[Not yet submitted for publication by the Royal Statistical Society.]

Challenging findings about the funding formula for Clinical Commissioning Groups from recently published NHS data

Mervyn Stone

University College London, London, UK

Summary. For over two decades statisticians and health economists have inconclusively crossed swords over the validity of the capitation funding formulas used by the Department of Health to guide the annual *per capit*a allocation of Primary Care Trusts (PCTs). In 2013, PCTs became Clinical Commissioning Groups (CCGs) and NHS England was made responsible for implementing the latest formula. Over the years, each year's current formula has been treated as the target *per capit*a allocation T towards which the actual allocation A should be moved when feasible. The 'distance from target' variable D = A − T is of particular concern to finance directors of below target CCGs. The National Audit Office (NAO) found that whether or not a CCG was below target was significantly associated with the financial surplus on its first-year account. My revision of the NAO analysis has revealed the statistical insignificance of D in a bivariate regression of surplus on A and D. This paper strengthens that finding with a combined analysis of the surpluses in the two CCG accounts for 2013-2016. NAO was unable to find any association between a CCG's above-below target status and three other measures of CCG performance. The paper introduces the concept of a 'statistical pirouette' and finds that NAO was not simply unlucky but that the dearth of association holds for all but a few of the 63 performance indicators for CCGs in the NHS Atlas of Variation 3.0. (The exceptional associations can be explained by features of indicator construction or by financial management respect for the funding formula.) The values of A and T in any particular year can be viewed as the labels on a couple of slowly-moving dancers in a historically extended pirouette around their average F = (A+T)/2 (the metaphor is feasible since A = F+½D, T = F−½D and the population-weighted average of D = A−T is zero in order to balance the books). The dearth of associations is based on significance tests of the OLS coefficient of D in 63 regressions of Atlas performance indicator I on the first-year values of F and D, in which I and F are regarded as fixed and the only assumption needed to justify the conventionally-calculated P-value is that the values of D have been, in effect, randomly allocated to the CCGs (whose number exceeds 180 for all 63 regressions). The findings of the paper are consistent with the harsh judgement of the current CCG funding formula by the statistical profession.

[Additionally: In 2014, the author was provoked to study the National Audit Office report on the end-of-year financial balance of 211 clinical commissioning groups for 2013-14 by how the report was interpreted by the Chair of the Parliamentary Accounts Committee. To his surprise, he found that the *per capita* target allocation did not help to 'explain' the size of the balance when the other 'explanatory' variable in a multiple regression was the distance of actual allocation from target. If statisticians are right to view the PBRA3-based target formula as largely nonsensical, the target variable ought to have helped to do just that, if only in the shape of a corrective sum with actual allocation. The dilemma for statisticians was tentatively resolved by a conjecture that the size of balance may be a management artefact (Stone, 2015). The second year of data on CCG financial performance strongly confirms the conjecture, strengthens it by exposing features of the scatter-gram of the two year-surpluses that look particularly artefactual and, finally, reveals that the quadrant frequencies of the scatter-gram are not significantly associated with whether or not CCGs have an above-median target allocation.]

Introduction

The third edition of NHS England's impressive Atlas of Variation[1, 2] (www.rightcare.nhs.uk) has 63 maps for England's 211 CCGs, with five shades of blue to represent the CCG score on each indicator. According to the Right Care report[1], variation in any *per capita* indicator of a state of health or health-care activity is acceptable if it reflects *patient-centred care and clinical responsiveness, based on the assessed need for the population served*. What is unwarranted is *variation in the utilisation of health care services that cannot be explained by* *variation in patient illness or patient preferences.* This note is an exploratory data analysis (EDA) of the relationships between the 63 indicators and two funding variables—the actual CCG allocation (£ per head), A say, for the financial year 2013-14 when CCGs began operating, and the formula-based target allocation (£ per head), T say. The indicator data were collected between 2012 and 2014. The relationship (expressed as a statistical correlation between an indicator, I say, and the best-fitting linear combination of A and T) may come from either a direct or indirect influence of I on the factors that determined the level of CCG funding, or from the direct or indirect influence of A and T on the level of I, or on a combination of the two influences.

In his Foreword to the Atlas, David Goodman drew attention to the funding question, even though the Atlas does no more than engage in general econometric speculation about value for money. That is for the good reason that *costs are complicated to identify in the NHS*—the earmarked funding of an indicator in individual CCGs depends not only on the nature of the particular indicator and its level, but also on the quality of financial record-keeping and possibly on management concern[3,4] about the annual CCG surplus. So it makes sense to try to extract any signals there may be in the associations with the overall funding variables A and T—which are so precisely determined that the population-weighted average over CCGs of the *per capita* 'distance from target' variable D (the arithmetic difference A−T) is required to be zero by Treasury fiat. Goodman encourages such exploratory data analysis of the fruits of the Atlas project:

> *England has risen to a singular position in the growing worldwide effort to understand health-system performance. Unlike many other countries, England is awash with healthcare data and measures for tracking over time and place, making the NHS the most thoroughly measured healthcare system in the world. But data, by itself, is neither information nor intelligence.*

Despite its final caveat, this comment may be too respectful of the role that 'big data' played in the Department of Health and now in NHS England—especially in formula-making, size too often displaces sense. But Professor Goodman is surely right to suggest that the simplicity of the Atlas maps and their accessibility will deepen understanding of their underlying logic. This note will argue and speculate about what that logic is. On a number of questions it is grossly under-informed, so I hope that readers who take issue with my arguments may thereby be provoked to find better ones.

