Comments on the CC's revised price concentration analysis (PCA)

Peter Davis, Erik Langer, Andy Parkinson and Zita Vasas[1]

11 November 2013

Non-Confidential Version

[1] The authors are respectively Executive Vice President, Economist, Senior Economist and Economist at Compass Lexecon, an economic consulting firm.

Table of contents

Section 1	Executive summary	1
	The CC's OLS regressions	1
	The CC's instrumental variable estimation	3
	The CC's lack of standard specification testing	4
	Summary and implications	5
Section 2	Introduction	7
Section 3	The actual scope of the CC's PCA findings	10
Section 4	Changes to the CC's baseline model	13
	The CC's revised data cleaning process	14
	The four focal treatments	14
	LOCI concentration measure	14
	Regional indicator variables	15
	The local area characteristics variables	17
Section 5	The CC's baseline OLS results	21
	The CC's baseline results	22
	The CC's results show that their explanatory variables do not achieve their aim of controlling for local demand and cost conditions	25
	The CC's approach cannot distinguish between problematic and non-problematic variation in self-pay prices	30
Section 6	The CC's dataset contains significant errors	33
	Missing data	33
	Incorrect data	35
	The implications for the CC's results	37
	The CC's baseline OLS regressions	38
	The CC's BMI-only regressions	39
	The CC's instrumental variables estimation	40
	Concerns arising from the data errors and omissions	40
Section 7	The CC uses invalid instrumental variables	41
	The validity of the CC's instrumental variables based on distance	44

	The validity of the instrumental variables based on concentration in a comparator market (insured patient LOCI)	54
Section 8	The CC's tests of its instrumental variables	58
	Tests of the IVs' relevancy	58
	Tests of the covariates' exogeneity	59
	Tests of the instruments' exogeneity	60
Section 9	The CC incorrectly imposes one LOCI coefficient across all operators	62
	There is evidence available to the CC suggesting important differences in price-concentration relationships between the hospital operators	63
	The CC's approach to pooling the data	64
	The appropriateness (or otherwise) of assuming a single LOCI coefficient across hospital chains and treatments	66
	Testing the CC's implicit parameter restrictions on the pooled data	68
	The CC's comments on the econometric analysis submitted on behalf of BMI	71
Section 10	The CC disregards standard tests for model misspecification	73
	The validity of the CC's stated concern about RESET tests	75
Annex A	Construction of local area characteristics variables	79
	Reconstructing the CC's local area characteristics variables at the NUTS3 level	79
	Computing the population density and average age for BMI hospitals' catchment areas	80
	Computing the ACORN measures for BMI hospitals' catchment areas	81
	Results for population density and average age	81
Annex B	A brief literature review on the RESET test	83
	Conclusion on the CC's concerns around RESET tests	86
	Data Room addendum	86

Section 1

Executive summary

1.1 This report comments on the price-concentration analysis ("PCA") that forms part of the CC's Provisional Findings. We find that the CC's analysis suffers from a number of significant flaws that would individually be likely to give rise to bias and a lack of reliability in the CC's PCA findings; cumulatively, the flaws in the CC's work render it profoundly unreliable as a basis for establishing the relationship between price and concentration in local markets. In our opinion, the CC simply cannot rationally place any evidential weight on the econometric results presented in the PCA in the Provisional Findings.

1.2 We mainly focus our attention on the PCA itself, but first examine the interpretation which is placed on the PCA in the Provisional Findings. The PCA is conducted over four inpatient self-pay treatments, but the Provisional Findings seek to extrapolate the reported finding of the PCA to other self-pay treatments (including outpatient and daycase treatments) and indeed to market power in national negotiations between PMIs and hospital groups. The PCA uses only a very small subset of the data on patient episodes available to the CC. In its Provisional Findings, the CC then effectively extrapolates back to the entire industry on the basis of its claims around the relationship between price and concentration for these (fewer than) 12,304 patient episodes, with the 3,349 BMI episodes therein accounting for less than 0.8% of BMI's revenues over the period 2009-2012. The CC does not present a clear and reliable basis for such a large extrapolation beyond the four inpatient self-pay treatments actually studied.

1.3 We then turn to the technical details of the PCA itself.

The CC's OLS regressions

1.4 First, we consider the changes made by the CC to the PCA between its March working paper[2] and the Provisional Findings. In particular, the CC moves from preferring a specification which allows for local differences in demand and cost conditions to vary between 34 regions in the UK, to a specification which instead only allows for 11 such regions and additionally includes "local area characteristics variables".

[2] Competition Commission, "Price-concentration analysis for self-pay patients", published on 28 March 2013.

1.5 These local area characteristics variables measure demographic and other factors in an attempt to control for local demand and cost conditions. However, the variables are constructed for "NUTS3" regions, which bear no relation to the actual catchment area of each hospital. We show in two ways that these variables cannot adequately capture the demand and cost conditions faced by individual hospitals. First, we note that some NUTS3 regions are relatively large, and therefore hospitals in (say) Reading and Windsor are assumed to face the same local demand and cost conditions, which is unlikely to be the case. Second, we construct alternative measures of local area characteristics on the basis of the catchment areas defined by the CC for BMI hospitals, and find that the variables that the CC uses in its PCA represent a poor proxy for what the CC logically should regard as the actual local area characteristics given its findings in relation to catchment areas.

1.6 This finding is borne out by the CC's estimation results. The CC's results show that the local area characteristics variables are entirely ineffective in explaining variation in prices between local areas, in that they generally have counterintuitive signs and are statistically insignificant. The cost measures used as a control by the CC are similarly ineffective. Yet effective controls are required in order to make a like-for-like comparison between hospitals in different local areas. These considerations imply that the estimation results that use the local area characteristics variables as controls (L3/FC3 and similar specifications) are highly likely to be unreliable.

1.7 Moreover, the only OLS regression (out of six reported) which finds a statistically significant relationship between price and concentration is L3 (and then only at the 10% level of significance); there is no such relationship found when the CC uses a larger number of regional controls. The CC's estimation results therefore clearly indicate that the econometric analysis cannot tell apart the two hypotheses that the CC would like to distinguish: (i) variation in price due to concentration and (ii) variation in price due to local differences in demand and cost conditions.

1.8 The CC's preference for specification L3 is ultimately a choice to ascribe variation in prices between local markets to concentration, rather than to local demand and cost conditions. This choice is made in a manner that is clearly contrary to a conventional econometric approach: statistical tests reject the restrictions imposed by L3 preferring those imposed by L2. A dispassionate interpretation of the evidence would instead conclude that there is no robust evidence of a relationship between concentration and price, as the data is consistent with two hypotheses (that price variation is due to concentration, or it is due to local differences in demand and cost conditions) and the CC's PCA cannot distinguish between them.

1.9 Moreover, the dataset that underlies all of the CC's estimation results suffers from serious shortcomings. Missing data artificially and significantly restricts the sample used in regressions: most of the results presented in the PCA (including all of the CC's preferred specifications) exclude all hospitals in Scotland as well as certain other regions. Given that the CC's "main focus throughout the [PCA] appendix is [...] on the broad relationship that is representative of the industry in general" (paragraph 4), it is surprising that the PCA's conclusions are based on a sample which excludes (in particular) all hospitals in Scotland. Moreover, the costs variable is incorrectly constructed. Once these errors are corrected, we find no evidence of a statistically significant relationship between price and concentration in the CC's OLS regressions.

1.10 This result is particularly striking since the CC argued that such OLS estimates were the appropriate results for it to rely on in its Annotated Issues Statement (AIS). In doing so, it reported many (what appeared to be) statistically significant effects and argued (much as it currently does in Provisional Findings) that the relationship it was finding was "robust" and indicated the "causal" effect of concentration on prices. The reality however is unarguably the opposite once we fix the errors and omissions in the CC's current sample.

The CC's instrumental variable estimation

1.11 A new development in the PCA over the March working paper is the increased emphasis on the relevance of "instrumental variable" ("IV") estimation, which is now the CC's preferred specification. Whilst in principle we welcome the increased attention devoted to potential endogeneity concerns, we emphasise that the reliability of IV estimates entirely rests on the validity of the instruments used. Unfortunately, the CC's discussion of the validity of the instruments used in the PCA is entirely unconvincing on any conventional economic analysis. As a result, and in accordance with our earlier submissions, we believe that the CC's IVs are invalid. In particular:

a. In order for the IV based on distance to the nearest rival hospital (IV1) to be valid, it must be the case that distance to the nearest rival hospital tells us nothing about a hospital's market power over and above the information provided by the LOCI. This is exceptionally unlikely to be the case: all concentration measures are imperfect proxies for market power, and basic economic considerations imply that LOCI is a particularly poor one. Moreover, the CC's proposition that hospitals with the same LOCI have the same market power irrespective of the distance to the nearest rival hospital is inconsistent not only with economic intuition and benchmark economic models, but also with the CC's own approach to local competitive conditions elsewhere in the Provisional Findings. Such an illogical proposition must therefore fail. On any reasonable economic analysis, IV1 is invalid.

b. The second instrument considered by the CC, the distance to the nearest hospital under

common ownership, is acknowledged to be irrelevant by the CC (in that it is not conditionally correlated with LOCI and therefore cannot serve as an instrument).

c. The validity of the instrument based insured LOCI (IV3) rests on whether the insured LOCI would tell us anything more about local demand and cost conditions than the control variables. We find that this is highly likely, since cost conditions are essentially the same between the insured and self-pay markets; local markets with high insured demand are likely to be affluent and therefore also have high self-pay demand, which is not adequately controlled for by local area characteristics and regional dummies; and the presence of heterogeneity across hospitals, for example due to quality or range of services, would affect both insured and self-pay demand for those hospitals. In addition, insured and self-pay LOCI are very highly conditionally correlated, suggesting that it is unlikely that insured LOCI excludes the factors that render self-pay LOCI endogenous; accordingly, it is unlikely that insured LOCI is exogenous if self-pay LOCI is not (and the CC accepts self-pay LOCI is not). Again, IV3 is very likely to be invalid on any reasonable economic analysis.

1.12 These are not small concerns. Whilst IV estimation promises a solution to potentially serious endogeneity issues, the difficulty of finding "good" instruments is well-known in econometrics. Unfortunately, the use of "bad" (invalid) instruments entirely vitiates IV estimation, and patently "bad" instruments have been used in the PCA. Accordingly, all of the CC's IV estimates can only be regarded as highly likely to be unreliable. No reasonable economist applying a conventional economic and econometric approach would find otherwise.

The CC's lack of standard specification testing

1.13 The standard approach to ensuring that the chosen econometric model specification captures relevant features is well established. Unfortunately, the CC ignores that approach.

1.14 The CC estimates just one parameter on LOCI across all hospital operators and treatments, arguing that it represents a "broad relationship that is of primary interest". Our review of the March working paper showed that the "broad relationship" that the CC claims exists was entirely driven by one operator (and even by certain individual hospitals). The standard approach (which we believe the CC should adopt here) is to statistically test the validity of parameter restrictions imposed, for example by testing if the parameter on LOCI is the same across operators. Such tests show that the CC's restriction imposing that the parameter on LOCI is the same across operators are rejected by the data. Nor do these tests require using only a subset of the data, as the CC appears to suggest.

1.15 The CC also dismisses the relevance of a standard test for model misspecification referred to as the RESET test (despite having used it in previous investigations). We review the literature on this statistical test and find that the CC's stated concerns around this test appear to have no basis in the econometric or statistics literature. We find that the tests that the CC does perform are inadequate and that the CC's preferred specifications OLS specifications L3 and FC3, and the CC's preferred IV specification for LOCI (L7), fail appropriate versions of the RESET test. We note that the code provided in the Data Room did include RESET tests for OLS regressions, so the CC's dismissal of the RESET test in Provisional Findings was presumably in the deliberate but unreported knowledge that its specifications failed the test.

Summary and implications

1.16 We note first that there is a significant contrast between the actual scope of the PCA results (which relate to four inpatient treatments for self-pay) and the way in which the PCA results are used in the rest of the Provisional Findings. Specifically we note that the CC's results for four self-pay inpatient treatments are extrapolated to apply to all self-pay inpatient, outpatient and day-case patients and also, at times, to all insured patients as well.

1.17 With regard to the PCA we find that:

a. The CC's control variables for local conditions are demonstrably not serving the purpose they are intended to serve; the CC's new local area characteristic variables are statistically insignificant and generally have implausible signs in the CC's baseline (OLS and IV) regressions, i.e. L3 to L7 and FC3 to FC7.

b. The CC's cost variable is a very poor proxy for local and for treatment specific cost conditions; it is statistically insignificant in all of the CC's baseline OLS and IV regressions and has the wrong sign in some of the CC's OLS and IV regressions.

c. The CC's preference to include only a small number of regional indicator variables amounts to a choice by the CC to assign all of the regional variation in prices onto the market concentration variable. In reality the CC's model cannot distinguish problematic variation in prices (due to concentration) from non-problematic variation in prices (due to regional variation in quality or costs). The CC's choice is not based on economic science since the CC's model restrictions are actively rejected by the data.

d. The results of the CC's OLS regressions are in any case driven by substantial shortcomings in the data, in that missing data artificially restricts the sample and the cost variable is incorrectly constructed.

e. The CC's preferred specification is now based on IV estimation. The reliability of IV estimates depends critically on the validity of the instruments used. However, all of the CC's instruments are invalid or irrelevant. In particular, IV1 is patently invalid on any

reasonable economic analysis; IV2 is, as the CC acknowledges, "irrelevant"; and IV3 is highly likely to be invalid due to common demand and/or cost shocks not accounted for in the regression equation. Thus the CC's IV results are not likely to be reliable on any reasonable economic analysis.

f. The CC's choices over model specification testing defy any conventional approach to such questions:

   i. The CC presumes a "broad" price-concentration relationship and does not test whether the data support such a relationship. When we do test the relationship it becomes clear that the data does not support it.

   ii. The CC refuses to adopt conventional model specification tests (particularly the Regression Error Specification Test (RESET)) for reasons that are not grounded in any known valid statistical concern.

1.18 Whilst the PCA's results taken at face value support the proposition that there is a systematic relationship between price and the CC's concentration measure, the analysis contains a range of very significant problems that are driving the results the CC obtains. Each of the problems described above would individually be likely to give rise to bias and a lack of reliability in the CC's PCA findings; cumulatively, the flaws in the CC's work render it profoundly unreliable. In our opinion, the CC simply cannot rationally place any evidential weight on the econometric results presented in the PCA.

1.19 Given the technical nature of many of our concerns, we respectfully submit that before placing any evidential weight on the PCA, the Group must ensure it has put itself in a position to properly understand and weigh these technical concerns. Should there remain any doubt as to the unreliability of the CC's PCA after consideration of our submission, we consider the Group should appoint a reputable independent expert econometrician to help the Group evaluate the technical econometric evidence submitted by its staff team and also our criticisms of that work. That independent expert's report should be made public, and we should be provided with an opportunity to comment on it.

Section 2

Introduction

2.1 The CC's Provisional Findings[3] outline the results of a price-concentration analysis ("PCA"). The CC states that it *"tested statistically whether prices charged to self-pay patients are higher in areas where private hospitals face fewer competitive constraints, using a technique known as price-concentration analysis (PCA)"*. The CC then concludes that its analysis *"showed that there is a causal relationship between self-pay prices and local concentration" and that "[p]rivate hospital operators, on average, currently charge somewhat higher prices in local areas where they face fewer competitive constraints"*.[4] In this report we comment on the CC's PCA.

2.2 The CC published an earlier version of its PCA in a working paper in March 2013 ("PCA1").[5] The CC then organised a Data Room at its offices to provide the parties' economic advisers access to the underlying data and analysis. Compass Lexecon was instructed by BMI to consider the robustness of the CC's PCA1 results on the basis of the information provided in the Data Room and in the PCA working paper.

2.3 Following the Data Room process, we submitted a report to the CC in May 2013 setting out our comments on the CC's PCA. We concluded that "our review of the CC's analysis suggests that the CC's baseline model is not well specified, the finding of a market wide statistically significant relationship between price and concentration is not in fact robust and the CC has done little of the substantive work that would be required to convincingly come to a view that higher prices are actually caused by high concentration".[6] We do not consider that the CC has fully addressed our criticisms of PCA1 in its revised analysis and accordingly will cross-refer to our earlier submission where necessary.

---

3 Competition Commission, "Private healthcare market investigation: Provisional findings report", 28 August 2013, hereafter "Provisional Findings".

4 Provisional Findings, paragraph 36.

5 Competition Commission, "Price-concentration analysis for self-pay patients", published on 28 March 2013.

6 P. Davis, S. Holbrook and E. Langer, "Comments on the Competition Commission's Price Concentration Analysis", 20 May 2013, paragraph 2.2.

2.4 The CC's revised PCA analysis is set out in Section 6 of the Provisional Findings ("Market outcomes self-pay prices", starting at paragraph 6.190) and in Appendix 6.9 (titled "Price-concentration analysis for self-pay patients"). We denote this revised analysis "PCA2".

2.5 The CC organised a second Data Room which took place at its offices in October and November 2013. This report includes the results of analyses performed in the Data Room. We note that a number of results presented in the Data Room differ from those published in Appendix 6.9; we understand that the CC corrected an error in its analysis after publication of the Provisional Findings, although we have not been able to examine whether this change underlies all of the differences. In addition, we note that certain underlying data was not provided in the Data Room, which impeded our ability to examine the CC's data cleaning, data processing, and construction of the dataset underlying the PCA.

2.6 This report provides detailed comments on PCA2, specifically:

- In Section 3 we note that there is a significant contrast between the actual scope of the CC's PCA2 – which relates to four inpatient treatments for self-pay patients – and the conclusions drawn from the PCA in the rest of the Provisional Findings – which relate at times to all inpatient treatments for self-pay patients and indeed are further extrapolated at times to all self-pay patients (including day-case and outpatient) and even insured patients.
- In Section 4 we describe the changes in the CC's baseline specification, comparing the baseline specification in PCA1 with that now used in PCA2. We note that the CC has reduced the number of focal treatments used to estimate the PCA; changed the LOCI measure; and changed the way in which it controls for local demand and cost conditions. We find in particular that the variables used to control for local demand and cost conditions in the CC's preferred specification are likely to be inadequate.
- In Section 5 we consider the degree to which the CC's baseline specifications in PCA2 can distinguish price variation that is problematic (due to concentration) from that which is not problematic (due to factors such as variation in local demand and cost factors). We find that the CC's preferred specification does not adequately control for local demand and cost factors; when more reasonable controls are used, no effect of concentration on prices is found.
- In Section 6, we explain that the CC's dataset contains significant errors. We find that certain variables have significant amounts of missing data, which artificially restricts the estimation sample, and that the cost variable used by the CC is incorrectly constructed. We correct these errors and find no evidence of a statistically significant relationship between price and concentration in the CC's OLS regressions.

- In Section 7, we consider the CC's use of instrumental variables based on distance and insured LOCI. Using a conventional approach to such questions, we find that the CC's instrumental variables are invalid, rendering the results unreliable. The use of instruments based on distance to the nearest rival hospital or to the nearest hospital under common ownership directly conflicts with standard economic models and with the CC's own statements elsewhere in the Provisional Findings. The instrument based on insured LOCI is also likely to be invalid since uncontrolled-for demand and cost conditions are very likely to be common between the insured and self-pay markets.
- In Section 8 we consider the CC's tests of its instrumental variables specification. We note that whilst these tests are superficially reassuring, in fact they do not provide any reassurance that the IV results are valid.
- In Section 9, we consider the CC's position on whether the CC should allow for different LOCI coefficients by operator and/or by treatment. We note that the standard approach is to test the imposed parameter restrictions, which the CC does not do. We find that the parameter restrictions imposed by the CC are rejected by the data (as they were in PCA1). We also note that, for similar reasons, the CC's interpretation of the solus analysis submitted by Compass Lexecon on behalf of BMI is incorrect. Contrary to the CC's assertion in Provisional Findings, the results are not consistent with those of the CC's PCA. In particular, there is no indication of a general effect of market concentration on prices – or even one which is systematic across solus hospitals. Rather statistical tests indicate that the data forcefully reject such an assertion. (As an aside, we note also that to our knowledge the CC has never requested any of the underlying data or computer codes to properly test or consider this evidence. The CC has been unwilling to allow us to present the results of our investigations to even the economics team at the CC working on the case, contrary to the CCs approach in some other cases. Indeed the CC has also never considered most of the substantive results in the paper – for example those indicating that solus hospitals tend to have lower nearby populations, lower capacity utilization, and lower margins.)
- In Section 10, we consider the CC's refusal to recognise that the consequence of its L3, FC3 and L7 specifications failing standard RESET tests is that these specifications are misspecified and are likely to suffer from misspecification either in functional form or (most likely) from omitted variable bias. We note that the CC has used RESET in previous investigations. We can find no basis for the CC's expressed concern about the use of these tests in the statistics or econometrics literature. Rather we find the opposite – that published academic work considers that the RESET tests basically work well and should be used. In fact, the only applicable concern expressed in the literature that we can find is that RESET will not always detect problems in a regression specification even if there are, in fact, omitted variables. This is a concern a RESET test will pass too easily – i.e., the <u>opposite</u> from the CCs stated concern that the test is too stringent and may reject a well specified model.

Section 3

The actual scope of the CC's PCA findings

3.1 The CC reports its conclusions in paragraph 6.290 of the Provisional Findings that "evidence from the PCA […] found a relationship between price and concentration for self-pay patients". Subsequently the CC relies on its PCA findings in coming to the view that there is market power derived from high concentration when hospital operators compete for insured patients: "Overall, on the basis of the considerations in 6.290 to 6.292, in relation to insured patients, [the CC] therefore concluded that HCA, BMI and Spire, have market power in negotiations with PMIs arising from high concentration and an insufficiency of competitive constraints at the local level" (paragraph 6.293).

