



Explicating 'wrong' or questionable
signs in England's NHS funding
formulas: correcting wrong
explanations

By Mervyn Stone (December 2013)
University College London, UK

with a foreword by Nigel Williams,
Civitas statistician

Address for correspondence: nigel.williams@civitas.org.uk

Online Report: December 2013

© Civitas 2013

55 Tufton Street
London SW1P 3QL

Civitas is a registered charity (no. 1085494) and a company limited by guarantee, registered in England and Wales (no. 04023541)

email: info@civitas.org.uk

Independence: Civitas: Institute for the Study of Civil Society is a registered educational charity (No.1085494) and a company limited by guarantee (No. 04023541). Civitas is financed from a variety of private sources to avoid over-reliance on any single or small group of donors. All the Institute's publications seek to further its objective of promoting the advancement of learning.

Foreword by Nigel Williams

This paper calls for an answer. On 17th December 2013, the NHS England Board meets to review the policy of how to allocate funding. The implications are potentially serious. In 2008, Stafford General was petitioning its local Primary Care Trust, South Staffordshire, for more money to close the shortfall in staff levels.

Paragraph 1.236 of the Francis report reads:

A report dated 28 May 2008, prepared by Mr Griffiths of SSPCT, noted that the Trust requested an additional £775,000 funding for medical and nursing staffing and support. It was also reported that the Trust had identified a £2.5 million shortfall in the necessary funding to correct nursing capacity and skill mix issues. The Trust had stated that it could not afford more than £1.15 million out of its own resources. The report stated that the PCT supported the view that nursing levels should be increased but was not sympathetic to the request to fund the gap¹.

South Staffordshire PCT's allocation was at the lower end of the per-person funding range, calculated using ancestors of the present formula. In 2009-10, South Staffordshire was allocated £1,351 per head by funding formula. If the hospital had approached North Staffordshire instead, there would have been £1,504 per head. Tower Hamlets received £1,852 per person.² Those would be big differences in a situation of plenty, when we would just hope that the unspent resources would be frugally returned. When margins are tight to the point of frozen posts and suspected rationing, differences of that scale need to be openly justified. Staffordshire did not receive the lowest allocation. Other areas with lower capitation provoked no such scandal. Nevertheless, many of the safeguards that will help avoid future failings in care depend on getting enough money through NHS England's system of allocation.

David Buck and Anna Dixon for the King's Fund³ made the point that 'Pace of Change' requirements, that no provider should receive a substantial real-terms cut from what they had received in the year before, were making any complicated means of distributing money increasingly irrelevant. They make a strong case for removing much of the baggage from the

¹ Francis, R. QC (chairman), (2003) *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Volume 1: Analysis of evidence and lessons learned (part 1)*

<http://www.midstaffpublicinquiry.com/sites/default/files/report/Volume%201.pdf> (retrieved 12/12/2013)

² Calculated from Department of Health, (2009) *Practice based commissioning budget guidance for 2009/10*

http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_094393.zip and 2009-10 and 2010-11 PCT Revenue allocations

http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091447.pdf (retrieved 12/12/2012)

³ Buck, D. and Dixon, A. (2013) *Improving the allocation of health resources in England*,

<http://www.kingsfund.org.uk/publications/improving-allocation-health-resources-england> (retrieved 12/12/2013)

calculation process; for getting back to a situation in which both the managers that account for the money and the public that provide it can tell why it is being given and what is meant to be done with it.

For over a decade, the misuse of multiple regression formulas for resource allocation has concerned Professor Stone and other statisticians who expect the formulae to be properly understood, especially by anyone using them to take decisions on behalf of others. A big problem is that, given a large set of data and a large number of explanatory variables, it is very easy to manufacture plausibly causal relationships that do not exist. An example is the penalty of £1,143 for 'other conditions originating in the perinatal period', specifically for the over-65s. It does not in itself prove that the allocation formula is wrong, but it does suggest that it is unlikely to match the true need for health care.

Buck and Dixon picked up on a paper Professor Stone wrote for Civitas early in 2013, that revealed the absence of logic in some striking features of the current formula⁴. Today's paper refines the earlier analysis and supplies a couple of corrections. It starts from the observation that several illnesses are associated in the funding model PBRA3 with reductions in predictions of the required funding. Among other ailments, if a practice puts a patient in hospital for dementia, they can expect a reduction in funding in the next round. By returning to the algebra underlying regression, Professor Stone has set out what these negative coefficients really mean: on average, people with those ailments receive less funding than people without them but with the same referral history. The details are in the paper.