Key findings

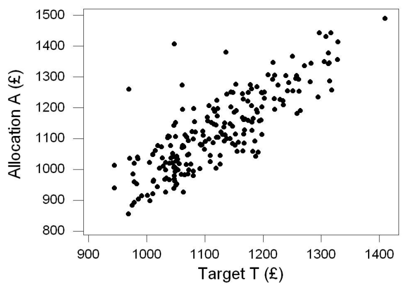
1. The overall finding is that, despite the manifestly high levels of unexplained variation, least-squares estimation is able to combine the latent information from over 200 CCGs to unveil robust, message-bearing models for the majority of the 26 indicators with the strongest associations.
2. The messages can be most easily read in just two numbers—the coefficients of the CCG funding variables (A, the actual allocation per head, and T, the formula-based target allocation)—when the basic model is expressed as a linear combination of their average F = (A+T)/2 and their difference D (the contentious 'distance from target' A−T). For why it is F and not A, see the Annexe.
3. The Map 2 variation of key antibiotic percentage (significantly decreasing with funding) may need joint analysis with the Map1 variation for an explanation and fuller understanding.
4. For Map 6 of directly-standardized cancer mortality, the 'T-party' (those who trust the target as a valid measure of fair funding) has a problem in the significant positive association with D for fixed T.
5. For Map 9A of the colonoscopy & flexisigmoidoscopy indirectly-standardized rate, the T-party may see the same association as a reflection of the benefit of funding for good performance.
6. The Map 15 variation of seizure-free epilepsy percentage (significantly decreasing with funding) needs Map 14 for fuller understanding.
7. For Map 18 variation of directly-standardized cataract surgery rate for over-65s, the association is dominated by distance-from-target D—probably as an artefact of the age-standardization and the management pressure.
8. When jointly analysed, Maps 30 and 32 for diabetes may be revealing evidence of the benefit of extra funding on the quality of diabetes care.

9. There is little or no place for the formula-based target and its derived distance-from-target variable in the interpretation the signals extractable from the information in over 200 CCGs.
10. The overall finding supports the stance of the ‘S-party’ of statisticians who see T as an unworthy concoction.

The statistical pirouette of allocation and target

The values of *per capita* allocation A and target T in any particular year can be viewed as a snapshot of a couple of slowly-moving dancers in a historically extended pirouette around their average F = (A+T)/2 (the metaphor is feasible since A = F+½D, T = F−½D and the population-weighted average of D = A−T is zero in order to balance the books):

Figure 1



The four CCGs in the top-left of Figure 1 are those with values of the D = A–T in excess of £200─Camden (£213), Isle of Wight (£244), Central London (£290) and West London (£359). These CCGs have here been excluded on the reasonable grounds that they may be misleadingly influential. The unweighted average of the remaining 207 values of D is −£1.26. The following three figures show how D relates to F and to each of their components A and T.

Figure 2

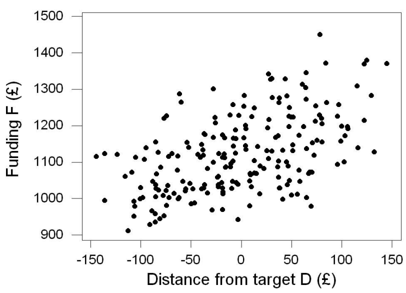


Figure 3

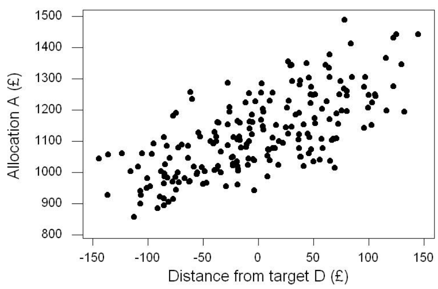
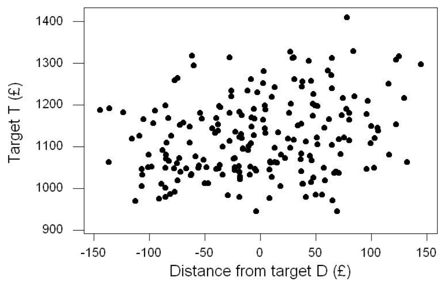


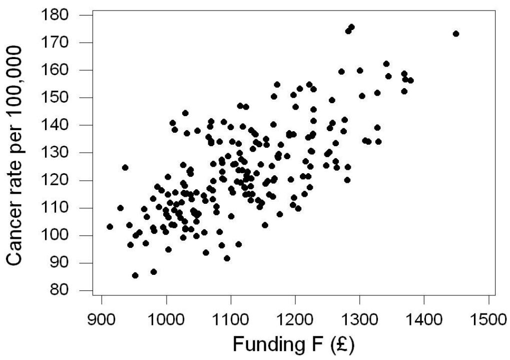
Figure 4



The clear-cut case of cancer mortality

Figure 5 exhibits the clearly positive association between the cancer mortality indicator (I_6) and the pooled measure of *per capita* funding F=(A+T)/2.