3.2 Thus the CC seeks to expand its PCA self-pay findings from the actual four self-pay treatments it covers: first to provide a general conclusion with regard to self-pay patients, and then further to a general conclusion about where the market power rests in national negotiations with PMIs. A significant question is therefore whether any of this "extension" of the PCA results is reasonable.

3.3 In the PCA, the CC studies self-pay episode prices for four inpatient treatments: hip replacement (CCSD code W3712), knee replacement (W4210), prostate resection (M6530), and gallbladder removal (J1830). The CC reports these are the top four treatments by self-pay inpatient numbers in CC's sample dataset.[7] Specifically, the CC reports these four treatments correspond to 57% of self-pay inpatient visits and 64% of self-pay inpatient revenues in CC's dataset, amounting to revenue of £91,207,964 over the period 2009-2012.[8]

7 Appendix 6.9, paragraph 22.

8 Figures taken from the version of Table 2 provided in the Data Room, which differs from that presented in the published Provisional Findings Appendix 6.9.

3.4 To view this in context, consider that the [✂] BMI episodes within the data sample on which the PCA is based (amounting to [✂]) relates to less than 0.8% of BMI's total net revenue of [✂] over the period FY2009-2012.[9] We also note that the value of revenue considered by the CC over the four-year period 2009-2012 (£91.2m) is less than the CC's initial estimate of consumer detriment (based on the alleged excess profitability of BMI, HCA, and Spire) of "between £173 million and £193 million a year".[10].

3.5 Our view is that the CC has not undertaken the work required to answer the question of whether an extension to draw wider conclusions is justified and so has no reasonable evidential basis to do so. In particular, the CC has no reasonable evidential basis to make any findings or draw conclusions based on the PCA about the price-concentration relationship for either (i) other treatments relating to self-pay inpatients, self-pay day-cases or self-pay outpatients, or (ii) insured patients. While the CC clearly seeks in Provisional Findings to claim generality in its results, it undertakes none of the tests or analysis required to support such a claim.

3.6 In particular we note that Table A1[11] shows that:

a. with respect to self-pay patients, the CC had access to data on total of 3.89 million self-pay patient episode prices from across the five main hospital operators, and within this around 1.4 million self-pay patient episodes from BMI.

b. with respect to insured patients, the CC had access to data from a total of just over 14.5 million patient episodes.

3.7 In contrast the CC's PCA in the Provisional Findings uses a maximum of 12,304 patient episodes (and fewer in most parts), corresponding to 0.3% of the total number of self-pay patient episodes on which the five main hospital operators provided data.

---

9 The [✂] episodes [✂] relates to the full sample used by L1/FC1. L3 and subsequent specifications use only a subset of this data (10,874 observations in total), amounting to [✂] BMI episodes and [✂] BMI revenue (less than [✂] of BMI's total net revenue). BMI total net revenue figures are for Acute ex-NUK; BMI Response to the Financial Questionnaire, Annex 4, Q4, plus FY 2012 revenue of [✂].

10 Competition Commission, Private healthcare market investigation: Notice of possible remedies under Rule 11 of the Competition Commission Rules of Procedure, 28 August 2013.

11 Appendix 6.9, page 48.

3.8 As the CC describes, its PCA excludes "outpatient or day-case episodes, episodes relating to specialities outside of the 16 specialities and oncology, episodes for non-acute treatments, episodes outside of the period 2009-2012, and episodes at hospitals outside of the 219 selected hospitals", referring to such exclusions as "irrelevant data".[12] While it will always be the case that some data cleaning is required to perform analyses, it is important to note that the majority of this difference does not arise because the CC is cleaning the data, but rather because the CC considers the vast majority of the data on both the self-pay and insured pricing side "irrelevant data" for its analysis. For example, the CC considers 13.9 million observations in the Healthcode data, and a little over 3.0 million observations relating to self-pay patient episodes, to be "irrelevant data".

3.9 The CC accepts at paragraph 4 of Appendix 6.9 that "the heterogeneous nature of the private healthcare industry – in treatments, in providers, and in regions – suggests that there are likely to be differences and nuances in the price–concentration relationship for particular segments of the industry" and yet the CC goes on to argue in the same paragraph that "these are not of direct interest". We agree that there is no reason that the CC should focus on "nuance" – but believe that it is not at all clear that such differences are indeed "nuances". We certainly disagree that the CC can make such a statement legitimately from the outset. (We discuss in Section 9 whether the CC can safely ignore differences in the price-concentration relationship across the five operators and four treatments actually studied by the CC.)

3.10 In our view, it is inappropriate to decide that differences are in fact "nuances" early on without studying whether they are more than nuances. It is also, in our view, equally inappropriate to decide early on in an investigation that data is "irrelevant" and then, instead of performing the analysis using the data collected, subsequently argue "by analogy" that the problems the CC believes it finds with respect to four treatments performed for self-pay inpatients can be presumed to apply far more generally.

3.11 To give just one illustration, the CC notes in its findings in respect of insurer bargaining that the range and nature of guided referrals and service line tendering is increasing.[13] However, the CC has not considered whether those treatments that are excluded from the PCA (i.e. treatments not in the top four self-pay inpatient treatments) are particularly susceptible to such attempts at directionality. For example, at paragraph 215 of Appendix 6.11 the CC explains that Bupa has a service line network for cataract surgery, but cataracts are one of the treatments not included in the CC's PCA in Provisional Findings. In extrapolating the results of the PCA across other treatments and to insured patients, the CC does not consider whether such service line tendering would affect local competition.

---

12 Appendix 6.9, Annex A, paragraph 5.

13 See Appendix 6.11, paragraphs 170-208 and 209-220 respectively.

Section 4

Changes to the CC's baseline model

4.1 The CC published its initial PCA (PCA1) in a working paper in March 2013.[14] The CC organised a Data Room at its offices to enable the parties' economic advisers to replicate the CC's economic analyses in the PCA working paper in order to test the robustness of the results. Compass Lexecon was instructed by BMI to consider the robustness of the CC's PCA results on the basis of the information provided in the Data Room and in the PCA working paper.

4.2 Following the Data Room process, we submitted a report to the CC in May 2013 setting out our comments on the CC's PCA1. We concluded that "our review of the CC's analysis suggests that the CC's baseline model is not well specified, the finding of a market wide statistically significant relationship between price and concentration is not in fact robust and the CC has done little of the substantive work that would be required to convincingly come to a view that higher prices are actually caused by high concentration".[15]

4.3 The CC's revised PCA (PCA2) contains several very significant revisions. These include:

i. Changes to the cleaning process generating a smaller number of observations in the "cleaned hospital dataset";

ii. basing the PCA on four focal treatments instead of the previous eight, reducing the number of patient episodes included in the regression by around 50%;

iii. replacing the LOCI concentration measure with one based on self-pay data instead of insured data;

iv. limiting the regional indicator variables to reflect much larger geographic areas, allowing regional differences to exist across just 11 areas instead of 34; and

v. introducing five wholly new control variables into the PCA.

4.4 In this section we consider in principle the desirability of each of these changes to the PCA.

---

14 Competition Commission, "Price-concentration analysis for self-pay patients", published on 28 March 2013.

15 P. Davis, S. Holbrook and E. Langer, "Comments on the Competition Commission's Price Concentration Analysis", 20 May 2013, Executive summary, paragraph 2.2.

The CC's revised data cleaning process

4.5 As noted above, the CC uses only a very small subset of the data provided to it by the hospital operators and Healthcode. Whilst most of the data is discarded due to the CC deeming it "irrelevant", the CC also discards over 80% of observations in the cleaned hospital dataset (moving from 119,101 observations to 21,406 observations).[16] The 21,406 observations for all treatments in the cleaned dataset represents a significant decrease from the equivalent figure in the March working paper (46,681).

4.6 The files required to examine this cleaning process have not been provided in the Data Room.[17] Accordingly, we cannot comment on the merits or otherwise of this data cleaning process.

The four focal treatments

4.7 The CC has moved from considering eight focal treatments in PCA1 to considering only four in PCA2. The CC does not explain why it has made this change. At paragraph 22, the CC seeks to justify its choice of focussing on four treatments but, in fact, only presents arguments for focussing on a relatively small number of treatments: not on just four as opposed to (say) five, six, seven or eight treatments.

4.8 In addition, the CC does not examine whether its results are sensitive to this choice of four treatments. Nor does the CC present any reasoning why competitive conditions for these four treatments should be representative more generally. Significant robustness checks and convincing reasoning would be required to conclude that, even if the PCA were representative of competitive conditions for the four treatments considered, the PCA is representative of competitive conditions more generally.

LOCI concentration measure

4.9 The CC has changed from using a LOCI measure calculated using insured patient episodes to a LOCI measure calculated using self-pay patient episodes. If the CC is to rely upon its LOCI concentration index, then we agree this is a sensible change since self-pay prices are now related to the self-pay LOCI (the variable of interest) rather than a potential proxy for the variable of interest (the insured LOCI).

4.10 Of course, this does not overcome the fundamental problems with using the LOCI measure of concentration as an indicator of market power:

---

16 Annex A to Appendix 6.9, paragraphs 7 and 11.

17 In particular, create_data_selfpay.do and create_data_pca.do cannot be run in the Data Room.

- LOCI is a non-standard measure of market share, in the sense that it does not relate to a conventional market share calculated for a well-defined geographic or product market (except to the extent that it implicitly takes the area from which the hospital draws patients at a particular moment in time as its market definition);
- LOCI by construction systematically down-weights exactly the consumers who are most likely to indicate the presence of a competitive constraint across hospitals – those in areas where significant numbers of patients choose another competing hospital. The CC acknowledges, that "the weighting scheme implicit in the LOCI measure typically assigns more weight to a hospital's share of patients in local areas that are nearby [to the hospital]".[18]

4.11 The change in concentration measure from insured LOCI to self-pay LOCI does not change the fact that the CC should, properly, specify a model which relates price levels to (i) market concentration and (ii) other determinants of prices: in particular, factors other than concentration which would indicate that a particular hospital has more (or less) market power so that all else equal it would be expected to have higher (or lower) self-pay prices. We develop this important point further in Section 7 below.

Regional indicator variables

4.12 The CC uses regional indicator variables based on NUTS regions. The NUTS system of territorial classification was created by Eurostat as a single hierarchical classification of spatial units used across the EU. The hierarchy of three NUTS levels is as follows:

- NUTS1 is described as "major socio economic regions" and divides the UK into 12 very broad regions;
- NUTS2 is described as "basic regions for the application of regional policies" and divides the UK into 37 regions;[19] and
- NUTS3 is described (albeit from the point of view of European wide socio-economic analysis) as "small regions for specific diagnoses", with 139 regions in the UK.[20]

4.13 In PCA1 the CC's analysis focussed on specifications using the NUTS2 regional indicator variables, while in PCA2 the CC uses two sets of regional indicator variables depending on its specification:

---

18 See paragraph 18, Appendix 6.4, Provisional Findings.

19 The hospitals in the PCA dataset are located in 34 of these 37 regions: footnote 28 of Appendix 6.9 states that "NUTS1 contains 11 categories and NUTS2 contains 34 categories".

20 See http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction and http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/eurostat/index.html

- some specifications include the so-called "NUTS2" regional indicator variables; and
- other specifications include the so-called "NUTS1" regional indicator variables (in these specifications the four new "local area characteristics" that are discussed further below are also included).

4.14 To make the difference concrete in the UK context, a region in NUTS1 is for instance "South East (England)" while NUTS2 would divide this area into the four large areas:

- Berkshire, Buckinghamshire and Oxfordshire;
- Surrey, East and West Sussex;
- Hampshire and Isle of Wight; and
- Kent.

4.15 The NUTS3 level then divides each NUTS2 region into smaller regions. For instance, the NUTS2 region "Berkshire, Buckinghamshire and Oxfordshire" is further divided into four NUTS3 regions:

- Berkshire;
- Milton Keynes;
- Buckinghamshire CC; and
- Oxfordshire.[21]

4.16 To understand the role these different indicator variables play in the CC's price-concentration analysis, it is worth noting that when controlling for regional variation at the NUTS2 level in PCA1, the CC introduced a total of 34 minus 1[22] = 33 regional indicator variables into its model of self-pay inpatient prices. Doing so allowed the average prices that patients paid for their treatment to be different according to the 34 NUTS2 regions in which they were treated. In contrast, the CC's new baseline specification in PCA2 only allows for NUTS1 regional indicator variables so that the CC has introduced a total of just 11 minus 1 = 10 regional indicator variables into its model. One way to describe this difference is that it has dropped a total of 23 variables from its analysis – variables that were previously deemed to be important.

---

21 See http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/eurostat/south-east--england-/index.html

22 The "minus 1" occurs because one region must always be taken to be the baseline in the regression analysis.

4.17 The difference between these specifications is potentially significant not only in terms of simply the "number" of control variables included in the CC's specification, but also more generally in terms of whether the CC is successfully addressing standard concerns which must be addressed in order for the results to be reliable.

4.18 In PCA1 the CC argued that it did not need to worry about issues such as endogeneity bias because it was of the view that its inclusion of NUTS2 regional variables was sufficient to address such concerns.[23] We did not agree in our previous submission on PCA1 (for reasons we discuss further below), instead we argued the CC should worry more about the issue, not less.[24] Paragraph 36 of Appendix 6.9 indicates the CC now shares that view and accordingly in PCA2, the CC has now more explicitly recognized the issue that demand and cost differences across hospitals must be controlled for properly, recognizing for example that the issue can lead to estimation results that suffer from bias.[25]

4.19 However, the CC's choice to include regional indicator variables at only the NUTS1 level in its baseline model specification has gone in the opposite direction – actively reducing the extent to which the CC's model allows for price variation across regions. Specifically, in PCA2 the CC has moved to using just 10 NUTS1 regional indicator variables in its baseline specifications instead of 33 in PCA1.[26]

4.20 Perhaps in order to compensate for the removal of 23 regional indicator variables, the CC has simultaneously introduced four new control variables in PCA2. We next introduce those variables and show that they should not be expected to properly compensate for variation across hospitals in demand or cost conditions. In Section 5 we also show that these newly introduced control variables are, in fact, demonstrably failing to do a good job of controlling for variation in regional prices in the CC's PCA2 baseline specification.

The local area characteristics variables

4.21 The final change we discuss in this section relates to the four new variables the CC introduces into its price-concentration regression specification that are collectively described under the heading of "local area characteristics".

---

23 See paragraph 46 of PCA1.

24 P. Davis, S. Holbrook and E. Langer, "Comments on the Competition Commission's Price Concentration Analysis", 20 May 2013, paragraph 2.10 et seq.

25 Specifically, the CC states in paragraph 46 of Appendix 6.9 that "depending on the nature of the endogeneity – the cause, the interrelationship between price and the covariates, and the degree of endogeneity – the resulting bias may be upwards, downwards or of a negligible magnitude."

26 The CC prefers specifications L3 and FC3 of the OLS regressions, and uses these specifications for IV estimation: see paragraph 47 of Appendix 6.9.

4.22 The local area characteristics variables record the population average age, population average Gross Household Disposable Income, population density, and the average NHS waiting time at the NUTS3 level. That is, these variables are constructed as averages across NUTS3 regions, so that the local area characteristics variables take the same values for all hospitals situated in (say) Berkshire. The CC then includes these averages as explanatory factors for local price levels for individual hospitals.

4.23 The question is whether such an approach is realistically likely to capture local demand and cost conditions facing individual hospitals (and which may differ in a manner that affects prices for reasons unconnected to market power). We believe that this is unlikely to be the case, for the simple reason that the local area characteristics variables do not adequately capture the local market characteristics facing each individual hospital.

4.24 First notice that many NUTS3 regions, whilst more disaggregate than NUTS1 or NUTS2 regions, still cover large areas and so are unlikely to represent well the local demand and supply conditions facing an individual hospital.[27] Consider a simple example. Both BMI Princess Margaret (in Windsor) and Spire Dunedin (in Reading) lie within the NUTS3 region "Berkshire". Therefore, the CC's proposition is that the local demand conditions facing both hospitals can safely be assumed to be the same, as the "Average Age (NUTS3)" variable and the other local area characteristics variable will take the exact same values for both BMI Princess Margaret and Spire Dunedin.[28] Clearly, this is not likely to be the case.

4.25 Given that the CC's NUTS1 regional variable would clearly take on the same value for both hospitals, the only way that the CC's model could explain a difference in average price between Spire's Dunedin in Reading and BMI's Princess Margaret in Windsor would be on the basis of other variables included in the model. Yet none of the other variables included in the model relate to demand conditions at the individual hospital. Whilst the CC states[29] that "the intended role of the regional dummies and the local area characteristics is primarily to control for differences in local levels of self-pay demand," in fact, the CC constructs its model so that differences in hospital prices are very unlikely to be explained by either its regional variables or its included local area characteristics variables, since neither are related to demand conditions that are, in fact, local to hospitals.

---

27 NUTS3 regions vary in area from under 50 sq km (such as UKD42 Blackpool and UKH21 Luton) to over 10,000 sq km (such as UKM61 Caithness & Sutherland and Ross & Cromarty and UKM63 Lochaber, Skye & Lochalsh, Arran & Cumbrae and Argyll & Bute). Source: Eurostat, table demo_r_d3area.

28 Of course the NUTS1 dummy will not vary between these two hospitals either.

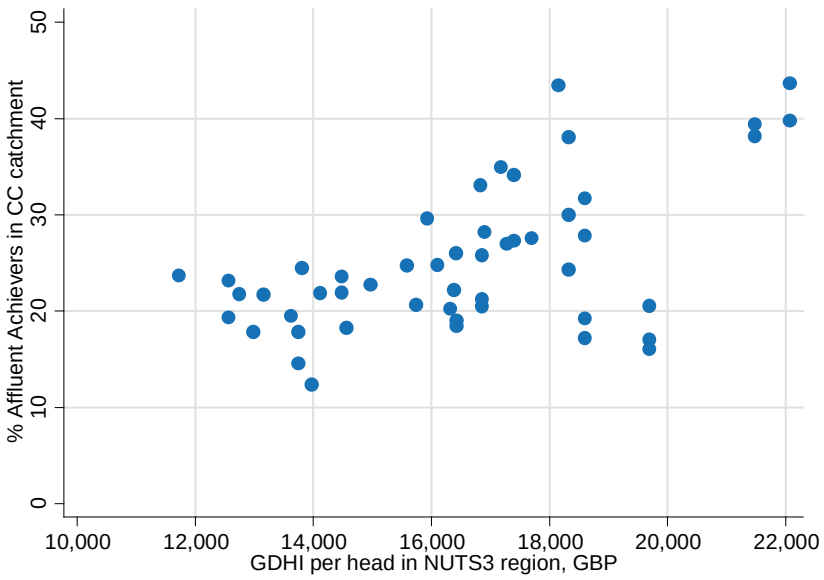
29 Appendix 6.9, paragraph 35.

4.26 A second way to consider whether the CC's NUTS3 variables adequately capture local demand conditions is to compare those variables to similar variables computed for the hospitals' individual catchment areas (as defined by the CC in terms of drive distance). Annex A presents a comparison for BMI hospitals[30] of population density and average age between those hospitals' catchment areas and the NUTS3 region in which they are located.

4.27 Given that GDHI is not available at more disaggregate geographies than NUTS3, we examined ACORN consumer classifications, which are used by BMI in assessing consumer characteristics.

4.28 BMI's analysis using CACI's Insite system shows that ACORN Category 1, "Affluent Achievers", has a much higher propensity to purchase self-pay treatments than other ACORN categories; this category is more than twice as likely as the population as a whole to have a self pay episode.[31] Figure 1 compares the GDHI within NUTS3 regions to the percentage of "Affluent Achievers" within each hospital's catchment areas. It is clear that GDHI at the NUTS3 level is not a good proxy for ACORN Category 1 within the hospitals' catchment areas; the R-squared is just 0.33.

Figure 1: Affluent Achievers in hospital catchment and GDHI in NUTS3 region



Source: Compass Lexecon analysis. See Annex A for further details.

30 We only did this for BMI hospitals as the catchment areas for non-BMI hospitals are only available in the Data Room or in a narrow confidentiality ring. The analysis covers 50 BMI hospitals; some hospitals are excluded because of data availability issues. See Annex A for full details.

31 Source: ACORN profile report for FY12 self-pay inpatient & daycase episodes.

4.29 Given these considerations, it is clear that the local area characteristics variables do not adequately capture the demand characteristics of each hospital's catchment area, and are unlikely to form adequate controls. Indeed, it is entirely unclear why the CC has chosen to compute "local area characteristics variables" at the NUTS3 level. Relevant "local area characteristics" variables should in principle reflect the local demand and supply conditions facing each individual hospital. Given that the data to construct relevant measures for hospitals' catchment areas is publicly available or could be readily supplied by the parties, the CC could (should) have brought that data into its analysis.

4.30 To put it even more simply, nowhere in the Provisional Findings other than in its PCA2 does the CC contend that a relevant local market is either "Berkshire" (NUTS3) or "Berkshire, Buckinghamshire and Oxfordshire" (NUTS2) or "South East (England)" (NUTS1). Yet the CC contends that controlling for indicators of "local demand conditions" in "South East (England)" and/or "Berkshire" is sufficient for its analysis to address the concern that local demand conditions vary between hospitals. We consider that the CC's analysis is unconvincing in this regard.

Section 5

The CC's baseline OLS results

5.1 In this section we comment on the CC's baseline PCA2 model, the results of which are included in Table 4 and Table 5 of Appendix 6.9.

