally, our model explains the majority of the variation in episode prices. However, the R-squared statistic is only a descriptive figure that summarises the extent to which a regression model explains the variation in the dependent variable (prices in our case).

117. We have also reviewed the regression results to confirm that the explanatory variables that relate to patient characteristics – age, gender and length of stay – produce reasonable estimates or 'coefficients'. We have examined both the value of the coefficients which capture the relationship between each explanatory variable and the episode prices, and whether these estimated coefficients were statistically significant. For example, in our regressions the coefficient on length of stay is our estimate of the relationship between the length of stay for patients who have had a specific treatment and the episode price charged for that treatment. A positive coefficient indicates that an additional night as an inpatient is associated with a higher episode price being charged to the insurer.

[78] As set out in Appendix C, R-squared figures in the region of 50% could be considered relatively high in the context of cross-sectional data, while for cross sections of individual data (as we use here) an R-squared figure of 20% may be noteworthy.
[79] As set out in footnote 109 in Appendix C, recent empirical work using comparable data and published in prestigious academic journals report R-squared figures of between 7% and 25% (Fang, Keane and Silverman, 2008) and 41% (Gowrisankaran, Nevo and Town, 2015).

118. Looking in detail at the results from these approximately 700 treatment-level regressions, we have found that for length of stay we generally estimate positive and statistically significant coefficients, as we would expect. However, many of the treatment-level regressions reported age and gender effects that were not statistically significant. While this is not wholly unexpected – for some treatments we would not expect age or gender to drive cost differences[80] – we are using these patient characteristics to control for differences in patient complexity, so we would generally expect them to play a role in explaining price differences for many treatments. As such, we considered whether we had adequately modelled the relationship between these patient characteristics and the episode prices.

119. Having considered the issue, we set out a number of reasons why our methodology is still a robust way to model this relationships:[81]

*(a)* First, as set out in paragraphs 154 to 164, our single regression approach[82] provides a slightly different way of modelling the relationship between these patient characteristics and prices. As outlined in paragraph 164, the result of the single regression approach indicate that there are statistically significant relationships between age and gender, and price for the majority of treatments. As one of our sensitivity tests for the IPA approach, we change our IPA treatment-level regressions to reflect our preferred model from the single regression approach. This still produces many statistically insignificant coefficients on age and gender, but we are confident that we are applying the correct specification, so we consider that we can rely on the coefficients from these regressions as being reasonable estimates of the relationships between age and gender, and prices. As set out in paragraphs 151 to 153, this sensitivity produces broadly similar results to our 5-episode IPA methodology, with an overall price difference of [✂]%.

*(b)* Second, even if the coefficients on age and gender are not statistically significant, these are still our best estimates of the relationship between these variables and episode prices.

*(c)* Third, if these patient characteristics were poor predictors of episode prices then they would potentially have zero (or near-zero) coefficient

[80] For example, whether a cataract patient is a man or woman may not affect the level of costs involved.
[81] This issue was raised by Nuffield in response to Provisional Findings and dealt with in the Final Report at footnote 18 of Appendix 6.12.
[82] Our single regressions approach uses one regression to estimate the relationship between episode prices and whether HCA or TLC provided the treatment, while controlling for patient characteristics, treatment-specific effects, as well as other factors which may influence the price. In contrast, the treatment-level regressions here form part of the IPA methodology and involves estimating over 700 regressions – one for each treatment for each insurer for each year – and uses the results to construct price indices.

estimates and therefore would not affect the estimated prices that form the price indices in any case. We have checked this point and found that excluding these coefficients from the regressions does not substantially affect our estimated price differences. This indicates that where these coefficients are statistically insignificant they are not introducing significant biases or distortions into our results.

## Testing the statistical significance of the price differences

### *Introduction*

120. We conducted statistical significance testing for the price differences in the insurer-year-specific price indices to understand whether the price differences between HCA and TLC reflects a genuine price difference or whether the price differences are the result of random variation or statistical 'noise' in the data. To calculate the standard error we used a bootstrap approach, which we programmed in our statistical software, Stata. The code that we used contained an error identified in the DRR, which led to the miscalculation of the standard errors for the insurer-year specific price indices. This error ultimately resulted in an overstatement of the statistical significance of the calculated price differences. Subsequently we have corrected this error and in this section we present the results for the corrected statistical significance testing, as well as a number of other improvements we have made to the testing procedure.

121. To test the statistical significance of the price difference we need to calculate the standard error of the price differences between the two hospital operators – for a given insurer and year. While we can readily calculate the standard error at the treatment level, it is not straightforward when we come to test the difference in the insurer-year price indices. We therefore employ a generally accepted statistical technique called a bootstrap, which allows us to calculate the standard error of the difference in the price indices.[83]

122. The error in the bootstrap is technical in nature and we include a detailed discussion of this in Appendix D. In summary, the coding error resulted in our using the variation generated for one treatment for the calculation of the standard error of other treatments as well. We have corrected this error and present the corrected results below.

[83] The bootstrap is an established statistical method to calculate the variance of an estimator, and thus its standard error. See, for example, Wooldridge, J.M., (2010), Econometric Analysis of Cross Section and Panel Data, section 12.8.2.

123. In the rest of this section we detail the additional changes we have made to the bootstrap to improve the robustness of the statistical significance testing further. This additional work took into account the correction of the error in the bootstrap code.

***The composition of the common basket in the bootstrap***

124. In the results presented in the Final Report, the weights of the treatments were kept fixed, while the number of patients per treatment was allowed to be random. The aim was to reflect the fact that hospital operators do not know in advance how many patients for each treatment they will treat in a given year. However, we have reconsidered our approach to calculating the weights within each iteration of the bootstrap.

125. Our revised approach is to keep the number of observations per treatment, and hence the weight of the treatments, fixed. In particular, we restrict the re-sampling of the bootstrap to the treatment-hospital operator level. In other words, in each iteration of the bootstrap, we allow the computer program to re-sample such that the number of patients for a given treatment and for a given hospital operator are constant in each iteration of the bootstrap, rather than being randomly drawn in each bootstrap iteration.

126. The reason for this approach is that it corresponds better with our economic approach, where the hospital operator and the insurers bargain over the overall price. Our research into the market shows that hospital operators and insurers bargain over the total sum paid in a year, rather than the price paid by treatment. Therefore, hospital operators and insurers pay less attention to the composition of the costs with respect to the treatments. In particular, they do not take into consideration how many patients were treated for each treatment. It is therefore reasonable to assume that a hospital would expect to treat the same number of patients each year. This thinking is reflected in the assumption that hospital operators and insurers assume fixed weights, ie the same number of patients, within a treatment.[84]

127. For its work in the DRR, KPMG used a different approach to the composition of the common basket. While it also fixed the number of treatments in the basket, the number of observations and the weight each treatment receives varies with each iteration of the bootstrap. For the reason set out above, we do not pursue this approach.

[84] This is also in line with the bootstrapping principle that the number of observations of the sampled distribution should be constant.

128. Table 9 summarises the differences in the three approaches outlined above – the CMA's original approach, KPMG's methodology used for the DRR and our current view. We report the results of the statistical significance tests based on the current approach, and a comparison to the previous approach, in Table 11 below.

**Table 9: Summary of approaches to the 'bootstrapping' methodology**

| | *Original approach in the FR* | *KPMG DRR approach* | *Updated approach* |
|---|---|---|---|
| Treatments in basket | Fixed in bootstrap | Fixed in bootstrap | Fixed in bootstrap |
| Observations per treatment | Random in bootstrap | Random in bootstrap | Fixed in bootstrap |
| Weights of treatments | Fixed in bootstrap | Random in bootstrap | Fixed in bootstrap |

Source: CMA analysis.

### *The number of insurers considered*

129. In the Final Report, we produced a statistical significance test for 25 insurer-year specific price indices. This approach was taken in order to focus the statistical significance testing on the larger insurers only. For the insurers that were not considered, we did not have observations for all years, with the exception of [✂].[85] This was a deliberate choice made when testing the statistical significance. For the sake of completeness we now report the results of the statistical significance testing for all insurer-years (in Table 11, below), although, as pointed out above, a number of smaller insurers have patient volumes that fall below the 30-episode threshold and so are not included in these results.

### *Statistical significance testing of the overall price difference*

130. We have also tested the statistical significance of the overall price difference between HCA and TLC across all insurers and all years. Our results show that overall HCA's prices are [✂]% higher than TLC's (using the 5-episode threshold) and [✂]% higher (using the 30-episode threshold) and that these price differences are statistically significant at the 99% confidence level. We provide an overview of the overall price differences in Table 10.

**Table 10: Overall price differences, KPMG, Final Report and our updated approach**

%

| | *Updated approach (30-episodes)* | *Updated approach (5-episodes)* | *KPMG DRR (5-episodes)* | *Final Report (5-episodes)* |
|---|---|---|---|---|
| Simple average | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis, KPMG, Final Report.

[85] The omitted insurers were [✂] (2010, 2011), [✂] (2011) and [✂] (2009-2011), as well as [✂].

131. Table 10 also compares our updated results with those presented by KPMG in DRR and to those in the Final Report. For the 5-episode threshold, our updated results are similar to both the DRR and Final Report results, while the 30-episode results from our updated approach show a price difference that is higher by [✂] percentage points.

***Our revised results for statistical significance testing***

132. In Table 11 below, we present the percentage price difference between HCA and TLC. We indicate statistical significance of our results at the 99%, 95% and 90% confidence levels in the table. We also present the results of the statistical significance test conducted by KPMG. Because they do not report their estimated price differences, we are only able to present their findings in relation to statistical significance at the 95% confidence level. The results are reported for all 36 insurer-year-specific price indices for the 5-episode threshold, while for the 30-episode threshold, we report results for 23 insurer-year-specific indices, as low patient volumes mean that some smaller insurers are not included in this analysis.

133. For comparison, in column C we also report the results of the statistical significance testing if we had not adjusted the weighting scheme of the different treatments within the bootstrap. While we do not attach weight to this weighting scheme, we include it to illustrate the impact of this change on our results.