Figure 5



For the straight-line regression of I_6 on F, the value of R^2 (the fraction of the total variance of the 207 values of I_6 accounted for by its fitted values on the line) is 48.7%. That is significantly better than the 45% for I_6 on A, which suggests that the historical pirouette of the couple A & T around their average (A+T)/2 may truly have resulted in a better measure of funding over the population of CCGs. The fitted regression on both F and D is $I_6 = -9.9+0.119F-0.0177D$. The P-value of the D-coefficient −0.0177 is non-significant at 0.26, equivalent to R^2 at 49.0% not being significantly larger than the value 48.7% for regression on F alone. To drop D from this regression would be to express an unbiased preference for an equally satisfactory model with just one rather than two fitting variables. However, before concluding that D can be dropped as of little explanatory value, a caveat is in order.

The use in EDA of the conventional formula $\alpha+\beta F+\gamma D$ for large variations of F and D (about their averages aveF & aveD) does not have a theoretical justification, but is widely used because it is readily fitted by least-squares and has least-squares mathematics to call in aid. There could well be other formulas that would give better fits and allow D to show some influence on cancer rate, and that can now be easily fitted by computer. One way of being content with an empirical model (at least in a large exploratory study) is to examine the residuals for any obvious pattern. There has to be a limit to such examination before discussing or applying the chosen model to in some way. Here is the result of testing the residuals of the I_6 on F&D regression by adding the variables QF = $(F-aveF)^2$, QD = $(D-aveD)^2$ and the interaction variable QFD = (F−aveF)(D−aveD). In a sample of size 207, these should be able to pick up as 'significant' any important residuals pattern with a quadratic component, in such a 'Q-test' of the basic model. The Map 6 entry for I_6 in Table 3 below gives the P-values 0.14, 0.65, 0.38 for the coefficients of QF, QD, QFD, respectively. These non-significant values show that the F & D model passes the Q-test, strengthening the case for dropping D from the F model on the grounds that its coefficient −0.0177 is not significantly different from zero. This simplification would not have been so easily spotted if either A or T had been used instead of F. For the model $\alpha+\beta A+\gamma D$, the least-squares estimate c of γ in the fitted formula a+bT+cD has a highly significant t-value of −4.2 and, with T in place of A, the t-value of the corresponding coefficient +3.0 is also significant.

Results for all indicators and for an objective selection

Tables 1, 2, 3 give the EDA results for the 26 indicators with the strongest relationship to funding. Let M_1, M_2 and M_3 denote the models for the regressions of I on F & D, of I on F, and of I on T&D, respectively. The 26 were those where the R^2 for M_1 exceeded 6%.

Table 1. Selected indicators of particular interest

Map	Indicator	G / P / ?	Average of I	R-squared: M1	M2
2:	Percentage of primary care antibiotic prescriptions that are for key antibiotics	G	11.2	8.0	6.4
6:	Directly age-sex standardized cancer deaths per 100,000	P	114	49.0	48.7
9A:	Indirectly age-sex-IMD standardized colonoscopies (etc) per 10,000	G	149	26.5	25.5
11 :	Crude rate of barium enemas per 100,000	P	77.8	9.1	8.5
14:	Directly age-sex standardized epilepsy emergencies per 100,000	P	128	17.6	17.5
15:	Percentage of seizure-free epilepsy for a year (% of treated cases)	G	61.7	23.7	23.2
18:	Directly age-standardised over-65 cataract surgery cases per 100,000	?	3100	7.8	4.6
21:	Percentage of COPDs that had a flu jab	G	82.1	10.3	8.0
22:	COPD emergency admissions per 100,000	?	270	25.0	24.5
26:	Hypertension-controlled percentage of CKD cases	G	76.2	7.1	5.9
27:	Observed/model-expected CKD cases (%)	G	0.710	10.5	10.5
30:	Type 1& 2 cases: percentage treated by NICE criteria	G	60.0	10.4	9.0
31:	Type 1& 2 cases: percentage with acceptable clinical tests	G	36.3	14.0	12.3
36:	Observed/model-expected hypertension (%)	G	0.560	14.3	11.0
38:	Directly age-standardized CHD death rate per 100,000	P	45.2	13.3	13.3
45:	(Mental cases SMR) /(National SMR) (%)	G	0.844	11.3	11.2
50:	Dementia: percentage diagnosed	G	0.559	13.2	13.2
55:	Dementia: over-65 admission rate x 100,000	?	3360	6.8	6.8
79:	Dental caries: under-5 admission rate x 100,000	?	353	7.0	6.7
83:	Child asthma emergency admissions per 100,000	?	223	15.7	15.4
86:	Child mental health: 3 or more admissions per 100,000	?	78.5	36.9	36.8
87:	Children in A & E: attendances per 1000	?	380	16.7	16.7
88:	Under-18 tonsillectomy: standardized rate per 100,000	P	257	18.7	18.2
92:	Adult learning disability per 1000	G	4.89	31.9	31.3
93:	Indirectly age-sex standardized A&E attendance per 1000	?	339	15.4	15.2
95:	Directly age-sex standardized emergency ambulatory cases per 100,000	P	833	26.1	26.0

For Table 2, the fitting formulas for M_1 and M_2 were expressed so that b_1, c_1 and b_2 are formula increments for an additional £1000 per head in both F and D.