5.2 The CC's PCA2 attempts to establish whether there is a correlation between price and concentration controlling for other potential determinants of prices charged to self-pay patients. Since it is rightly accepted by the CC that hospitals' prices may vary with respect to local demand and cost conditions, the CC accepts it should control for factors which may vary across hospitals in order to make a "like-for-like" comparison across markets. Put simply, suppose we find that market A is both more concentrated and has higher prices than market B. A naïve analysis would conclude there is a positive association between price and concentration. However a proper analysis would consider whether such a finding was misleading – as it would be if for example, in reality, the costs of doing business in market A were simply higher.

5.3 The PCA therefore seeks to control for factors which may cause prices to vary across hospitals but which are unrelated to concentration.

5.4 Unfortunately, our consideration of the CC's analysis makes clear that the CC's model cannot distinguish problematic variation in prices (relatively higher prices due to high concentration) and non-problematic variation in prices (relatively higher prices due to local differences in demand and cost conditions).

5.5 This section proceeds in three steps:

- First, we show that the CC's choice of whether and how to control for regional variation significantly affects its conclusions.
- Second, having previously established in Section 4 that the NUTS1 and local area characteristics variables are highly likely to be inadequate controls for local demand and cost conditions, we next show that the CC's results indicate this is the case. Specifically, the CC's local area characteristics are all statistically insignificant and moreover have entirely implausible or counterintuitive signs.

- Third, we explain that the CC ultimately faces a choice between ascribing variation in prices to its concentration measure or to local conditions. It cannot distinguish between the two with the available data. The CC ultimately *assumes* in PCA2 that variation in prices is to be explained by its concentration measure, not by variation in local conditions. The CC's baseline model simply cannot tell apart problematic variation in prices (due to concentration) and non-problematic variation in prices (due to local differences in demand and cost conditions). Any dispassionate review of the evidence would properly conclude that, in fact, the PCA does not show a relationship between price and concentration.

5.6 In light of these findings, we do not believe that the CC's baseline econometric analysis can reasonably be considered to provide reliable evidence of an effect of concentration on self-pay prices for even the four inpatient treatments included in the PCA.

The CC's baseline results

5.7 Table 1 below summarises the results of the CC's baseline (OLS) regression analysis reported in Tables 4 and 5 of Appendix 6.9 of the PFs.

5.8 The CC reports results from three different specifications for each of its two measures of concentration (i) LOCI and (ii) fascia counts. The LOCI specifications are numbered with the prefix "L", so L1, L2 and L3 while the fascia count specifications are numbered with the prefix "FC", so FC1, FC2 and FC3. With the exception of the different concentration measure used, L1 is a specification with the same variables as FC1, L2 the same variables as FC2 and so on. Of course L1/FC1 are different from L2/FC2 in other respects.

5.9 In particular, note that for each type of concentration measure:

- The first specification, FC1/L1, does not include any regional dummies or local area characteristic variables.
- The second specification, FC2/L2, includes regional dummies at the NUTS2 level.
- The third specification, FC3/L3 includes regional dummies at the NUTS1 level as well as what the CC terms "local area characteristic variables", despite being more accurately described as "NUTS3 region characteristic variables".

5.10 The results show that the CC does not find any statistically significant results using 95% level of significance in either Table 4 (using self-pay LOCI as the concentration variable) or Table 5 (using facia count as the concentration variable). However, the final LOCI specification L3 is (marginally) statistically significant at the weaker 90% level.

5.11 Given that the CC's specifications FC1 and L1 do not find any statistically significant effect of fascia count LOCI on self-pay prices, and that those specifications do not control for local demand conditions, we do not discuss them further here. Instead we focus our discussion in turn on the results presented in L2 and L3.[32]

[32] Many of our comments however apply equally to FC2 and FC3 where the CC's results do not indicate any statistically significant effect of its concentration variable on prices.

Table 1: Summary of the CC's results

	CC's Table 5			CC's Table 4		
	FC1	FC2	FC3	L1	L2	L3
Year Dummies						
2010	0.0123*	0.0135**	0.0139*	0.0127**	0.0134**	0.0125*
	(0.0063)	(0.0060)	(0.0072)	(0.0063)	(0.0060)	(0.0070)
2011	0.0524***	0.0553***	0.0559***	0.0529***	0.0552***	0.0549***
	(0.0080)	(0.0070)	(0.0082)	(0.0081)	(0.0071)	(0.0081)
2012	0.0684***	0.0753***	0.0631	0.0685***	0.0755***	0.0639
	(0.0090)	(0.0256)	(0.0462)	(0.0091)	(0.0269)	(0.0476)
Operator Dummies						
HCA	n/a	n/a	n/a	n/a	n/a	n/a
Nuffield	n/a	n/a	n/a	n/a	n/a	n/a
Ramsay	n/a	n/a	n/a	n/a	n/a	n/a
Spire	n/a	n/a	n/a	n/a	n/a	n/a
Treatment Dummies						
Prostate resection	0.0500***	0.0482***	0.0576***	0.0492***	0.0470***	0.0573***
	(0.0127)	(0.0123)	(0.0128)	(0.0129)	(0.0124)	(0.0128)
Hip replacement	0.8595***	0.8624***	0.8733***	0.8604***	0.8631***	0.8744***
	(0.0153)	(0.0134)	(0.0143)	(0.0155)	(0.0135)	(0.0141)
Knee replacement	0.9371***	0.9403***	0.9478***	0.9377***	0.9406***	0.9493***
	(0.0173)	(0.0158)	(0.0170)	(0.0175)	(0.0160)	(0.0168)
Patient Characteristics						
Patient sex	–0.0076**	–0.0071**	–0.0097***	–0.0078**	–0.0076**	–0.0099***
	(0.0034)	(0.0029)	(0.0034)	(0.0035)	(0.0030)	(0.0034)
Patient age	–0.0003*	–0.0003*	–0.0003*	–0.0003*	–0.0003*	–0.0003*
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
Episode number of patient nights	0.0055***	0.0053***	0.0051**	0.0054***	0.0053***	0.0050**
	(0.0020)	(0.0018)	(0.0021)	(0.0020)	(0.0019)	(0.0021)
Concentration variables						
Self-pay LOCI				–0.0437	–0.0612	–0.0943*
				(0.0387)	(0.0450)	(0.0540)
Fascia Count (0-9) miles	–0.0088	–0.0097	–0.003			
	(0.0081)	(0.0082)	(0.0093)			
Fascia Count (9-17) miles	0.0011	0.0024	0.0035			
	(0.0021)	(0.0025)	(0.0028)			
Fascia Count (17-26) miles	0.0004	0.0024	0.0011			
	(0.0020)	(0.0025)	(0.0029)			
Hospital Characteristics						
ln(average direct costs)		–0.0014	0.0033		–0.0016	0.0024
		(0.0074)	(0.0140)		(0.0080)	(0.0144)
CCL3 provision dummy		0.0583***	0.0529*		0.0536**	0.0500*
		(0.0220)	(0.0277)		(0.0215)	(0.0269)
Local area characteristic variables						
Average Age (NUTS3)			0.0025			–0.0021
			(0.0052)			(0.0056)
Average GDHI (NUTS3)			–0.0036			–0.0004
			(0.0052)			(0.0047)
Average NHS wait (NUTS3)			-0.0000			-0.0000
			(0.0004)			(0.0004)
Average population density (NUTS3)			–0.0009			–0.001
			(0.0009)			(0.0009)
Regional dummy variables		**NUTS2**	**NUTS1**		**NUTS2**	**NUTS1**

*Source: Table 4 and Table 5 of Appendix 6.9. n/a implies that the results are not presented in the non-confidential version of the Provisional Findings. ***/**/* indicates statistical significance at the 1%/5%/10% level.*

5.12 Notice that the CC's specification L2 includes regional indicator dummies for 34 NUTS2 regions. That is, in specification L2, local prices are allowed to differ between these 34 subsets of hospitals located in the respective NUTS2 regions. The results show that allowing even for only broad regional differences in price levels at the NUTS2 level, the CC does not find a statistically significant correlation between concentration and prices. That is, when the CC estimates the specification which is analogous to the specification it has now discarded in the PCA1, it does not report in PCA2 a statistically significant effect of concentration on price – even measured at the 90% level of statistical significance.

5.13 In contrast, the CC's specification L3 includes regional indicator dummies for only the 11 NUTS1 regions – albeit along with the four new local area characteristic variables – and does find a statistically significant effect.

5.14 The contrast makes clear that the inclusion (or otherwise) of the regional indicator variables leads to a difference in the conclusion drawn from the CC's baseline specification. In L2 a conventional approach would conclude "no proven case" while the CC emphasises instead the statistically significant results in L3. Indeed, the CC then relies on specifications that include only the NUTS1 regional indicator variables and the four new NUTS3 regional area characteristic variables throughout the rest of its PCA2 analysis. The question is then whether the CC's preference for L3 over L2 is justified. We analyse below whether the NUTS3 regional area characteristic variables provide adequate control for local demand and cost conditions. We also examine the role of cost variables in the CC's regressions. In light of our findings, we then consider whether the CC's preference for L3 is appropriate.

The CC's results show that their explanatory variables do not achieve their aim of controlling for local demand and cost conditions

5.15 As we noted above, self-pay prices can vary across hospitals for a variety of reasons, one of which may be concentration. The PCA performed by the CC has sought to "evaluate the relationship between price and concentration […] while accounting for other factors so that a like-for-like comparison is achieved."[33] These other factors, for which the PCA should control, include differences in local demand and cost conditions.

5.16 The CC correctly recognises that price variation is not necessarily problematic and prices will vary, for example, due to differences in demand factors. It states that this is the reason it includes regional dummies and local area characteristic variables in its regressions:

[33] Appendix 6.9, paragraph 1.

> *"The intended role of the regional dummies and the local area characteristic variables is primarily to control for differences in local levels of self-pay demand. [Footnote omitted] All else equal we would expect higher prices in areas of high demand and thus we wish to control for this effect in our analysis. Regional dummies and local area characteristic variables are two ways to do this. The regional dummies have the advantage that they will pick up any differences in self-pay demand between regions regardless of the precise economic source of these differences, and thus do not rely on our ability to measure self-pay demand factors."*[34]

5.17 The results for the CC's L3 specification make clear that the four new variables the CC adds to allow for regional differences in supply and demand conditions do not help in explaining the variation in price levels across hospitals: these factors are statistically insignificant and either all or most of the coefficients appear to have counterintuitive (or unexpected) signs.

5.18 Specifically the CC's predictions in its specification L3 for the sign of the effects in its regressions are as follows:

- An **increase** in average age in the local NUTS3 region is predicted to **decrease** self-pay hospital average prices.
- An **increase** in average disposable income (GDHI) in the local NUTS3 region is predicted to **decrease** hospital average prices.
- An **increase** in the average population density in the local NUTS3 region is predicted to **decrease** hospital average prices.

5.19 All or most of the signs of these effects appear counterintuitive. For example, should not older, richer patients end up paying more, not less (all-else-equal), since they would tend to have greater demand for private healthcare than in those areas with a lower average age/income? Similarly local population density all else equal would intuitively be associated with a higher demand for private healthcare and so higher prices. The implication is that the control variables singularly fail to control for variations in local demand and supply conditions.

5.20 Of course, the effects of all of these variables on self-pay prices are statistically insignificant. In line with CC's view on statistical significance, "the estimation results cannot reject the possibility that there is no true relationship"[35] between these variables (i.e. between average age, income, etc. and self-pay prices).

---

34 Appendix 6.9, paragraph 35.

35 Appendix 6.9 of CC's provisional findings report, paragraph 38.

5.21 The sign on the average NHS wait time in the local NUTS3 region is almost indistinguishable from zero (it is zero to four decimal places) and is highly statistically insignificant. Again, this control variable is failing to control for local demand conditions.

5.22 The counterintuitive signs and lack of statistical significance show that it is very unlikely that the CC is successfully controlling for the variation in local demand and cost conditions in its analysis, even having recognised that doing so is important. This has significant implications for the validity of the results of specification L3.[36] Indeed, removing the average age, disposable income, population density and NHS wait time at NUTS3 level from specification L3 shows that these variables explain less than 0.02% of the variation in self-pay prices (conditional on the other variables in the specification). Removing the NUTS1 dummies as well shows that these variables – supposedly accounting for all variation in local demand and cost conditions – explain 0.26% of the variation in self-pay prices, conditional on other variables in the specification.[37]

5.23 Accordingly, the results of the CC's specification L3 makes clear that absolutely none of the regional market characteristics the CC includes are actually (conditionally) correlated with the level of self-pay prices. Similar observations apply to the CC's baseline IV regressions L4 to L7 and FC4 to FC7.[38]

5.24 It follows that the NUTS3 regional market characteristics do not explain the regional variation the CC says it aims to control for. Thus in specification L3 regional variation is effectively only being controlled for by regional dummies at the NUTS1 level (i.e. very big regions with the UK being divided into just 11 NUTS1 units which cannot capture this kind of variation at the local level). As a result, variation in local prices within each NUTS1 region is likely to be driven in significant part by omitted local regional demand conditions, which may well be correlated with LOCI. The CC's estimate of the effect of concentration on self-pay price appears correspondingly likely to be subject to a very serious concern that the results are biased and therefore misleading.

---

36 Similar considerations will apply to other regressions based on the same specification as L3. We note for example that all of the local area characteristics are statistically insignificant in L4 to L7 and FC4 to FC7 and the point estimate on age is always negative.

37 See do-file "Section 5 – LACs and L2 L3 restrictions.do", provided to the CC in the Data Room.

38 These results are reported in Tables 6 and 7. The CC does not report the full results in these tables; the full results were available in "Tables for PCA Appendix.xlsx" provided in the Data Room. In particular, all of the local area characteristics variables are insignificant; the coefficient on average age is always negative; the coefficient on population density is always negative for the LOCI regressions; and the coefficient on NHS wait is, at most, -0.0002.

5.25 The CC claims that "specification L3 is preferred to L1 and L2, on the grounds that it controls for more factors than L1 but does so in a more parsimonious way than L2."[39] In our view, the CC's proposition that specification L3 controls for regional variation in demand and cost conditions in a manner which is more parsimonious to L2 is demonstrably false. We note that the rest of the CC's PCA2 analysis then proceeds on what is clearly a false premise.

5.26 Specifically, we note that, whether specification L2 or L3 is preferred can be tested statistically by testing the implicit parameter restrictions imposed in each specification. Specifications L2 and L3 can be regarded as restricted versions of a more general model which includes both NUTS2 dummies and local area characteristics variables. From that general model, L2 imposes that the parameters on the local area characteristics variables are jointly zero (since they are not included in the specification), whereas L3 imposes that the parameters on the NUTS2 dummies do not vary within NUTS1 regions (since only NUTS1 dummies are included in the specification).

5.27 In order to test the validity of these restrictions, we estimate a general model which encompasses both L2 and L3 and then test the relevant parameter restrictions. We include the four local area characteristics variables (average age, average GDHI, average NHS wait, and average population density) as covariates. We also include NUTS1 dummies and NUTS2 dummies excluding one NUTS2 dummy per NUTS1 region (recall that NUTS2 regions are nested within NUTS1 regions). To take an example, the NUTS1 region UKE is split into NUTS2 regions UKE1, UKE2, UKE3, and UKE4. We create a NUTS1 dummy for region UKE and NUTS2 dummies for regions UKE2, UKE3, and UKE4. Then a test for the joint insignificance of the coefficients on the UKE2, UKE3 and UKE4 dummies indicates the validity of the restriction that only the NUTS1 dummy is required to capture the variation observed in the data.

5.28 Table 2 presents the results. We find that the four local area characteristics dummies are jointly insignificant but the NUTS2 dummies are not. This indicates that the restrictions imposed by L2 are valid (relative to the model that encompasses both L2 and L3) but the restrictions imposed by L3 are not. Accordingly, conventional statistical tests suggest that specification L2 should be preferred over specification L3, as it imposes only valid parameter restrictions on the more general model.

Table 2: Test of restrictions imposed by L2 and L3 on the general model

Test	p-value
Four local area characteristics variables are jointly insignificant	0.4664
NUTS2 dummies are jointly insignificant	0.0000

Source: "Section 5 – LACs and L2 L3 restrictions.do", provided to the CC in the Data Room.

[39] Appendix 6.9, paragraph 42.

The CC's hospital characteristic variables

5.29 The CC also includes two variables which aim to control for variation in hospitals' costs or quality. First we note that the new added variable, Critical Care Level 3 (CCL3) – which is included presumably in an attempt to control for a potentially significant difference in hospital quality – does *prima facie* appear to be playing a significant role.

5.30 Second, the results suggest that the CC's cost measure, the log of average direct costs, does not appear to be successfully controlling for cost differences across hospitals in the CCs OLS regressions. Specifically, the CC's cost variable is not statistically significant in any of its specifications in either Table 4 or 5, which strongly suggests that the relevant treatment level cost differences are not reflected in the CC's cost measure. The same observation applies to the CC's IV regressions L4 to L7 and FC4 to FC7.[40] In FC2 and L2 the cost variable actively has the wrong sign – suggesting prices decline as costs go up – although the parameters are statistically insignificant.

5.31 In our previous submissions we have noted that the CC is examining treatment-level prices at each hospital while the cost measure is only at the hospital level. As a result, it may not be surprising that the CC's cost measure appears to be doing a poor job of controlling for cost differences across the hospitals for these specific treatments.

5.32 The CC recognises this issue at paragraph 50, stating with respect to measurement error in the cost variable that the CC "agrees that a disaggregated cost measure would be preferable if it were to be available (it is not), but we [the CC] consider that in conjunction with the CCL3 dummy and regional dummy variables, the three variables are sufficient to account for the salient cost differences between hospitals." However, it does not recognize in its PCA that the results clearly indicate that this concern is a material one.

5.33 As we noted above, the regional dummy variables and local area characteristics variables do not adequately control for local conditions, and this observation applies equally to local cost conditions as to local demand conditions.

40 These results are reported in Tables 6 and 7. The CC does not report the full results in these tables; the full results were available in "Tables for PCA Appendix.xlsx" provided in the Data Room.

The CC's approach cannot distinguish between problematic and non-problematic variation in self-pay prices

5.34 The CC faces a choice between specification L2, using regional dummies at the NUTS2 level which allow for regional variation in prices across 34 regions, specification L3 using regional dummies at the NUTS1 level, allowing for regional variation across 11 regions, in conjunction with local area characteristics variables. This choice should be based on whether the regional dummies and local area characteristics variables adequately control for local variation in demand and cost conditions.

5.35 The CC is commendably clear that:[41]

> *"….the disadvantage of including regional dummies is that, if used at a very granular level, they can absorb much or all of the useful variation in prices between hospitals, and leave no between-hospital price variation for us to evaluate against local concentration.."*

5.36 The CC however applies this argument to justify its choice of specification L2 over specification L3 – arguing that "the regional dummies are effective [in controlling for differences in self-pay demand between regions], but may not be best suited to controlling for very local differences in self-pay demand."

5.37 First, we note that neither L2 nor L3 are using regional dummies at a "very granular level". The NUTS2 regional dummies used in L2 are not obviously attempting to control for "very local differences in self pay demand." Rather the CC's L2 results make it clear that, if the CC allows self-pay prices to vary across the 34 NUTS2 regions – i.e., the CC ascribes that price variation to normal regional differences in local market conditions – then the CC's baseline results indicate there is no statistically significant effect of local concentration on self-pay prices.

5.38 Second, recall that the CC's L2 specification is in effect only allowing prices to vary for reasons other than concentration between quite large regions – for example between NUTS2 region "Berkshire, Buckinghamshire and Oxfordshire" and NUTS2 region "Surrey, East and West Sussex" rather than only between the NUTS1 regions "South East (England)" (of which both of those NUTS2 regions are part) and "South West (England)".

5.39 Given that the CC is ultimately currently contending it finds fairly small price effects – that a 20 percentage point movement in LOCI is associated with a 3% self-pay price increase for these four inpatient treatments – it is easy to imagine that accounting for NUTS2 regional variation in cost and demand conditions instead of NUTS1 could more than account for the differences in prices the CC is finding.

---

41 Appendix 6.9 paragraph 35.

5.40 Third, we note that in this quotation the CC is rightly acknowledging that in its specification L2, it cannot distinguish between the hypotheses that the observed price variation is caused by (i) concentration and (ii) regional differences in demand and supply conditions at the NUTS2 level. Thus, in order to place weight on the results of the PCA, the CC must believe that its regional NUTS1 controls are sufficient when used in conjunction with the four local area characteristics variables – which we have already shown are both statistically insignificant and (at least mainly) have implausible signs in specification L3.

5.41 The CC goes on to acknowledge it actively faces a choice over how much variation in hospital prices to attribute to LOCI and how much to attribute to variables which attempt to capture differences in local demand conditions. The CC states:

> *"In contrast, the local area characteristic variables are direct attempts to measure the economic factors that we think will proxy for local levels of self-pay demand. Relative to the regional dummies, these variables have the advantage that they can be measured at a more local level and* ***do not absorb the price variation that we wish to compare with local concentration****." (emphasis added)*

5.42 In short, specification L3 does not "absorb the price variation that [the CC] wish[es] to compare with local concentration", in contrast to specification L2. In effect the CC has made a deliberate choice not to control for much regional variation in PCA2 (in effect it has controlled for much less regional variation than in PCA1) even while it acknowledges that:[42] *"A disadvantage of this approach, however, is that the measures are likely to be only imperfect proxies and thus may not reflect all of the local differences."*

5.43 The CC makes this choice in Provisional Findings. However, conventional statistical tests reject L3 in favour of L2. As a result of the CC's deliberate choice of L3, it is finding in specification L3 that regional variation in prices is attributed to its concentration variable, LOCI.