**Table 11: Summary of approaches to the 'bootstrapping' methodology**

| | | | | | % |
|---|---|---|---|---|---|
| | | Updated approach | | Original approach with errors corrected* | KPMG DRR approach |
| *Insurer* | *Year* | *30-episodes* | *5-episodes* | *5-episodes* | *5-episodes* |
| | | *A* | *B* | *C* | *D* |
| Aviva | 2011 | [✂] | [✂] | [✂] | [✂] |
| AXA | 2007 | [✂] | [✂] | [✂] | [✂] |
| AXA | 2008 | [✂] | [✂] | [✂] | [✂] |
| AXA | 2009 | [✂] | [✂] | [✂] | [✂] |
| AXA | 2010 | [✂] | [✂] | [✂] | [✂] |
| AXA | 2011 | [✂] | [✂] | [✂] | [✂] |
| Bupa | 2007 | [✂] | [✂] | [✂] | [✂] |
| Bupa | 2008 | [✂] | [✂] | [✂] | [✂] |
| Bupa | 2009 | [✂] | [✂] | [✂] | [✂] |
| Bupa | 2010 | [✂] | [✂] | [✂] | [✂] |
| Bupa | 2011 | [✂] | [✂] | [✂] | [✂] |
| Bupa Int'l | 2007 | [✂] | [✂] | [✂] | [✂] |
| Bupa Int'l | 2008 | [✂] | [✂] | [✂] | [✂] |
| Bupa Int'l | 2009 | [✂] | [✂] | [✂] | [✂] |
| Bupa Int'l | 2010 | [✂] | [✂] | [✂] | [✂] |
| Bupa Int'l | 2011 | [✂] | [✂] | [✂] | [✂] |
| Cigna | 2007 | [✂] | [✂] | [✂] | [✂] |
| Cigna | 2008 | [✂] | [✂] | [✂] | [✂] |
| Cigna | 2009 | [✂] | [✂] | [✂] | [✂] |
| Cigna | 2010 | [✂] | [✂] | [✂] | [✂] |
| Cigna | 2011 | [✂] | [✂] | [✂] | [✂] |
| Exeter | 2010 | [✂] | [✂] | [✂] | [✂] |
| Pruhealth | 2008 | [✂] | [✂] | [✂] | [✂] |
| Pruhealth | 2009 | [✂] | [✂] | [✂] | [✂] |
| Pruhealth | 2010 | [✂] | [✂] | [✂] | [✂] |
| Pruhealth | 2011 | [✂] | [✂] | [✂] | [✂] |
| Simplyhealth | 2009 | [✂] | [✂] | [✂] | [✂] |
| Simplyhealth | 2010 | [✂] | [✂] | [✂] | [✂] |
| Simplyhealth | 2011 | [✂] | [✂] | [✂] | [✂] |
| SLH | 2007 | [✂] | [✂] | [✂] | [✂] |
| SLH | 2008 | [✂] | [✂] | [✂] | [✂] |
| SLH | 2009 | [✂] | [✂] | [✂] | [✂] |
| SLH | 2010 | [✂] | [✂] | [✂] | [✂] |
| SLH | 2011 | [✂] | [✂] | [✂] | [✂] |
| WPA | 2010 | [✂] | [✂] | [✂] | [✂] |
| WPA | 2011 | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.
*This applies the methodology from the FR (see Table 9, above) with the coding errors and any data-related errors corrected.
Notes:
1. This table presents statistical significance tests for the percentage price differences between HCA and TLC. Statistical significance is presented at the 99% level by (a), 95% level by (b) and the 90% level by (c). Results for KPMG in column D show statically significance only, as the DRR did not report all of their estimated price differences.
2. A positive number means that HCA is more expensive compared to TLC.

134. As described above, we have tested the statistical significance of the price differences between the hospital operators, using a 30-episode threshold. For the 30-episode threshold, the results reported in column A of Table 11, above, show that all but one of the tested price differences are statistically significant at, at least, the 95% confidence level. The only exceptions is for the price index for [✂] in 2009. Price differences for smaller insurers were not tested as the patient volumes were too low.

135. For the 5-episode threshold the results reported in column B of Table 11, above, show that the majority of price differences are statistically significant at

the 95% confidence level.[86] For 22 of the 36 insurer-year pairs the price differences between HCA and TLC are statistically significant at the 95% confidence level.

136. Using the 5-episode threshold, the estimated price differences for [✂] in all years are not statistically significant, meaning that it is possible that the small price differences that we observe may be the result of random variation or statistical ‘noise’ in the underlying data, rather than representing actual price differences between HCA and TLC.

137. However, using the 30-episode threshold, the price differences for [✂] are statistically significant. We also observe that the variability in the data, ie the standard deviation, reduces (see Table 12, below) – in an extreme case from £[✂] to £[✂] in 2010. By increasing the threshold from 5 to 30 episodes per treatment, the price index is less influenced by extreme observations (ie outliers) – within treatments as well as across treatments. Consequently, the standard deviation of the price index decreases and so the precision of our estimate of the price differential is increased.

**Table 12: Comparison of results for [✂] prices charged to [✂]**

| | | | | *£* |
|---|---|---|---|---|
| | *30-episodes* | | *5-episodes* | |
| | *Price difference* | *Standard deviation* | *Price difference* | *Standard deviation* |
| | *A* | *B* | *C* | *D* |
| 2007 | [✂] | [✂] | [✂] | [✂] |
| 2008 | [✂] | [✂] | [✂] | [✂] |
| 2009 | [✂] | [✂] | [✂] | [✂] |
| 2010 | [✂] | [✂] | [✂] | [✂] |
| 2011 | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.
Note: The standard deviation is for [✂] only.

138. Overall, almost two thirds of the insurer-year price differences are statistically significant based on the 5-episode threshold, while for the 30-episode threshold over 90% are statistically significant.

[86] In the academic economics literature 95%, along with 99%, is the benchmark most commonly used. However, there is no absolute benchmark, and it is a somewhat arbitrary choice.

## Robustness checks

139. To assess the robustness of our results based on the price index methodology employed in the IPA, we have conducted a number of sensitivity tests. These tests make modifications to the methodology and/or data included in our analysis.

140. In this section we first describe the sensitivity tests that we have conducted on a number of elements of our IPA, and report the results of these. We then conduct a further robustness check where we adopt an alternative approach to calculating the price difference between HCA and TLC – the 'regression approach'.

### *Sensitivity tests*

141. We have looked at six changes to the underlying assumptions:

*(a)* Changing the definition of the representative patient from 'median' to 'mean'.

*(b)* Changing the definition of the representative patient from 'median' to 'mode'.

*(c)* Including episodes where multiple CCSDs have been recorded.

*(d)* Dropping those treatments that lead to 'irrational' price predictions.

*(e)* Applying a constant weighting for the episode numbers of each insurer across years.

*(f)* Expressing prices in log form.

142. Most of the sensitivity tests presented in this section are the result of our own review of the IPA methodology, with the exception of excluding treatments with 'irrational' price predictions and including those episodes where multiple CCSDs have been recorded. These two have been proposed by KPMG in the course of the analysis undertaken for the DRR.

143. Table 13 below presents a summary of the impact of these changes on our overall estimated price difference between HCA and TLC. As these results indicate, the estimates of the overall price difference between HCA and TLC are within the range [✂]% to [✂]% when we vary these assumptions, compared to the [✂]% to [✂]% range that we have calculated using the updated IPA methodology set out in this paper, for the 5- and 30-episode

thresholds, respectively.[87] The results on the insurer-year-specific price indices are presented in the corresponding sections below.

**Table 13: Sensitivity tests to the IPA (5-episode) methodology**

*%*

| | | *Change* | *Impact* | *Overall price difference between HCA and TLC* |
|---|---|---|---|---|
| Baseline | | | | [✂] |
| Price difference | *(a)* | Representative patient defined as mean, rather than median | Positive and negative changes in price differences for some insurer-years. | [✂] |
| Price difference | *(b)* | Representative Patient defined as mode, rather than median | Positive and negative changes in price differences for some insurer-years. | [✂] |
| Price difference | *(c)* | Including multiple-CCSD episodes* | Positive and negative changes in price differences for some insurer-years. | [✂] |
| Price difference | *(d)* | Dropping of specific treatments that lead to irrational price predictions† | Small impact on 6 out of 36 insurer-year price differences. | [✂] |
| Price difference and statistical significance thereof | *(e)* | Constant treatment weighting across years | Overall price difference reduces. | [✂] and is significant at the 99% level‡ |
| Price difference | *(f)* | Express prices in log form | Positive and negative changes in price differences for insurers and years. Overall price difference decreases. | [✂] |

Source: CMA analysis.
*See Table 4 above.
†See paragraph 94.
‡We report the statistical significance of the price differences in relation to *(e)*, as the changes that we make in relation to constant treatment weights affect both the calculation of the price differences and the specification of the bootstrap used in the statistical significance testing.

### *Alternative definitions of the 'representative patient' – (a) and (b)*

144. To estimate the prices for a specific treatment for a specific insurer in a given year we use a 'representative patient' to ensure that we are making a like-for-like comparison between HCA and TLC. The underlying idea is to compare the price that an insurer would be charged for an identical, and typical, patient (in terms of the characteristics that we observe – age, gender and length of stay) for that specific treatment at each of HCA and TLC.

145. The approach we have taken in the IPA is to define the representative patient as exhibiting the median characteristics for all of those patients in the data who receive that specific treatment either at HCA or TLC. The advantage of defining the representative patient using median values is that is gives us the characteristics that is central in the data. As part of our review of the IPA methodology, we have checked the sensitivity of the estimated price differences to two alternative assumptions about the representative patient, as we outline below.

[87] Note that the sensitivity test results presented in Table 13 are based on a 5-episode threshold and so are most comparable to the [✂]% price difference that we report for the 5-episode threshold.

146. The first sensitivity check we have conducted is to define the representative patient as having the mean characteristics of the patients for each treatment in the data set. The advantage of this definition is that it better reflects the distribution of patient characteristics in the data. The results of our comparison of using the median (our baseline approach that we use in the IPA) and the mean patient are presented in column B of Table 14, below, using the 5-episode threshold. The results show that the price differences increase for some insurers and years and decrease for others. However, overall there are still substantial positive price differences between HCA and TLC, for most insurers, with [✄] being the main exception to this. Looking at the overall price difference, across insurers and years, HCA is [✄]% more expensive compared to TLC when we use a mean representative patient, compared to [✄]% using a median representative patient.

147. We have then defined the representative patient as having the modal characteristics. The mode is defined as the value that appears most frequently in the data. Similar to the median, the advantage of the mode is that it is not affected by outliers. The results for this version of the representative patient (see column C of Table 14, below) again show that for some insurers and years the price differences increase while for others they decrease. Overall, there are still substantial positive price differences between HCA and TLC for most insurers, again with the exception of [✄], where most of the price differences are small relative to other insurers. Looking at the overall price difference across all insurers and all years, HCA's price are [✄]% higher than TLC's when we use the mode, compared to [✄]% when we use the median, as we do in the IPA methodology set out in this paper.

