THIS DOCUMENT IS MY WRITTEN EVIDENCE TO THE HOUSE OF COMMONS HEALTH COMMITTEE INQUIRY INTO NHS DEFICITS IN 2006. IF "ACCEPTED" IT WILL BE SUBJECT TO "PARLIAMENTARY PRIVILEGE" WHEN THE COMMITTEE REPORTS.

MERVYN STONE

May 2006

Written evidence to the inquiry into NHS deficits

Mervyn Stone is emeritus professor of statistics in the Department of Statistical Science at UCL. His pro bono interest in the PCT funding formula began in 2002 in response to the wider concerns of The Community Voice, the NW London & SW Herts organisation that represents the interests of NHS patients in at least three PCTs.

Summary

Graphical and tabular evidence is presented to suggest that unfairness in the current funding formula might account for some Primary Care Trust deficits. Evidence is tabled that the formula has been constructed by indefensible statistical methods. An attempt is made to communicate a wider understanding of that evidence in non-technical language.

1 Per capita figures

1.1 Figure 1 of Annex A plots data for the 301 PCTs out of the 303 listed in the Department of Health's document[1] for which figures are available for both axes. The y-axis is the (presumably cumulative) *surplus* (negative if deficit) by the end of 2004-05, expressed as a percentage of the 2004-05 *turnover*. The x-axis is the PCT's *target index* for the period 2003/6, defined as the (*per capita*) ratio of the formula-derived *unified weighted population* to the estimate of the *population* for which the PCT is responsible.

1.2 Table 1 gives the numbers and column percentages of PCTs in the six categories generated by two classifications of the points in Figure 1.

Table 1: Numbers and percentages of PCTs, 2004/05

2004/2005	Target index below average[a]	Target index above average[a]	Total
Surplus	98 (57%)	110 (86%)	208 (69%)
No surplus or deficit	4 (2%)	2 (2%)	6 (2%)
Deficit	71 (41%)	16 (12%)	87 (29%)
Total	173 (100%)	128 (100%)	301 (100%)

[a]Weighted by PCT population.

Table 2 does the same for the six-month forecasts of the surplus/deficit percentage for 2005/06 as rounded to 1 decimal place in the DoH table[2].

Table 2: Numbers and percentages of PCTs, 2005/06

2005/2006 Forecast	Target index below average[a]	Target index above average[a]	Total
Surplus	11 (6%)	20 (16%)	31 (10%)
In (-0.05%, 0.05%)	113 (65%)	90 (70%)	203 (67%)
Deficit	49 (28%)	18 (14%)	67 (22%)
Total	173 (100%)	128 (100%)	301 (100%)

[a]Weighted by PCT population.

1.3 Tables 1 & 2 and Annex A's Figures 1 & 2 reveal a strong positive association between surplus (whether actual or forecast) and target index. For Table 1, only 12% of PCTs have a deficit when their target index is above the national average, compared with 41% when their index is below the national average — the association is so clear that most statisticians would not bother to test it formally. For Table 2, it is statistically significant at $P < 0.001$. The association poses a sceptical question for anyone claiming that deficits of such magnitude are due to nothing more than mismanagement by the responsible chief executives.

1.4 Such a claim rests on the implicit assumption that the target index has helped to create a reasonably level playing field with respect to the variation of health needs between PCTs. Managers should therefore not complain about the up to two-fold differences in *per capita* funding (over the years during which deficits have been accumulated). If accepted, the "nothing more than mismanagement" claim implies that:

> Recruitment of managerial ability by individual PCT boards has been strongly influenced by factors correlated with an index whose values were first ascertainable in 2002 — so that PCTs getting below average target indices have (on the whole) recruited less competent chief executives than PCTs with above average indices.

Here I use the term ‘ascertainable’ rather than ‘published’ since I have not been able to find the *per capita* figures on the DoH website, where the unified weighted population and the estimated age-weighted population are listed in separate tables without bringing them together anywhere.

1.5 An alternative hypothesis is that the distribution of managerial talent has been effectively random across the PCTs of England and that we must look closely at the integrity of the allocation formula that generated the index and that has influenced and continues to influence the *per capita* funding of PCTs. A third possibility is that the formula *is* truly fair between PCTs and that there are historical pressures associated with differential funding that happen to be correlated with the largely innocent index.