5.44 In summary, the results show that the CC only finds an effect for LOCI in its baseline results in Table 4 when the CC deliberately chooses to very imperfectly control for regional differences in demand and cost conditions. The statistical testing indicates the CC's choice of L3 over L2 is not a valid one.

[42] Appendix 6.9, paragraph 35.

5.45 Fundamentally, we agree with the CC that it faces a difficult problem. The fact that the CC cannot distinguish the effect of concentration on price when it includes the modest set of NUTS2 regional control variables makes very clear that the CC's baseline specifications cannot in reality distinguish between the two hypotheses at issue: (i) that cost and/or local demand conditions lead to relatively higher prices in some areas than others but in a manner that is unaffected by concentration (i.e., there is unproblematic price variation), and (ii) that higher concentration leads to increased prices (i.e. there is problematic price variation).

5.46 However, the CC should in this situation acknowledge that it cannot distinguish between these two hypotheses: the PCA provides no robust evidence of a relationship between concentration and price.

Section 6

The CC's dataset contains significant errors

6.1 We examined the dataset used for the PCA for unusual or missing data and found a number of problems with the CC's dataset. In this section, we describe the missing and incorrect data in the CC's dataset, and then consider the implications of these data issues for the CC's results.[43]

Missing data

6.2 The CC's dataset contains variables with significant amounts of "missing data" whereby a value is not available for an observation or a set of observations:[44]

- The CCL3 dummy variable is missing for one hospital (BMI South Cheshire) within the sample used for the regressions, accounting for 30 observations. The information provided by BMI to the CC in respect of this hospital indicated that the hospital has an SLA with the NHS Trust in respect of CCL3 beds.[45]
- The measures of age and population density at the NUTS3 level are missing for hospitals located in Scotland (655 observations). This arises because the CC relies on census data which covers England and Wales only.
- The average NHS wait is missing for 1400 observations. The CC notes that "the NHS waiting time variable is not available for Scotland and certain NUTS3 regions in Wales (3 regions) and the East Midlands (2 regions)" (footnote 29).

43 All of the figures and results provided in this section are generated by the do-file "Section 6 – dataset cleaning.do", provided to the CC in the Data Room.

44 The figures below refer to the sample used for the regressions, i.e. for which the variable "exclude" is equal to zero.

45 See "Master hospital list 010713.csv" in the folder "Local analysis", provided by the CC in the Data Room.

6.3 It is important to note that missing data can restrict the number of observations used in a regression: where any of the variables used in the regression is missing for an observation, that observation is dropped from the regression. The missing data in these four variables accounts for the change in the number of observations between specifications L1, L2, and L3 (and similarly FC1, FC2, and FC3): introducing the CCL3 dummy into L2 results in a loss of 30 observations between L1 and L2, and the introduction of age, population density, and average NHS wait at the NUTS3 level results in the loss of 1400 observations between L2 and L3.[46]

6.4 This missing data results in a significant restriction on the sample. All of the CC's regressions with the same control variables as the L3 specification, i.e. L3, FC3 and all subsequent specifications apart from certain sensitivities in Table 12, exclude all hospitals in Scotland and in "certain regions in Wales and the East Midlands". This amounts to excluding ten hospitals from the analysis completely and excluding a further two hospitals from the analysis for certain years. Given that the CC's "main focus throughout the [PCA] appendix is [...] on the broad relationship that is representative of the industry in general" (paragraph 4), it is surprising that the vast majority of the analysis (and all of the CC's preferred specifications) is performed across a sample which excludes all hospitals in Scotland and certain other regions of the UK.

6.5 Obviously, regression analysis can only be performed over patient episodes for which sufficient data is available. However the choice whether to exclude 1400 observations (over 10% of the full sample) across twelve hospitals because the average NHS waiting time is not available should be carefully justified by reference to two considerations: first, whether suitable data or a suitable proxy is available for the full sample, and second, whether the variable is important to the analysis.

[46] For the 655 observations for which age and population density are missing, the NHS wait is also missing. Therefore there is no additional effect of the missing age and population density data over and above the missing NHS wait data.

6.6 On the first question, to our knowledge, further information is not available for the NHS wait time, although the CC may have been able to use adjacent years or regions in some cases to impute the data (similarly to the way in which cost data is imputed by the CC). On the second question, the regression results make it clear that the average NHS waiting time has no explanatory power in the CC's regressions: in L3 the estimated coefficient is zero to four decimal places and is highly statistically insignificant. Accordingly the average NHS wait time should, on the CCs analysis, be dropped from the specification and the sample thereby not unduly restricted.[47]

6.7 For the other missing data, we note that data on average age and population density at the NUTS3 level is publicly available for Scotland,[48] and that the dummy variable for CCL3 provision could be completed (although we acknowledge that a certain amount of judgement must be applied in this case). Accordingly we see no reason to exclude observations on the basis of missing data for these three variables (if the CC persists in using market characteristics measured at the NUTS3 level).

Incorrect data

6.8 We find that the cost variable used by the CC shows extremely wide variation due to incorrect construction. Whilst data for 2009, 2010, and 2011 is transformed to represent cost per patient episode, the data for 2012 is not. This variable is therefore measuring cost per patient episode in 2009, 2010 and 2011, and total cost in 2012.[49] The variable will therefore not act as a suitable control. Table 3 shows the resulting mean costs by year in the PCA dataset.

---

47 Of course, whether in truth there is an impact of NHS waiting times on the prices charged by hospitals would require a more serious analysis. In particular, the CC would need to consider whether there is sufficient variation either across time or across hospital locations in waiting times to be able to find the effect of a change in waiting times on hospital prices on a like for like basis. The CC may be finding a zero effect in its specifications because it is not sufficiently controlling for the demand and cost drivers of prices.

48 Data on population by age in NUTS3 regions is available from the General Register Office Scotland (http://www.gro-scotland.gov.uk/statistics/theme/population/estimates/special-area/nuts.html) and data on the area of NUTS3 regions is available from Eurostat, table demo_r_d3area.

49 See lines 170-174 of "create_data_pca.do", which replaces pro_dcost with a per-patient episode measure for 2009, 2010, and 2011, but not for 2012.

Table 3: Mean cost by year in PCA dataset

Year	Mean "hospital average direct cost per patient (£)" in dataset
2009	521
2010	486
2011	339
2012	8,413

Source: "Section 6 – dataset cleaning.do" provided to the CC in the Data Room, on the basis of data_pca_focal.dta.

Note: means taken by year, across observations for which exclude=0.

6.9 Moreover, the inconsistent cost data for 2012 is then used to impute values for other years. As footnote 25 of Appendix 6.9 explains, "Cost data was available for almost all hospitals in our analysis. For hospitals with missing cost data, we have imputed the data on the basis of hospitals owned by the same operator in the same region and year; if data for the desired year is not available, we use the average for the operator and region over years that are available." Accordingly, when the second method is used, the data for 2012 can enter the calculation of averages for other years. Table 4 provides an example of the result.

Table 4: Cost by year for Spire Murrayfield hospital in PCA dataset

Year	Hospital average direct cost per patient (£)
[✂]	[✂]
[✂]	[✂]
[✂]	[✂]
[✂]	[✂]

Source: "Section 6 – dataset cleaning.do", provided to the CC in the Data Room, on the basis of data_pca_focal.dta.

6.10 Given these observations, it is entirely unsurprising that the CC's cost variable is insignificant in all of its baseline (OLS and IV) regressions. In the CC's results, the cost variable is failing to control for cost variations between hospitals because it is incorrectly constructed.[50]

[50] That is not to say that once the cost variable is correctly constructed, it will necessarily control well for variations in cost or quality between hospitals.

The implications for the CC's results

6.11 In order to assess whether the CC's results are affected by the missing data and incorrect data outlined above, we constructed a dataset which corrected the errors, where possible, and ran specifications that avoided dropping significant parts of the sample.

6.12 In respect of the missing data:

- We assigned the CCL3 dummy for BMI South Cheshire to zero, given there is no CCL3 provision at the hospital.[51]
- We completed the average age (NUTS3) and population density (NUTS3) variables for Scotland on the basis of publicly available data.[52]
- We attempted to correct the cost variable, although our ability to perform this correction in the Data Room is limited by the fact that the CC did not provide us with access to the full underlying data and we were unable to run the do-file that computed the cost variable.[53] As an imperfect correction, we computed the average per patient costs using data on total costs per hospital and number of patient episodes per year, applying the CC's methodology to impute missing observations.[54] As data on the number of patient episodes in 2012 was not available to us in the Data Room, we assumed that cost per patient episode in 2012 was the same as in 2011.

6.13 Given the considerations above, we also ran specifications dropping the NHS wait time, as this variable holds no explanatory power in the CC's specifications and artificially restricts the sample.

---

51 We note that the broad conclusions are unchanged if we assign the dummy to be 1, on the basis that CCL3 provision is available under an agreement with the NHS Trust.

52 This data is available in nuts3_age_popden_GB.csv. We constructed this data outside the Data Room. Raw data and do-files that construct the data are available on request. There are also small differences in the average age and population density variables for England, as we match two Local Authority Districts that merge by name but not code (St Albans and Welwyn Hatfield: see the merges at lines 17 and 20 of create_lookup_demog.do, provided by the CC in the Data Room), and we compute average age as a weighted average by Local Authority District rather than an unweighted average (lines 25-26 of create_lookup_demog.do, provided by the CC in the Data Room).

53 The do-file "create_data_pca.do".

54 The figures on patient episodes and hospital costs are provided in the CC's files "lookup_hospitals.dta" and "lookup_cost.dta" respectively. The CC's methodology for imputing missing data is shown in "create_data_pca.do", lines 210-217, and described in footnote 25 of Appendix 6.9.

The CC's baseline OLS regressions

6.14 Table 5 presents the results of making these corrections and changes, individually and cumulatively, to the CC's OLS regressions. We present the effect of the corrections over the CC's specifications L2 and L3 (specification L1 is not affected by the corrections). In so doing, we present the effect of the corrections on specification L3 whilst still including the NHS wait time, and then separately present the corrections dropping NHS wait time from the specification. In the former case, we do not present the effect of completing the age and population density variables for Scotland, since the affected observations are in any case dropped from the regression due to the missing NHS wait time variable. It can be seen that once all of the corrections are applied (the final row of the table) the sample size is 12304 observations, i.e. the full sample used in L1 rather than the artificially restricted samples used in L2 and L3.[55]

6.15 We find that almost all of these corrections to the data or changes to the specification results in the coefficient on LOCI becoming insignificant and smaller in magnitude, as does combining these corrections and changes. The only correction which does not result in the coefficient on LOCI becoming insignificant is the isolated correction to the costs variable in specification L3.

[55] For brevity we do not present the results for FC2 and FC3. We find that the coefficients on the fascia count measures remain insignificant.

Table 5: OLS regressions with data and specification corrections

Base specification	Correction	Self-pay LOCI	Std error	N
L2	CC's results	-0.0612	0.0450	12274
	Complete CCL3 variable	-0.0550	0.0452	12304
	Clean costs variable	-0.0670	0.0443	12274
	CCL3 and costs	-0.0602	0.0446	12304
L3	CC's results	-0.0944*	0.0540	10874
	Complete CCL3 variable	-0.0853	0.0544	10904
	Clean costs variable	-0.0968*	0.0550	10874
	CCL3 and costs	-0.0874	0.0552	10904
L3 without NHS waiting time	Drop NHS waiting time	-0.0719	0.0489	11619
	Complete CCL3 variable	-0.0648	0.0491	11649
	Clean costs variable	-0.0778	0.0509	11619
	Include local area characteristics for Scotland	-0.0648	0.0450	12274
	CCL3, costs, and local area characteristics for Scotland	-0.0622	0.0466	12304

*Source: see do-file "Section 6 – dataset cleaning.do", provided to the CC in the Data Room. ***/**/* indicates statistical significance at the 1%/5%/10% level.*

6.16 In summary, we find that the CC's baseline OLS results are seriously affected by data errors and omissions. Once these errors and omissions are corrected for, no significant relationship between price and concentration is found in baseline OLS regressions. The CC cannot rationally rely on regression results that are driven by incorrect or missing data.

The CC's BMI-only regressions

6.17 The effect of the missing data is to exclude six BMI hospitals from the regressions reported in Table 13: Albyn, Fernbrae, Park, Ross Hall, and Werndale have missing data for NHS wait, and South Cheshire has missing data for the CCL3 dummy. These hospitals account for [✂] observations. This amounts to a significant exclusion given that the BMI-only regressions are performed over 43 hospitals and [✂] observations.

6.18 Applying the corrections outlined above, we find that the OLS point estimate is positive and statistically insignificant. This again indicates that there is no evidence that BMI charges higher prices in areas of lower LOCI. Table 6 presents the results for a specification excluding NHS waiting time, applying corrections to the CCL3 and costs variables, and including local area characteristics for Scotland.

Table 6: Effect of data and specification corrections on BMI-only regressions

Specification	Self-pay LOCI	Std error	N
CC's specification	-0.0980	0.0963	2820
With cleaned dataset and specification	0.0390	0.0905	3349

Source: "Section 6 – dataset cleaning.do", provided to the CC in the Data Room.

The CC's instrumental variables estimation

6.19 We find that correcting these data errors and omissions to the CC's instrumental variables specifications does not substantively change the results generated. However, this should not give the CC confidence in the IV results: by contrast, these results are not robust for other reasons, which we explain in the next section.

Concerns arising from the data errors and omissions

6.20 Before the Data Room opened, the CC identified and corrected a mistake in the cleaning process for the Healthcode data. We have not been able to check the cleaning process for the Healthcode data and for other raw data, because, the data files required to properly check the cleaning of the Healthcode data and other raw data have not been provided in the Data Room.

6.21 We have shown in this section that material errors and omissions still exist in the CC's dataset, and correcting these errors and omissions has a significant impact on the CC's PCA results.

6.22 Whilst errors can occur in all empirical analysis, it is concerning that two material errors have been identified in the CC's data even after the publication of Provisional Findings. As we noted in our response to the first PCA Data Room, there must be a concern that such errors are indicative of a deeper issue with the quality of the work carried out by the CC. Whilst the Data Room process has been helpful in identifying data errors and omissions, we are unable to review and confirm all of the data cleaning and processing performed by the CC, both because of restrictions on the data provided in the Data Room and because of restrictions on time available in the Data Room (a detailed reviewing and checking process would have required additional time).

Section 7

The CC uses invalid instrumental variables

7.1 The CC's PCA2 baseline results in Tables 4 and 5 of Appendix 6.9 use a technique known as Ordinary Least Squares (OLS). In PCA2, the CC now accepts that there is a significant concern around the potential impact of omitted variables on these estimates.[56] As the CC sets out, if one or more of the covariates is endogenous (for example, if there are factors directly affecting prices that are also correlated with concentration but not included in the covariates), then the OLS estimation results may be biased.

7.2 We have, in our previous discussion in this document, shown that these concerns are very likely to be well founded. In particular, that the CC's control variables for both costs and local market conditions are poor and evidently not doing the job the CC believes they should in the CC's baseline specification. We agree with the CC that it should take these concerns seriously and that the approach of using the technique of Instrumental Variables (IV) is the standard way to attempt to do so (and has a proud history in economics, summarized in Angrist and Kruger (2001)[57]). However, as we discuss further below, the often very significant challenge with applying IV techniques is to find one or more genuinely suitable "instrumental variable(s)" that satisfies the required technical assumptions. Angrist and Kruger discuss the challenge by reference to Maddala's (1977)[58] classic textbook, writing[59] "Maddala (1977, p. 154) rightfully asks, "Where do you get such a variable?" Like most econometrics texts, he does not provide an answer."

56 See for example paragraph 46 of Appendix 6.9.

57 Angrist, J. and Kruger, A. B. (2001) "Instrumental Variables and the Search for Identification", *Journal of Economic Perspectives*, Vol. 15, No. 4.pp 69-85.

58 Maddala, G. S., *Econometrics,* New York: McGraw-Hill, 1977.

59 See page 73 of Angrist and Kruger (2001), op. cit.

7.3 Tables 6 and 7 of Appendix 6.9 contain the CC's IV estimation results and these results are discussed in paragraphs 46-63. In concluding the section, the CC states in paragraph 63 that its results indicate that the "IV estimates are preferable to the OLS estimates", which we understand means that the CC prefers the results from this section to the baseline (OLS) specifications presented in Tables 4 and 5. Of the different IV specifications, the CC expresses a preference for its specifications labelled L7 in Table 6 and FC7 in Table 7.[60]

7.4 The CC uses three instrumental variables in its analysis:

a. the distance to the nearest rival hospital (IV1);

b. the distance to the nearest hospital under common ownership (IV2); and

c. the insured LOCI (IV3).

7.5 In order for the estimation results to be unbiased, the CC has accepted that the instruments must be valid. The CC sets out three conditions which a valid instrument must satisfy in paragraph 51:

a. The instruments should be conditionally correlated with the potentially endogenous variable (LOCI in the baseline specification)

b. The instruments should be uncorrelated with the unobserved term in the model (i.e., with anything that is missed out of the model that should in truth be included)

c. The instruments should themselves be excluded from the covariates in the price equation.

7.6 Note that if condition (c) is not satisfied (in that the candidate instruments should properly be in the price equation but are instead used as instrumental variables) then condition (b) will not be satisfied. That is, if the instruments should be in the regression equation but are omitted (so (c) is not satisfied) then their variation will be absorbed into the model's unobserved term and the unobserved term is correlated with the instruments.[61] Therefore, whilst our discussion below is made in terms of condition (c) – whether the instruments themselves should be excluded from the covariates in the price equation – the same arguments simultaneously imply that condition (b) will not be satisfied if (as is the case here) the instruments are not included as covariates in the price equations.

---

60 Appendix 6.9, paragraph 63.

61 More formally, suppose that the true Data Generating Process (DGP) is $y = x_1 \beta_1 + z \beta_2 + \varepsilon$ while the CC's model of the world is $y = x_1 \beta_1 + u$. Then the unobserved term in the model will be $u = z \beta_2 + \varepsilon$. If z and x_1 are correlated and we estimate our model using z as an instrument for x_1 it will be invalid according to criterion (b) precisely because of the failure of criterion (c).

7.7 We first note that none of these instrumental variables are of the form considered most likely to "mimic" a randomized trial as well as possible, which the economic literature has called the "natural experiment" approach to instrumental variables and has been emphasized by leading proponents of IV techniques such as Angrist and Kruger (2001). Second, we note that the CC presents no economic theory motivating the particular use of these instruments. By contrast, as we have noted before – and will discuss further below – there are good theoretical reasons to believe the CC's distance instruments in particular are invalid. Third, we note the pertinent warning Angrist and Kruger[62] provide:

> *"What can go wrong with instrumental variables? The most important potential problem is a bad instrument, that is, an instrument that is correlated with the omitted variables (or the error term in the structural equation of interest in the case of simultaneous equations). Especially worrisome is the possibility that an association between the instrumental variable and omitted variables can lead to a bias in the resulting estimates that is much greater than the bias in ordinary least squares estimates. Moreover, seemingly appropriate instruments can turn out to be correlated with omitted variables on closer examination."*

7.8 The rest of our discussion in this section proceeds in two parts.

- First, we consider the instruments based on distance (IV1 and IV2). The CC finds that IV2 is not a relevant instrument, so we focus on IV1. We show that the CC's "condition (c)", set out above in paragraph 7.5, is unlikely to hold for IV1. The reason is that LOCI is an imperfect proxy for market power and therefore distance to the nearest rival hospital is likely to (conditionally) affect a hospital's market power and hence prices.
- Second, we consider the instruments based on insured LOCI. We find that IV3 is also unlikely to be a valid instrumental variable, because it is highly likely that there are common demand and cost conditions across the self-pay and insured markets, and accordingly IV3 will be correlated with unobservable differences in demand and cost. In addition, the high conditional correlation of insured and self-pay LOCI is inconsistent with the notion that insured LOCI "misses out" the endogenous factors present in self-pay LOCI. Indeed, in PCA1 the CC considered it acceptable to use insured LOCI as its measure of concentration precisely because it considered insured LOCI was likely to be very closely related to self-pay LOCI, noting that "concentration measures based on insured patients and self-pay patients are expected to be highly correlated."[63]

---

62 Page 79, op. cit.

63 Competition Commission, "Price-concentration analysis for self-pay patients", published on 28 March 2013, footnote 15.

7.9 In summary, we find that the instruments IV1 and IV3 are unlikely to be valid instruments. Using these clearly invalid instruments can only reasonably be expected to give rise to bias in the CC's estimation results. To rely on such results would not be reasonable.

7.10 Whilst this section focuses on the validity of the CC's IVs, we note that the concerns raised in Section 5 around the inadequacy of the control variables included in the OLS regressions also apply to the CC's instrumental variable regressions – at least to the extent that the CC considers (as it argues) that its inclusion of regional dummies and local control variables help ameliorate the concern with omitted variables.[64]

The validity of the CC's instrumental variables based on distance

7.11 Two of the CC's chosen instrumental variables are:

- "distance to the nearest rival hospital" (IV1); and
- "distance to the nearest hospital under common ownership" (IV2).

7.12 At paragraph 60 of Appendix 6.9, the CC finds that IV2 is, in practice, not found to be a relevant instrument. We therefore focus our discussion on IV1.

7.13 Condition (c) for the validity of an instrument requires that the instruments should themselves be excluded from the covariates in the price equation. In the case of IV1 therefore, the question is whether the distance to the nearest rival hospital has explanatory power for market power over and above the LOCI measure.

7.14 The question of whether these variables satisfy assumption (c) is therefore simply the question of whether these variables should be expected to affect local self-pay prices. The CC accepts this at paragraph 58 of Appendix 6.9:

64 At paragraph 48 of Appendix 6.9 the CC argues that: "In the current case, factors that might cause endogeneity are, for example, omitted supply and demand factors. We have attempted to control for such differences through the control variables, and specifically the regional dummies and the local area characteristics variables."