Table 2: Least-squares estimates and t-values for models M_1 and M_2

Map	b_1	t	(or P-value)	c_1	P-value	b_2	t	Sign & t-value of c_3 in fit of M_3
2	- 4.67	- 2.2	(0.03)	- 6.48	0.06	- 6.72	-3.7	- 2.9
6	119.00	12.5		- 17.70	0.26	113	14.0	+ 3.0
9A	131.00	7.6		- 13.20	0.64	127	8.6	+ 2.1
11	369.00	3.1	(0.002)	211.00	0.28	436	4.2	+ 2.3
14	153.00	5.9		- 23.40	0.58	146	6.6	+ 1.4
15	- 28.10	- 6.1		- 8.67	0.25	- 30.8	-7.9	- 3.4
18	599.00	1.3	(0.18)	1950.00	0.01	1220	3.1	+ 3.5
21	- 4.40	- 2.5	(0.02)	- 6.60	0.02	- 6.51	-4.2	- 3.4
22	474.00	6.4		146.00	0.23	52.0	8.2	+ 3.6
26	5.96	3.9		- 4.09	0.10	4.66	3.6	- 0.5
27	0.49	4.2		- 0.01	0.98	0.487	4.9	+ 1.4
30	18.70	2.9	(0.004)	19.30	0.07	24.9	4.5	+ 3.1
31	13.90	5.7		- 8.10	0.04	11.3	5.4	- 0.3
36	0.16	5.8		- 0.21	0.01	0.123	5.0	- 1.2
38	44.40	5.0		- 5.49	0.71	42.7	5.6	+ 1.3
45	0.52	4.0		0.12	0.58	0.560	5.1	+ 2.0
50	0.28	4.8		- 0.01	0.89	0.272	5.6	+ 1.5
55	2190.00	3.5	(0.001)	- 330.00	0.75	2090	3.4	+ 0.8
79	606.00	2.6	(0.01)	298.00	0.44	700	3.6	+ 1.8
83	390.00	5.6		- 87.00	0.44	363	6.1	+ 1.1
86	216.00	9.1		19.40	0.62	223	10.9	+ 3.7
87	477.00	5.3		46.40	0.75	491	6.4	+ 2.2
88	285.00	5.2		97.00	0.28	316	6.8	+ 3.0
92	5.81	7.6		1.70	0.18	6.34	9.7	+ 4.2
93	359.00	4.8		75.40	0.53	383	6.1	+ 2.4
95	1170.00	7.4		- 76.10	0.77	1150	8.5	+ 2.2

Table 3. P-values for quadratic test variables QF, QD and QFD

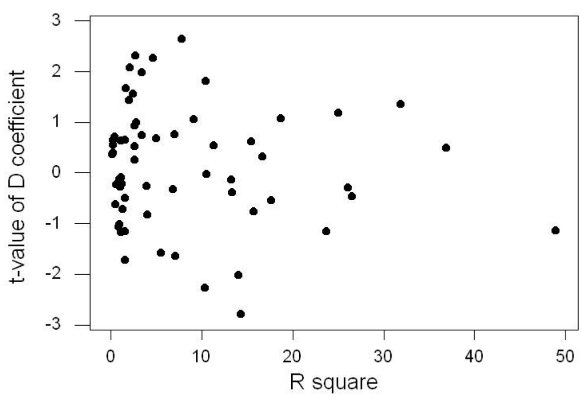
	QF	QD	QFD	
Map 2	88%	80%	81%	☺
Map 6	14%	65%	38%	☺
Map 9A	0.20%	18%	3%	☻
Map 11	0.30%	49%	17%	☻
Map 14	0.00%	2%	0.00%	☻
Map 15	11%	48%	20%	☺
Map 18	77%	42%	83%	☺
Map 21	57%	36%	27%	☺
Map 22	1%	3%	2%	☻
Map 26	58%	78%	99%	☺

Map 27	90%	30%	30%	☺
Map 30	34%	57%	20%	☺
Map 31	29%	84%	92%	☺
Map 36	4%	2%	4%	☻
Map 38	1%	2%	1%	☻
Map 45	47%	23%	42%	☺
Map 50	5%	0.20%	3%	☻
Map 55	1%	2%	3%	☻
Map 79	59%	4%	47%	☺
Map 83	63%	31%	51%	☺
Map 86	37%	80%	23%	☺
Map 87	0.00%	0.00%	0.20%	☻
Map 88	37%	41%	26%	☺
Map 92	72%	45%	74%	☺
Map 93	0.00%	0.00%	0.20%	☻
Map 95	3%	38%	6%	☻

Robust (almost assumption-free) interpretation of a non-significant D-coefficient

The t-value for a D-coefficient is the ratio of the coefficient to its least-squares-theory standard error: it is a measure of how far the coefficient is from zero. When, as here, the number of observations is over 200, the normality-based 95% confidence interval for a hypothetical true value does not cover the value zero when $t > 2$. Figure 6 plots the t-value for the D-coefficient in the regressions of all 63 indicators (when they are formulated as regressions of I on F&D)—against R^2, the conventional measure of the strength of the association between I and the two funding variables. It shows that only a handful of indicators have t-values appreciably outside the range (−2 to +2) of t-values corresponding to a non-significant association with D at the 5% level (consistent with a 95% confidence interval covering a zero value of a postulated 'true' coefficient). With 63 observations, the expected number outside the range would be 3 if the zero applied to all 63. The overall evidence is that discarding D would be broadly justified, although it will be of interest to look at the indicators that are outside the interval, especially those with a larger R^2 to strengthen any explanation.