1.6 That the formula can be taken to be fair between PCTs is hard to square with the independent analysis here reported, which suggests that fairness could only be fortuitous. I suggest that the reality may involve both a degree of administrative mismanagement in some fraction of PCTs and, almost certainly, some influential defect in the formula itself. Others will have to decide whether the Department of Health’s allegations of mismanagement are based on evidence that goes beyond the size of the deficit and is not just tautological: a good starting point would be to try to find a joint explanation of the interesting features of Figure 1 and the other graphs in Annex A. The next section will document the evidence that the current funding formula is indefensible.

2 A formula based on faith not science

2.1 The funding formula introduced in 2003 has undergone some changes for the 2006/7 and 2007/8 allocations. The changes are documented in the 5th edition[3] of DoH’s helpfully detailed explanation of the formula. The original formula has lasted three years and is the one that may well have influenced the 2004/5 deficits. It was constructed by DoH’s resource allocation team from a variety of recommendations whose details are richly documented in the report[4] from the Information and Statistics Division of NHS Scotland. That report was available for a while in electronic form as DoH’s RARP 26 which is how I will refer to the Scottish report. Copies are now obtainable only by personal application to ISD Scotland or, for personal use subject to copyright, from my own UCL webpage.

2.2 In 2003, in Written Answers for 8 January (pt 19), the Secretary of State for Health described the ISD Scotland research as “research into the new formula”. The Health Committee should be aware that there

appear to be only five[5−9] pieces of published research on the formula *in its integrated operational form*. I therefore request that these items be 'put on the (evidence) table' despite their sadly uniform provenance.

2.3 Ref. 5 is to a refereed paper that focuses on the acute & maternity services component of the formula. Ref. 6 focuses on the mental health component and is being prepared for publication by my co-author Jane Galbraith. Ref. 7 is a less academic piece that tries to communicate our critique to a wider readership (I am grateful to Dr David Green, Director of Civitas, for putting it onto the Civitas website). Ref. 8 provides access to an interactive worksheet that allows any Excel user to see the influence on the target index of the PCT's scores for the socio-economic proxy variables that are weighted together to give the target index. Ref. 9 corrects a mistake in the computer program that supports Refs 5 - 8. These papers are tabled as Annex C for the convenience of the Health Committee.

2.4 To the best of my knowledge there is only one other paper in the academic literature that is directly relevant to the matter in hand. That is the 'horse's mouth' account[10], by Professor Gravelle and the other members of the RARP 26 team, of the theoretical underpinning of one of the statistical models recommended in RARP 26 (a refinement of the 'basic model' for acute care). Since this particular model was not adopted by DoH Leeds, it plays no role in the formula, but I will try to provide a readable explanation of the underlying theory in Annex B. The same theory was used to justify the two models (for acute & maternity services and mental health services) that were pieced together by DoH Leeds to make the formula we now have. Together with the more readable sections of Refs 5 - 7 and the downloadable graphs from Ref. 8, Annex B offers those who can tolerate a little symbolic mathematics a route to a deeper understanding of the formula and its provenance than some of its defenders appear to have.

2.5 The whole critique of a PCT funding formula still based on the ISD Scotland research — and of the machinery of government that imposed it on NHS England — can now be placed in a broader context:

(i) England has been promised an NHS computer system that will bring all NHS GPs into a mutually accessible information data-base of continuously updated patient records that will encompass hospital and pharmacy data.

(ii) GPs are the gate-keepers to the bulk of health expenditure by PCTs and are likely to be the gatekeepers whatever fate awaits PCTs.

(iii) The Treasury may wish to devise new ways of funding that continue to respond to different needs for health resources in different areas of the country.

(iv) Is it conceivable that the Treasury would want to perpetuate any way that depended on the use of

socio-economic proxies rather than on some form of direct measurement of health needs?

(v) It is now four years since RARP 26[4] held that:

> The allocation of resources for health care across geographical areas in the NHS is based on the principle that *individuals in equal need should have equal access to care*, irrespective of where they live. To implement the principle it is necessary to measure need for health care in different areas. But those allocating resources do not have sufficient information to measure need directly [my italics].

(vi) The utilisation/proxy approach to the problem of inadequate information has failed to do more than placate political pressures — in different directions at different times. I suggest that a research programme be initiated that would aim, when implemented, to ensure that allocators *would* have enough information to measure most needs directly, to allocate funds fairly and to collect the evidence of that to satisfy the National Auditor.