**Table 14: Price index: alternative definitions for the representative patient (5-episode threshold)**

%

| Year | Insurer | Median | Mean | | Mode | | KPMG approach* | |
|---|---|---|---|---|---|---|---|---|
| | | A | B | | C | | D | |
| 2011 | Aviva | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2007 | AXA PPP | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | AXA PPP | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | AXA PPP | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | AXA PPP | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | AXA PPP | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2007 | Bupa | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | Bupa | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | Bupa | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Bupa | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | Bupa | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2007 | Bupa int'l | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | Bupa int'l | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | Bupa int'l | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Bupa int'l | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | Bupa int'l | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2007 | Cigna | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | Cigna | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | Cigna | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Cigna | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | Cigna | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Exeter | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | Pruhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | Pruhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Pruhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | Pruhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | Simplyhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | Simplyhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | Simplyhealth | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2007 | SLH | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2008 | SLH | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2009 | SLH | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | SLH | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | SLH | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2010 | WPA | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| 2011 | WPA | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |
| | *Overall* | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.
*We show results based on KPMG's approach set out in the DRR of dropping specific treatments that KPMG pointed out were causing 'irrational' price predictions.
Notes:
1. This table presents the price differences between HCA and TLC for different definitions of the representative patient.
2. Columns B, C and D also report differences compared to Column A as shown in parenthesis.

### *Constant insurer weights across years – (e)*

148. As a further sensitivity test we have investigated an alternative way of weighting the various insurer-year-specific price indices ([✂] in 2008, [✂] in 2011 and so on) when calculating the aggregate price index for each of HCA and TLC over the five years for all insurers. When calculating the overall price difference we aggregated across each insurer and each year by weighting the price difference for each insurer in each year by the number of patient episodes for that insurer in that year. In the data set we observed that the number of patients for a given treatment varies somewhat across years.

When calculating the price differences between HCA and TLC, this could potentially lead to a ‘compositional’ effect when calculating the price difference. In other words, changes in the price index could be driven by changes in the mix and number of different treatments (or ‘composition’) within the price index rather than by changes in actual prices. To test if this were the case, we have estimated the price differences by holding the treatment mix constant over time by fixing the weight each treatment receives across all years in our analysis.

149. In addition to potentially biasing the price index results in the IPA, this compositional effect might introduce additional variation into the data set that may also lead to a bias in the standard errors, which are used in the statistical significance testing. This in turn could affect the robustness of our price difference estimates. For these reasons, there may be an argument for keeping the treatment weights constant over time.

150. Therefore, as a sensitivity test, we have held the weights for each insurer constant across years to assess whether there is a compositional effect in the data. In particular, for a given treatment, we have used the average number of patients across all five years as the weight in the price index calculation. We have then calculated the price index for each insurer and aggregate across all years and all insurers. Finally, we have calculated the standards errors for the statistical significance testing, using our bootstrap approach. We have found that the overall price difference using these constant weights is [✂]%, which is statistically significant at the 99% confidence level.

***Expressing prices in log form – (f)***

151. In our treatment-level regressions in the IPA, we have expressed our dependent variable (price) and our explanatory variables in absolute values or levels, that is, in pounds. So far we have implicitly assumed that the absolute levels of the effect of each patient characteristic on the episode price are constant within a treatment. For example, for a given treatment, the implicit assumption is that an additional year of age adds £10 to the hospital price, whether the patient is 40 years old or 80 years old.

152. With respect to this issue, expressing the price as a logarithm (log) of the episode price means that the implicit assumption then changes so that all patients that are treated at HCA or TLC have the same percentage price difference. For example, if we estimate a 10% increase in the price for the age effect of a patient, patient with one additional year has a 10% higher price regardless of whether the patient is 40 or 80. Thus using the log specification allows us to assess the sensitivity of the price difference to our assumption of a constant absolute price difference.

153. The results are presented in Table 15 below. For some insurers in some years, there are notable price changes; for example, the results for [✂] in 2010 and [✂] in 2008 change sign, with the [✂] result implying that HCA is [✂]% more expensive in our baseline, but that TLC is [✂]% more expensive when we use the log specification. However, overall, the price difference is broadly similar for both specifications, with the log version having an overall price differences of [✂]% – that is, [✂] percentage point lower than for our 5-episode IPA result.

**Table 15: Price differences for log specification**

| | | | *%* | |
|---|---|---|---|---|
| *Year* | *Insurer* | *5-episodes baseline* | *Log specification* | *Difference* |
| 2011 | Aviva | [✂] | [✂] | [✂] |
| 2007 | AXA PPP | [✂] | [✂] | [✂] |
| 2008 | AXA PPP | [✂] | [✂] | [✂] |
| 2009 | AXA PPP | [✂] | [✂] | [✂] |
| 2010 | AXA PPP | [✂] | [✂] | [✂] |
| 2011 | AXA PPP | [✂] | [✂] | [✂] |
| 2007 | Bupa | [✂] | [✂] | [✂] |
| 2008 | Bupa | [✂] | [✂] | [✂] |
| 2009 | Bupa | [✂] | [✂] | [✂] |
| 2010 | Bupa | [✂] | [✂] | [✂] |
| 2011 | Bupa | [✂] | [✂] | [✂] |
| 2007 | Bupa int'l | [✂] | [✂] | [✂] |
| 2008 | Bupa int'l | [✂] | [✂] | [✂] |
| 2009 | Bupa int'l | [✂] | [✂] | [✂] |
| 2010 | Bupa int'l | [✂] | [✂] | [✂] |
| 2011 | Bupa int'l | [✂] | [✂] | [✂] |
| 2007 | Cigna | [✂] | [✂] | [✂] |
| 2008 | Cigna | [✂] | [✂] | [✂] |
| 2009 | Cigna | [✂] | [✂] | [✂] |
| 2010 | Cigna | [✂] | [✂] | [✂] |
| 2011 | Cigna | [✂] | [✂] | [✂] |
| 2010 | Exeter | [✂] | [✂] | [✂] |
| 2008 | Pruhealth | [✂] | [✂] | [✂] |
| 2009 | Pruhealth | [✂] | [✂] | [✂] |
| 2010 | Pruhealth | [✂] | [✂] | [✂] |
| 2011 | Pruhealth | [✂] | [✂] | [✂] |
| 2009 | Simplyhealth | [✂] | [✂] | [✂] |
| 2010 | Simplyhealth | [✂] | [✂] | [✂] |
| 2011 | Simplyhealth | [✂] | [✂] | [✂] |
| 2007 | SLH | [✂] | [✂] | [✂] |
| 2008 | SLH | [✂] | [✂] | [✂] |
| 2009 | SLH | [✂] | [✂] | [✂] |
| 2010 | SLH | [✂] | [✂] | [✂] |
| 2011 | SLH | [✂] | [✂] | [✂] |
| 2010 | WPA | [✂] | [✂] | [✂] |
| 2011 | WPA | [✂] | [✂] | [✂] |
| | *Overall* | [✂] | [✂] | [✂] |

Source: CMA analysis.

## *Regression approach*

154. The IPA sought to control for differences in patient mix between HCA and TLC using regressions that took account of patients' age, gender and length of stay. To control for differences in the mix of treatments that HCA and TLC

provided, we have conducted the analysis separately for each treatment. This approach produced a set of price indices for each insurer for each year for each of HCA and TLC. We have then compared these to determine whether HCA and TLC charged insurers different prices. In order to establish whether any price differences were statistically significant, we have used a bootstrap approach as discussed from paragraph 120 above onwards.

155. In this section we present an alternative approach to calculating the overall price difference between HCA and TLC and testing the statistical significance of the estimate. This approach allows us to assess the robustness of our results from the price-index approach that we have used in the IPA and the statistical significance testing that we have conducted using a bootstrapping approach, as part of that methodology. The approach is based on a regression analysis where we model the price of each episode as a function of the hospital where the treatment is provided (HCA or TLC), while controlling for patient characteristics and any other factors that are specific to the insurer, treatment or year, as outlined below. In this section we explain this approach and discuss the results. A discussion of the technical details is set out in Appendix E.

156. While the price-index approach we have followed in the IPA establishes the extent to which there are price differences between hospital providers, testing whether these price differences are statistically significant is not straightforward. In particular, calculating the standard error from the predicted prices in the IPA, and aggregating them to the insurer-year level is not straightforward, because of potential interdependencies between the price indices which mean that these cannot be aggregated or simply ‘added up’ to calculate the overall standard error of the price differences that we calculate.[88] We have used a bootstrapping approach to calculate the standard error of the price difference between HCA and TLC and to test the statistical significance of the price difference. In particular, the bootstrap relies on a number of assumptions. This alternative regression approach allows us to estimate the price difference and test the statistical significance of that price difference in a single regression, whereas the IPA involves a number of different steps to construct various price indices in order to calculate the overall price difference between HCA and TLC and then a further step in order to test the statistical significance of that price difference.

[88] For a more detailed discussion, please see Appendix D.

157. The regression approach models the price for a given episode of a given treatment at a given hospital in a given year as a function of:

*(a)* whether that patient has been treated at a HCA or TLC hospital;

*(b)* the patient's characteristics – specifically, age, length of stay, gender and a day-case indicator;

*(c)* an indicator for the treatment received;

*(d)* the insurer the patient is insured with; and

*(e)* the year in which a patient was admitted to the hospital.

158. The estimated relationship between the HCA variable and the episode price is the main result of interest. It tells us the price difference between HCA and TLC, averaged across all treatments, years and insurers, while controlling for the above mentioned factors. Based on the results of the IPA price-index approach, we would expect a positive price difference for the HCA effect in this regression approach. A negative price difference here would suggest that TLC is more expensive than HCA.

159. HCA in its submissions to the CMA claimed, that the difference in prices is (partly) justified by the different patient mix. In the regression approach, we take into account the potential difference in both the treatment mix and the patient mix of the hospital operator. For instance, bypass surgery is usually more expensive relative to a colonoscopy. If hospital operator A performs relatively more bypass surgery and relatively fewer colonoscopies as compared to hospital operator B, then we may wrongly conclude that hospital operator A is comparatively more expensive. Hence, it is important to take into account the treatment mix of each hospital operator when estimating the price difference.

160. The approach we have taken in this regression analysis to account for the potential differences in the treatment mix between the two hospital operators is to use indicator or 'dummy' variables for the treatment a patient received in order to control for treatment-specific differences in costs.[89] This effectively strips out the treatment-specific element from the episode prices charged. For example, a heart bypass will always cost more than a cataract operation regardless of the hospital providers, the age of patient, the insurer and so on.

[89] We adopt a similar approach in calculating the price-indices for the IPA by estimating the price difference for each treatment separately.