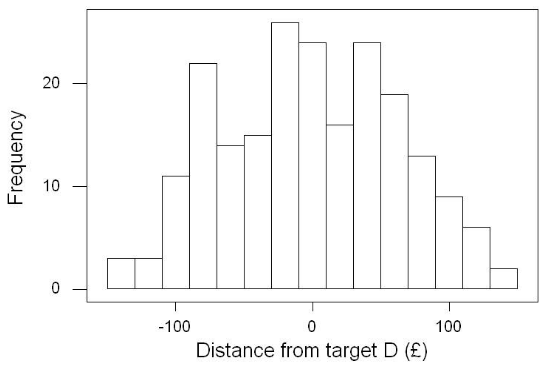
Figure 6



An analogous plot of P-values would show a roughly uniform distribution over the unit interval. The least-squares theory correspondence between t and P involves the number n of CCGs in an indicator (but not the number p of parameters in the fitted model, which is a constant 3) but the correspondence is almost a 1-1 functional relationship because n is always larger than 180 (a statistical infinity!). These values of P can be given a robust interpretation as the P-values of a *pure significance test* in which a null hypothesis has been suggested by some data set and is then considered without considering explicit alternative.

A good example comes from Daniel Bernoulli's observation of the closeness of the orbits of the six then-known planets. Bernoulli used three measures of closeness to test and reject the null hypothesis of a uniform distribution of the 'north poles' over the celestial sphere. Here, the observation is that the 63 P-values are almost uniformly distributed over the unit interval and our interest is to see whether, taken together, they accept (rather than reject) a null hypothesis that entails (at least approximately) the separate uniform distribution of each P-value—namely, the hypothesis H_0 that the 211 components of D are approximately randomly distributed over the 207 CCGs. Consider just one of the 63 indicators. The informal proof of the entailment is rather delicate: it represents the variations-from-average I−aveI, F−aveF, D−aveD as vectors **I**, **F**, **D** in R^{206}. First consider **I** and **F** as fixed, but **D** as uniformly oriented with respect to the unit sphere in R^{207} (idealizing the assumption of the random distribution of D-values). If, in this picture, we choose to consider **D** as fixed and the plane of **I** and **F** as random, the P-value of the D-coefficient for the least-squares regression of I on F and D is robustly interpretable as the P-value for the coefficient of **D** in the regression of **I** on **F** and **D**. The only distributional assumption required to entail its uniform distribution over the unit interval is the now-spherical distribution of **I**. Figure 6a shows that the roughly normal shape of the distribution of D-values should enhance the approximations here involved:

Figure 6a



General comments on the Tables

There are two main parties to the technical debate about the CCG funding formula (i.e. the target allocation T). Those parties are the 'T-party' (consisting of the health economists and Department of Health finance officers who devised the formula and take it seriously) and the 'S-party' (mainly statisticians who see T as a statistically-unprincipled and untrustworthy concoction). This note provides more than circumstantial evidence that those interested in the honesty of resource allocation for healthcare should join the S party.

There are five directly-standardized and two indirectly-standardized indicators in Table 1. For direct standardization, the data on individual CCG rates had to be available and there was a choice between the two methods. The Right Care report[1] does not give any reason for making the choice that was made in favour of either method, although the issue of standardization turns out to be highly-relevant to explanations of the signs of the coefficients b_1, b_2, c_1 and c_3 in Table 2. A statistically-significant value of c_1 may represent a breakdown of the basic F-model. Only four of the 26 indicators have a P-value for c_1 less than 0.05: Maps 18, 21, 31, 36 with values 0.01, 0.02, 0.04, 0.01 respectively. The 0.04 is not even conventionally significant when allowance is made for it being the smallest value in a sample of 23 which leaves only three representing breakdown. For the A&D formulation, however, there are eight t-values greater than a highly significant 2.9. In Table 3, if we let Map 95 pass, 15 of the 26 indicators passed the Q-test of the A&T (or F&D!) model. Only a handful of indicators have t-values appreciably outside the range (−2 to +2) of t-values corresponding to a non-significant association with D at the 5% level (consistent with a 95% confidence interval covering a zero value of a postulated 'true' coefficient). With 63 observations, the expected number outside the range would be 3 if the zero applied to all 63. The overall evidence is that discarding D would be broadly justified, although it will be of interest to look at the indicators that are outside the interval, especially those with a larger R^2 to strengthen any explanation. All of which adds to the case for taking the basic F-model as a benchmark from which to explain offending exceptions.

S and T parties vie to explain results for some individual indicators that pass the Q-test

The statistical relationship between an Atlas indicator and CCG funding is one that necessarily involves A and T in some joint analysis. The precise values of A and T were made public well before the start of the financial year and are therefore known by finance directors well before they face the problem of managing the finances for that year. So explanations are free to invoke the management artefact [3, 4, Annexe] revealed by data on the annual financial surplus plotted against D.