Today's paper is on the technical side even for Civitas' Statistics Corner. However, any Civitas reader should be concerned with the sensible use of public funds. If policy makers are using machinery that requires explanation beyond general public understanding, that is itself a cause for concern. An obvious place for publication would be the Royal Statistical Society's 'Statistics in Society', where it did not get past the single referee, who felt that it was no more than an exercise in explanation. Instead, it is offered here in the hope that others will join in to explain or simplify. RSS referees and authors of the funding formula that Professor Stone criticises are very welcome to contribute to the discussion.

⁴ Stone, M., *Plain Explanation or Special Pleading* (2013), Civitas, <http://www.civitas.org.uk/nhs/download/PBRA%20Pleading.pdf> (Retrieved 12/12/2013)

Explicating ‘wrong’ or questionable signs in England’s NHS funding formulas: correcting wrong explanations

By Mervyn Stone

Summary

The ‘coefficients’ of variables in formulas devised by health economists for the current funding of Clinical Commissioning Groups or for possible future commissioning of health-care by England’s GP-practices have actual or potential financial consequences that are far from negligible. The paper exploits a simple ratio expression for the estimate of the coefficient of any ‘dummy’ variable. The numerator is then employed to explicate inadequate or illogical explanations of questionable coefficients in person-based resource allocation formulas.

1. Background

In September 2013, those attending NHS England workshops on a Fundamental Review of Allocations Policy were informed that a new formula with the acronym PBRA3 (the most recently developed resource allocation formula) may be used for at least two years as a major component of the funding of Clinical Commissioning Groups of GP-practices. The formula has been fitted by least squares in an additive linear model.

Unexpected signs of coefficients in least-squares-fitted models, often described as ‘wrong’ by the formula constructors themselves, have been influential for decades in the design of resource allocation formulas for England’s National Health Service. Two Series A papers (Stone and Galbraith, 2006; Galbraith and Stone, 2011) together cover the chequered history between 1994 and 2008 of ‘wrong-sign’ coefficients that may or may not have been indicators of ‘unmet’ need for healthcare in a proportion of England’s 8000 electoral wards. Since 2008, a formula with the acronym CARAN (Morris *et al*, 2009) that failed to accommodate such coefficients has been implemented—but only with supplementary allocations designed to meet elsewhere-conjectured unmet need.

Since 2006, the Department of Health’s Advisory Committee on Resource Allocation has been guiding the development of a different sort of formula, based on direct measurement of health-care usage by large random samples of individuals. These are the ‘person-based resource allocation’ (PBRA) formulas from the Nuffield Trust (PBRA Team, 2009; Dixon *et al*, 2011; PBRA3 Team, 2011) whose primary goal has been the optimal prediction (highest R^2) of personal total annual tariff-based hospital cost, by explanatory variables that are known at the start of the costed financial year. Although predictive performance, as measured by R^2 , was the main driver, the choice of variables was constrained by subjective considerations of the

plausibility of the signs of coefficients estimated in the many models tested. The choice was also subject to principled exclusion of powerful predictors that might offer perverse incentives, such as previous years' hospital costs.

2. The necessary language of formula construction

An NHS resource allocation formula is typically *additive*. It can be viewed as a national average followed by a string of adjustments, up or down, one for each of a very large number of empirically assembled *variables*—as in some sort of naive accountancy even though individual terms usually have no monetary formulation. For each *item* to which a formula is to be applied (for PBRA3, an individual, a general practice or a clinical commissioning group), each adjustment is the product of an estimated *coefficient* and the amount (plus or minus) by which the item's value of the corresponding variable is above or below its national average. The coefficients are typically estimated by the evergreen statistical technique of *least-squares*—that fits the formula as closely as possible to a *dependent variable* (e.g. annual cost) for a class of items.

'Closeness' is determined by the sum of the squares of the *residuals*—the differences (observed minus fitted) between the dependent variable and its fitted value. The residuals are the bits that can be said to be 'unexplained' by the fitted formula—by those who use the expression 'explanatory variables' even when there is no explanation beyond mere description. A standard measure of the closeness of a least-squares fit is what some, including Dixon *et al* (2011), call the 'coefficient of determination' (R^2) and others the 'square of the multiple correlation coefficient'. The *model* for a formula simply means the (often baroque) choice of explanatory variables and how they were strung together: an *additive model* is one where the stringing is of the accountancy type already specified. A *dummy variable* is an explanatory variable that takes the value 1 or 0 for whether or not an item has the property that defines the variable—for example, a dummy variable for a diagnosis of dementia has the value 1 if the item (an individual!) has had the diagnosis but 0 otherwise.