The indicator variable for each treatment follows the CCSD code that is used to classify the treatments on the invoices in the Healthcode data.[90]

161. In addition, we have included patient characteristics as we consider these are likely to capture the effects of the patient mix on the price difference between the HCA and TLC.[91] The aim of including patient specific demographic characteristics in the regression approach is to take into account severity or complexity of the condition of a particular patient. For example, for a given treatment, a patient with a more severe or complex condition might have to stay in hospital longer. This affects the price of the treatment received regardless of which hospital the patient was treated at. If, as HCA claimed in its submissions to the CMA, HCA treats more complex patients, who may or may not stay longer, the observed episode price for a given treatment should be higher. However, we have already controlled for this higher costs required to treat more complex patient by including the length of stay explicitly in our regression. A similar argument can be made for age (as older patients tend to have more co-morbidities[92] and hence are more complex to treat) and gender (for those treatments where gender affects costs).

**Table 16: Regression results: the baseline model**

| | *A* | *B* | *C* |
|---|---|---|---|
| | *Baseline* | *Day-case dummy* | *Treatment interactions* |
| HCA | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] |
| Male | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Age (log) | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Length of Stay | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Day Case | [✂] | [✂]<br>[✂] | [✂] |
| R-squared (adjusted) | [✂] | [✂] | [✂] |
| No. of observations | [✂] | [✂] | [✂] |

Source: CMA analysis.
Note: Standard errors are in parentheses. Statistical significance of the coefficients is represented at the 99% level by (a), 95% level by (b) and the 90% level by (c).

162. We present the results of the baseline regression approach in Table 16, above. The results of the main specification are presented in column A. The coefficient on HCA suggests that HCA is on average [✂]% more expensive

[90] We have used episode CCSD fixed effect to control for the treatment mix. For more detail see Appendix E, which sets out in detail our estimated equations.
[91] In doing so we follow the recent academic literature. For example, Gowrisankaran, Nevo and Town (2015), "Estimating the Price Impact of Hospital Mergers: An Analysis of Inova's Proposed Acquisition of Prince William Hospital", *American Economic Review*, January, Vol. 105, No. 1.
[92] Co-morbidities are other conditions that a patient has alongside the main diagnosis that they are being treated for, eg high blood pressure, a heart condition, asthma or diabetes.

than TLC.[93] Also, the price difference is statistically significant at the 99% level. The length of stay has the anticipated positive effect on the price an insurer is charged for an episode, with longer length of stay being associated with higher prices.

163. In addition we have tested several other regression specifications. First we have included a variable that indicates whether a patient is a day-case to take into account potential price differences between day and in-patients. We present the results in column B of Table 16. Our results suggest that HCA is [✂]% more expensive compared to TLC and that this price difference is statistically significant. The effect of the day-case dummy is negative, which suggests that the price charged for a day-patient is lower compared to an inpatient.

164. Patient characteristics are likely to have different effects depending on the treatment a patient receives. For example, for some treatments older patients are likely to be more expensive than younger patients, eg an 80-year-old patient is likely to have more co-morbidities than a 60-year-old patient, whereas some paediatric treatments may be more expensive to perform on newborn babies than on 5-year-old children. We address this issue by estimating a regression that allows the effects of these patient characteristics to differ for each treatment.[94] We present this in column C of Table 16, above. The results suggest that there is a very small effect on the HCA coefficient.[95] This implies that the way in which we control for patient characteristics in our regression does not substantially change our estimate of the price difference between HCA and TLC.

*Inclusion of King Edward VII Hospital*

165. In addition to the results presented above, we have also included the King Edward VII Hospital (KEVII) in the comparison. The DRR raised the issue that the inclusion of the KEVII changed the relative price differences between HCA and TLC and gave results that showed that TLC was more expensive than HCA in three of the five years in our data set.[96] Furthermore, averaging over

---

[93] The coefficient presented in the table is almost equivalent to a percentage difference, but must be converted to an exact percentage difference by using $per\ cent\ \ Price\ Difference = (e^{HCA} - 1) * 100$ .
[94] Note that this is moving towards a more similar specification to the treatment-level regressions that we estimated in constructing the price indices for the IPA. The closest specification would be to interact each patient characteristic with treatment, insurer, year and hospital operator.
[95] We checked the statistical significance of the coefficients on the treatment-patient characteristic interaction terms. Across the various treatments about 75% of those coefficients are statistically significant.
[96] DRR, Table 12.

all years, the DRR showed that TLC is more expensive than HCA, while KEVII is the least expensive hospital.

166. We considered that addressing the issue of the three-operator comparison using the regression approach was most appropriate. The reasons are twofold. First, the regression approach allows us to use a wider basket. Specifically, in the IPA, when moving from a HCA-TLC comparison to a three-operator comparison for a minimum 5-episode threshold the number of treatments in the common basket reduces greatly from 704 to 147.[97] In the regression approach we are able to use a larger common basket, which should increase the precision of our estimated price differences.[98] Second, in the regression approach we are better able to take into account common effect across all hospital operators, for example effects across years, as compared to the price index approach.[99] Both arguments point towards the regression approach in the three-way comparison between HCA, TLC and KEVII.

167. Using our regression approach, presented in column A of Table 17, we have included data for patients treated at the KEVII.[100] We have included an indicator for patients treated at KEVII, while all other variables included in the regression approach are the same as discussed above in paragraphs 157 to 161. The results suggest that HCA is more expensive relative to TLC by [✂]% and that this difference is statistically significant. These results suggest that our finding that HCA is more expensive than TLC is robust to the inclusion of KEVII in the regression analysis. While the estimated effect for KEVII hospital suggest that it charges, on average, a higher price compared to TLC, this price difference is not statistically significant, so we cannot conclude from these results which of these two operators actually charges higher prices.

168. In addition, we have used a specification in which we include a day-patient indicator variable. We report the results in column B of Table 17. The results of this specification suggest that HCA is [✂]% more expensive compared to TLC. In column C of Table 17 we present our results when we make the patient characteristics treatment specific, as we did in column C of Table 16, above. The results of this specification suggest that HCA is [✂]% more expensive than TLC.

---

97 These numbers are for an analysis based on all years and all insurers applying the 5-episode threshold.
98 In the data cleaning for this three-way regression specification, we first restrict the data to include a minimum of 2 patient-episodes for all hospital operators and then further restrict the data to include treatments common between HCA and TLC in the common basket. The latter point implies that King Edward VII does not provide all treatments in the common basket.
99 Related to the example, we take into account time effects by using time fixed effects.
100 Note that the results presented in column C of Table E3 are calculated on a common basket of HCA and TLC, unlike KPMG's results, which are based on a common basket between HCA, TLC and King Edward VII.

**Table 17: Three-operator regression results**

| | *A* | *B* | *C* |
|---|---|---|---|
| | *KEVII baseline* | *Day-case dummy* | *Treatment interactions* |
| HCA | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] |
| KEVII | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] |
| Length of stay | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Gender | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Log age | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Day patient | [✂] | [✂]<br>[✂] | [✂] |
| R-squared | [✂] | [✂] | [✂] |
| No. of observations | [✂] | [✂] | [✂] |

Source: CMA analysis.
Notes:
1. Standard errors in parentheses. All standard errors are clustered at the treatment level.
2. Statistical significance of the coefficients is represented at the 99% level by (a), 95% level by (b) and the 90% level by (c).

*Conclusions on the regression approach*

169. In summary, based on the results of the regression approach we have concluded that HCA, on average, charges a higher price than TLC, taking into account different factors, and regression specifications, that might influence the price difference. In addition, our regression results have shown that the estimated price differences are statistically significant. We have also included the KEVII in the regression approach. Our results suggest that HCA remains the more expensive hospital operator out of the three hospital operators considered.

170. These regression results indicate that the estimated price difference between HCA and TLC is robust to using this alternative regression approach, as it yields broadly similar price differences between HCA and TLC to those calculated as part of our IPA approach set out in this paper.

# Appendix A: Other data-related issues

1. This appendix sets out the other data-related issues raised by KPMG in the DRR and explains how we addressed these issues.

2. The DRR highlighted a number of ‘errors’ in CMA’s cleaning and processing of the Healthcode data. These are set out in Section 3, and Annex 2 of the KPMG report. Section 3.1 of Annex 3 presents KPMG’s price index differences results (overall averages and insurer-specific averages) after it had corrected these ‘errors’. As explained below, we do not agree with all the points made by KPMG in this regard.

3. KPMG have raised a number of very minor points relating to the highly granular nature of the Healthcode data. We were aware of some of these points, and held detailed discussions with Healthcode (in particular, in relation to ancillary items, consultant fees and duplicated line items). Through those discussions, we were informed that the Healthcode data set contains various minor imperfections and nuances and there is often no perfect solution to dealing with them. These issues are typical for a large and complex data set of this nature. We proposed an approach to each of these issues that was discussed with Healthcode. While there may be alternative approaches in some cases, taking one reasonable approach relative to another cannot be construed as an ‘error’. In either case, the issues relate to a very small minority of the vast Healthcode data set, and thus in practice these do not make any material difference to our analysis.

4. Each issue raised is discussed below. We start by referring to those errors that we identified prior to the opening of the Data Room. We then address other ‘errors’ identified in the DRR. The last section explains the impact that the appropriate corrections to the Healthcode data cleaning and processing has on the overall price differences that we calculate using the IPA approach.

## Errors identified prior to the Data Room

5. When preparing the August-September 2014 Data Room, we identified a number of minor errors in our cleaning and processing of the Healthcode data for the IPA. We corrected all but one of those, put both the corrected and the uncorrected results (that is, as presented in the Final Report) in the Data Room.

6. We uncovered three errors in our computer code prior to opening the Data Room, which were:

*(a)* mistakenly pooling patients from King Edward VII's Hospital (KEVII) into the calculation of the representative patient for outside central London;

*(b)* mistakenly classifying patients with a missing insurer name as self-pay patients. This error only affected the comparison of insured prices to self-pay prices; and

*(c)* obtaining differences in the way line items were aggregated to episodes every time the computer code was processed.

7. The impact of the first two errors on the IPA results was negligible, and we noted that were not able to solve the third issue prior to opening the data room, but explained that the data discrepancies caused by this issue were negligible as well. KPMG accepted our corrections of the errors described at 6(a) and 6(b), above. This is acknowledged in paragraphs 27-29, and paragraph 31, of the DRR.

8. In paragraph 32 of the DRR, KPMG suggest a solution to the issue described in paragraph 6(c), above. We agree that this solution is appropriate, and have now implemented this in the processing of the Healthcode data.