> *"Condition (c) will hold if the distance instruments or the insured LOCI instrument are not thought to directly affect prices in Equation 1. For the distance variables, this would hold if the concentration measures we use capture all of the pricing power possessed by a hospital, and the distance measures did not themselves reflect another dimension of local concentration. Several parties have argued that the distance instruments are themselves relevant to local concentration and patients' price-sensitivity, and are thus not valid instruments because the variables should feature in the pricing equation. While we recognize that distance does play a role in differentiating hospitals, we consider that because the self-pay LOCI incorporates geographic relationships between hospitals in its calculation (see Appendix 6.4 for a discussion of the weighting scheme in LOCI and how it relates to distance), it is reasonable to exclude the distance variables from Equation 1 and assume condition (c) holds."*

7.15 In this section we show that it is, in contrast to the CC's view, not reasonable to exclude the distance variables. We do so in three stages. First, we repeat our position that LOCI is not a good proxy for market power in and of itself. Second, we repeat our position that the distance to the nearest rival hospital is likely to reflect a dimension of local competition over and above that captured by LOCI and note that that position is accepted by the CC elsewhere in the Provisional Findings. Third, we summarise the implications for the validity of IV1.

LOCI as measure of market power

7.16 We have previously noted that the only known model where LOCI is a good proxy for a hospital's market power is based closely on the LOGIT model, and that the LOGIT model is an unrealistic foundation for a measure of market power.[65] The CC has now removed its discussion of the theoretical underpinnings of LOCI and argues instead that LOCI does not rely on any specific model.[66] This is true in one specific sense, namely that the CC can clearly calculate the LOCI (and interpret it as one minus a weighted average market share) without relying on a specific underlying theoretical model. It is not however true in the sense that the CC wishes to rely upon. Specifically, the only known economic model where the LOCI (a measure of market share) is a good proxy for a hospital's market power is built from the LOGIT model. The economic foundations show that in essentially every other economic model of price determination, the CC would not be able to rely on LOCI as a reliable proxy for the determinants of a hospital's market power. Indeed, it is well-known that HHI would be the relevant one if the world were characterized by the Cournot model.

[65] See Dr Peter Davis, *Comments on CC's annotated issues statement: The CC's approach to measuring concentration*, 17 April 2013. We note that at paragraph 5 of Appendix 6.4 the CC seeks to address the concern that LOCI is far from widely accepted as a measure by the competition economics community. In that regard we note that the CC has now indeed found a single published reference –

7.17 We agree with the CC that it is common, as a first screen, to consider possible indicators of market power such as market shares and concentration (e.g. HHI). We agree that in its PCA the CC can reasonably use measures of market share or concentration as a proxy. However, we submit that the CC needs to accept fully that its LOCI measure is just that – a hugely imperfect proxy indicator of market power (as indeed HHI would be) and adjust the econometric analysis accordingly.

7.18 Specifically, the CC accepts that LOCI over-weights volume/revenue shares in postcode areas near the hospital,[67] while economics makes clear that it is the marginal consumers – and their willingness and ability to switch – which are likely to impose competitive constraints on price levels. Thus, by construction, LOCI is inherently likely to be a poor proxy of a hospitals' market power. This is true even aside from the fact that, for example, patients will also use hospitals near work while the CC (perhaps necessarily) use only patients' home (or reported) postcodes in constructing the measure.

7.19 The implication is that the CC should consider the likely determinants of a hospital's market power and, where only imperfectly observed proxies are available, seek to control where possible for the other relevant factors. Indeed this is the approach the CC at least attempts to take for local variation in demand and cost factors.

Other determinants of market power

7.20 The CC's PCA presumes that a measure of weighted market share (1-LOCI) is a sufficiently good proxy for market power that all other likely determinants of hospital substitutability in the eyes of consumers can be omitted from the price equation.

---

albeit one which does not do anything like endorse the concept but rather just notes it as "an alternative model" (see the quote in paragraph 8 of Appendix 6.4). In this regard, we note that the paper is, in fact, a "self-citation" since the Antwi, Gaynor and Vogt's (2006) unpublished paper has a co-author in common, Professor Gaynor, with the book chapter by Professors Town and Gaynor (2012).(See footnote 7, paragraph 5, Appendix 6.4). The 2007 US work the CC refers to in paragraph 5 is the same authors as the 2006 paper and remains unpublished. The CC also cites some work by the NZa in Netherlands. As we understand it, the NZa provides only advisory opinions to the NMa and it is therefore not at all clear that LOCI has been relied on in any actual decision of the NMa, let alone an actual significant competition intervention. The other paper cited by the CC is an unpublished paper by two Dutch economists - one of whom, Misja Mikkers, is based at NZa – while the paper is written about Ireland. It remains the case that the CC has not yet found a single peer reviewed academic article that relies on LOCI, although it has clearly put some effort into a literature search. This is hardly evidence of its widespread acceptance among the professional economist community. We also note that contrary to the assertion at paragraph 6, there is nothing "healthcare specific" about LOCI - it could equally be applied in retail markets; but has not been to date.

66 See paragraph 7, Appendix 6.4 where the CC states "We would emphasize that our interpretation of LOCI…does not rely on the assumptions of any particular economic model holding." The network LOCI does not even correspond to any specific model of competition, as the CC now accepts at paragraph 29 of Appendix 6.4 (and in particular footnote 23).

67 See Appendix 6.4, paragraph 18

7.21 Economists do not generally believe that concentration is a sufficient proxy for market power.[68] Rather economists typically consider that the degree of market power will depend on substitution possibilities. That is, market power depends on how much consumers at the margin are willing and able to switch away from a given hospital to nearby rival hospitals. Irrespective of the merits of LOCI in particular as a measure of concentration, as with essentially all measures of concentration, LOCI is not likely to provide – on its own – a good measure of the competitive constraint on a given hospital.

7.22 The CC is, in effect, arguing that its measure of market share is the only relevant proxy for a hospital's market power:[69]

> *While we recognize that distance does play a role in differentiating hospitals, we consider that because the self-pay LOCI incorporates geographic relationships between hospitals in its calculation (see Appendix 6.4 for a discussion of the weighting scheme in LOCI and how it relates to distance), it is reasonable to exclude the distance variables from Equation 1 […]*

7.23 The CC appears to wish to rely upon considerations in its consideration of LOCI in Appendix 6.4, paragraph 18:

> *Third, research in the health literature has consistently shown distance to be an important element of patients' preferences and thus a driver of hospital choice. [Footnote omitted] Given that hospitals are differentiated by geographic location, hospitals that are nearer are likely to represent a stronger constraint than hospitals that are further away. The weighting scheme implicit in the LOCI measure typically assigns more weight to a hospital's share of patients in local areas that are nearby—as a result, the preferences of patients with regard to geographic differentiation are directly reflected in the concentration measure.*

7.24 However, the proposition that LOCI perfectly reflects geographic differentiation, without any additional role for the distance to the nearest rival hospital, is entirely inconsistent with both economic theory and the CC's own statements elsewhere in the Provisional Findings.

---

68 This belief is reflected in competition authorities' guidelines: for example see the European Commission's Guidelines on the Assessment of Horizontal Mergers, paragraph 14, and the CC's Merger Assessment Guidelines (CC2), paragraph 5.3.2 and footnote 63.

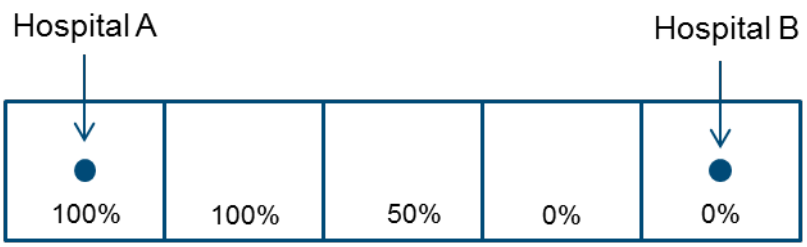
69 Appendix 6.9, paragraph 58.

7.25 Consider first economic theory. A simple example shows that LOCI does not and cannot capture all dimensions of market power, and that distance between hospitals is likely to also be important. Consider two competing hospitals which, together, draw all of their patients from five postcode areas. Assume that each of the postcode areas has equal numbers of patients and, within a postcode area, there is a uniform distribution of patients. Finally, assume that each patient chooses the hospital closest to their home (i.e. travel distance is important to patients in choosing a hospital).

7.26 Figure 2 shows one possible configuration of resulting patient demand. Each box represents a postcode area. The hospitals are located at the centre of the first and fifth boxes and each postcode area has by assumption one-fifth of the overall population living in it. The percentages written within each box then show the proportion of patients living in that particular postcode area which choose hospital A. As we assume that each patient chooses the hospital closest to their home, all patients in the leftmost two boxes choose hospital A and half of all patients in the central box choose hospital A. The remainder of patients choose hospital B, as it is closer to their home.

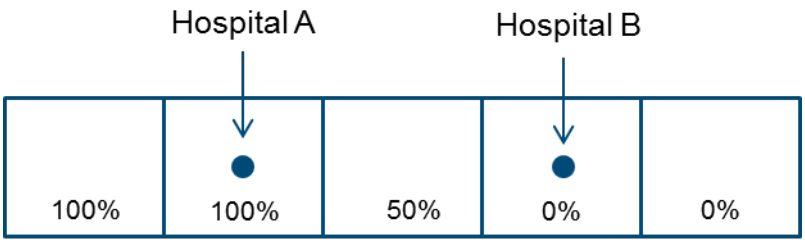
7.27 The LOCI of hospital A is equal to 0.1 (1 minus 0.9), since 40% of its patients come from each of the leftmost two boxes (where it has a 100% market share), and 20% of its patients come from the central box (where it has a 50% market share); 0.4*1 + 0.4*1 + 0.2*0.5 = 0.9.

Figure 2: Stylised representation of market 1



7.28 Now consider a slightly different configuration, in which the hospitals are instead located at the centre of the second and fourth boxes. Again the percentages within each box show the proportion of patients living in that postcode area which choose hospital A, under the assumption that each patient chooses the hospital closest to their home. It is simple to see that the LOCI is again equal to 0.1.

Figure 3: Stylised representation of market 2



7.29 The CC's proposition is that, because the LOCI is the same in each market, the market power of hospital A is the same in each of these markets, and that the distance between hospital A and hospital B has no predictive power for price. But a simple analogy to the benchmark Hotelling model shows that this is not true. With quadratic transport costs, it is simple to show that the Nash equilibrium price is higher when firms are located at the extremes of the line as opposed to when they are located closer together: see, for example, the exposition in section 7.1.1.1 of Tirole (1989).[70] The intuition is that the closer together are the two hospitals, the better substitutes they are from the point of view of a patient in any given location. When the hospitals are located in the same position, they are perfect substitutes (since for all patients the transport costs to hospital A and hospital B are equal).

7.30 In other words, in order for the CC's IV1 to be valid, it must be the case that hospital A has the same market power in the cases illustrated in Figures 1 and 2. This is against economic intuition and simple formalisations such as the Hotelling model. For the avoidance of doubt, the example is not specific to the Hotelling model – rather it captures the simple intuition that when hospitals have better substitutes they are likely to have less market power.

7.31 The CC's proposition is also inconsistent with the approach it adopts elsewhere in the Provisional Findings. First, the CC makes abundantly clear that in its local competition assessments it uses both LOCI and fascia counts to identify "hospitals of potential concern" (see paragraphs 6.89-6.101 of Provisional Findings.)

7.32 Second, in describing its actual competitive assessment (excluding central London) at paragraph 6.105 the CC states:

> *"In order to assess the extent of any competition faced by each hospital of potential concern, we have taken into consideration several factors, including:* ***(a) results of the different concentration measures;*** *(b) the hospital's own individual characteristics as well as the characteristics of the nearby private hospitals and PPUs, either competitor hospitals or hospitals under the same ownership…****(c) characteristics of the local area in which the hospital is situated****; (d) documentary evidence submitted by the parties; and (2) the views of the parties. To help our analysis, we have also used maps of catchment areas and population density by local authority." (emphasis added).*

7.33 At paragraph 6.107 the CC describes in more detail the second of these relevant factors:

[70] Jean Tirole, *The Theory of Industrial Organisation*, Cambridge, MA: The MIT Press, 1989.

> *"As for hospital individual characteristics, in respect of each hospital of potential concern and nearby private hospitals and PPUs […] we have considered the factors listed below:*
>
> *(a) range of specialities offered […]*
>
> *(b) availability and type of ICU […]*
>
> *(c) hospital size […]*
>
> *(d) hospital patient mix […]*
>
> *(e)* ***location and distances between hospitals****; and*
>
> *(f) size of the catchment area in miles and the extent of any overlap between catchment areas."*

7.34 The CC clearly accepts that distances between hospitals are relevant factors for its competition assessment. As a result, it cannot legitimately argue (as it currently does) that they are simultaneously not relevant as a dimension of market power (over and above LOCI) when conducting its quantitative analysis in the PCA. In particular the CC is explicit in paragraph 6.106 that its local competition assessment – in which it attempts to consider which hospitals have market power – takes into account the "location and distances between hospitals" in a manner which is over and above its concentration measures in its competition assessment. The CC's approaches in its local competition assessments and in the discussion of the validity of IV1 are entirely inconsistent: both cannot simultaneously be correct.

Implications for the validity of IV1

7.35 In the section above, we noted that economists do not generally believe that concentration is a sufficient proxy for market power; that LOCI is in itself a poor proxy for market power, and that distance to the nearest rival hospital is likely to be an important determinant of hospital substitutability (and therefore hospitals' market power) over and above LOCI. We also noted that the CC has accepted the additional role of distance to own and rival hospitals in its local competition assessments.

7.36 In terms of PCA2, the CC recognises that the relevant test for the validity of IV1 is:[71]

> *Condition (c) will hold if the distance instruments or the insured LOCI instrument are not thought to directly affect prices in Equation 1. For the distance variables, this would hold if the concentration measures we use capture all of the pricing power possessed by a hospital, and the distance measures did not themselves reflect another dimension of local concentration.*

[71] Appendix 6.9 paragraph 58.

7.37 Therefore the discussion above has important implications for the CC's econometric analysis. In particular, it unambiguously implies that the CC's instrumental variables based on distance are highly likely to be invalid.

7.38 Distance is not a measure of local concentration. However, distance is clearly a factor that is likely to affect a hospital's market power over and above the element captured in LOCI.[72] The effect is likely to be beyond its LOCI since two hospitals with the same LOCI may have very different degrees of market power, depending on, for instance, how close rival substitute hospitals are.

7.39 Thus the CC in fact has every reason to believe that LOCI is not going to capture competitive constraints well and that, to the extent it does have other measures relevant for allowing substitution between the hospitals, those should be included in the regression analysis if the result is to be robust. If they are not, the CC will need to accept that those factors will be in the unobservable of the model and so are very likely to cause biases when we try to use them as instrumental variables for LOCI (i.e. a failure of condition (b)). In either case, the distance variables are not valid instruments.

7.40 As this is an important point, we also note that the CC further considers including distance and LOCI in the price equation in footnote 36 to Appendix 6.9:

> *"The argument that distance should feature in the price equation would result in an equation with two distinct concentration measures. The same argument would also imply that we should include fascia count measures in the same equation, as well as LOCI and the distance variables, since all are measures of concentration. In order to keep the model simple and coherent, and following standard practice, we think it is reasonable to test one concentration measure at a time. Even if distance is considered a measure of local concentration, we do not think it is preferable to either LOCI or fascia count."*

7.41 This discussion is confused. Distance is not a concentration measure. Distance to rival hospitals is however likely to affect whether prices are high or low for any given level of market share (or LOCI). This is a proposition the CC has accepted elsewhere in the Provisional Findings. Distance is (contrary to the CC's position in PCA2) highly likely to be relevant for local pricing: IV1 (the distance to rival hospitals), for example, would ordinarily affect the substitutability of the hospitals, while IV2 (the distance to hospital under common ownership) will affect the diversion that is internalized by common ownership and so should be expected to affect price levels. If there are efficiencies from common ownership, then the sign (i.e. directionality) of the latter effect is unclear.

---

[72] The immediately analogous point also of course applies to each of the fascia count specifications that the CC presents in for instance Table 7 of PCA2. We present the discussion in terms of just LOCI in order to simplify the presentation but the point is not specific to a particular concentration measure.

7.42 Of course, in principle if the CC prefers to keep the model "simple and coherent", it could choose not to include the distance in the price equation. But if it does so, it would need to accept that those factors will then be in the unobservable of the model and so are very likely to cause biases when the CC tries to use them as instrumental variables for LOCI (i.e. a failure of condition (b)).

7.43 It is in no way "standard practice" to ignore the implications of substitution possibilities for endogeneity and what are/are not likely to be valid instruments for measures of concentration. See for example Mazzeo (2002)[73], who allows for product differentiation in his price-concentration regressions. Mazzeo estimates a regression that allows the effect of same-type and different-type competitors to vary (in the current case, competitors are differentiated geographically, but the analogue is obvious). Or see Davis (2005)[74], who allows geographic differentiation to play an important role in price-concentration regressions, with competitors with different degrees of differentiation impacting prices differently.[75] Nor is it standard practice to artificially insist on only one measure of concentration; for instance, Mazzeo incorporates four parameters (by competitor type and for first/additional competitors). Of course, even if it were standard practice, that would not be a rational reason to take a mistaken approach. Indeed, the CC's own fascia count model specifications (e.g. those labelled FC1-FC3) allow for three different concentration measures depending on geographic differentiation, contradicting the CC's own argument. It is not, in our view, the approach a reasonable expert body would take.

7.44 In our view, the clear and unambiguous implication of the arguments set out in this section is that the CC's distance variable IV1 cannot logically be considered to be likely to be a valid instrument. And since it is not valid, the CC's results in this regard are not reliable and indeed may well be wholly misleading.

---

73 Mazzeo (2002) "Competitive Outcomes in Product-Differentiated Oligopoly," *Review of Economics and Statistics* 84(4), pp.716-728.

74 Davis (2005) The Effect of Local Competition on Retail Prices in the US Motion Picture Exhibition Market", *Journal of Law and Economics*, Vol. 48, No. 2.

75 Authors (and competition authorities) have sometimes been able to circumvent this issue to some extent by a choice of a careful identification strategy. For example, some papers use an exogenous (to local market conditions) change in concentration that affects only a subset of markets to facilitate a "differences-in-differences" estimator. See for example Hastings, J. (2004). "Vertical relationships and competition in the retail gasoline markets: empirical evidence from contract changes in Southern California" *American Economic Review* 94:317-28. Such known exogenous variation can be used to construct 'difference in difference' estimators. The CC's identification strategy is, in contrast to all good papers in the academic literature, poorly thought through.

7.45 While the CC argues in the PCA that the distance to rival hospitals is not relevant, the CC argues the opposite in its local competition assessments. The approach in the competition assessment in this respect is plainly the correct one and the consequence is that the instrumental variables based on distance used in PCA2 cannot be valid instruments. Accordingly, estimation results using these instruments are not reliable.

7.46 A further method of seeing that the CC's assumptions in relation to IV1 are not correct is to test them directly. The CC's contention is that IV1 satisfies condition (b), IV1 is exogenous, and condition (c), IV1 is excluded from the covariates in the price equation. A simple test of this contention is to add IV1 to the covariates of the price equation whilst using IV3 (a valid and relevant instrument, according to the CC) as an instrument for self-pay LOCI or fascia count.[76] If the CC's contention were true, we would expect the coefficient on IV1 to be (not statistically significantly different from) zero precisely because of conditions (b) and (c).

7.47 We run two specifications, the first including IV1 and the second including IV1 and IV1 squared. Using LOCI as the measure of concentration, we find that the coefficient(s) on IV1 (and IV1 squared) are positive and (jointly) significant at the 10% level (as we would expect, price increases in distance to closest rival) and the coefficient on self-pay LOCI is statistically insignificant. Using fascia count, in specification (1) we find the coefficient on IV1 is positive but insignificant and the coefficient on fascia count is insignificant; in specification (2) we find that the coefficients on IV1 and IV1 squared are jointly significant at the 10% level. Table 7 presents the results.

[76] In short, we add iv_dist_compf to the covariates in specifications L6 and FC6.

Table 7: IV1 in price equation

	Self-pay LOCI		Fascia count	
	(1)	(2)	(1)	(2)
Self-pay LOCI	-0.0923	-0.0998		
Fascia count (0-9 miles)			-0.0348	-0.0457
Distance to nearest rival	0.0013*	0.0001	0.0004	-0.0045
Distance to nearest rival squared		0.0000		0.0001
R-squared	0.91	0.91	0.91	0.91
N	10874	10874	10874	10874
Test of null hypothesis that instruments are irrelevant (F-statistic)	523.25	479.59	13.19	8.99
Test of null hypothesis that the covariates are exogenous (p-value)	0.10	0.10	0.08	0.08
Test of null hypothesis that distance variable(s) are (jointly) insignificant (p-value)	0.0870	0.0979	0.7880	0.0913
Instruments	IV3	IV3	IV3	IV3

*Source: "Section 7 - ins and sp LOCI correlation and IV1 in price equation.do", provided to the CC in the Data Room. ***/**/* indicates statistical significance at the 1%/5%/10% level.*

7.48 These results show that the CC's hypothesis that IV1 satisfies conditions (b) and (c) is rejected by the data even on the CC's assumption that insured LOCI is a valid instrument.

The validity of the instrumental variables based on concentration in a comparator market (insured patient LOCI)

7.49 The third instrumental variable (IV3) the CC proposes to use is the insured patient LOCI. This is a new proposal for an instrumental variable, which was not used in PCA1 - not least because IV3 was included in the PCA1 regression specification.