The Atlas indicators are based on data closely related to the health care activities in the year 2013-14 for A and T. They cannot be consistently described as *performance* indicators, as the Right Care report[1] frankly acknowledges: *The classification is shaded from light blue (lowest value) to dark blue (highest value) on the column charts and maps. The ranges and shading do not indicate whether a high or low value represents good or poor performance.* The report classifies each indicator as *activity, outcome*, or *quality*. However, the Map 86 indicator, *proportion of children with three or more admissions to hospital per year for mental health problems*, is labelled both *activity* & *quality*. So the following explanations will not make much use of such abstract and therefore unreliable concepts (Table 1 merely lists the Good/Poor/Questionable classifications of a cooperative senior clinician and a consultant dermatologist doing their joint best).

The basic stance of the S-party will, I suggest, be broad acceptance of the privileged status of F stemming from the reality of the historical pirouette, together with an open mind as to whether D may be having an influence on I via the well-established management concern about D for the financial surplus indicator studied by the National Audit Office [3, 4, Annexe]. For the T-party, however, the basic stance is non-historical (with no pirouette nonsense), in line with what the latter study took for granted. It holds that, as a driver of A, it is T that has a privileged status as a trustworthy measure of a CCG's health care need, so that for the T-party there is an expectation that any additional association between D and I from the A term in A−T will be easily explained. Here are my suggestions for how the parties might deal with a selection of the 26 indicators in Table 1 that passed the Q-test:

Map 2: Percentage of all antibiotic prescription items in primary care that are for key antibiotics

Prescription cost is a sizeable fraction of the current target formula (incidentally, creating a perverse incentive?). The management concern about D may therefore have a greater influence on I_2 than on other indicators, so it cannot be expected that it might be showing itself as suggested in the Annexe (through F alone). The neutral regressions are those of I_2 on A alone and the basic regression expressed as I_2 on A & D: these have the fitted formulas $I_2 = 18.1-0.0062A$ and $I_2 = 16.4-0.0047A-0.0041D$ where the t-value for D is only 1.0, giving the option of dropping D in favour of the first formula in which the t-value for A is a highly significant 4.1

The denominator of the percentage is the Map 1 indicator, *Mean number of defined daily doses (DDD) of antibiotics prescribed in primary and secondary care per day per 1,000 population by NHS area team.* The two maps appear to have an almost-complementary shading, suggesting that the indicators themselves are strongly negatively correlated. The Right Care report[1] gives 19.2 to 25.6 (per 1000) for the range of I_1 for the 25 area teams and 4.5% to 18.0% for the 63 CCGs for I_2, but I do not have the stamina to extract the Map 2 area team averages for the five levels of Map 1 colouring or locate the Map1 figures for individual CCGs, needed for proper comparison of the ranges. That data would also be needed to see the sign and size of the association between funding and I_1.

David Goodman's foreword[1] comment on Map 1 does not implicate Map 2:

> *High rates of antibiotic prescribing in primary and secondary care are hard to explain by population differences in bacterial infection incidence. Over-use is a more likely explanation, although it could be deemed a costly misuse, given the absence of benefit and a greater likelihood of antibiotic resistance and allergies.*

This note may throw some light on the residual uncertainty of that. The well-determined negative coefficient of A in the fitted regressions calls for an explanation of why, with extra funding, good quality prescribing (if that is what it is) represented by I_2 is going down. For primary care prescriptions, the product of Map 2 and Map 1 at CCG level would be the CCG's rate of key prescriptions per 1000. Plotting the product against Map 1 would partly address the question raised by Goodman, while their individual plots against A would almost complete the address: there is a potentially interesting relevant variable without an Atlas map—the proportion of refusals to a patient requesting an antibiotic (which could be economically estimated by an anonymized questionnaire to a stratified random sample of GPs in each CCG). The S and T parties would be well-advised to await the results of this further research before trying to explain the Table 2 findings or those of the above regressions of I_2 on A and A&D of Map 1. Here we need not heed such advice and can engage in grossly-uninformed speculation on behalf of the S-party. Fig. 3 shows that D is positively correlated with A, so that management will be pressing for overall reduction of costs by CCGs with large D and A. The key antibiotics, as those of last resort, will be largely immune to such pressure which will be directed at the other antibiotics, constituting the bulk. The reduction in the total number of prescriptions (from what it would have been without pressure) results in a smaller denominator for I_2 and an explanation for the otherwise puzzling negative coefficient of A in the I_2 on A regression.

Map 6: The directly age-and-sex standardised rate of mortality from cancer in under-75s

When D is dropped (to the satisfaction of the S-party), the fitted formula is $I_6 = -3.55+0.113F$ from which an interesting assumption-dependent statistic can be derived from the precise definitions of I_6 and F. As a directly-standardized indicator, I_6 (before being multiplied by 100,000 to be free of decimals) is the weighted sum of the age-sex rates with weights given by the national age-sex profile. For a CCG with the national age-sex profile, the sum is the number of cases per head, and the numerator of F in pounds then has £100,000/0.113 or £88,000 for each case. Could this figure realistically reflect the average cost of a cancer death, or is it that large because funding per head is historically much higher for CCGs where cancer mortality happens to be higher? The target formula T, that drives and is closely correlated to F, has variations in which Age and Deprivation are major factors. So is Prescription Rate, but my information is that the prescription cost for cancer mortality is not met by the CCG which is thereby excluded from any explanation.