## Measurement of patient age

9. KPMG found that there was an error in CMA's computer code which meant that patient's age – one of the control variables in the regression analysis – was calculated incorrectly (paragraph 34 of the DRR, and paragraphs 12-16 in Annex 2 of the DRR). KPMG explained that the CMA incorrectly subtracted the year of patient's birth from 2012 in order to calculate patient's age, whereas it should have subtracted the year of birth from the date when the episode took place.[101] KPMG noted that this error affected both the regressions, where age was one of the three control variables, and the characterisation of the representative patient.

10. We agree that this was an error in the computer code. The code calculated each patient's age as of 2012, and not in the year that the patient was treated. This error meant that all patients' ages were overestimated by a constant amount of years for each year (by two years for the 2010 analysis). However, since our baseline analysis was conducted for each year separately, and

[101] KPMG gave an example of a patient born in 1980 and treated in 2011: the CMA calculated the age of such patient to be 32, whereas it should have been 31.

since it is patients' age relative to each other that matters in the analysis, this error does not affect our analysis or results in any way.[102]

11. We have now corrected this error such that patient age is correctly calculated by subtracting the year of birth from the year in which the patient was discharged from the hospital for the particular episode.

## Inclusion or exclusion of episodes at certain HCA hospitals

12. KPMG made two points in relation to the inclusion or exclusion of episodes at HCA's hospitals. First, it noted that the CMA failed to exclude non-central London HCA hospital episodes from the analysis. HCA operates, in partnership with the NHS, two hospitals outside of central London, and KPMG claimed that data related to these non-central London hospitals should be excluded from the IPA for central London (and it has done so in the analysis presented in its DRR).[103]

13. We do not agree with the approach suggested by KPMG to exclude the non-central London HCA episodes from the analysis. We chose to include these hospitals in the comparison of HCA's and TLC's prices because hospital operators negotiate prices with insurers for their complete portfolio of hospitals (see paragraph 6.292 of the Final Report). Footnote 237 in the Final Report (as well as footnote 13 in Appendix 6.12 of the Final Report) acknowledges that we have included HCA's units outside London in the analysis, and notes that 'these facilities accounted for less than 1% of the price data that we analysed and are therefore unlikely to have a material effect on our results'.

14. Second, KPMG submitted that, for a small number of episodes, the Healthcode data identified the operator as HCA but did not identify the specific hospital where the episode took place. KPMG has excluded these [✂] episodes[104] from the analysis it presented in the DRR.

15. We disagree with this approach as it unnecessarily removes useful information that can be reliably used in the analysis. We did not exclude these episodes from the data because the IPA compares prices charged by hospital operators, and as such the identity of the specific hospital where an episode took place is not important. As explained above, we deliberately included HCA's patient episodes in the central London analysis even if these took

[102] KPMG have acknowledged that this error gave rise to a mismeasurement of the age variable in a non-systematic way (footnote 7 in Annex 2 of the DRR).
[103] KPMG also stated that including observations for these HCA hospitals slightly increase the common baskets of treatments between HCA and TLC for some insurers and in some years. See paragraph 38 of the DRR, and paragraphs 22-23 in Annex 2 of the DRR.
[104] See footnote 18 to paragraph 24 in Annex 2 of the DRR.

place in one of the hospitals HCA operate outside of London (this is, in any case, a small proportion of its overall business).

16. Thus, we are of the view that we should still follow the approach adopted in the Final Report – ie we have included both the non-central London HCA hospital episodes and the unknown HCA hospital episodes in the insured price comparison between HCA and TLC for central London for the reasons explained above.

## Duplicate line items

17. KPMG claimed that it found ‘duplicate line items’ in the Healthcode data and decided to exclude such line items (see paragraph 33 in Annex 2 of the DRR). KPMG describes that:

> ‘the duplicate line items have the same invoice ID, Industry Standard Code ... and line item price. They only differed in the diagnosis code associated with them’.

18. We have queried this issue with Healthcode, which clarified that those line items with the same invoice ID, Industry Standard Code and line item price, but with a different diagnosis code were, in fact, likely to be duplicates. They explained that sometimes patients have more than one diagnosis recorded for the same treatment and so an extra line is added in the Healthcode data set to record this. Based on this, we accepted KPMG’s suggested change to the data set and excluded from our analysis those line items that appear to be duplicates within the same episode. We have not assessed the impact of this change, although, given KPMG’s results, we do not expect it to be material given it only affect a small number of episodes.

## Ancillary fees

19. KPMG stated that, contrary to what paragraph 12 of Appendix 6.12 in the Final Report suggests, the CMA had not removed ancillary fees from the data when calculating episode prices (paragraphs 31-32, Annex 2, DRR). KPMG noted that it had already queried this error with the CMA, and the CMA had confirmed, in the course of the CAT appeal, that it did not remove such fees from episode prices. The CMA also noted that it ‘has reviewed the data and considers that only a negligible number of charges included in this data relate to ancillary items’.[105]

[105] Paragraph 43, DRR.

20. While inaccurately described in the Final Report, the fact that we have not removed ancillary fees from our episode prices is not an error, as such, and either approach could be taken. KPMG stated that it was not able, in the time provided, to exclude any such charges from the data (paragraph 44 of the DRR, and paragraph 32 in Annex 2 of the DRR) and determine the materiality of this error. We have kept any ancillary fees in the data (ie we have followed the same approach as originally in the Final Report) on the basis that ancillary fees arise in a negligible number of charges and so are extremely unlikely to materially impact our results.

## Consultant fees

21. For the IPA in the Final Report, we excluded consultant fees when calculating episode prices.[106] KPMG submitted that the CMA only excluded consultant fees where the industry standard code for a given line item was present within the 'Specialist and Practitioner fees' industry standard category. KPMG claimed that it had identified a number of other industry standard codes that related to consultant fees, and implied that these should have been excluded.[107] However, KPMG was unable during its period of review to provide a list of these other industry standard codes and an explanation of how these categories can be consistently identified in the data. KPMG stated in its DRR that it had not excluded any such further codes in the course of its analysis.

22. We accept the possibility that the invoices in the Healthcode data may have included a small number of consultant fees we did not identify. However, we did remove consultant charges to the extent that this was possible, and KPMG did not provide an explanation how any remaining charges could be identified. Further, based on KPMG's results, we do not anticipate that removing additional consultant fees would have a material impact on the price difference between the hospital operators.

23. In relation to consultant fees, KPMG also claimed that the CMA failed to exclude consultant fees correctly where there were multiple line items having both the same industry standard code belonging to the same invoice and the same price.[108] As set out at paragraph 25(c) below, we have looked into this issue and did not find any errors that required correction.

[106] The CMA noted that "for the majority of episodes, the Healthcode data does not include the consultant fee. In cases where the consultant fee is included (eg because a hospital operator bills on behalf of the consultant), we have subtracted this from the episode price" (Appendix 6.12 to the Final Report, Annex A, footnote 2). The CMA, therefore, excluded all consultant fees to the extent that these were clearly identified in the data.
[107] See paragraphs 35-36 of the DRR, and paragraphs 18-19 in Annex 2 of the DRR.
[108] See paragraphs 20-21 in Annex 2 of the DRR and paragraph 37 of the DRR

## Impact of these data cleaning issues on the IPA results

24. KPMG presented the cumulative impact of their data corrections in Table 1 of the DRR, and noted that they 'do not have a large impact on the average price indices'.

25. As explained above, we do not agree with a number of the changes KPMG made to the data. However, we agree with the following changes:

*(a)* KPMG's suggested solution to the issue described in paragraph 6(c) of this Appendix above.

*(b)* Correction to the calculation of patient age (paragraph 34, DRR).

*(c)* We have checked the invoices where the subtraction of consultant fees may not have been done correctly, as set out in paragraphs 21 and 23, above.

26. We have summarised the impact on the average price index differences between HCA and TLC of corrections *(a)* and *(b)* above in Table A1 below along with the results from the Final Report (Table 2 in Annex B to Appendix 6.12 of the Final Report) and KPMG's 'Data Error Correction IPA' (Table 1, DRR). Because point *(c)* has not uncovered any mistakes as such, there is no correction which has affected the results.

**Table A1: Average price index differences between HCA and TLC with the data corrections**

| | | | *%* |
|---|---|---|---|
| *Year* | *CMA Final Report†* | *KPMG data corrections‡* | *CMA updated data set and approach* |
| | *A* | *B* | *C* |
| 2007 | [✂] | [✂] | [✂] |
| 2008 | [✂] | [✂] | [✂] |
| 2009 | [✂] | [✂] | [✂] |
| 2010 | [✂] | [✂] | [✂] |
| 2011 | [✂] | [✂] | [✂] |
| *Average* | [✂] | [✂] | [✂] |

Source: Final Report, CMA analysis, and KPMG Data Room Report.
Note: The results presented are for based on the 5-episode threshold.
†Appendix 6.12, Annex B, Table 2.
‡DRR Table 1.

27. Table A1 shows that correcting for two of the three corrections that CMA agrees with has a negligible impact on the results if compared to the Final Report; the average difference between HCA and TLC increases from [✂]% to [✂]%.

28. Comparing columns B and C also shows that the changes to the data that KPMG made and that we do not agree with did not have a material impact on the overall results, but did introduce some changes to average price

differences in certain years. With KPMG's data corrections, its results show an average price difference between HCA and TLC of [✄]%, which is [✄]% higher than our revised results.

29. In addition to the issues raised in the DRR, we have conducted our own review of our Stata coding and of the data-cleaning process. This has led to some modifications to the processing of the raw data in forming the final episode-level data set used in the IPA. These changes have not had a material effect on the results of IPA. For example, the raw data from Healthcode came in different files for different time periods, which needed to be appended. As some patient episodes begin in one time period and end in the next, some episodes were erroneously 'split in two' by our data processing. We have now rectified this error. Only a very small fraction of episodes were affected.

## Rebates and 'shortfalling'

30. HCA's May submission on the IPA raised two data-related issues. It pointed out that the Healthcode data set may not fully reflect the payments made by insurers to HCA and TLC.[109] In particular, it pointed out that:

*(a)* some PMIs receive rebates from HCA, so the invoiced amounts may not be reflective of HCA's revenues; and

*(b)* some PMIs 'shortfall' their patients, that is they pay only part of the invoiced amount and the hospital operator may not receive the full amount invoiced.

31. In relation to rebates, this issue was dealt with in the Final Report, where information received from insurers indicated that rebates accounted for [✄] of the revenues that went to the hospital providers. We stated in the Final Report that 'the value of a rebate is typically less than [✄]% of the total expenditure between a PMI and a hospital operator during the year, and no rebate exceeded [✄]%'.[110]

32. Given the above, we consider that making an adjustment to take account of this issue would make no material difference to the results.