7.50 Here the question is once again whether insured patient LOCI (IV3) satisfies the three conditions for a valid instrument set out by the CC in paragraph 51 of Appendix 6.9. The CC at paragraph 57 argues:

> *"For [IV3] to be exogenous, we require that a hospital's insured LOCI [...] be unrelated to the local level of self-pay demand. Given [the arguments presented by the CC], namely that a hospital location may be determined by the local insured demand rather than the local self-pay demand… [the CC] also think it reasonable to assume that a hospital's local strength in the insured market is unrelated to the local levels of self-pay demand. Put differently, even if the insured LOCI and self-pay LOCI are closely related, knowledge of a hospital's insured LOCI does not necessarily provide information about the level of demand for self-pay treatment in the local area."*

7.51 In terms of this specific argument, it seems far more likely that a hospital would consider both the likely PMI and self-pay demand in a particular area than just one of the two, since that may be economically irrational. Indeed, BMI's internal documents studying the advantages of entry or expansion at a particular location [✂].

7.52 Moreover, the CC's consideration of the validity of IV3 is notably sparse. When we consider an instrument derived from another market, we need to consider whether there are any commonalities in the markets in terms of either demand or cost conditions. Such commonalities (if not adequately controlled for) would imply that the instrument is correlated with unobserved demand or cost factors, rendering the instrument invalid.

7.53 On the cost side, the cost conditions in the insured and self-pay markets clearly have substantial commonalities: insured and self-pay patients are treated in the same hospitals by the same clinical staff (supported by the same non-clinical staff), using the same supplies (such as drugs and prostheses). If a hospital is located in a high-cost area, it will face high costs for treating both insured and self-pay patients.

7.54 The CC argues at paragraph 50 that the cost variable and CCL3 and NUTS1 regional dummies will pick up any cost differences between geographic regions. In Section 4 we showed that the NUTS1 regional dummies cover very wide geographic areas and are unlikely to pick up local cost conditions, and in Section 5 we showed that the CC's cost variable does not adequately control for local cost conditions. Accordingly we find it likely that there are unobserved cost components in the error term of the model.

7.55 High-cost areas will attract fewer entrants (all else equal) and therefore have lower insured and self-pay LOCIs. Given that insured LOCI will therefore be correlated with the unobserved cost component in the error term of the model, IV3 will therefore fail condition (b).

7.56 On the demand side, the CC acknowledges at paragraph 55 of Appendix 6.9 that "there may be within-region and unobservable differences in demand that substantially affect prices charged and are not included in the regression." Given that the CC accepts there will be unobserved demand factors in the error term of the model, we consider whether there will be commonality in demand conditions between insurance and self-pay markets.

7.57 There are at least two ways in which commonality in demand conditions may arise. First, it is likely to be the case that, across the cross-section of local markets, areas of high self-pay demand are also areas of high insured demand. Demand in both insured and self-pay markets is driven by relatively affluent individuals: insured patients either pay for PMI themselves (requiring affluence) or have PMI provided by their employer (and such benefits tend to be associated with relatively high remuneration). Self-pay patients pay for treatment themselves, which again requires affluence. Again, given that high-demand areas will attract more entrants and result in a high LOCI, these considerations imply that insured LOCI will be correlated with the unobserved term of the model, and IV3 will fail condition (b).

7.58 Second, there may be heterogeneity of hospitals' individual demand due to, for example, hospital-level differences in quality, expertise, or range of services, which is not sufficiently controlled for in the regressions. If patients have a preference for "quality", then high-quality hospitals will (all else equal) have a low insured LOCI. Self-pay patient prices will also tend to be higher at hospitals which are high quality (all else equal). "Quality" is not observed and therefore appears in the unobserved term of the model (the only attempt that the CC makes to control for such factors is through the CCL3 dummy and the cost variable). This implies that the insured LOCI will be correlated with the unobserved term in the model, which will induce a failure of condition (b) for IV3.

7.59 In sum, we established above that the evidence in PCA2 indicates that the CC's attempts at correcting for local demand and cost conditions demonstrably do not succeed. The implication of the substantial commonality in demand and cost conditions between the insured and self-pay markets is that IV3 is likely to be correlated with the unobserved term in the model, as that term includes unobservable differences in demand and cost. This makes it likely that condition (b) will fail for IV3.

7.60 The validity of insured LOCI as an instrument is also questioned by the CC's reported estimation results. The CC's test of IV3's relevancy indicates that the self-pay and insured LOCI variables are very highly conditionally correlated. The test statistic (F-statistic) for L6 reports a value of 643.73[77] while the CC states that "[a] common benchmark for this test that indicates the instruments are relevant is an F-statistic of 10 or higher" (paragraph 60). Clearly the self-pay and insured LOCI variables are very highly conditionally correlated indeed (much more so in particular than IV1 in Table 6 or IV1 or IV3 in Table 7). Indeed, the (non-conditional) correlation between the self-pay and insured LOCI is 0.90 across the sample used in L6.[78]

---

77 Figure reported in the PCA provided in the Data Room; Table 6 of Appendix 6.9 gives a slightly different figure.

78 See do-file "Section 7 - ins and sp LOCI correlation and IV1 in price equation.do", provided to the CC in the Data Room.

7.61 In the extreme, if self-pay LOCI and insured LOCI were perfectly correlated, then insured LOCI would definitely not be a valid instrument for self-pay LOCI. In order for insured LOCI to be a valid instrument, the CC's hypothesis is that the insured LOCI misses out factors that affect self-pay LOCI but which are correlated with the omitted variables in the regression must be true. The higher the correlation of self-pay LOCI and insured LOCI, the less likely it is for the CC's results to be consistent with this hypothesis.

7.62 For the reasons set out above, the insured LOCI (IV3) is in our view unlikely to be a valid instrument for the CC's analysis. Since it is unlikely to be a valid instrument, in our view the CC's regression results using IV3 are not likely to be reliable.

Section 8

The CC's tests of its instrumental variables

8.1 The CC reports three tests of its instrumental variables in Tables 6 and 7 of Appendix 6.9. These tests seek to consider whether:

a. the instrumental variables are irrelevant, i.e. whether condition (a) for the validity of the instrumental variables is satisfied;

b. The covariates are exogenous, i.e. whether instrumenting is required because of endogeneity of covariates; and

c. The instruments are exogenous, i.e. whether condition (c) for the validity of the instrumental variables is satisfied.

8.2 In this section we consider in turn the implications of these tests. The CC shows that IV2 is irrelevant in the sense that it is not conditionally correlated with LOCI or the CC's fascia count (0-9 miles) measure. It therefore fails condition (a) of the CC's three conditions for IV validity. In addition, IV1 is redundant in L7, in the sense that it is not conditionally correlated with self-pay LOCI once insured LOCI is taken into account. This is a natural finding given the high correlation between self-pay LOCI and insured LOCI.

Tests of the IVs' relevancy

8.3 The first test considers whether the instrumental variables are irrelevant by considering whether condition (a) of paragraph 51 – that the instrument is correlated with the potentially endogenous variable – is satisfied. More specifically it considers a regression of self-pay LOCI on the proposed instrument and the other covariates included in the CC's regression and then reports whether the instrument is (or a combination of instruments are) statistically conditionally correlated with self-pay LOCI. The CC finds that IV1 and IV3 pass the test but IV2 does not.

8.4 This motivates the use of IV1 and IV3 as instruments in the CC's preferred specification L7/FC7. (We note that Tables 6 and 7 of Appendix 6.9 incorrectly indicate that the CC uses IV1 and IV2.)

8.5 We agree that in principle the test undertaken by the CC is a sensible test to consider, although to the extent that other variables should be included in the analysis, i.e. to the extent that condition (c) is not satisfied, the CC's approach does not amount to a valid test of the relevance of the instruments.

8.6 We examined the first stage regression results provided in the Data Room. We find that IV1 and IV3 are of the expected sign in the first stage regressions. However, the results for L7 show that IV1 is a redundant instrument, as it is insignificant in the first stage regression. This is unsurprising: IV3 is highly correlated with self-pay LOCI and is capturing all of the relevant variation.

8.7 The notes to Tables 6 and 7 in CC's Appendix 6.9 state that the specifications include control variables that are the same as those in specifications L3/FC3 respectively. That means, for example, that the specifications all only include NUTS1 regional dummy variables. The CC acknowledges that the specification likely omits a variety of relevant variables, so the power of these tests may not be high, but on the face of it, these results do provide an undoubtedly helpful reassurance that IV1 and IV3 pass the first of the CC's three tests set out in paragraph 51 of Appendix 6.9. Of course, this is not sufficient for an instrument to be valid, and we showed in Section 7 above that IV1 and IV3 are unlikely to be valid for other reasons.

8.8 However, as we noted at paragraph 7.60 above, this test indicates that self-pay and insured LOCI are very highly conditionally correlated which casts doubt on the validity of IV3.

Tests of the covariates' exogeneity

8.9 We understand that the second test is a Hausman type test[79] comparing the difference between the OLS and IV parameter estimates. The F tests that the CC reports appear to be performed across all the parameters in the model, not just the LOCI coefficient. When performing these tests across all parameters in the model, it is clear that the null hypothesis (that the OLS and IV parameter estimates are equal) is rejected for many specifications.

[79] As help ivreg2 notes, "the endogeneity test implemented by ivreg2, is, like the C statistic, defined as the difference of two Sargan-Hansen statistics […] under conditional homoskedasticity, this endogeneity test statistic is numerically equal to a Hausman test statistic".

8.10 However, when considering such results it often makes sense to also consider the impact of using IVs on the estimate of the parameter of interest alone. Indeed we find that the coefficients on LOCI in specification L3 in Table 4 and (say) specification L7 are actually not statistically significantly different from one another.[80] The CC's test is presumably therefore rejecting exogeneity on the grounds primarily that at least some of the other parameters in the equation are collectively significantly different between specifications L3 and L7.

Tests of the instruments' exogeneity

8.11 The CC describes at paragraph 62, Appendix 6.9 that the "third statistical test in the tables ('Test of null hypothesis that the instruments are exogenous (p-value)')… can only be performed for the specifications L7 and FC7…"

8.12 This is because only those specifications introduce two instrumental variables rather than one and as a result the "extra" instrument can be used to test whether using one or the other makes a difference to the parameter estimates. It is efficient to use both if they are both valid instruments, but if one of them is not a valid instrument then the specification using both instruments would be invalid.

8.13 The test implemented is a Sargan-Hansen test is a test of overidentifying restrictions; for the efficient GMM estimator, the test statistic is Hansen's J statistic.

8.14 As the CC states, "the test assumes that at least one of these instruments is valid". Given the arguments presented in Section 7 above, this assumption is highly likely to be untrue and accordingly the test is not likely to provide any useful reassurance.

8.15 In non-technical terms, the test considers whether using both instruments in L7 and FC7 makes a big difference to just using one of them alone. Under the null hypothesis to be tested, using both instrumental variables together should not make too much difference because the estimates using just one instrument are assumed to be valid ones.[81]

8.16 Given that IV3 is very highly conditionally correlated with self-pay LOCI and IV1 is insignificant in the first stage regression for L7, it is unsurprising that the results of L7 are very similar to those of L6 because the addition of IV1 makes very little difference to the first stage regression.

---

80 The coefficient on LOCI in specification L7 is -0.1549 with a standard error of 0.0574, while the coefficient on LOCI in specification L3 is -0.0943 with a standard error of 0.054. Therefore, the coefficients on LOCI appear to be within approximately 1 standard deviation of each other.

81 More technically "consistent" estimators.

8.17 It is however striking that there is a large difference between the LOCI coefficient in specifications L4 and L6, whereas the CC's proposition - that both of its instruments IV1 and IV3 are valid – would suggest that these two LOCI coefficients should in fact be similar to one another.

8.18 Therefore whilst the test result does not reject that the instruments are exogenous, this appears to be primarily due to the fact that L6 and L7 give very similar results (as IV1 is redundant in L7), despite the fact that the results of L4 and L6 are very different.

The operator-level regressions

8.19 Table 13 of Appendix 6.9 presents IV estimates of operator-level regressions but does not show the test statistics discussed above. The files provided in the Data Room show that the test for covariates' exogeneity rejects the null hypothesis that the instruments are exogenous in the regression for BMI (p-value=0.046) even while the CC finds no statistically significant relationship between self-pay LOCI and prices. If the CC is to take the fact that, for L7, the same test does not reject the null as an indication that the instruments are valid, for consistency it would need to accept that the rejection of the null in the BMI regression indicates that, in this regression, the instruments are invalid and the coefficient estimates are not reliable. The CC should also consider the implications of such a finding for its across-operator regression analysis.

Section 9

The CC incorrectly imposes one LOCI coefficient across all operators

9.1 In this section we consider the CC's response to concerns that were raised by the parties in PCA1 regarding the imposition of the restriction to have only one LOCI coefficient. These concerns are described by the CC at paragraph 84 of Appendix 6.9 thus:

> *"The parties have made several arguments in relation to the operator-level results. In particular, they have argued that: our results are not robust when considered at the operator-level; our results are driven by only one operator; and, for conclusions to be reached regarding the general price-concentration relationship, the operator-level estimates should be statistically significant. In the Data Room, the parties have also re-run our analysis but excluded all episodes from certain operators, and argued that our results are not robust to such exclusions."*

9.2 The CC goes on to explain why it disagrees with these arguments in paragraphs 85 to 87. In summary:

- The CC considers that "the estimated relationship when pooled across operators is an estimate of the price-concentration relationship at a general level." (paragraph 85)
- The CC considers that it has "not received evidence to suggest that there would be meaningful differences in the price-concentration relationship between operators." (paragraph 86)
- While the CC "agree[s] that the operator-level analysis could in principle be used to assess potential differences between operators in the price-concentration relationship, this is a more ambitious task than the one [the CC] set out to achieve." (paragraph 87)
- And the CC considers that "estimates at the operator level (from the 'operator approach') are always likely to be less precise than our main results based on all operators (the 'pooled approach')." (paragraph 87)

9.3 Thus, in paragraphs 68-96 of Appendix 6.9, the CC sets out its extensive views as to whether the LOCI coefficient should be allowed to vary across hospital operators and treatments or, alternatively, to impose the restriction that the LOCI coefficient is the same across chains and treatments. In particular, the CC argues at length (but ultimately unconvincingly) that it should only be interested in a single LOCI coefficient across all treatments and operators.

9.4 In this section we consider these claims. We do not believe the CC has made anything approaching a convincing case. Rather the CC has expended significant effort in attempting to defend a position that is ultimately indefensible as a matter of the principled application of statistics and econometrics.

9.5 This section proceeds in four parts.

- First, we consider the CC's claim that it has "not received evidence to suggest that there would be meaningful differences in the price-concentration relationship between operators" (paragraph 86). We show that both our own previous submissions, and the CC's own findings, contradict this claim.
- Second, we consider the CC's approach to pooling the data. We explain that the CC's approach is wholly contradictory – the CC attempts to cross-check its model using subsets of the data while critiquing the parties for interpreting the results of those cross-checks in an entirely conventional manner.
- Third, we express some sympathy with the CC's view that it should pool the data (i.e. analyse the data for all operators and all treatments together). But we go on to note that the usual approach to model specification involves clear statistical testing of the implicit parameter restrictions imposed when pooling the data – tests which the CC has not reported. If the CC is to pool the data it must test those restrictions to check they correctly represent the patterns actually in the data.
- Fourth, we show that the restriction imposed by the CC – that the coefficient on LOCI is the same across all hospital operators – is roundly rejected by the data.

There is evidence available to the CC suggesting important differences in price-concentration relationships between the hospital operators

9.6 At paragraph 86 of the Appendix 6.9, the CC states that:

> *"[the CC has] not received evidence to suggests that there would be meaningful differences in the price-concentration relationship between the operators. Thus any attempts to estimate separate relationship for each operator (or to exclude certain operators from the analysis) are not based on any expectation, intuition or economic rationale."*

9.7 First, we do not consider this to be true in terms of our own submissions. In particular, we note the CC has received evidence (for example in the form of our significant submission following the PCA1 Data Room) that its own econometric analysis was suggesting that there are very significant differences in the price-concentration relationship between the operators. Specifically, in our submissions on PCA1 following the Data Room we were very clear that the CC's data wholly rejected the idea that the coefficient on LOCI was the same across the operators when tested using the full data sample. We discuss this issue further below.[82]

9.8 Second, this claim plainly contradicts central elements of the CC's Provisional Findings. Specifically, the CC provisionally concludes at paragraph 6.282 that BMI, HCA and Spire have been earning excess returns while the CC makes no such finding with respect to either Nuffield or Ramsay.[83] [✂]

9.9 Third, as we understand the position, it is in any event for the CC to put itself in a position where it can answer the statutory question; it is not for the CC to require the parties to provide a positive case in their own defence.

The CC's approach to pooling the data

9.10 The CC prefers a regression model which pools all the data (for all operators and treatments) and forces the parameter on the concentration variable (i.e. self-pay LOCI) to be the same across operators and treatments because the CC considers that it is "*this broad relationship that is of primary interest*" (Appendix 6.9, paragraph 64). Analogously, the CC considers that "the estimated relationship when pooled across operators is an estimate of the price-concentration relationship at a general level." (paragraph 85)

---

82 In addition, we note that the results in Table 13 in PCA2 very clearly indicate – once again – that there is prima facie evidence from the operator level regression results that the LOCI coefficients may be different across hospital operators. In particular, the results suggest a finding of a very strong and statistically significant relationship for Nuffield, and not for the other operators. The CC does therefore have clear evidence from its investigation in general and also the results and consultation on both PCA1 and now PCA2 that there is at least a concern that the relationship may not be the same across the hospital operators.

83 See in particular paragraph 6.282.

9.11 We agree that there is some significant force in the CC's argument at paragraph 89 that "*excluding hospitals or looking only at a subset of hospitals (e.g. belonging to one operator) reduces the useful information contained in the data for our analysis."*

9.12 However, the CC's wider argument fails for three reasons. First, some of our submissions in response to PCA1 showed that dropping very small parts of the data (e.g. a single Nuffield hospital) removed the statistically significant relationship between price and concentration found by the CC. We believe that such checks are very useful even though they discard some data because they made very clear that the CC's PCA1 results were not at all robust to very minor changes in the sample – albeit ones that the CC had not considered in presenting its results in the PCA1 Working Paper.

9.13 Second, the CC's argument does not in any way explain why the CC appears to be entirely ignoring our central submissions on PCA1, which used statistical tests based on the whole data sample.

9.14 Specifically, during our review of the PCA1, our statistical test results overwhelmingly suggested that the only hospital operator which was driving the CC's PCA results was Nuffield. Far from proving the CC's case, the CC's results went in exactly the opposite direction: the one operator which the CC found does not make excessive profits was found in the PCA to be the one operator whose data was driving the average results where the CC was finding a statistically significant relationship between self-pay prices and concentration. The average coefficient estimate was therefore wholly misleading as the data did not support the presence of the "broad relationship" the CC contends exists. Rather the statistical tests strongly suggested that there was (at most) a narrow relationship between self-pay prices and concentration for one hospital operator - Nuffield.

9.15 Third, we note that it was the CC which first presented numerous regression results which relied on only subsets of the data as appropriate robustness checks in PCA1. Moreover we further note that in PCA2 it continues to present its own robustness checks in terms of the subsets of the data.

9.16 It is important to note that we essentially agree with one aspect of the CC's statement that pooling the dataset together can be appropriate. However, we believe that when considering which model to estimate on that pooled dataset, it is appropriate to examine whether the statistical tests support the implicit restrictions being imposed. If the CC's statement that there is no reason "to suggest that there would be meaningful differences in the price-concentration relationship between operators" is correct, then such statistical tests will be accepted by the full dataset. To preview our findings, they are not. We discuss the matter further in the next section.

The appropriateness (or otherwise) of assuming a single LOCI coefficient across hospital chains and treatments

9.17 A proper statistical approach to model specification is not a matter of controversy in general. It is better to use all the data where possible, but to allow the model to capture differences where the data clearly indicates significant differences exist. As a result, the usual statistical approach is to consider on the basis of the full sample whether the implicit parameter restrictions that the CC wishes to impose can, in fact, legitimately be imposed.

9.18 Specifically, the CC can – and in our view must – test whether the data "accepts" the restrictions on the model specification that the CC is imposing. The CC can do so using the full sample dataset as we showed in our response to PCA1 following the Data Room.[84] We note that the CC makes no reference to our findings in that regard and, so far as we can see, does not discuss the approach of testing on the full dataset the "market wide" relationships claimed to exist in its Provisional Findings. Rather it continues to take the approach it took in PCA1 of examining subsets of the data, while at the same time critiquing the parties for making standard statistical interpretations of the CC's results thus obtained.

9.19 To illustrate the matter further, consider that when the CC estimates its equation with a single LOCI parameter (and all the other parameters in the model) on the pooled dataset (i.e. including data for all operators), the CC is explicitly assuming that the coefficient on LOCI for BMI for treatment W4210 is actually the same as the coefficient on LOCI for (say) Spire on treatment W3712. This is an assumption that concentration has the same relationship with self-pay prices for these operators and treatments.

9.20 A proper analysis of the question would report explicit tests down from the general model (where the parameters by treatment and operator were different) to the model which imposes the restrictions on the parameter values (like the CC's preferred model). If the parameter restrictions are statistically "accepted" on the full dataset, this would suggest that the CC's approach of applying restrictions on the parameter values could be justified. However, if restrictions are not "accepted" on the dataset, the CC's approach of applying restrictions on the parameter values would not be a valid approach. In our response to PCA1, we showed that the CC's approach of restricting the LOCI coefficient to be identical across operators was simply not supported by the data. Rather the restrictions (and hence the CC's model) were unambiguously rejected by the data.