(vii) Such a programme would have to throw light into some dark corners, just one of which would be the use by DoH of the phrase “equal health care for equal need” (echoed by RARP 26) which neglects the question of how to judge the priorities of widely different and competing types of health care — a question that an ill-comprehended formulaic approach may have served to conceal.

2.6 For some of us, statistical puzzles are better resolved by verbal rather than written explanation. I would therefore be willing to give a complementary oral account of the evidence.

ANNEX A

Table A1: Variables in the scatter-plots

Variable	*Definition or source*
Target index	(Unified weighted population)/(PCT population): Table 4.3, col. G[11]; Table 5.2, col. H[11]
2004/05 surplus (% of turnover)	DoH document for Health Committee[1]
2005/6 forecast surplus (%)	DoH document for Health Committee[2]
2003/04 *per capita* allocation (£000s)	Table 4.1, col. F[11]
2003/06 allocation increase (%)	Table 4.1, col.U[11]

There are 301 points in each of the appended scatter-plots — one point for every PCT with data for all the variables in Table A1. Such graphs suggest questions but do not resolve them. Curiosity about the possible interaction of target index and the variation in the large increase in allocations between 2003 and 2006 (in their relationship to deficits) leads one to Table A2 — a further breakdown of the surplus/deficit numbers of Table 1 but excluding the six PCTs with neither surplus nor deficit.

Table A2: Numbers and percentages of PCTs

2004/2005	Low[a] target Low[a] increase	Low[a] target High[b] increase	High[b]target Low[a] increase	High[b] target High[b] increase
Surplus	77 (63%)	21 (45%)	42 (88%)	68 (87%)
Deficit	45 (37%)	26 (55%)	6 (12%)	10 (13%)
Total	122 (100%)	47 (100%)	48 (100%)	78 (100%)

[a]Below average target index or 03/06 allocation %-increase

[b] Above average target index or 03/06 allocation %-increase

Looking at the percentages in the first two columns of the table, there is an indication — no more than that — of an association, among PCTs with below-average target index, between the deficit and the 2003 to 2006 allocation %-increase (high increase goes with deficit!). The association is formally significant at $P < 0.03$ not allowing for selection of feature.