33. In relation to 'shortfalls', as outlined in the Final Report, this occurs where the insurers are unwilling to pay the full consultant fee and so either the consultant accepts a lower fee than had been invoiced or the patient may be

[109] KPMG, 'A Submission on the Analysis of Insured Prices', see paragraphs 47 to 50.
[110] See footnotes 484 and 485, Final Report.

liable for the balance.[111] The issue of who pays the consultant fees is not relevant to our analysis, as we are focussing on the prices charged by hospital operators net of any consultant fees, as outlined in some detail above.

[111] See paragraph 7.68, Final Report.

# Appendix B: The R-squared statistics

## What does an R-squared statistic show?

1. R-squared is a statistic that describes, for a single regression, the proportion of variation in the dependent variable that is explained by the explanatory variables. In other words, how well do the explanatory variables in the econometric model explain the outcome of interest? In the present context, the statistic can be interpreted as the proportion of variation in episode prices (the dependent variable) collectively explained by age, gender and length of stay (the explanatory variables). There is one R-squared statistic with this interpretation for each regression, with each regression relating to single treatment, for a particular insurer and year.

2. In theory, the R-squared statistic always lies between zero and one (or, equivalently, between 0% and 100%).[112] An R-squared value of zero implies that the explanatory variables explain none of the variation in the dependent variable, while an R-squared value of one or 100% implies that the explanatory variables explain all of this variation. An R-squared value that lies between these two extremes indicates that the explanatory variables go some way to explaining the variation in the dependent variable.

3. There is no hard threshold for what constitutes a good or bad R-squared value. As one popular postgraduate textbook (Greene) describes:

> 'the value of R-squared we obtained [of between 0.46 and 0.95] for the consumption function in Example 3.2 seems high in an absolute sense. Is it? Unfortunately there is no absolute basis for comparison. In fact, in using aggregate time-series data [such as in the example given], coefficients of determination this high are routine. In terms of the values one normally encounters with cross sections, an R-squared of 0.5 is relatively high. Coefficients of determination in cross sections of individual data as high as 0.2 are sometimes noteworthy'.

4. According to this textbook, then, one possible benchmark for a 'noteworthy' R-squared is 0.2 (or 20%) for the type of data used in the IPA (a cross-section of

[112] Note that in practical applications negative R-squared values are possible. This is a result of specific computations of the R-squared of a statistical programme. The implications of a negative R-squared for the estimated coefficients is immaterial.

individual data). It is routine to see R-squared values of a similar magnitude to those quoted above in peer-reviewed and published academic work.[113]

5. Even if in some cases the values of R-squared may seem low (that is close to zero), it does not mean that such models are deficient. The reason is that, regardless of the R-squared value, a regression can still produce estimates that are close to the true relationship between an explanatory variable and the dependent variable, ie the estimates are unbiased. Several undergraduate textbooks make this point very clearly and even caution against placing too much weight on the R-squared statistic.[114]

6. Note that KPMG reported R-squared statistics in the DRR,[115] while the figures reported in the Final Report were adjusted R-squared statistics. Adjusted R-squared statistic of any regression is typically lower than the unadjusted one. The distinction between these two measures is not important in the current context, and it is irrelevant to the nature of error described by KPMG.[116] We report adjusted R-squared statistics when presenting our corrected results below.

7. We agree with KPMG's statement that there was a coding error in our Stata codes that led to us reporting R-squared statistics for treatment-level regressions that are too high.[117] The regressions that form part of our IPA

---

[113] Looking at the healthcare literature, which uses comparable data, shows that (when it is reported) the adjusted R-squared values is often relatively low. For example, a publication from one of the most prestigious academic journals that uses data of a similar nature to the CMA (a cross-section of individual-level insurance data) and similar techniques (basic regression analysis) reports R-squared values between 0.07 and 0.25. Source*:* Fang, H., Keane, M.P. and Silverman, D. (2008), 'Sources of Advantageous Selection: Evidence from the Medigap Insurance Market', Journal of Political Economy, Vol. 116, No. 2. Another recent paper which uses comparable data, reports an average R-squared of 0.41. See: Gowrisankaran, Nevo and Town (2015**),** "Estimating the Price Impact of Hospital Mergers: An Analysis of Inova's Proposed Acquisition of Prince William Hospital", American Economic Review, January, Vol. 105, No. 1.
[114] See, for example, Wooldridge (2003), *Introductory Economics: A Modern Approach*, Thomson South-Western, Edition 2e, p.196: 'We have not focused much on the size of R-squared in evaluating our regressions models, because beginning students tend to put too much weight on R-squared. As we will see shortly, choosing a set of explanatory variables based on the size of R-squared can lead to nonsensical models. […] Nothing about the classical linear model assumptions requires that R-squared be above any particular value'. The classic linear model is what we use in our analysis.
[115] Table 9, DRR.
[116] The adjusted R-squared takes a similar approach to the unadjusted R-squared but takes account of the number of explanatory variables in the model, so that adding extra explanatory variables does not automatically increase the adjusted R-squared. The adjusted $R^2$ is generally lower (or, at least, equal to) the unadjusted R-squared.
[117] In practical terms, the error occurred as follows. Stata assumes by default that it should include a 'constant term' in any regression it is estimating, and makes this inclusion automatically. The way we formulated our regressions led to it supplying Stata with its own constant terms (the hospital dummy variables). This meant that Stata did not need to include its own constant term as it normally would. We instructed Stata accordingly by using an option known as 'nocons' (meaning Stata should add 'no constant'). While this option succeeded in telling Stata not to include a constant term, it did not inform Stata that the R-squared statistics should be adjusted, since the formula is different for models with and without constant terms. Stata then calculated the R-squared statistics as if the model had no constant term. The appropriate option, which would have succeeded on both counts (told Stata not to include a constant term, and told Stata to adjust the R-squared calculation), is known as 'hascons' (meaning the regression being supplied already 'has a constant').

methodology involve estimating a regression for each treatment in the relevant common basket for each insurer for each year. These regression model episode prices for each treatment as a function of patient characteristics and a constant term.[118] It is important to note that this error in the Stata codes affected only the calculation of the R-squared statistics for the regressions, and did not affect any other calculation or results.

[118] The core output of a regression contains a list of the estimates associated with each variable, and the statistical significance of each estimate. Alongside this core output there are often reported a range of other statistics related to the regression. These statistics offer auxiliary information or tests associated with the regression but are often given less attention than the core output. R-squared is one of these pieces of auxiliary information.

# Appendix C: 'Irrational' price predictions

1. One of the main issues in relation to the IPA raised by HCA and its advisers, KPMG, relates to the 'irrational' price predictions from the regressions that we used to estimate comparable prices for each treatment. KPMG, in its DDR, states that it uncovered 'errors in the CMA's regression analysis that lead to irrational or incorrect price predictions…'.[119] In this section we discuss the three alleged errors specified in the DRR. We first refer to HCA's view and then we look at the impact of corrections proposed by KPMG on the price differences.

## HCA's views

### *Negative price predictions*

2. The DRR, paragraph 58, stated that:

> 'Our review of the CMA's price predictions for individual treatments showed that on a number of occasions the HCA's methodology resulted in negative prices being predicted for certain treatments'.

3. In addition, the DRR, paragraph 59, stated that:

> 'These predicted prices are the result of the CMA's regression analysis estimating negative impacts for at least one of its control variables in a treatment'.

Paragraph 60 continues:

> 'Depending on the characteristics of the representative patient for a given treatment, these negative impacts could lead to predicted episode prices that are negative'.

4. In the DRR, KPMG identified four treatments for which negative prices are observed (see Table 2 on page 17, DRR). In addition, in Annex 4 to the DRR, KPMG presents an example of the reason for negative price predictions. In summary, they identify three outlier patients, which stay, on average, over four times longer than the other patients for those same treatments. KPMG demonstrated that these outliers have a strong impact on the estimation of the effect of patient age on the episode prices for the relevant treatments. In particular, removing the three patients from the estimation results in the effect

[119] DRR, paragraph 49.

of age on the price being positive for those specific treatments, as we might expect. This results in positive price predictions.

5. KPMG's approach to these 'irrational' price predictions was simply to discard the treatments that exhibited such predictions. In the DRR, paragraph 76, it stated:

> 'We have adjusted the CMA's analysis in such a way that allowed us to identify the treatments characterised by negative predictions, zero price predictions, and out of sample price predications and excluded them from the calculation of the insurer-specific and average price indices'.

6. KPMG concluded that:

> 'These errors caused the CMA to overestimate the difference in the insurer-specific price indices for certain insurers and years, as well as in the average price indices for 2007, 2010 and 2011'.

### *Zero price predictions*

7. The DRR, at paragraph 64, states that:

> 'These zero prices were predicted where the CMA was unable to estimate the impact of all of the factors in its model. Sometimes, when there is little variation across episodes, the Stata software package is unable to estimate the impact of one or more factors on episode prices. […] it [the Stata software package] set the price for that operator to zero. In other words, the CMA assumed that the operator would offer the treatment for free'.

8. KPMG had 'identified four instances in the CMA's corrected output where zero prices were predicted',[120] and these are presented in Table 3 on page 18 of the DRR. KPMG state that these price predictions 'have an important impact on the results…'.[121]

### *Out-of-sample predictions*

9. In the DRR, KPMG stated that:

[120] DRR, paragraph 65.
[121] DRR, paragraph 66.

> 'More generally, the CMA did not appropriately account for "out-of-sample predictions", ie where, due to the lack of variation in the episodes observed, the CMA's model could not identify the impact of each of its control variables on the episode charge'.[122]

10. In the DRR, KPMG identified two treatments where this issue arises, which are presented in Table 4 on page 19 of the DRR, although they also state that 'the impact on the differences in insurer-specific price indices from these out of sample predictions is immaterial'.[123]

***Approach taken by HCA/KPMG to address these issues and their results***

11. In response to the pricing prediction issues that KPMG identified, it:

> '…adjusted the CMA's analysis in such a way that allowed us to identify the treatments characterised by negative predictions, zero price predications, and out of sample price predications and exclude them from the calculation of the insurer-specific and average price indices'.[124]

12. In addition, KPMG implemented a command in the Stata code that removes variables which are collinear in a specific way. Additionally, in its code, after it had implemented the above fix, KPMG took any treatments that still resulted in negative or out-of-sample predicted prices and dropped these from the analysis.

13. KPMG stated that:

> 'fixing these errors [irrational price predications] leads to a decrease in the estimate of pricing differences between TLC and HCA in all years except [✂] and [✂]. The average difference across the five years also decreases'.[125]

In Table 8 on page 21 of the DRR KPMG reports that across the years the price difference between TLC and HCA decreases by [✂] percentage points if the treatments with irrational price predictions are dropped. The price difference falls from [✂]% to [✂]% overall.[126]

[122] DRR, paragraph 67.
[123] DRR, paragraph 71.
[124] DRR, paragraph 76.
[125] DRR, paragraph 79.
[126] The price difference reported without taking into account the irrational price predications is [✂]%, while the reported price difference taking into account the irrational price predictions is [✂]%.