84 P. Davis, S. Holbrook and E. Langer, "Comments on the Competition Commission's Price Concentration Analysis", 20 May 2013, paragraph 4.4 et seq.

9.21 For the avoidance of doubt, if the statistical tests on the model estimated on the full sample suggest that the parameter restrictions being imposed (to enforce the same relationship between prices and concentration across treatments and operators) are not affecting the true patterns in the data, then we would agree with the CC that it is better to use all the data and restrict the coefficients on the grounds of efficiency. However, if the statistical tests are not accepted, then that would suggest that the CC is choosing to disregard the actual patterns in the data and imposing *a priori* restrictions which the data are actually telling us should not be imposed.

9.22 Irrespective of the CC's desire for there to be a single broad relationship between self-pay prices and concentration (LOCI) in all treatments and for all operators, the CC has to accept the fact that the data may simply reject that idea. In truth, the relationship between LOCI and self-pay prices may well not be the same for all operators. If the CC's conclusions are to be genuinely led by the evidence, the CC must not be considering results based on averages (and relationships based on averages), unless the statistical tests on the model tell the CC that it is legitimate to restrict the model in this way. A desire for "simplicity" should not over-ride the reality of complexities in the data.

Statistical tests of parameter restrictions do not require only using subsets of data

9.23 In response to submissions about the earlier PCA analysis, the CC states in its Provisional Findings (at paragraph 6.200) that:

> *"[the CC] disagreed with the parties' interpretation of the results of [the CC's] analysis applied separately for each operator. As [the CC has] explained in Appendix 6.9, [the CC] considered that the analysis applied separately for each operator is a weaker approach compared with the analysis pooled across operators, and that it is likely to produce less precise and less reliable estimates. When applied separately for each operator, [the CC] found that the results of [the CC's] analysis did lack precision and, moreover, [the CC] did not consider these results to contradict the results described above [from the pooled model]."*

9.24 This paragraph is, on our analysis, setting out a false argument and then dismissing it. There is no reason whatsoever that the CC should only consider this issue by looking at only subsets of the data. It can be reassuring to look at subsets of the data and get the same results, but in this case, the CC does not get the same results when considering the subsets of the data.

9.25 Whether these assumptions - embodied as parameter restrictions - are valid is testable using statistics applied on the full dataset. The right question then is whether the full dataset statistically "accepts" these restrictions, or whether the data is telling us that the relationship is different across hospital operators/treatments. The answer is that the CC can and should test its assumptions using the full dataset. Our criticism does not require each operator to be looked at individually using subsets of the dataset only.

Testing the CC's implicit parameter restrictions on the pooled data

9.26 As noted above, the CC's preferred specifications include a restriction requiring the LOCI coefficient to be equal across operators and treatments. In this section we test whether this restriction is valid.

9.27 We first consider whether the LOCI coefficient should be restricted to be equal across the five operators. We do so by "interacting" the LOCI variable with dummy variables for each of the five main operators, creating a general model from which the parameter restrictions can be tested. We applied this change to the CC's preferred specifications L3, L7, FC3, and FC7.[85] Table 8 presents the results for regressions using LOCI as the concentration measure, and Table 9 presents the results using fascia count as the concentration measure (for brevity, we present only the coefficients on fascia count 0-9mi for FC3).

Table 8: Self-pay LOCI by operator

Self-pay LOCI by operator	L3		L7	
	Coeff	Std Error	Coeff	Std Error
BMI	0.1966	0.1427	0.0376	0.1685
HCA	1.4991	1.3109	3.3477**	1.3948
Nuffield	-0.1926***	0.0724	-0.2208***	0.0766
Ramsay	-0.081	0.1368	-0.0106	0.1318
Spire	-0.1442	0.0892	-0.2195***	0.0823
R-squared	0.92		0.91	
N	10874		10874	

*Source: do-file "Section 9 – LOCI parameter restrictions.do", provided to the CC in the Data Room. ***/**/* indicates statistical significance at the 1%/5%/10% level.*

85 In order to estimate the IV regressions, we interacted the relevant instruments with the operator dummies in order to ensure identification.

Table 9: Fascia count 0-9mi by operator

Fascia count 0-9mi by operator	FC3		FC7	
	Coeff	Std Error	Coeff	Std Error
BMI	0.0183**	0.0088	0.0059	0.0190
HCA	0.0644***	0.0113	0.2579***	0.0403
Nuffield	-0.0098	0.0183	-0.0431**	0.0181
Ramsay	-0.0243	0.0249	-0.0082	0.0459
Spire	-0.0211	0.0137	-0.0428**	0.0186
R-squared	0.92		0.91	
N	10874		10874	

*Source: do-file "Section 9 – LOCI parameter restrictions.do", provided to the CC in the Data Room. ***/**/* indicates statistical significance at the 1%/5%/10% level.*

9.28 The results for L3 show that, consistent with our findings in PCA1, Nuffield is the only operator that has a negative and statistically significant coefficient on the self-pay LOCI variable. The coefficients for the other hospital operators are not statistically significant.

9.29 In L7, we again find that Nuffield has a negative and statistically significant coefficient, as does Spire. The point estimates of the coefficients for BMI and Ramsay are close to zero and statistically insignificant, implying that there is no evidence for a relationship between price and concentration for these operators. The coefficient for HCA is positive and significant, although we note that HCA has only four hospitals in the estimation sample.

9.30 The fascia count results are similar, although with positive and significant coefficients appearing for BMI and HCA and no statistically significant effect for Nuffield in FC3.

9.31 In sum, these estimation results are consistent with a problematic price-concentration relationship for Nuffield and Spire but not for BMI and Ramsay. In order to test whether a broad price-concentration relationship applies across the industry, we test whether the coefficients on the individual operators' LOCI variables are jointly equal. Table 10 presents the results. We find that the data rejects the notion that the coefficient on LOCI (or fascia count) is the same across operators. Of course the finding that Nuffield and Spire have a problematic relationship between price and concentration should be interpreted very cautiously given the other flaws we have identified with the CC's Instrumental Variable approach.

Table 10: Test of joint equality of LOCI/fascia count coefficients by operator

Base specification	Test that operator coefficients are jointly equal (p-value)
L3	0.0910
L7	0.0316
FC3	0.0000
FC7	0.0000

Source: do-file "Section 9 – LOCI parameter restrictions.do", provided to the CC in the Data Room. Note: we present the results of the test that the coefficient on fascias within 0-9 miles is equal across operators. The result of the more restrictive test that can be applied to FC3, that the coefficients on each of the three fascia measures are equal across operators, is rejected with a similar p-value.

9.32 We next consider whether the coefficient on LOCI should be allowed to vary by treatment. Adopting a similar approach to that above for operator, we find that this restriction is not rejected by the data for the specifications using LOCI as the concentration measure but is rejected for the specifications using fascia count. For brevity, we present in Table 11 only the results of the test of equality of coefficients.

Table 11: Tests of joint equality of LOCI/fascia count coefficients by treatment

Base specification	Test that treatment coefficients are jointly equal (p-value)
L3	0.6486
L7	0.1834
FC3	0.0000
FC7	0.0097

Source: do-file "Section 9 – LOCI parameter restrictions.do", provided to the CC in the Data Room. Note: we present the results of the test that the coefficient on fascias within 0-9 miles is equal across treatments. The result of the more restrictive test that can be applied to FC3, that the coefficients on each of the three fascia measures are equal across treatments, is rejected with a similar p-value.

9.33 In sum, the results of the tests presented above show that the CC's desire for a "broad price-concentration relationship" to exist is rejected by the data.

The CC's comments on the econometric analysis submitted on behalf of BMI

9.34 At paragraphs 97-103 of Appendix 6.9, the CC discusses the econometric analysis submitted by Compass Lexecon on behalf of BMI. The conclusion of that paper authored by Compass Lexecon economists ("Do private healthcare providers have market power in solus hospital markets?", dated 11 January 2013) was that there is no systematic effect of solus status of a hospital on self-pay prices; rather, the evidence indicated that self-pay prices are sometimes higher, sometimes lower, and sometimes no different in specific solus hospitals compared to an average non-solus hospital.

9.35 However, the CC misrepresents and misinterprets the conclusions of the analysis.

9.36 First, coming to its view that self-pay episode prices at solus BMI hospitals are, on average, between [✂] than at non-solus BMI hospitals for three of four treatments, the CC appears to be looking across different specifications without discriminating between them. The [✂] figure appears to be derived from the results presented in Table 7, in which a statistically significant effect is found for two treatments, whereas the [✂] figure appears to be derived from the robustness checks presented in Tables 32-35.

9.37 Second, the CC considers only the regressions that do not allow for the effect of solus status to vary by hospital in order to conclude that solus hospitals systematically charge higher prices than non-solus hospitals. However, the regressions reported in Table 8 show that the solus effect varies by hospital (both in terms of sign and significance). Indeed, the hypothesis that the solus effect is uniform across solus hospitals is forcefully rejected by the data. Accordingly the CC's contention that there is a general price-solus relationship must also be roundly rejected by the CC. Moreover, there is no systematic effect of solus status on self-pay prices across treatments either (in a pooled regression with solus-treatment interactions, the restriction that the parameters on solus indicator are equal for each treatment is rejected by the data).[86]

9.38 Third, the CC ignores the fact that even among the statistically significant results, for one of the four treatments prices are [✂] lower at solus hospitals, which is entirely inconsistent with the CC's conclusions that there is a positive relationship between solus status and prices.

[86] The analysis showing the rejection of parameter restrictions is available to the CC on request.

9.39 Finally, as an aside, we note also that to our knowledge the CC has never requested any of the underlying data or computer codes to properly test or consider this evidence. The CC has been unwilling to allow us to present the results of our investigations to even the economics team at the CC working on the case, contrary to the CCs approach in at least some other cases. Indeed the CC has also never considered most of the substantive results in the solus paper; for example CC the has not considered the results indicating that solus hospitals tend to have lower nearby populations, lower capacity utilization and lower margins than non-solus hospitals. We do not consider that such an approach to economic evidence represents the CC at its best. Indeed, more generally, many of the mistakes and issues in the CC's work that we identify in this report could undoubtedly have been discussed and addressed, had the CC chosen to take a more open and less defensive approach early on in the investigation.

Section 10

The CC disregards standard tests for model misspecification

10.1 This section explains that the CC is inappropriately disregarding standard statistical tests for model misspecification, in particular Ramsey's "Regression Specification Error Test" [RESET], a widely used test of model specification error. We find that the CC's preferred OLS specifications, L3 and FC3, both fail the RESET test, and the CC's preferred IV estimation for LOCI (L7) also fails an IV implementation of the RESET test. Whilst the CC does not report these test results in Appendix 6.9, it invests significant effort (at paragraphs 79-81 of Appendix 6.9) in arguing why it should not rely on the test.

10.2 At paragraph 80 in the Appendix 6.9 of the CC's Provisional Findings report, the CC states:

> *"The parties have applied the RESET test and focused on whether the test result is a 'pass' or 'fail'. In our view, this approach does not address the issue at hand-that is, whether our main results are robust to the consideration of more flexible and complex specifications, and/or whether our specification can be improved. We also note that the RESET test is an exhaustive and data-driven test that is very demanding of the data in this case. By this, we refer to the way in which the test includes additional covariates that are squared, cubic and quartic in the original covariates, as well as interacted versions of these covariates and their higher powers. The parties have not argued why such complex relationships are to be expected, or, moreover, why not accounting for these potential complexities would bias the price-concentration relationship that we have estimated. Finally, we find that the RESET test cannot be applied to specifications L7 and FC7. In these cases, the software program returns an error message indicating that the specification contains too many interrelated variables. This is a likely consequence of the exhaustive nature of the test noted above."*

10.3 The CC has in saying this clearly rejected the longstanding standard model specification test used by econometricians for now more than 40 years, the Ramsey (1969)[87] RESET test which is a standard misspecification test.

10.4 Indeed, the CC has used the RESET test in previous investigations. For instance:

- In the Local bus services market investigation, the CC applied the RESET test in econometric analysis of asset betas, noting that "[…] these results should be taken with caution for the following reasons: [two reasons are presented]. In addition, we rejected the hypothesis that the model has no omitted variables at a 10 per cent significance level (RESET test). This implies that there is statistical evidence that the OLS estimator may be biased."[88] In addition, when considering parties' comments on the "performance-concentration analysis", a number of parties noted that certain specifications failed the RESET test. The CC's response to these comments does not contend that the test is inappropriate.[89]
- In the Rolling Stock Leasing market investigation, the CC applied the RESET test to econometric analysis of capital rentals and competition, finding that "The Ramsey regression specification error test (RESET) is used to test that we have specified the correct functional form for our regression model, in particular testing for the presence of omitted variable bias. [Footnote omitted] Low values (less than 0.1) of the p-values presented imply that we may not have correctly specified our model, often due to the omission of relevant variables from our model. Our preferred model did not display any evidence of misspecification or of omitted variables."[90]
- The CC has also reported the results of the RESET test on estimation results in a number of other enquiries.[91]

87 J. B. Ramsey (1969) "Test for Specification Errors in Classical Linear Least Squares Regression Analysis" Journal of the Royal Statistical Society, Series B, Vol. 31, pp350-371.

88 Local bus services market investigation final report, Annex B to Appendix 10.2, paragraph 12.

89 Local bus services market investigation final report, Annex B to Appendix 7.1, paragraphs 60 and 66.

90 Rolling Stock Leasing market investigation final report, Appendix 6.2, paragraph 32.

91 See Groceries market investigation (2000), Appendix 10.2; *Napier Brown Foods plc / James Budget Sugars Ltd*, Appendix I; and the Bristol Water plc price determination, Annex 6.

The validity of the CC's stated concern about RESET tests

10.5 Long and Trevedi (1992)[92] write: *"If the functional form is incorrect, estimates from the incorrect model are at best approximations to those from the correct model and interpretation of the estimates should be avoided."*

10.6 However, the CC takes issue with this basic proposition and instead argues that "In our view, this approach [of using the RESET test] does not address the issue at hand - that is, whether our main results are robust to the consideration of more flexible and complex specifications, and/or whether our specification can be improved." (paragraph 80).

10.7 The CC then presents in Table 12 the LOCI coefficients estimated on a variety of specifications. First we note that many of these specifications simply exclude variables that the CC is currently including so they do not appear to be in the spirit of a test with power relative to omitted variables. As such these sets of results can hardly constitute a coherent test for (e.g.) omitted variables.

10.8 The CC does present three results which include additional functions of variables included in the CC's regression specification. These three specifications are described as:

- "include additional interactions";
- "include additional squared terms"; and
- "Include additional interactions and squared terms".

10.9 The CC reports in the notes to Table 12 that "Additional interaction terms include operator and treatment interactions, treatment and length of stay interactions, and treatment and age interactions. Additional squared terms are for patient age, number of nights, average direct cost, and each local area characteristic variable."

10.10 This is a very limited set of additional interactions and squared terms and does not approach the power of the RESET test. Moreover, the CC does not interpret the results of these additional specifications in the normal manner. The regression results indicate that at least some of the interactions and squared terms are statistically significant, and as such may be evidence of variables omitted from the CC's preferred specifications.

92 Long, J. S. and Trevedi, P. K. (1992) "Some Specification Tests for the Linear Regression Model" *Sociological Methods and Research*, Vol. 21, No. 2.

10.11 In addition we note that in Table 12, the additional of non-linear terms reduces the magnitude of the coefficient on LOCI from -0.15 to -0.136 or -0.137 in two specifications. Since the CC seeks to rely in the Remedies Notice on the magnitude of the price effects from the PCA2 analysis, the addition of suitable non-linear terms can potentially have a material effect on the proportionality assessment. In particular a 0.2 change in LOCI would be associated with a 3.0% price effect for the CC's baseline model but only a 2.72% price effect in the final specification of Table 12. Whilst the change in the estimated price effect is small in absolute terms, it could be material to the CC's calculations at the remedies stage since it reflects a 10% reduction in the magnitude of the computed gain from intervention to reduce the LOCI by 0.2.

10.12 At paragraph 80 the CC "also note that the RESET test is an exhaustive and data-driven test that is very demanding of the data in this case. By this, we refer to the way in which the test includes additional covariates that are squared, cubic and quartic in the original covariates, as well as interacted versions of these covariates and their higher powers. The parties have not argued why such complex relationships are to be expected, or, moreover, why not accounting for these potential complexities would bias the price-concentration relationship that we have estimated."

10.13 First we note that the CC appears to be misunderstanding the RESET test. RESET works when the additional variables indicate that the error term in the base model is correlated with an omitted variable. (If the additional variables are not correlated with an omitted variable then the test may not have power.) The idea is not to propose that the correct model involves every squared, cubed and quartic term; rather that adding such variables may (sometimes) help detect misspecification including omitted variables.

10.14 Second, we note that this appears to suggest the CC has a particular implementation of the RESET test in mind. RESET need not, as the CC claims, be an "exhaustive" test – there are different variants. In fact the basic RESET test is quite a restrictive test because, while in principle a regression specification test could include all squares, cubes and interaction terms for every variable in a regression's specification, the basic RESET test actually restricts heavily the way those terms can enter the test – contrary to the CC's description of the test.[93]

10.15 In our response on PCA1 we for example reported that the CC's specification rejected a very parsimonious version of the RESET test based on only the square of the fitted value from the CC's baseline regression. Yet the CC's model still failed this restrictive implementation of the test.

[93] Thus the RESET test only allows inclusion of linear combinations of the regressor variables where the linear combination relies on the estimated coefficients (using standard notation, calculated as $\hat{y} = x'\hat{\beta}$.)

10.16 Third, we note that the CC's stated reasoning – that RESET is "an exhaustive and data-driven test that is very demanding of the data" – amounts to a concern that the "size" of the RESET test is higher than it should be for some reason not given. Recall that the "size" of a statistical test is defined as the probability that the test rejects a null hypothesis (here of a correct model specification) when the model is in truth well specified. A statistical test might, for example, be applied using a 95% level of significance – which corresponds to a test with a 100-95% = 5% size because 5% of the time an analyst would reject the null hypothesis even if it is true.

10.17 We have considered this issue by looking at the statistical and econometric literature on the topic. However, our review of the literature has found no grounds that would justify the CC's concern. Instead, we have found a range of academic work which suggests that the size properties of the RESET test are good in the kind of sample sizes that the CC has available. A brief review of the literature we have explored is provided in Annex B.

10.18 To summarize our conclusions from Annex B, our brief review of the literature examining the effectiveness of the RESET test has not provided any grounds for the CC's apparent concern for ignoring the standard regression specification test in economics. On the contrary, rather we find support for the proposition that the size properties of the RESET test are generally good – certainly in samples larger than 250 – and that the power of the test to detect specification errors increases with sample size as expected and also with the severity of the specification errors. We also find published academic papers in peer reviewed journals that positively advocate the use of multiple variants of the RESET test.

10.19 Our review of the econometric literature on the RESET test has found no coherent basis for the CC's stated concern with using the RESET tests. We have however found a variety of published papers which clearly recommend its use over the space of four decades. Moreover we have found papers in the literature which suggest that while the *size* properties of the basic RESET test are basically good, the *power* (the ability to reject a false hypothesis of correct specification) is greater for the variants of the test which include all of the squared, cubic and quartic terms (the variant of the RESET test that the CC critiques in particular).

10.20 Long and Trevedi (1992) state that "if the functional form is incorrect, estimates from the incorrect model are at best approximations to those from the correct model and interpretation of the estimates should be avoided," and go on to recommend using the RESET test for specification testing. The CC ignores this recommended approach.

10.21 The CC's results presented in the Data Room show that the CC did in fact run the RESET test on its preferred specifications L3 and FC3 and found that these specifications failed the RESET test. The CC did not however report that its specifications fail the RESET test.

10.22 The CC does remark: "Finally, we find that the RESET test cannot be applied to specifications L7 and FC7. In these cases, the software program returns an error message indicating that the specification contains too many interrelated variables. This is a likely consequence of the exhaustive nature of the test noted above." It is common for statistical package implementations of tests not to be robust to every situation an analyst comes across and there is no compelling reason to interpret this error as indicative of a generic problem with the RESET test. We found in particular that this error message is specific to the CC's choice of implementation of the RESET test.

10.23 The command used by the CC, *ivreset*, provides the option of reporting a "C-test statistic" instead of the default Pesaran-Taylor (1999) option, which returns an error. All of the CC's IV regressions using LOCI (i.e. L4 to L7) fail this test.[94] Alternatively the CC could have implemented the RESET test by testing a more limited set of quadratic, cubic and quartic variables. The literature suggest that quartic is the most powerful approach but other implementations would use just quadratic and cubic for example. As we describe in Annex B, Wooldridge (2010)[95] is generally supportive of the RESET test but does note that it can "consume many degrees of freedom" (page 137) and suggests that the original Ramsey (1969) implementation can be used in such an eventuality.

94 See do-file "Section 10 – IV RESET tests.do", provided to the CC in the Data Room.

95 Wooldridge (2010) "Econometric Analysis of Cross Section and Panel Data".

Annex A

Construction of local area characteristics variables

A.1 This annex provides full details of the analysis outlined at paragraph 4.21 et seq. The underlying data and analysis is available at the CC's request.