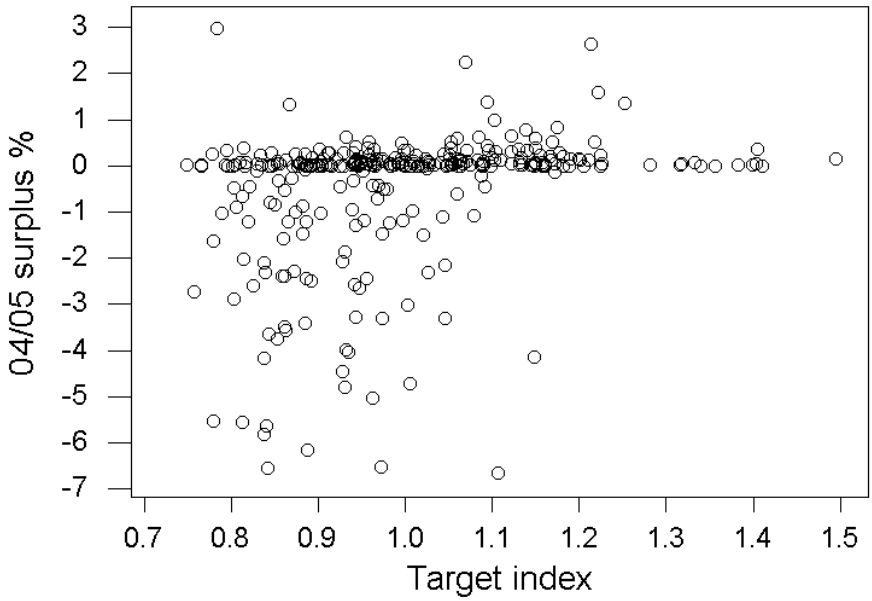

Figure 1: The association between deficit and target index

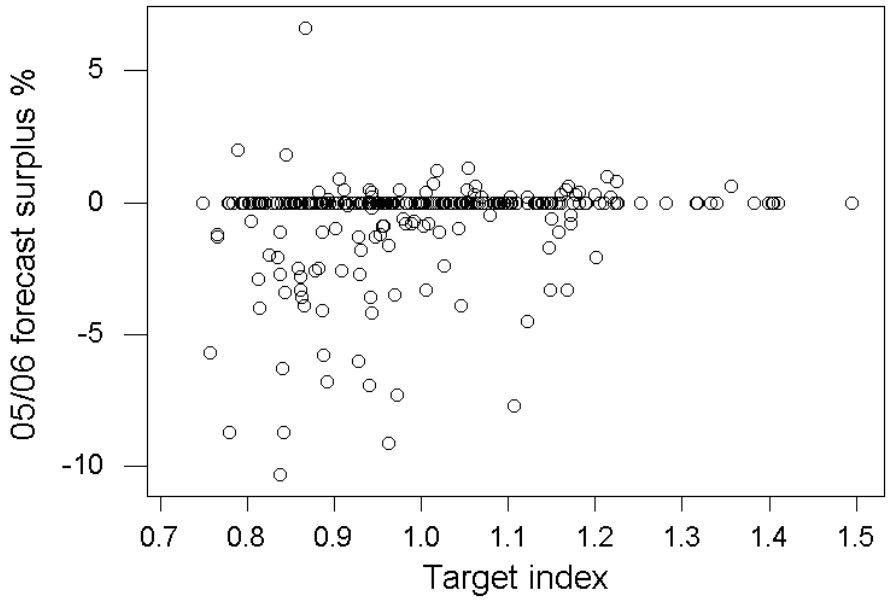

Figure 2: Forecast surplus versus target index

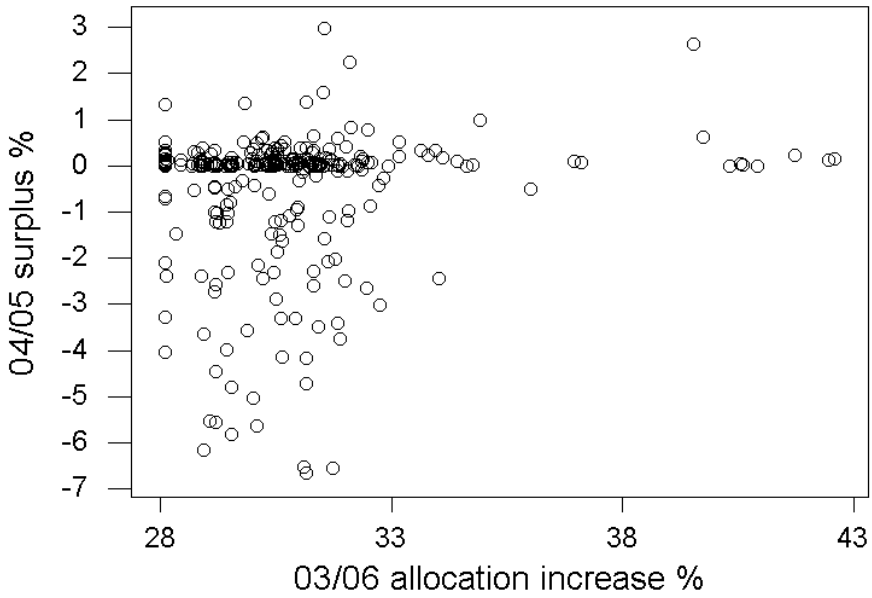

Figure 3: Surplus versus 03/06 allocation increase

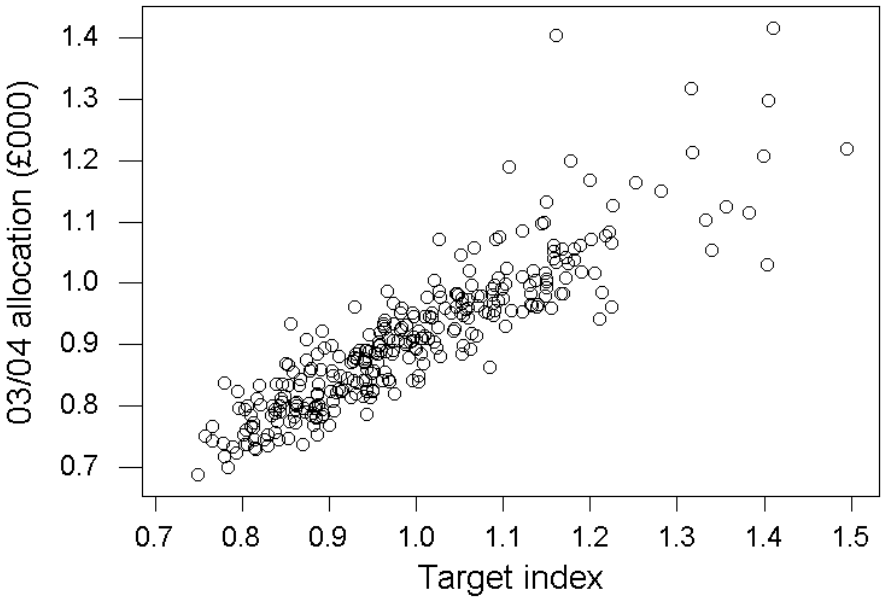

Figure 4: 03/04 allocation per head (£000s) versus target index

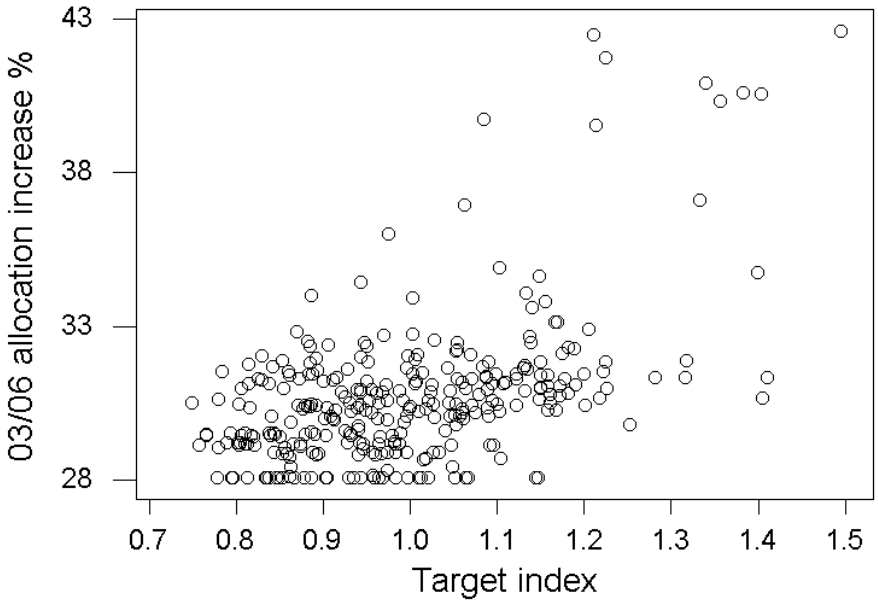

Figure 5: 03/06 allocation increase versus target index

ANNEX B

Understanding the theoretical underpinning

This attempt to express a verbal understanding is necessarily discursive. It uses minimal mathematical symbolism to bridge the gap between airy verbalism and statistical calculation.

B1 The first half of the summary of Gravelle *et al.*[10] reads well, even when broken down by sentence:

1. "Many health-care systems allocate funding according to measures of need."
2. "The utilisation approach for measuring need rests on the assumptions that use of health care is determined by demand and supply and that need is an important element of demand."
3. "By estimating utilisation models which allow for supply it is possible to isolate the socio-economic and health characteristics which affect demand."
4. "A subset of these variables can then be identified by a combination of judgement and further analysis as needs variables to inform funding allocations."
5. "We estimate utilisation models using newly assembled data on admissions to acute hospitals, measures of supply, morbidity and socio-economic characteristics for 8414 small geographical areas in England."

The 8414 areas here are the electoral wards for which values were found to be available for all of the variables thought relevant to the RARP 26 study.

B2 Fine words butter no parsnips! The abstract concepts of "demand", "need" and "supply" have to be related somehow to more concrete quantities. For demand and need, however, it seems that nothing more is required than that we postulate a single quantity D that varies from ward to ward and expresses the demand for some particular category of health service (acute admissions in Gravelle *et al.*[10]) by individuals in a ward, and then a single quantity N that similarly expresses the need for that service. We do not have to define the quantities involved in order to express the quantitative sense of part of sentence B1.2 as the simple equation

$$D \;=\; N \;+\; N_{not}$$

where N_{not} is the part of demand that is not need.
The theoreticians we are trying to understand are happy for us to leave both D and N at this less than concrete level. In contrast "supply" can be represented quite concretely by a number of immediately quantifiable variables. Table 1 of Gravelle *et al.*[10] lists the eight supply variables from "mean waiting time for elective admissions" to "accessibility score for private beds" that were selected from a much larger number. The multidimensional character of supply calls for a distinctive notation for which bold $\mathbf{S}$ will serve.