For the T-party, it is the highly significant positive coefficient of D in the T&D regression that requires explanation. The T-party would probably decline to explain it as a regrettable outcome of management pressure on CCGs heading for an embarrassing financial surplus when A is well above target (therefore ignoring the Pulse report[5] of such pressure) but how then would it explain why additional funding above a fixed target value is associated with a larger cancer rate?

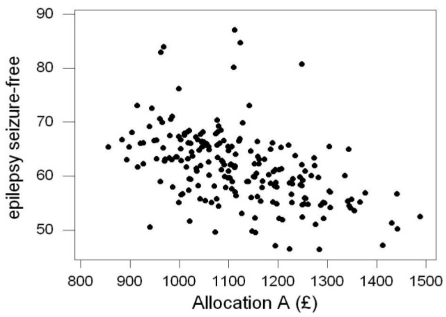
Map 9A: Indirectly-standardized rate of colonoscopy and flexisigmoidoscopy procedures per 10,000

Indicator I_{9A} fails the Q-test for model specification by only a small margin, so is here included. The fitted formula has a well-determined F-coefficient of 0.131: the D-coefficient in the F& D regression is not significant (to the satisfaction of the S-party). As an indirectly-standardized indicator, I_{9A} (before being multiplied by 10,000 to be free of decimals) is the product of an observed/expected ratio and the national rate, where the ratio is the number of cases to the expected number if the CCG had national rates. For a CCG with the national age-profile, the numerator of F (£ per head) therefore has £10,000/0.131 = £76,000 for each case. Could this figure realistically reflect the average cost of a colonoscopy flexisigmoidoscopy case, or is it that large because funding per head is historically much higher for CCGs where such investigations happen to be higher? The case for the latter explanation is as for Map 6. The T-party might see the just statistically-significant positive coefficient of D as a crumb from the funding table that favours more such clinical investigations.

Map 15: Seizure-free proportion of epileptics: percentage of persons aged 18 years and over, receiving drug treatment for epilepsy recorded on GP practice registers and seizure-free for the preceding year.

Like Map 6, the D-coefficient is not statistically-significant, so the S-party need only explain the pronounced negative correlation with A shown in:

Figure 7



The clinicians classified this indicator as ‘good’, so an explanation is called for. Both parties can find one in the fact that I_{15} was calculated as a percentage, with a numerator given by the *number of patients aged 18 years or over on drug treatment for epilepsy, who have been seizure free for the last 12 months recorded in the preceding 12 months* and a denominator given by the *number of people aged 18 years and over receiving drug treatment for epilepsy recorded on practice register including exceptions* [2]. In all likelihood, the denominator is strongly and positively correlated with the *rate of epilepsy emergency admissions to hospital by over-18s*, whose Map 14 can be seen to have nearly complementary shading to that of Map 15. So better-funded CCGs (higher A values) may be admitting a much higher proportion of epileptics being treated by their GPs but whose number may not be much affected by funding, thereby reducing the percentage. Confirmation of this explanation would need a joint analysis of the Map 14 and Map 15 data.

Map 18: Directly age-sex standardized admission rate of over-65s for cataract surgery per 100,000

The fitted formula in the F&D regression is $I_{18} = 2430+0.599\,F+1.95D$ with standard errors 0.45 and 0.73 for the F and D coefficients respectively. The S-party has to explain the gross anomaly of the results for this indicator—the only one with a non-significant value of the positive F-coefficient in model M_1 and a statistically-significant value of a thrice-as-large D-coefficient. Here is a tentative explanation based on belief in (i) the reality of management pressure[3, 4] for providers to cut expenditure on high-frequency and expensive procedures when D is large (and overall activity is heading for a financial surplus in the annual statement), (ii) the CCG’s age-profile does not vary much with D when funding F is fixed and (iii) that, even within the over-65s, the cost of cataract surgery increases appreciably with age. The suggested explanation then depends on the direct age-standardization of surgery rate and comparison of two hypothetical CCGs with the same F and the same age-profiles—CCG^- with a large negative D, and CCG^+ with a large positive D that makes it subject to management pressure to reduce the cost of the cataract surgery cases. To simplify the necessary algebra, consider division of the over-65s into two age subgroups: ‘young’ and ‘really old’ with corresponding population proportions $p_1, p_2,$ and fractions f_1, f_2 of the number of surgery cases. Before multiplication by 100,000, I_{15} is $f_1(P_1/p_1)+f_2(P_2/p_2)$ where P_1, P_2 are the national population proportions. The cost per head is proportional to $C_1p_1r_1+C_2p_2r_2$ (say, where $C_1 < C_2$) which equals $(C_1f_1+C_2f_2)(p_1r_1 + p_2r_2)$. Comparing CCG^- and CCG^+, I_{15} would be larger for CCG^+ (other things being equal) and cost would be lower (for the same unstandardized surgery rate $p_1r_1 + p_2r_2$) if, unlike CCG^-, it gives a larger percentage of its cataract surgeries to the ‘young’ when p_2 is large (with $P_1/p_1 > P_2/p_2$) than when it is small.

For the T&D regression, the fitted formula is I_{18} = 2430+0.599T+2.25D with standard errors 0.45 and 0.64 respectively. If the T-party rejects the S-party's assumptions (i) and (ii) that management may be involved with a distorting influence on the variation of certain indicators, how would they explain the 2.25, which is no more than a reflection of the otherwise-unexplained sign inequality between the coefficients A and T in the equivalent basic formula I_{18} =2430+2.25A−1.65T?