3. The Nuffield Trust formulas PBRA1 and PBRA3

The first PBRA research team assembled data for over 50 million individuals registered with an England GP on April 1 2007. From a random sub-sample of over 5 million individuals, they constructed a resource allocation formula (PBRA1 here) that they were happy to recommend. The huge and labour-intensive data-base was justified on the grounds that it gave enough room for two equally-large 'validation' samples—one for randomly-selected individuals and another one for individuals registered in a 10% random sample of GP-practices. What the PBRA Team report did not say was that, to test the robustness of the many subjective judgements in the construction, a genuine cross-validation would have required replication of the whole construction trail by a completely independent team.

PBRA1 is here taken to be the 'parsimonious model' formula described by Dixon *et al* (2011): there is a slightly different version in PBRA Team (2009). The dependent variable fitted by the 'explanatory' formula was personal hospital cost for the 2007-08 financial year, levied on Primary Care Trusts (PCTs) by hospitals according to national tariffs for 'completed consultant episodes' coded as detailed ICD10 diagnoses. The explanatory variables were enrolled

additively (as if in some trustworthy financial accountancy) as a string of 38 age/sex dummy variables (for whether or not a person was in a particular age/sex category), 150 grouped-diagnosis dummies D_1, \dots, D_{150} (each for whether or not the individual had a subsumed tariff-determining hospital diagnosis in the two years prior to April 1 2007), 151 PCT dummies (for whether or not a person's GP was in a particular PCT)—plus 10 covariates whose t -values for a test of non-zero coefficient exceeded 2.58 in magnitude (the 1% significance value) and whose selection was judged to be plausible as well as helpful to predictive performance. One of the covariates was actually a personal dummy rather than an area-based covariate—it was for whether or not a person had a privately-funded inpatient episode of care provided by the NHS in years 2004/05 or 2005/06. Another (true) covariate was the percentage of disability allowance claimants in the small ONS-defined area in which the person lived. Table 2 of Dixon *et al* (2011) lists the 10 covariate coefficients—the coefficients for the 38 age/sex and 150 diagnosis dummies can be found on the London School of Hygiene & Tropical Medicine website.

It is not surprising that, even with 349 explanatory variables, PBRA1 has an R^2 of only 13%: pure chance is the main factor in predicting the cost of the health-care an individual needs in the next year of his/her life. What I find surprising is the confidence expressed in Dixon *et al* (2011)—in the recommendation that the model '*could be used for allocations to practices for commissioning*'. That is because the choice of model is both empirical and subjective, and 87% (i.e. 100% minus 13%) of the variation of cost between individuals remains 'unexplained'. One would have to claim that most of that residual variation is purely random to exclude the probability that unexplored features of *terra incognita* could be moulded by other researchers into an even better predictor of personal cost—one that would make quite different allocations when personal fitted costs are summed, cancelling out much of the individual variation and giving an aggregated R^2 at GP-practice level higher than PBRA1's 77% at that level. The confidence in PBRA1 appears to stem from the achievement of this value—perhaps in the mistaken belief that the 77% establishes it as a leader in the whole field of less-narrowly-constructed formulas.

PBRA3, the formula recommended in PBRA3 Team (2011), is a rich development of PBRA1, using new data from 2007/08 and 2008/09 to improve predictive performance (i.e. do better than PBRA1's modest 13% for R^2) and also to reduce the number of what are described as "incorrect/unexpected" (but statistically-significant) negative coefficients. The improvement was achieved by separate sub-models for three age-bands (under-15s, 15-64s and over-65s)—with covariates selected from a battery of over 300, and with 32 'co-morbidity interaction' dummies for some pairs of the 150 diagnoses. One novelty was five 'morbidity dummies' for whether or not a person had 2, 3, 4, 5 or 6+ different diagnoses from D_1, D_2, \dots, D_{150} during the two years 2007/08 and 2008/09. The recommended age-band formulas join to make a much longer string than the 349 terms of PBRA1. What it achieved was an increase in R^2 from 13% to 15%, which was seen as a worthwhile improvement in performance. Aggregation to GP-practice level gave 86%, an improvement on PBRA1's 77%—but the caveat about over-confident interpretation applies as much to the 86% as to the 77%.

There is one ground, however, on which PBRA1 and PBRA3 cannot be criticized: no-one can say they are 'over-fitted' thereby giving an over-optimistic value of R^2 . The huge sample sizes made the ratio of the number of adjustable coefficients to the number of fitted costs so small that R^2 decreased only slightly between construction and 'validation' (from 12.7% to 12.3% for PBRA1, from 15.3% to 15.2% for PBRA3)—an only-slight reduction is what the statistical theory of 'large numbers' could have predicted without having a computer show it.