## Appendix D: Testing the statistical significance of price differences

1. Our IPA methodology begins by constructing a number of price indices in order to estimate the price differences between HCA and TLC for each insurer in each year. We are then interested in the statistical significance of these price differences. We therefore have to calculate the standard error of the price differences at the insurer-year level, as a high standard error would indicate that our price estimates are imprecise. While we can readily calculate the standard error at the treatment level, it is less straightforward when we come to test the statistical significance of the difference in the insurer-year-specific price indices.[127]

2. The idea behind the representative patient is to compare prices between the hospital operators if the representative patient had to choose between these hospital operators for a specific treatment. We use regression analysis to estimate the effect of the different patient's characteristics on the episode price per treatment. We then use the representative patient to predict the price this patient would face choosing between the hospital operators. Because the representative patient has no variation in the patient characteristics we are unable to calculate the standard deviation of this estimated price directly.[128] We therefore employ a statistical technique called a 'bootstrap', which allows us to calculate the standard deviation of the difference in the price indices.

3. The 'bootstrap' follows a simple logic: randomly choose a patient episode from the data set, record the relevant information, including patient characteristics and the episode price, and generate a dataset based on a subset of the existing patient episodes.[129] From this newly generated dataset, we re-calculate the price difference, and repeat the first two steps (ie the re-sampling and price calculation) a very large number of times. Using this logic, we are able to use the repeatedly calculated prices differences to establish a standard deviation of the price difference, which we subsequently use for statistical significance testing. While we do not go further into the statistical theory underlying the bootstrap methodology, we note that the bootstrap is a recognised and regularly-used method for computing the standard deviation for this type of statistical significance testing.

[127] Note that the same argument holds for the annual and overall price difference.
[128] Note, that it is desired that the representative patient has constant characteristics. The variation is coming from the differences in patients characteristics used in the regression analysis. If the representative patient also had variation in its characteristics it would not be possible to compare like for like and hence the price indices.
[129] The re-sample method is with replacement. This means that once we have recorded the patient's characteristic, we 'put' the episode back into the sample that we use for re-sampling. This means that an episode can be re-considered for generating the bootstrap sample.

4. To generate the bootstrap we are able to use a built-in command in our statistical software package*,* Stata*.* The program contains an algorithm which automatically re-samples the data, and carries out the specified statistical calculation. We wrote the computer code to carry out the price difference calculation, which was used by Stata's bootstrap command. After repeatedly calculating the prices for HCA and TLC, Stata returns the price difference and the standard deviation of the price difference.

## The error

5. In its Re-amended Notice of Appeal (paragraph 115) HCA sets out its view on the coding error in the bootstrap program and on its impact on the statistical significance testing:

> '…a computer coding error had the consequence that the statistical significance tests for each [price] index comparison were performed in relation to the price for only one treatment in the common basket in the insurer-year pair in question rather than for the entire basket… That is, for each insurer-specific price index, the CMA took the estimated price variation of one treatment and interpreted it as the variability of the entire insurer-specific price index, which was in fact composed of multiple treatments differing in nature and price…'

6. This statement is supported by the findings in section 5.1 of the DRR. Here KPMG identified the error in the code in more detail. In particular, the DRR states that:

> 'due to an error in the CMA's writing of its bootstrapping code, [however], the CMA performed its bootstrapping analysis for each insurer-specific price index making use of the episodes associated with only one treatment'.

Further,

> 'the CMA took the estimated price variation of one treatment and interpreted it as the variability of the entire insurer-specific price index, which was composed of multiple treatments…'.

This led to an underestimation of the standard deviation of the price indices and thus an overstatement of the statistical significance of the insurer-year price index.

7. We agree that there is a coding error in the bootstrap program. As a result, our estimates of the statistical significance of the price differences were

incorrect. We corrected the program, by including the ‘nodrop’ option as suggested by KPMG in its DRR.

8. The nature of the error is the result of a peculiarity of Stata’s bootstrap program. This program was used in calculating the statistical significance of the price differences. As mentioned above, the program repeatedly re-samples the data and performs the calculation of the price index. To avoid statistical problems that can sometimes occur in this process, the bootstrap program drops all missing values from the dataset.[130] If missing values are dropped, Stata relies on examining the most recent treatment. In our program we rely on more than one treatment and Stata only considers the most recent one, and does not consider any treatment that may have preceded it. Consequently, Stata deleted all of the data except for a single treatment (the last one in an alphabetical list) and ended up computing the statistical significance for only one treatment.

9. In addition, HCA stated that the error made in the bootstrap was compounded by erroneously multiplying the incorrectly estimated standard error with the weight of the single treatment in the basket. We agree with this point. This was a direct outcome of the way our code was written, but not a separate error. Therefore, when correcting the coding error in the calculation of the bootstrap, the aforementioned problem disappears.

[130] We removed all missing values from the dataset prior to running the bootstrapping algorithm.

# Appendix E: Regression approach to estimating a price difference

1. In this appendix we provide more detail on the regression approach. Our approach is based on modelling the price of each episode as a function of the hospital where the treatment is provided (HCA or TLC), while controlling for the patient mix, using the patient characteristics and any other factors that are specific to the insurer, treatment or year. The aim of this approach is to assess the robustness of the price-index approach to calculating price differences between the two hospital providers as well as the bootstrapping approach to testing the statistical significance of those price differences.

2. One advantage of the regression approach is that we do not have to rely on a representative patient, which we have used in the price index approach.

3. While the price index approach establishes a reliable answer to the question of price differences between hospital providers, testing whether these price differences are statistical significant is not straightforward. In particular, calculating the standard error from the regressions on the insurer-year-treatment level, and aggregating them up to the insurer-year level presents a challenge. The reason is that treatment prices might be correlated, which has to be taken into account when calculating the standard error. Hence, we use a bootstrapping approach to calculate the standard error of the price difference between HCA and TLC in the Final Report, as well as in this working paper. We then use the standard error to test the statistical significance of the price differences. However, the bootstrap approach relies on a number of assumptions. For example, the bootstrap approach implicitly relies on, and is sensitive to, the minimum episode threshold that is applied for a given treatment. For example, looking at the price difference for [✂] in 2009: using a threshold of 5 episodes, the price difference ([✂]%) is not statistically significant. Moving to a 30-episode threshold, the price difference ([✂]%) is statistically significant at the 99% confidence level.

4. Notwithstanding the limitations identified above with the bootstrapping approach, we consider that it produces robust results. However, it is important to understand to what degree the bootstrap might be affected by changes in the assumptions. A simple way to test the results of the bootstrap is to use the regression approach, which provides a test of the statistical significance of the coefficient estimates as part of the regression output.

5. In comparing the regression approach and the price-index approach used in the IPA, particularly relevant points include the following:

   *(a)* While the two approaches aim to answer the same questions, there are a number of differences:

(i) Depending on the exact specification of the regression equation estimated, the effect of patient characteristics may be estimated in aggregate (across all treatments) rather than for each treatment and for each of HCA and TLC separately.

(ii) In the regression approach we estimate the effect of being treated by HCA on episode prices directly while simultaneously controlling for patient characteristics and any treatment-, insurer- and year-specific effects, rather than constructing a series of price indices in order to do this.

*(b)* While in the regression approach we avoid having to make the same assumptions as we do in implementing the IPA and the bootstrapping approach – in relation to the representative patient, statistical distributions, independence of variables and so on – estimating a regression is, of course, based on a number of assumptions about the relationship between prices and patient characteristics, the relationship between prices and HCA's bargaining power, among others.

## Data

6. We have used the same cleaned data set for the regression approach as we did for the IPA, although the regression approach uses more of the data. In the IPA we restricted the minimum number of patients treated at a hospital operator for a given treatment, insurer and year – using both 5- and 30-episode thresholds. We did this because we estimated regressions at the treatment-insurer-year-hospital-operator level, including three explanatory variables, and we therefore needed a sufficient number of episodes, ie 5 or 30, in order to be able to estimate the individual treatment-level regressions in the IPA approach. While we still consider patients within the common basket only, in the regression approach we are able to reduce the minimum episode threshold for each treatment to at least two episodes.[131] As a result of being able to include more treatments in our analysis the number of observations for the regressions approach is about 85,000 compared to around 68,000 for the 5-episode IPA.

[131] Lower patient numbers per treatment means that we are not able to estimate the respective treatment fixed effect. Note that all treatments for which this is the case are subsumed in the constant.

## Regression specification

7. The baseline regression equation we estimate is

$$\ln p_{tij} = \beta + \beta_1 HCA + \beta_2 X_{tij} + \gamma_t + \gamma_i + \gamma_j + u_{tij},$$

where X is a matrix containing the patients' (logarithm of) age, length of stay and gender. The $\gamma$s are treatment (t), insurer (i) and year (j) fixed effects, respectively.[132] HCA denotes a HCA dummy, indicating whether a patient received the treatment at a HCA hospital.

8. The academic literature demonstrates that controlling for the patient mix is important when estimating a price difference between two hospital operators.[133] We follow Haas-Wilson and Garmon (2009) and use treatment fixed effects, denoted by $\gamma_t$, to control for the patient mix. In addition we include patient characteristics in the regression. The former takes into account all factors that are constant within a treatment group. The latter controls for the severity of the individual patients.

9. Similar to the academic literature on healthcare we use the CCSD code to identify the treatment that a patient receives.[134] The aim of the CCSD codes it to provide a standardised way of recording medical procedure, ie treatments, to hospital-operators and insurers.[135] The CCSD codes provide a fine grained indication of complexity of the patient. For example, within the chapter of Chemotherapy (Chapter 18) the CCSD codes are subdivided into 0-7 days, 1-14 days, 1-21 days and 1-28 days. However, there could still be differences in patient's severity with a treatment group. We therefore control for additional patient characteristics, such as age, length of stay and gender.