Map 30: Percentage of people in the National Diabetes Audit (NDA) with Type 1 and Type 2 diabetes who received all eight NICE-recommended care processes

In his foreword, David Goodman highlights this as an example of 'under use' of funds:

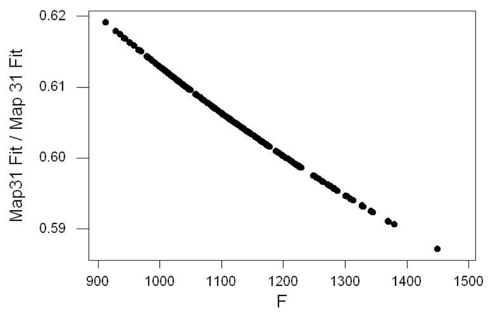
> *... the percentage of patients with diabetes receiving evidence-based processes of care shows the potential for CCGs to provide care that is likely to improve patient' health and well-being. The metric helps to identify where existing resources should be directed, and the specific interventions are straightforward and within the current capacity of the NHS.*

The formula fitted by the broadly-acceptable model M_2 is I_{30} = 32.0+0.0249F where the F-coefficient has an estimated standard error of 0.0056 and there is an extra 1% in I_{30} for an increase of F of £40 per head.

Map 31: Percentage of people in the National Diabetes Audit (NDA) with Type 1 and Type 2 diabetes who met HbA1c, blood-pressure and cholesterol targets in three of the eight processes (I_{31})

The formula fitted by the broadly-acceptable model M_2 is I_{30} = 23.5+0.0113F where the F-coefficient has an estimated standard error of 0.0021. For an extra 1%, you need an increase of £88 per head in F. The following plot shows how the fitted proportion (of those meeting all NICE criteria who also met three particular targets) fell from about 62% for the CCG with the smallest F to under 59% for the CCG with the maximum F. (Although small, this drop is statistically significant given the precision with which the underlying fitted lines are estimated. The perturbations from D in the I-on-F&D regression are negligible (down to the first-decimal place) when the fitted values are plotted against F.)

Figure 8



This feature would not have been detectable from the regression on F of the raw ratio Map 30/Map 31. The Atlas could have used that ratio, instead of Map 31, as an indicator of the quality of CCG's care for diabetes. One explanation of the feature is that it is a statistical artefact of the estimates, 32.0+0.0249 and 23.5+0.0113F—an outcome of the particular shape of the indirect associations between F and the two indicators. Another is that it could be a reflection of a causal relationship between the funding level and that quality. Both explanations are available for both S and T parties.

Conclusion

Despite the huge unexplained variation in indicators, there are readable signals in the cumulative information of over 200 CCGs—a statistical infinity! The main message is that distance-from-target D can be dropped as an explanatory factor except where, as for Map 18, it expresses management concern over embarrassingly large over-target values of D (thereby adding to the evidence that such concern can be a significant factor). When D is dropped, target plays little part in the explanations—its main role was in influencing the level of actual funding A, the minor role was in the minor change from A to F as what appears to be a simplifying specification of a single funding variable (see Annexe). The import of this broad message in the above EDA is more than agreeable to the S-party! By dragging or pushing a CCG's A-value to where it is currently is, T is making A complicit in its own irrational formulation and thereby undermining the status of A as an appropriate level of *per capita* funding for the CCG's particular health-care needs, whatever these are.

Annexe

There were fruitless exchanges about the CCG funding formula at a lively Parliamentary Accounts Committee hearing on a report, in which the National Audit Office quietly expressed reservations about the formula. My statistical analysis [3, 4] of the PAC/NAO hiatus enlarged the NAO reservations and almost dictated the conjecture that CCG finance directors engage in unwarranted influence on the end-of-year surplus on their accounts. The conjecture has since been confirmed by the 2014-15 CCG data. The influence comes from embarrassment at large over-target values of A−T, presumably acting as a pervasive pressure on spending by such CCGs—equivalent to them having been given a somewhat smaller allocation A. The second conjecture is that, on the average, this reduction is fortuitously equivalent to the otherwise puzzling emergence of $F = A - \frac{1}{2}(A-T)$ as the operative expression of funding level. The PAC hearing was largely a response to an NAO graph that plotted financial surplus against D, for which T was taken to be a trusted lighthouse in a sea of swirling allocations. The absence of any major role for D in the present associations explains why NAO could not find any non-financial indicator with a significant association with D.

References

1. Right Care (2015), The NHS Atlas of Variation in Healthcare: Reducing unwarranted variation to increase value and improve quality. London: Public Health England (www.rightcare.nhs.uk/atlas/Atlas3.Data_150915.xlsx).
2. Right Care (2015), Indicator Metadata. (http://www.rightcare.nhs.uk/atlas/)
3. Stone, M. (2014), National Auditing beats Parliamentary Accounting. London: Civitas (www.civitas.org.uk/pdf/CivOctRev5).
4. Stone, M. (2015), New development: The remarkable insignificance of NHS England's CCG funding formula, *Public Money & Management*, **35**, 4, pp. 311-314.
5. Matthews-King, A. (2015), GP practices offered 'ethically questionable incentives' to cut urgent cancer referrals, *Pulse Newsletter, 1 Oct*.