4. Effect of ‘freezing’ supply variables

The empirical (theory-free) approach that delivered the apparently impressive, aggregated R^2 of 86% for PBRA3 has to accept losses when implemented, if it has to accommodate the econometric theory that claims to remove the influence of supply variables, to comply with the goal of ‘equal access to health care’. When the values of supply variables used in the implemented formula are ‘frozen’ at national average values, any variations in those variables between GP-practices no longer affect prediction, so that only variations in need matter. The econometric assumption that justifies freezing is that, apart from random variation, the fitted accountancy formula at individual level is truly the sum of a ‘supply’ component and a ‘need’ component and that freezing delivers the latter, so that, when aggregated over individuals, the frozen formula would give a GP-practice the ‘level playing field’ funding it requires to meet the actual need of the individuals on its list—once the ‘playing field’ supply variations have been levelled.

Table 6 of PBRA3 Team (2011) gives a measure of the serious consequences of freezing: the 86% goes down to 73%, which means that the complementary measure of the gap between actual and predicted costs almost doubles from 14% to 27%. (Statisticians know that, because the measure is based on the squares of the gaps, it is weighted by the size of the gap itself towards the larger, perhaps more consequential, gaps.)

The most influential supply variables are the 151 PCT dummies (the 152nd PCT is the ‘dummies all 0’ case). The PBRA3 report concedes that including the PCT dummies ‘*has a large redistribution effect, greatly affecting the shares allocated to practices*’. The effect can take a curious form. Consider the case of the Hillingdon PCT that became a CCG overnight, whose GP-practices thereby inherited a common value 1 for the PCT dummy in the formula for individuals on their lists. The identity of CCG and former PCT means that the *aggregate* allocation to Hillingdon CCG would be uninfluenced by the other 348 carefully-documented variables for the individuals in its care! The same holds for any CCG that is coterminous with a former PCT.

5. Getting (or not getting) satisfactory explanations

To give a fair hearing to explanations of questionable coefficients in PBRA1 and PBRA3 requires a willing suspension of any disbelief in the corresponding models—by anyone who can see that the formulas are (literally) expressions of econometric hubris and who does not expect the coefficients to make much sense. Suspension means accepting the PBRA stance that coefficients can be explained without rejecting the model—and without questioning the assumption that, with the comprehensive inclusion of 150 diagnosis dummies, the formula is a broadly unbiased estimator of 3rd-year cost of the sub-coded hospital referrals.

At the heart of logical assessment of questionable coefficients, there is a useful *concept* and a related *statistic* that tie the assessment to the way the computer calculates the coefficient from the data that was inputted—rather than treating it as a solitary output that can be freely discussed without reference to the chain that connects input and output.

The ‘concept’ is the *penultimate model* in which the dummy variable (D say) corresponding to the coefficient has been omitted from the full model used for the ultimate formula. Without loss of generality, let the dummy play the role of the last variable to be fitted in the full model in

order to update what the penultimate model (based on all the other variables) has already calculated without knowledge of the values of the dummy. At that stage of the least-squares calculation (before the dummy variable is introduced in the software), the computer has already calculated the *penultimate formula* that best fits the 3rd-year costs—and the ‘related statistic’ is simply the average of the fitted values for individuals with $D = 1$ (*AvePen1* for short). Thus, for the dementia dummy in PBRA1, *AvePen1* is the average of the values fitted to dementia cases by the formula based on all the variables except the dementia dummy variable, at the stage when the dementia cases have not been identified in the calculation.

The Appendix specifies an expression C , for the finally-estimated coefficient of any dummy variable, as the ratio of a *numerator* given by the difference, $AveCost1 - AvePen1$, between actual and penultimately-fitted averages of cost for individuals with $D = 1$ —to a positive *denominator* that does not come into the arguments below and does not involve 3rd-year costs. The numerator is therefore, on its own, an assumption-free platform for explanation of any questionable sign. For negative coefficients, we simply have to give reason why *AvePen1* has turned out to be significantly larger than *AveCost1*.

Here are four examples—three for dummy variables and one for a ‘covariate’ that is, strictly speaking, not a dummy variable. The first example was not thought (on largely intuitive grounds) to be questionable by the PBRA teams. For the second example, the numerator in one of the two explanations is left as the average of the penultimate residuals (3rd-year costs minus penultimate fits) for individuals with $D = 1$. The third example develops a historical rationality that contrasts with the insubstantiality of the PBRA3 explanation. The last example is included because the covariate’s population distribution (in its variation from area to area) has a U-shape that may approximate that of a scaled dummy variable—well enough for explanation to be based on $AveCost1 - AvePen1$.