10. In selecting the patient's characteristic, we were guided by the recent academic literature. We included age, gender, length of stay and a day-patient dummy.[136] The academic literature additionally suggests the use of the patient's race, diagnosis and comorbidities in the regression as well. However, we are not able to do this because the data we have does not

---

[132] Note that some treatment fixed effects are dropped if there is an insufficient number of patients for that treatment.
[133] We use similar control variables in the treatment-level regressions that we estimate as part of the IPA as well.
[134] For details on the CCSD codes please see the CCSD website.
[135] A CCSD code does not provide any guidance on the price a hospital operator is able to charge for the medical procedure, which is determined by the insurer and hospital operator in their price negotiations. The CCSD code also does not provide any indication about costs of a specific medical procedure relative to another medical procedure. This is unlike the Diagnosis Related Group used in academic publication focusing on the US-healthcare market. Because the relative weights are not available to us, we rely on a fixed effects approach in controlling for the treatment
[136] We add the day-case dummy to take into account potential difference in hospitals treating patient over-night or not. We do not use this indicator variable in the IPA approach because we use a more flexible regression form, specifically at the hospital-treatment-insurer level.

contain information on the patient's race and comorbidities.[137] The patient's diagnosis is available, however, in our conversations with Healthcode, who provided the data, we were told that the diagnosis code is unreliably coded and we therefore did not use this information in our regressions.

11. The standard errors are clustered at the treatment level.[138] The reason for clustering at the treatment level is that the error term across patients might be correlated for patients receiving a particular treatment. The standard response in the academic literature is to adopt a clustering approach to estimating the standard error.[139] In addition we explore alternative clustering, for example at the treatment-hospital level. We are conservative in our choice of clustering, reporting clustering with the largest standard errors compared to other reasonable approaches.[140]

12. We are mainly interested in the sign of the HCA specific effect, $\beta_1$, and whether the coefficient is statistically different from zero. A positive coefficient suggests that HCA is charging a higher price relative to TLC. Based on the price-index approach we employ in the IPA, we would expect a positive price difference.

13. From the discussion above, we would expect that the length of stay of a patient has a positive effect on the price in the regression approach. The reason is that an additional night in the hospital increases the costs for the hospital operator and is likely to be reflected in the price. We do not have a prior expectation on whether age and gender are likely to have positive or negative effects on the price charged for an episode. For example, for paediatric services a younger patient (eg, a newborn baby) might be more difficult to treat than an older child, while an older patient might make the treatment more costly, for example, for a hip replacement operation.

14. We also used two further sets of control variables:

*(a)* We took into account different effects of insurers on the price. Each insurers may possess a degree of bargaining power, which is likely to lead to different prices being charged to for different insurers' patients.

[138] By clustering the standard errors, ie, the standardised deviation from the mean, we take into account the correlation within a group, here, we cluster at the treatment level.
[139] In addition we explore alternative clustering, for example at the treatment-hospital level, for our baseline approach. We chose to report the clustering at the treatment level only, because the standard errors are the largest compared to other reasonable approaches.
[140] We do not explore clustering at the hospital operator level because of too few clusters. Additional clustering we explored were at the treatment-hospital-operator level and heteroscedasticity-robust standard errors.

(b) We took into account factors that vary across years but do not have different impacts on different insurers, treatments or providers, eg inflation in input costs.

## Results

15. We present the results of the baseline regression approach, defined as in paragraph 7, in Table E1 below. The results of the main specification are presented in column A. The coefficient on HCA suggest that HCA is on average [✂]% more expensive than TLC. Also, the price difference is statistically significant at the 99% level.

16. With respect to the patients' characteristics, only the length of stay and the day-case dummy have effects on the episode price which are statistically significant.[141] For length of stay we would expect a positive effect on the price of the treatment. Our regression approach suggest that for each additional night a patient stays in hospital, on average, the price increases by [✂]%. The effect of a patient's gender indicates that male patients incur lower prices, while for patient age the effect is positive, but, as both effects are statistically insignificant we cannot conclude that these variables have explanatory power for the price differences in this specification of the regression.

17. In column B we also use a day-case dummy to understand whether controlling for the differential costs involved in treating day cases affects our estimate of the price difference between HCA and TLC. For day-case patients we would expect the treatment to have a lower price because a day-case patient does not have the added costs needing overnight care. We find that a day-case patient has a lower price for a treatment by about [✂]%.[142] Compared to the results presented in column A, the HCA effect increases. The results suggest that HCA charges a [✂]% higher price compared to TLC and that this difference is statistically significantly.[143]

18. In the columns C and D of Table E1 below, we explore different variations to the main specification to check the robustness of our results. First, in column C, we focus on day-cases only. Day cases make up about three-quarters of the episodes in our data set and we therefore want to understand potential price differences compared to inpatients. The results suggest that for day cases only, HCA charges [✂]% more than TLC. Similarly, in column C we focus on in-patients only. The results suggest that for inpatients HCA is more

---

[141] Which indicates those episodes where the patient was treated as day cases with no overnight stay.
[142] Again, the coefficient on 'day-case' is roughly equivalent to a percentage difference, but not exactly so, as in footnote 93, above.
[143] Note that the f-test for joint significance of the HCA dummy and the HCA x Day-Patient coefficient is [✂] (p-value: [✂]).

expensive by [✂]% compared to TLC. Overall we can conclude that our estimated price difference from the regression is robust, ie does not change meaningfully, in response to changes in the control variables that we include.

**Table E1: Regression results**

| | *A*<br>*Baseline* | *B*<br>*Day-case dummy* | *C*<br>*Day-cases only* | *D*<br>*In-patients only* | *E*<br>*Treatment patient interaction* |
|---|---|---|---|---|---|
| HCA | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] |
| Male | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Age (log) | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂]<br>[✂] | [✂] |
| Length of stay | [✂]<br>[✂] | [✂]<br>[✂] | [✂] | [✂]<br>[✂] | [✂] |
| Day case | [✂] | [✂]<br>[✂] | [✂] | [✂] | [✂] |
| R-squared (adjusted) | [✂] | [✂] | [✂] | [✂] | [✂] |
| Number of observations | [✂] | [✂] | [✂] | [✂] | [✂] |

Source: CMA analysis.
Notes:
1. This table presents statistical significance tests for all estimated coefficients. Statistical significance to a: 99% level is indicated by (a); to a 95% level by (b); and to a 90% level by (c). Standard errors are in parentheses. All standard errors are clustered at the treatment level. All specifications are with treatment, insurer and year fixed effects.
2. A positive number on the HCA coefficient means that HCA is more expensive than TLC.

### *Comparison of the price-index IPA and the regression approach*

19. We have also compared the price difference between HCA and TLC for the IPA price-index approach and the regression approach. The overall price difference between HCA and TLC in the price-index approach set out in this working paper is [✂]%, based on a minimum of 5 episodes, while for the 30-episode threshold, it is [✂]%. This compares with the estimate of [✂]% that we find in our regression approach.

### *Interaction between patient characteristics and treatments*

20. As mentioned above, patient characteristics are likely to have different effects depending on the treatment a patient receives. For example, for some treatments older patients are likely to be more expensive than younger patients, eg an 80-year-old patient is likely to have more co-morbidities than a 60-year-old patient, whereas some paediatric treatments may be more expensive to perform on newborn babies than on 5-year-old children. While in our baseline regression we found that gender and age were statistically insignificant, this might be due to the restriction that these two variables have an effect on the aggregate level only – that is we force them to have the same effect averaged across all treatments. We addressed this issue by estimating

a regression that allows the effects of the patient characteristics to differ for each treatment, as is likely to be the case for age and gender.

21. In the second specification we estimate, we take into account that patient characteristics may have a different impact for different treatments. We therefore augment our initial regression by interacting the patient characteristics with the treatment. The regression equation then becomes:

$$\ln p_{tij} = \beta + \beta_1 HCA + \beta_2\, T * X_{tij} + \gamma_t + \gamma_i + \gamma_j + u_{tij},$$

22. The additional term $T * X_{tij}$ is the interaction between treatment and patient characteristics. Again, our main interest is in sign and statistical significance of the HCA dummy. As before, we cluster the standard errors at the treatment level.

23. We present the results on this specification in column E of Table E1 above. The results suggest that there is no significant change in the HCA coefficient. This implies that the way in which we take into account patients' characteristics in the regression approach does not substantially change our estimate of the price difference between HCA and TLC.

## Three-operator comparison – King Edward VII Hospital

24. In addition to the results presented above, we have also included the King Edward VII Hospital (KEVII) in the comparison as well. KPMG in its DRR has raised the issue that the inclusion of the KEVII changed the relative price differences between HCA and TLC and gave results that showed that TLC was more expensive than HCA in three of the five years.[144] However, averaging over all years, KPMG in its DRR showed that TLC is more expensive compared to HCA, while KEVII is the least expensive hospital.

25. We therefore include a dummy variable for KEVII as well. The regression equation now reads:

$$\ln p_{tij} = \beta + \beta_1 HCA + \beta_2 KEV + \beta_3\, X_{tij} + \gamma_t + \gamma_i + \gamma_j + u_{tij},$$

where KEV is the King Edward VII dummy. As before, we cluster the standard errors at the treatment level.

26. In the Final Report we conducted a comparison of the prices charged by HCA and KEVII using the price-index approach. In the regression approach we now also take into account TLC. Furthermore, we also change the common basket

[144] DRR, Table 12.

to cover KEVII, HCA and TLC, such that it includes the common treatments provided by the three hospital operators.

27. We present our results in Table E2, below. The results suggest that HCA is, on average, more expensive relative to TLC by [✄]% and that this price difference is statistically significant at 99% confidence level. While the results suggest that the KEVII hospital charges, on average, a higher price compared to TLC, this price difference is not statistically significant, so the price difference is likely to be due to random variation or 'noise' in the data.

28. It is important to note that our analysis here is different to KPMG's. The three-operator analysis in the DRR is based on a small common basket of those treatments that all three of HCA, TLC and King Edward VII provide. And where each provider treats at least five patients for that treatment in a given year. In contrast, our regressions approach includes all treatments that more than one of HCA, TLC and KEVII provide and it includes all treatments where any episodes of those treatments were provided.

**Table E2: Regression results – three-operator comparison**

| | *A* | *B* | *C* |
|---|---|---|---|
| | *Log price* | *Log price* | *Log price* |
| HCA | [✄]<br>[✄] | [✄]<br>[✄] | [✄]<br>[✄] |
| KEVII | [✄]<br>[✄] | [✄]<br>[✄] | [✄]<br>[✄] |
| Length of stay | [✄]<br>[✄] | [✄]<br>[✄] | [✄] |
| Gender | [✄]<br>[✄] | [✄]<br>[✄] | [✄] |
| Log age | [✄]<br>[✄] | [✄]<br>[✄] | [✄] |
| Day-patient | [✄] | [✄]<br>[✄] | [✄] |
| R-squared | [✄] | [✄] | [✄] |
| Number of observations | [✄] | [✄] | [✄] |

Source: CMA analysis.
Notes:
1. Standard errors are in parentheses. All standard errors are clustered at the treatment level.
2. This table presents statistical significance tests for all estimated coefficients. Statistical significance to a 99% level is indicated by (a); to a 95% level by (b); and to a 90% level by (c).