Reconstructing the CC's local area characteristics variables at the NUTS3 level

A.2 We use ONS data on Gross Disposable Household Income (GDHI) per head at current basic prices published at the NUTS3 level.[96]

A.3 We construct the population density and average age variables from 2011 Census data published by NOMIS.[97] We obtain the data at the Local Authority District level and aggregate to the relevant NUTS3 region using the lookups provided by the ONS.[98] We confine the analysis to England and Wales only in line with the availability of census data at the postcode sector level (see below).[99]

A.4 Population density is provided in population per hectare. The data on age provides population by single year of age (from zero to an open-ended category of 100+). We assume

96 Source: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-298694

97 Sources: http://www.nomisweb.co.uk/census/2011/qs102ew and http://www.nomisweb.co.uk/census/2011/qs103ew.

98 Available from https://geoportal.statistics.gov.uk/geoportal/catalog/content/filelist.page. Note that two NUTS3 regions had Local Authority District codes that were mismatched with the lookups chart. The census data specified E07000100 for St Albans and E07000104 for Welwyn Hatfield, whereas the lookup identified the districts as E07000240 and E07000241, respectively.

99 Our analysis is intended to illustrate the difference in local area characteristics between hospitals' catchment areas and NUTS3 regions and so we confine the analysis to data that is readily available. This does not diminish the importance of performing the PCA across the full dataset, including Scotland.

the actual ages of the population are uniformly distributed within each age category[100] in computing the average age in each NUTS3 region.

A.5 To map BMI hospitals to the appropriate NUTS3 region we use the ONSPD.[101]

Computing the population density and average age for BMI hospitals' catchment areas

A.6 We construct the BMI hospitals' catchment areas using information on drive distance from each hospital to postal sector supplied by BMI, and use the data given in Appendix 6.6 of the CC's Provisional Findings. We assume that a catchment distance of (say) 5 includes all postal sectors up to 5.5miles driving distance. This gives a list of postal sectors within the catchment area for each BMI hospital. Information on catchment drive distance is not presented in Appendix 6.6 for twelve BMI facilities: eight outpatient facilities (9 Harley Street, City Medical, Evesham Outpatient, Heath Lodge Clinic, Paddocks Clinic, Southend, Sutton Medical Centre, and Syon Clinic) and four hospitals (Blackheath, Fitzroy Square, London Independent, and Weymouth Hospital). A limitation of the drive distance data leads us to exclude McIndoe from the analysis.[102]

A.7 We retrieve census data at the postal sector level from NOMIS and combine with the catchment area information. We note that census data at the postal sector is not yet available for Scotland. We find that certain postal sectors are not matched. We believe that this is due to (i) non-geographic/special postcodes, e.g. B99; (ii) confidentiality thresholds, whereby data is anonymised for very small units. We note that the total population recorded in England and Wales in the census data by postal sector matches other sources and reassures us to the completeness of the data at the postal sector level. As an additional check, we compare the census data to the ACORN measures (described below), for which the match is exact and which includes "base population". We find that almost half of the postal sectors which are not found in the census data have zero population according to the ACORN data; that the remainder have very small populations (a mean of just 73); and that the ACORN and census measures of population are within 1% of each other for all catchment areas. This provides reassurance that the census data is complete.

---

100 So that we assume that the "age 30" group have an average age of 30.5. We assume that persons with ages of 100+ have an average age of 100.5.

101 Available from https://geoportal.statistics.gov.uk/geoportal/catalog/content/filelist.page.

102 The drive distance data supplied by BMI provides the drive distance from each postal sector in the UK to the nearest 100 private medical facilities. In the London area, there can be 100 facilities within 35 miles of a given postal sector and we have no information on the drive distance to facilities beyond this. Given the wide catchment of McIndoe, it is possible that there are postal sectors which lie within McIndoe's catchment area but for which we do not have information on the drive distance to McIndoe.

A.8 This enables us to compute average age[103] and population density for the catchment areas.

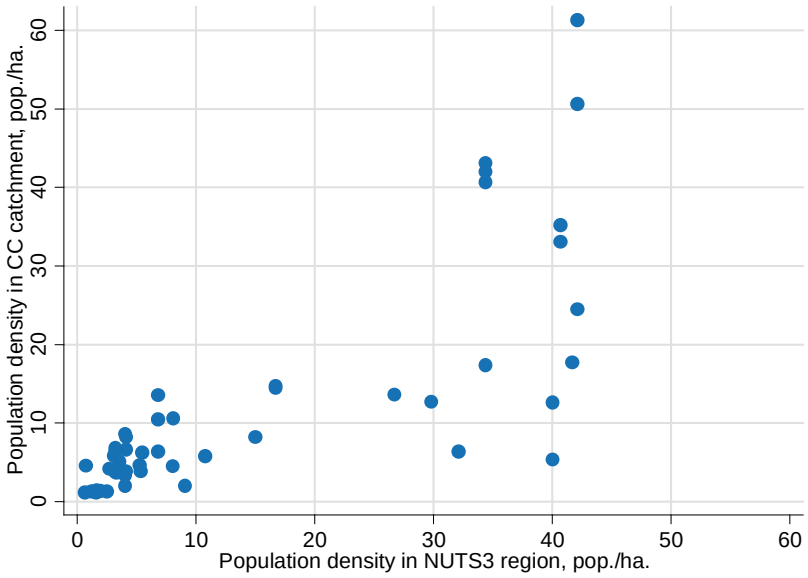
Computing the ACORN measures for BMI hospitals' catchment areas

A.9 BMI provided us with the base population and number of persons falling within each ACORN category for each postal sector in the UK. We used the catchment area data detailed above to compute the proportion of the population falling into each ACORN category within each BMI hospital's catchment area.

Results for population density and average age

A.10 Figure 4 and Figure 5 present comparisons of population density and average age respectively between the CC's catchment area around BMI hospitals and the NUTS3 regions within which the hospitals are located.

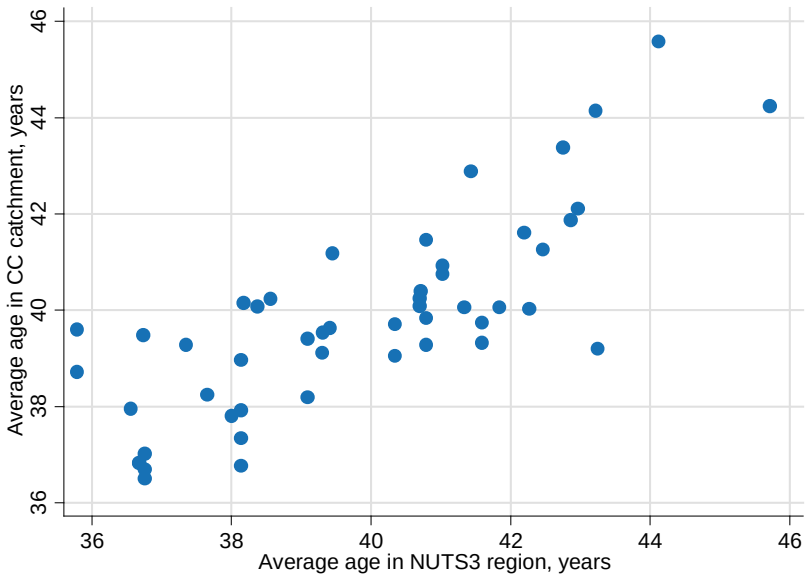
Figure 4: Population density in CC's catchment area and NUTS3 region



Source: Compass Lexecon analysis.

103 Note that we assume that a person with age at last birthday of (e.g.) 0 has an average age of 0.5, and that persons with ages of 100+ have an average age of 100.5.

Figure 5: Average age in CC's catchment area and NUTS3 region



Source: Compass Lexecon analysis.

Annex B

A brief literature review on the RESET test

B.1 In response to the CC's concerns around the use of the RESET test, we have sought to examine the econometric literature to consider whether we can see any concern being expressed around the *size* (the probability that the null hypothesis of no specification error is rejected even though the null is true) of the range of RESET tests in single equation models. In the time available we have found a variety of papers which undertake Monte Carlo studies on the RESET test but we can find nothing in the academic literature that would appear to justify the CC's concern.

B.2 We begin our summary with the older studies.

B.3 First we do find some discussion of the power of the test – that is the likelihood that a false null hypothesis of a correct specification will be correctly rejected. There is for example a discussion in Ramsey and Gilbert (1972)[104] who study the power of the basic RESET (and also two other tests) to pick up forms of misspecification including (i) omitted variables, (ii) incorrect functional form, (iii) endogeneity, and (iv) heteroskedasticity. Those authors conclude that RESET is the most powerful of the three tests against misspecification of the first order properties of the model. Thursby and Schmidt (1977)[105] discuss situations when the power of the RESET test may decline and propose using joint tests of squared, cubed and fourth powers of all explanatory variables instead of a RESET test to address the potential concern. These kinds of results motivated the variants of RESET that are used which include all of the powers of the variables in a regression. However, their concern in doing so is expressly about low power – too few rejections – rather than the concern the CC raises in paragraph 80 about the size of the test. To improve the power of the test, these authors actively suggest using up to fourth powers of the explanatory variables rather than the restricted version of just using squares and/or cubes of the fitted values (which was Ramsey's original 1969 proposal).

---

104 J. Ramsey & R. Gilbert (1972) A Monte Carlo Study of Some Small Sample Properties of Tests for Specification Error, *Journal of the American Statistical Association*, 67:337, 180-186

105 Jerry G. Thursby & Peter Schmidt (1977) Some Properties of Tests for Specification Error in a Linear Regression Model, *Journal of the American Statistical Association*, 72:359, 635-641

B.4 Thursby (1979)[106] extends the Monte Carlo study by Thursby and Schmidt (1977) to compare the RESET test with a number of tests for autocorrelation noting that such tests had (at the time) "often been used as a means of testing overall model specification as well as a means of detecting autocorrelation." He concludes positively: *"Our results indicate that RESET has substantial power in detecting a nonzero disturbance mean and is more powerful than a variety of alternative tests. In addition, it is robust to autocorrelated disturbances."* There appears to be nothing in the Thursby (1979) paper that would justify the CC's approach.

B.5 We next consider the study by Long and Trevedi (1992)[107] who note that standard form of the RESET test assumes homoscedasticity (and so can be sensitive to heteroskedasticity) but note that several authors have proposed a RESET type test known as Robust RESET tests. They find in their Monte Carlo experiments that Godfrey's (1988) suggestion of a Robust RESET test sometimes has poor properties but that the LM RESET test following Davidson and MacKinnon (1985) has much better properties.[108] More specifically:

- Under a correct model specification without heteroskedasticity, the size properties of the standard RESET test are "quite good" (page 184) and while they note issues with the LM test with a sample size of less than 100, they conclude that by the time the sample includes 100 observations the size properties of the LM test approximate those of the standard RESET test, i.e. "quite good". (See their discussion of their Figure 2.)
- Next they consider the performance of the RESET tests in the presence of heteroskedasticity. In this case, the LM RESET test outperforms the standard RESET test as expected and the authors note: "The robust LM version of the test is not affected by heteroskedasticity and has reasonable size properties beginning with n=50."
- With non-normal errors, the standard RESET test is found to have similar properties to as when no misspecification is present, and by n=1,000, the empirical alphas [sizes] are almost exactly equal to the nominal alphas.

106 Jerry G. Thursby (1979) "Alternative Specification Error Tests: A Comparative Study", *Journal of the American Statistical Association*, 74:365, 222-225

107 Long, J. S. and Trevedi, P. K. (1992) "Some Specification Tests for the Linear Regression Model" *Sociological Methods and Research*, Vol. 21, No. 2.

108 Godfrey, L. G., *Misspecification tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. New York: Cambridge University Press, 1988; and Davidson, R. and J. MacKinnon (1985), "Heteroskedasticity-Robust Tests in Regression Directions", *Annales de l'INSEE* 59/60:183-218.

- In the presence of small amounts of non-linearity, the authors find their empirical alphas are "similar to those for data without misspecification" (page 188) and while they find some differences between the standard RESET test and the LM RESET test, "by n=100 the differences are small." (page 188) And with greater non-linearity "the tests behave similarly, with the ability of the test to detect nonlinearity increasing steadily with sample size." Again the RESET test has larger empirical alphas [size] than the LM RESET test but "by n=100 the percentage difference is small."

B.6 In short, there appears to be nothing in Long and Trevedi's (1992) study which should provide the CC with a reason for concern that the size of the RESET test is a problem – i.e., that there is any reason to consider that either the standard or LM versions of the RESET test should over-reject the null hypothesis of correct specification if it is in fact true that the model is correctly specified.

B.7 Long and Trevedi's (1992) study is also enlightening with regard to the effectiveness of the RESET tests to detect an omitted variable. They conclude in particular that: "Both tests are extremely powerful by n=250, with the LM version again being somewhat less powerful for all samples" (page 190). They conclude that "the standard RESET and the robust LM RESET tests are very effective for detecting first order specification error. In the absence of heteroskedasticity the standard RESET test has slightly better size properties and is slightly more powerful. However, it is affected by heteroskedasticity while the LM version of the test is not. Given the likelihood of heteroskedasticity in real world data and the relatively slight disadvantages of the LM test, the robust LM RESET test appears to be the preferred test."

B.8 Finally we note that a small literature has fairly recently considered the small sample properties of the RESET test in systems of equations. For example Shukur and Edgerton (2002)[109] considered the size properties of the Rao (1973)[110] F-test variant of the RESET test and concluded their discussion of size on page 917 that "the RAO test is clearly superior to all the other alternatives, with only one result (of 30) outside the 95% confidence interval. The next best test is the Edgeworth adjusted likelihood ratio test, LRE, which performs well in systems up to 3 equations, but then gradually begins to deteriorate in small samples." Clearly since the CC has one equation being estimated, these results should be reassuring for the size properties of the RESET test. Mantalos and Shukur (2007)[111] similarly study the size and power of RESET tests in systems of equations (ranging from 1 to 10). Looking at their Table 5, which reports their "Estimated size for the alternative RESET tests at 5% nominal size", in the column that corresponds to only one equation all of the actual sizes calculated for their tests are in the range 4.7%-5.1% for the Rao (1973) implementation of

109 Ghazi Shukur & David Edgerton (2002) The small sample properties of the reset test as applied to systems of equations, Journal of Statistical Computation and Simulation, 72:12

110 Rao, C. R. (1973). Linear Statistical Inference and its Applications, 2nd ed. Wiley, New York.

111 Mantalos and Shukur (2007) "The Robustness of the RESET Test to Non-Normal Error Terms."

the RESET test. These authors find their LRE and LRT-C tests have similar good size properties for the single equation case even in very small samples. While some of the discussion in the paper is critical and describes the outcomes as "bad", these concerns all relate to estimating large numbers of equation systems using very small sample sizes. With more than 75 degrees of freedom (i.e. still a pretty small sample), all of the RESET tests report a size for the 1 equation case in the range 4.8%-6%. In short, there appears to be nothing in this paper that should cause the CC concern about the size properties of RESET tests, at least in this particular single equation context.

B.9 Long and Trevedi (1992) conclude their paper on misspecification with their recommended approach to specification testing: "The following strategy for testing should provide a useful assessment of one's model. First the standard RESET test and the robust LM RESET test…should be applied. If both tests pass, suggesting an adequate conditional mean specification, the IM test of higher order moment misspecification should be used. If the standard RESET test fails, but the robust version passes, correct functional form is indicated but higher order misspecification seems likely."

Conclusion on the CC's concerns around RESET tests

B.10 In summary, our brief review of the literature examining the effectiveness of the RESET test has not provided any grounds for the CC's apparent concern for ignoring the standard regression specification test in economics. On the contrary, rather we find support for the proposition that the size properties of the RESET test are generally good – certainly in samples larger than 250 – and that the power of the test to detect specification errors increases with sample size as expected and also with the severity of the specification errors. Published academic papers in peer reviewed journals positively advocate the use of particular variants of the RESET test, although there is some discussion of exactly which variant is best.

Data Room addendum

B.11 In the Data Room the CC provided a number of standard econometrics textbooks. We reviewed what those textbooks said about the RESET test.

Wooldridge (2010) "Econometric Analysis of Cross Section and Panel Data"

B.12 Wooldridge (2nd Edition) begins quite positively, stating at page 137 that: "Sometimes we need a test with power for detecting neglected nonlinearities in models estimated by OLS or 2SLS. A useful approach is to add nonlinear functions, such as squares or cross-products, to the original model."

B.13 However he also notes that "Putting in squares and cross-products of all exogenous

variables can consume many degrees of freedom" (page 137) so he advocates "An alternative is Ramsey's (1969) RESET, which has degrees of freedom that do not depend on K [the number of regressors in the model]". Thus Wooldridge suggests using the original Ramsay (1969) approach rather than the Thursby and Schmidt (1977) approach of adding all of the squares, cubes and quartic terms. This of course is a choice about the variant of the RESET test being used – not a critique of using the RESET test itself.

B.14 More critically he also goes on at page 138 to say: "There is some misunderstanding in the testing literature about the merits of RESET. It has been claimed that RESET can be used to test for a multitude of specification problems, including omitted variables and heteroskedasticity. In fact RESET is generally a poor test for either of these problems. It is easy to write down models where an omitted variable, say q, is highly correlated with each x, but RESET has the same distribution that it has under H0 [the null hypothesis of correct specification]." He then provides a specific example: "A leading case is seen when E[q|x] is linear in x. Then E[y|x] is linear in x [even though E[y│x]≠E[y|x,q]], and the asymptotic power of the RESET equals its asymptotic size."

B.15 The important point about Wooldridge's critique (with respect to omitted variables) is that Wooldridge is saying the RESET test will only have *power* against (in particular) omitted variables only under potentially strong assumptions. That means that the RESET test may unfortunately have very low power sometimes – that is it will only detect a problem when there truly is a problem some of the time. Of course this critique should not provide the CC reassurance that its critique in Provisional Findings is a valid one, since the CC's critique is a claim that the RESET test will find a problem too much of the time, whereas *Wooldridge is concerned with the opposite problem to the CC* (that RESET will not find omitted variables problems that truly do exist).

B.16 The second aspect of Wooldridge's critique is that RESET may not be a good test for heteroskedasticity – the situation when the variance of the error depends on the value of the regressor(s) in the model. Fortunately there are alternatives which test more directly heteroskedasticity and there appears to be consensus in the literature that it also makes sense to test for heteroskedasticity – once we have a model where we are reassured that the conditional mean assumptions are satisfied. Long and Trevedi (1992) for example propose this two-step approach. Wooldridge notes at page 138 that "for both OLS and 2SLS, heteroskedasticity does not affect the consistency of the estimators, and it is only a minor nuisance for inference."

Cameron and Trivedi (2005) "Microeconometrics: Methods and Applications"

B.17 This textbook considers the RESET test at page 277-8. The authors write "A common functional form misspecification may involve neglected nonlinearity in some of the regressors. Consider the regression y=x'β+u, where we assume that the regressors enter linearly and are asymptotically uncorrelated with the error u. To test for nonlinearity one straightforward approach is to enter power functions of exogenous variables, most

commonly squares, as additional independent regressors and test the statistical significance of these additional variables using a Wald test or an F test. This requires the investigator to have specific reasons for considering non-linearity, and clearly the technique will not work for categorical x variables." Thus Cameron and Trivedi do have some sympathy for the CC's view that for a functional form test there should be some reason to test for non-linearity. Of course, there is always such a concern and the CC do actually find (but do not report) that there should be some non-linearity in the model - so this is somewhat a moot point ultimately. Note that they also flag that the CC's approach in Table 12 cannot be applied to categorical variables such as CCL3 in the CC's model.

B.18 The next paragraph continues "Ramsay (1969) suggested a test of omitted variables from the regression that can be formulated as a test of functional form. The proposal is to fit the initial regression and generate new regressors that are functions of the fitted values $\hat{y} = x\hat{\beta}$ such as $w = [(x'\hat{\beta})^2, (x'\hat{\beta})^3 .., (x'\hat{\beta})^p]$. Then estimate the model $y = x'\beta + w'\gamma + u$, and the test of nonlinearity is the Wald test of the p restrictions, $H_0: \gamma = 0$ against $H_a: \gamma \neq 0$. Typically a low value of p such as 2 or 3 is used. This test can be made robust to heteroskedasticity." Thus this amounts to a description of the original Ramsey (1969) version of the test – and finally a reference to the variant that is robust to heteroskedasticity. We cannot see anything in this discussion to support the CC's concerns with the RESET test.

Green, W. (2003) "Econometric Analysis" (5th Edition)

B.19 The edition of the Green textbook available in the CC Data Room does not refer to RESET in its index. It does however discuss the topic of 'Specification analysis and model selection' at chapter 8. He notes in particular that: "There has been a shift in the general approach to model building in the last 20 years or so [...]. With an eye towards maintaining simplicity, model builders would generally begin with a small specification and gradually build up the model ultimately of interest by adding variables. But [...] we can surmise that just about any criterion that would be used to decide whether to add a variable to a current specification would be tainted by the biases caused by the incomplete specification at the early steps. Omitting variables from the equation seems generally to be the worse of the two errors. Thus, the **simple-to-general** approach to model building has little to recommend it. Building on the work of Hendry [e.g., (1995)] and aided by advances in estimation hardware and software, researchers are now more comfortable beginning their specification searches with large elaborate models involving many variables and perhaps long and complex lag structures. The attractive strategy is then to adopt a **general to simple**, downward reduction of the model to the preferred specification. Of course this must be tempered by two related considerations. In the "kitchen sink" regression, which contains every variable that might conceivably be relevant, the adoption of a fixed probability for the type I error, say 5 percent assures that in a big enough model, some variables will appear to be significant, even if "by accident." Second, the problems of pretest estimation and stepwise model building also pose some risk of ultimately misspecifying the model. To cite one unfortunately common

example, the statistics involved often produce unexplainable lag structures in dynamic models with many lags of the dependent or independent variables." (pages 151-152)

B.20 Thus for Green the CC's approach of simplicity first "has little to recommend it" while the general testing down approach is an "attractive strategy" – albeit one with acknowledged risks. Again, in the round, there is nothing here that is obviously supportive of the CC's approach to model specification.