(i) The –£556 in PBRA1 for privately-funded NHS care —‘going-private’.

Dixon *et al* (2011) expresses the intuition that there has to be a deduction for going private:

‘The negative impact of use of privately funded care on NHS costs means that practices with registered patients who used privately funded inpatient care in NHS facilities will have a lower target allocation than practices whose patients do not (all other things being equal). This is intuitively correct, as demands on the NHS budget for inpatient care will be lower in the former practices.’

The ‘all other things being equal’ condition is postulating a subset of the $D = 0$ class whose usage of NHS facilities in 2005/06 or 2006/07 matches that of the $D = 1$ class, and then asserting that, for some unexplained reason, the corresponding full formula averages (‘target allocations’) would be different and thereby deliver a negative coefficient. What the condition is doing, however, is to obscure the difference in usage for $D = 1$ and all $D = 0$ cases that should (trusting the model) have allowed *AvePen1* to match *AveCost1* and thereby rule out a statistically-significant negative.

Two years earlier, the same intuition was given a different slant in PBRA Team (2009):

‘Having accessed private health care’ has a negative coefficient reflecting the likelihood that these patients will tend to be relatively better off and more likely to use private care again in the future as a substitute for NHS care.’

To make any sense, the ‘more’ in this quotation has to be a steady year-by-year comparison of the better off with the rest, in which case the explanation is no more than an ineffectual tautology.

A model-consistent explanation is still awaited for why *AveCost1* should be significantly less than the *AvePen1*—an average that is truly based on the costed non-private usage by the better off. The temptation to use intuition can be resisted by looking at things from the standpoint of the software writer—the data presented to the least-squares algorithm is analysed in the same colourless way whatever labels are attached to it. The going-private dummy could have been accidentally labelled D_{151} , as if going private (at some time during 2005/06 and 2006/07) were a hospital referral. But it would be a notional referral without any cost implications, and *AvePen1* would be a broadly unbiased estimate of *AveCost1* (assuming no significant change from year to year in the going-private usage). Note that, if the PBRA explanation were correct, the model could be said to have failed to predict demand in affluent areas!

(ii) The –£436 for all-age dementia in PBRA1 and –£250 for over-65s dementia in PBRA3.

We will see that the argument just given for the going-private dummy can be applied to dementia but, first, this is what PBRA Team (2009) said about the –£436:

‘The negative coefficient on the dementia prevalence rate might reflect the fact that people with dementia tend to live in a supported environment and this might help to reduce hospital costs. Note that the dependent variable does not [include] mental health costs.’

The ‘hospital costs’ here must be referring to costs of tariffed ICD10 codes for diagnoses other than dementia. Given that dementia cost is excluded from the dependent variable, the least-squares machinery can make no distinction between dementia cases and other cases making similar use of the NHS in 2005/06 or 2006/07 for the diagnoses corresponding to dummies D_1, \dots, D_{150} other than dementia. Another way of putting it is to say that the non-inclusion of the cost of dementia from the dependent variable is equivalent to a zero tariff for dementia referrals, so that (suspending disbelief in the model) we would expect the dementia coefficient to be not significantly different from zero. If the ‘supported environment’ has some (as yet unexplained) effect, it can find expression only via the 2007/08 hospital costs.

Dementia is the sole ‘mental health’ ICD10 code in any formula, so the PBRA3 team must have intended the cost exclusion of dementia referred to in this paragraph to account for the –£250 without further argument:

‘A small proportion (around 3%) of the morbidity flags [coefficients] were negative i.e. associated with lower future costs in hospital care. These can be explained either as conditions that are associated with death; some conditions whereby treatment reduces the likelihood of further problems e.g. removal [of] appendix, and some infectious diseases or **conditions where costs from the dependent variable may have been excluded, mental health, specialist care.**’ (Added emphasis)

This all-embracing paragraph is more assertive than explanatory, and the dementia negatives are still unexplained in the literature. So here are two suggestions that politely suspend disbelief in the model and that could be verified or refuted by data in the full PBRA data-bases:

Cheaper sub-codes:

A UCL colleague has suggested that the negatives may represent a difference in the distribution of dementia referrals over the ICD10 sub-codes that fix the hospital tariffs, for any of the non-dementia ICD10 codes in the formula. If, for some reason that could presumably be documented, dementia cases referred for a non-dementia diagnosis are treated for cheaper sub-codes, we may expect the difference $AveCost1 - AvePen1$ to be negative. That is because $AveCost1$ is based on the sub-codes whereas $AvePen1$ is based on the coefficients of diagnosis dummies in the formula, that are optimised for all 5 million persons of which only 1% are dementia cases.

Correlation with a 'bent' covariate in PBRA1:

Imagine plotting the 5 million penultimate residuals for the dementia dummy (as y-axis) against the covariate 'area percentage of disability claimants' (on the x-axis)—one of the non-dummy covariates in the PBRA1 model. Least-squares fitting ensures that both the mean penultimate residual and the correlation coefficient in the plot are zero, but there could still be a hill-shaped pattern in which the residuals are predominantly negative for the lower and upper quartiles of the disability percentage.

The average magnitude of all 5 million residuals in PBRA3 is £520 and the 50,000 dementia cases have above average disability. It is therefore quite possible that the penultimate residuals for those cases have an average of around -£436 — which gives around -£436 for the coefficient itself since the denominator can be taken to be unity in this case (see Appendix).

(iii) Counting different morbidities in PBRA3.

Table 1 reproduces the 15 coefficients in PBRA3's three sub-models of the morbidity dummies for whether or not a person had c different diagnoses (morbidities) from D_1, \dots, D_{150} in the two years 2007/08 and 2008/09. Repetitions of the same diagnosis do not count, although they necessarily count in the dependent variable of 3rd-year personal cost.

Table 1. PBRA3 coefficients for five 'morbidity count' dummies

Morbidity count C	Age 0-14 (group 1)	Age 15-64 (group 2)	Age over 65 (group 3)
2	-£230	-£238	-£9
3	-£384	-£409	-£98
4	-£514	-£566	-£186
5	-£624	-£666	-£269
6+	-£653	-£863	-£311

PBRA3 Team (2011) has a ready 'explanation' for the 15 negatives—perhaps seeing them as some sort of benefit of scale as in the adage that 'two can live cheaper than one':

'The impact of the variables that count different morbidities are [sic] negative indicating that the additional costs of having more than one condition is less than the sum of cost of the two individual components.'

There is no indication here of where the saving shows itself, in any formulated and interpretable expression of a 'coefficient', and there is no expression of surprise that the coefficients in Table 1 (plotted in Fig. 1) can be so regular for such a socially interactive process.

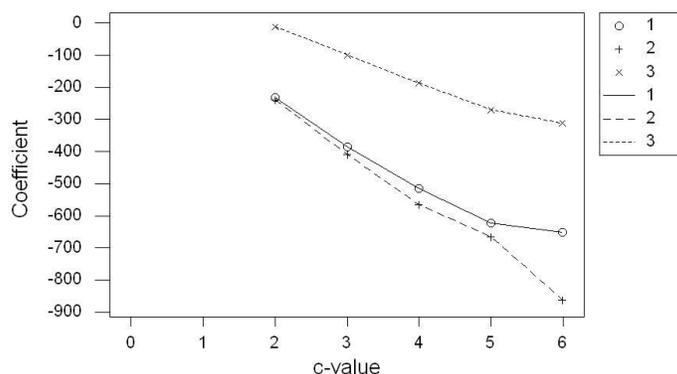


Fig. 1. Plotting the three age-groups (the 6 should be 6^+) with \circ for under-15s, $+$ for 15-64s and \times for over-65s.

So why was the PBRA3 team confident that it had found good reason for debiting GP-practices by £863 for a 60-year old, diabetic alcoholic with six different D_1, \dots, D_{150} referrals in 2007/08 and 2008/09? Fig. 1 shows that, at least for $c \leq 5$ (the unmixed cases), dependence of coefficient on c is remarkably linear in each age-group. (The points that are out of line might be due to small numbers of 15-64s with 5 different diagnoses and of the others with 6 or more.)

To explain Figure 1, it is necessary to go back over a century to the ‘regression to the mean’ concept that Francis Galton developed from his observation that the sons of tall (short) fathers were on average shorter (taller) than their fathers. In this century, the phenomenon had to be taken into account in adjusting the reduction in fatalities that follows the introduction of speed cameras at traffic black spots.

To apply the concept to any of the 15 groups of individuals in Table 1, the only substantial assumption needed is that disease incidences and the average yearly hospital cost (as used in the coefficient numerator) did not change appreciably over the three years of the database. Hospital cost is a variable that, like the number of accidents at black-spots, must be regressing to its mean from year to year—people getting on the sickness ladder (babies), getting off it (the dying) or generally moving up and down. Any large group identified as having a high general sickness in the 2nd-year with a high average hospital cost in that year will be expected to have a lower average cost in the 3rd-year. The regression may be expected to be reflected (even amplified) in the average change between the group’s average cost of the first *two* years and the 3rd-year average (which is *AveCost1* for a dummy that identifies the group), while the two-year average may be well-approximated by *AvePen1*—whence an *AvePen1* larger than *AveCost1*.

Consider the dummy D for 15-64s and $c = 5$, so that $D = 1$ identifies a group that has shown an exceptionally high disease incidence in years 1 and 2. This incidence will regress (unless there is little renewal of the individuals that constitute the group by those with lower values of c) thereby generating a negative numerator (and therefore a negative coefficient) that will be statistically significant if the group is large enough. If D were the dummy for 15-64s with $c = 0$, the ‘less than’ would be (and has to be!) a ‘greater than’ and the coefficient would be positive. The PBRA3 team did not include any dummy for $c = 1$, but extrapolation of the lines in Fig.1 suggests that the inequalities switch between 0 and 1 for under-15s and 15-64s, and between 1

and 2 for over-65s (precisely as expected if the switch point is around the age-band average of c corresponding to the average cost). Moreover, the decrease in the magnitude of the coefficient as c approaches the switch reflects the usual shape of any regression-to-the-mean phenomenon.

The Galton concept has here explained the negatives, and gone some way to explaining other features of Fig. 1—but not yet their striking linearity (quantitative modelling of the numerators of C might be considered, if the denominators could be extracted from the PBRA3 data-base so that ‘observed values’ could be calculated for the numerators). The success of the explanation should not, however, be interpreted as validating the whole PBRA3 formula: its ‘by necessity’ quality contrasts with the *ad hoc* quality of the alternative ethereal ‘explanation’ that the negatives are simply analogous to the benefit of scale in the saying ‘two can live as cheaply as one’.

(iv) Freezing a non-supply covariate to meet ‘differentially met’ need.

One of the 15 PBRA3 covariates for two of its three sub-models was the ‘area BME proportion’—a covariate that is not a dummy variable. The coefficients were both significantly different from zero at the 1% level but of opposite signs: namely, +£30 for under-15s and –£30 for 15-64s if the ‘proportion’ in Table 9 of PBRA3 Team (2011) was a percentage, otherwise only 30p. According to the PBRA3 report, the negative coefficient

‘should be interpreted as an indication of differentially met need and that its effect should not be included when calculating needs-based predictions at practice level.’

Although the covariate here is not a dummy variable, there is a generalisation of the Appendix expression C for smoothly-varying covariates, on which an analogous explanation of a questionable sign can be based. (Continuous covariates may, in any case, have a population distribution approximating that of a scaled dummy.) The PBRA3 justification for freezing the ‘area BME proportion’ adjustment is therefore subject to the same objection as the ‘explanations’ in (i) and (ii) above—namely, that (suspending disbelief in the model) the coefficient should not be significantly different from zero (unless usage by high BME proportion areas has fallen in the 3rd-year for some reason or other). It was, however, the failure of the CARAN report (Morris *et al*, 2009) to find such negative coefficients that led a Minister of Health to change the formula.

6. Conclusion

The primary objective of PBRA research was (manifestly) high-performance prediction of a large component of GP-commissioned health care, but the provenance and structure of the formula to do that seem to have been of secondary interest. The question remains whether an empirical (theory-free) formula constructed with an R^2 as low as 15%, and with an aggregated R^2 at GP-practice level of only 73% after freezing of supply factors, can be an acceptable financial instrument for allocating tens of billions to CCGs and their constituent GP-practices. Should the recommendation of such a formula be respected?— when it is accompanied by explanations of secondary features of the formula (those pesky negative coefficients!) that are demonstrably short of statistical logic and straight-thinking, and when some of those features could

substantially affect the financial balance between a GP-practice and its CCG.

Appendix: Assumption-free derivation of an interpretable expression for the least-squares-estimated coefficient C of a dummy variable D .

Theorem: C is the ratio of

(i) a numerator that is $AveCost1 - AvePen1$, and therefore also the average, $AveRes1$, of the penultimate residuals (observed minus fitted costs) for individuals with $D = 1$ to

(ii) a denominator that is the product $P_0(1 - R_D^2)$ of P_0 , the proportion of individuals with $D = 0$ and R_D^2 , the R^2 for a re-run of the original least-squares computation in which (just as for the penultimate model) D is omitted from the explanatory variables, but with the difference that each of the individual costs is replaced by the corresponding value of D .

The numerator is also the product of P_0 and $AveRes1 - AveRes0$ (the average residual for individuals with $D = 0$), which leads to an alternative numerator $AveRes1 - AveRes0$, for expressing the coefficient as a ratio with denominator $1 - R_D^2$.

Proof: Let \mathbf{d} , \mathbf{d}_c , \mathbf{d}_o , \mathbf{r} , \mathbf{r}_p be n -vectors for, respectively, the dummy variable D , its mean-corrected version, its component orthogonal to the penultimate model variables, and the full and penultimate model residuals. Then

$$n_1 AveRes1 = \mathbf{d}^T \mathbf{r}_p = \mathbf{d}_c^T \mathbf{r}_p = n P_1 P_0 (AveRes1 - AveRes0) \quad \text{and}$$

$$\mathbf{d}_c^T \mathbf{r}_p = \mathbf{d}_c^T (C \mathbf{d}_o + \mathbf{r}) = C \mathbf{d}_c^T \mathbf{d}_o = C |\mathbf{d}_c| |\mathbf{d}_o| \cos \varphi = C |\mathbf{d}_o|^2 \cos^2 \varphi = C n P_1 P_0 \cos^2 \varphi,$$

where φ is the angle between \mathbf{d}_c and \mathbf{d}_o , so that $\cos^2 \varphi = 1 - R_D^2$,

whence $C = AveRes1 / P_0 (1 - R_D^2) = (AveRes1 - AveRes0) / (1 - R_D^2)$, since $P_1 AveRes1 + P_0 AveRes0 = 0$.

The statistic R_D^2 is a measure of how highly the dummy variable is correlated with the other variables (as a whole). Unless R_D^2 is appreciably different from zero, the denominator takes the simpler form of just P_0 , which will itself be close to 1 if the proportion of individuals with $D = 1$ is close to zero. If both conditions apply, the denominator can be taken to be unity, and the coefficient C is given by the value of the numerator alone. It is likely, for example, that the dementia dummy does just that. Although most dementia cases are aged, most of the aged do not have a hospital referral for dementia in a 2-year period and only 50,000 referrals were made during 2005/06 and 2006/07 (a 1% of the 5 million that makes $P_0 = 0.99$)—which means that the R_D^2 for the $-\text{£}436$ coefficient is likely to be close to zero and the numerator itself close to $-\text{£}436$. The alternative numerator is a comparison of complementary residuals in the full model and does not involve the penultimate model. However, it is less easily interpreted since, when expanded for analysis, it involves four rather than two averages.

The main concern for those who trust the model is the fact that the coefficients considered are all significantly different from zero. The fact that the denominator of C in both ratios is a *positive* function of numbers in the database of explanatory variables ensures that a statistically-significant negative coefficient means that (whatever the denominator does to influence the magnitude of C) the $AvePen1$ in the numerator is significantly larger than $AveCost1$ —which is what has to be explained to alleviate the concern.

References

- Dixon, J., Smith, P. C., Gravelle, H., Martin, S., Bardsley, M., Rice, N., Georgiou, T., Dusheiko, M., Billings, J., De Lorenzo, M., Sanderson, C. (2011) *A person based formula for allocating commissioning funds to general practices in England: development of a statistical model*. British Medical Journal, 22 November <http://www.bmj.com/content/343/bmj.d6608> (retrieved 12/12/2013)
- Galbraith, J. and Stone M. (2011) *The abuse of regression in the National Health Service allocation formulas: response to the Department of Health's 2007 'resource allocation research paper' (with Discussion)*. J. R. Statist. Soc. A, **174**, 517-528.
- Morris, S., Carr-Hill, R., Dixon, P., Law, M., Rice, N., Sutton, M., Vallejo-Torres, L. (2009) *Combining age-related and additional needs (CARAN)*. London: Department of Health http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_093169.pdf (Retrieved 12/12/2013).
- PBRA Team (2009) *Developing a person based resource allocation formula for allocations to general practices in England*. London: The Nuffield Trust, http://www.nuffieldtrust.org.uk/sites/files/nuffield/document/Developing_a_person-based_resource_allocation_formula_REPORT.pdf (Retrieved 12/12/2013)
- PBRA3 Team (2011) *Developing a PBRA formula for allocations in 2012/13: Update for ACRA* (draft). London: Nuffield Trust https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213335/ACRA201119-GP-Commissioning-General-and-Acute-PBRA-update.pdf (retrieved 12/12/2013)
- Stone, M. and Galbraith, J. (2006) *How not to fund hospital and community health services in England*. J. R. Statist. Soc. A, **169**, 143-164.