B3 The remainder of B1.2 claims that "utilisation" or "use" (U say) is "determined by demand and supply". If the word "determined" means anything, it means that there must be a some formulation of U,

$$U \;=\; f(D, \mathbf{S}, \mathbf{R})$$

that, as a function f of *three* arguments, matches the ward's utilisation. The idea is that the variation of U would be maximally explicable by the variation of D and $\mathbf{S}$ from ward to ward, but that a possibly multidimensional element of pure randomness $\mathbf{R}$ is needed to make the function fit perfectly. Being strictly random, $\mathbf{R}$ is quite unpredictable, and the equation here would represent "truth" in the sense that there is no closer or more truthful "determination" of U than would be provided by $\mathbf{S}$ and, if we knew them, D and f — leaving $\mathbf{R}$ in the lap of the Gods.

B4 We need even more general theory before we can get to grips with B1.3! Let $\mathbf{P}$ stand for the "socio-economic and health characteristics" of an electoral ward that are taken to affect the demand in that ward. The symbol $\mathbf{P}$ stands for proxies — socio-economic factors are manifestly proxies for health need (such as the percentage of 17-year olds not going to university) and even the "health characteristics" that find their way into the Gravelle *et al.*[10] formula, such as the census-based 'standardised illness ratio' and 'percentage of babies of low birthweight', are proxies for any *direct* measures of ill-health. Table 1 of

Gravelle *et al.*[10] lists the 17 proxies that were initial candidates for inclusion in the formula that finally emerged.

B5 The theory then expresses the not-necessarily causal relationship between the proxies $\mathbf{P}$ and need N:

$$N \;=\; g\,[\mathbf{P}, \mathbf{r}]$$

where $\mathbf{r}$ is the random element that, when combined with $\mathbf{P}$ in the function g, matches the hypothetical value N of the ward's need. Since N can remain undefined, so can the function g!

B6 We can eliminate the nebulous N between the equations in B2 & B5 and put the resulting expression for the equally nebulous D into the B3 equation. After all that, we still have a mystery wrapped in an enigma:

$$U \;=\; f(\; g\,[\mathbf{P}, \mathbf{r}] + N_{not},\; \mathbf{S},\; \mathbf{R}).$$

This equation may seem to take us a little closer to applicability because it does contain the observables U, $\mathbf{P}$ and $\mathbf{S}$. But the unspecified functions f, g and the nebulous entity N_{not} forcibly remind us that such general theory cannot dictate what reality may have in store. We may have got rid of the fine words but we are no nearer to having edible parsnips. The real point in setting out such nebulous mathematics, is to add force to the question — how has the NHS been able to make the transition from enigmatic to pragmatic? How has the formula that is claimed to measure need been able to emerge from something so insubstantial?

B7 This is the stage where, in the view of many independent statistically savvy observers, the RARP 26 investigators have gone well beyond the "judgement" referred to in B1.4 and have substituted a faith that works wonders for a science that may be unattainable, when pursued along the lines of the general theory just formulated.

B8 The first act of faith was to replace the nebulous expression on the right-hand side of the equation in B6 by a simple combination of the variables $\mathbf{S}$ and $\mathbf{P}$, calculable for each ward. The combination involves numerical coefficients that weight the individual contributions of those variables. The coefficient values were adjusted so that the surrogate formula came as close as it could to reproducing the variations of different standardisations of U from ward to ward *within* each of the 95 health authorities in existence when the building blocks of the funding formula were constructed. (The latest formula continues to use, in the two blocks selected by DoH Leeds, the original values of the adjustable coefficients.) Most of the adjustments were drastic — namely exclusion of the corresponding variable either because it was not statistically significant on some criterion or because its inclusion was judged to be unreasonable in the context of the surviving variables.

B9 The second act of faith was to go beyond $isolating$ (in B1.3) and $identifying$ (in B1.4) a subset of the proxy variables that might merely "*inform* funding allocations" (in B1.4). At the end of the Gravelle *et al.*[10] summary and in the body of the paper, there is more than a suggestion that a fitted formula can usefully influence directly the construction of a utilisation-based allocation formula. As just noted, the suggestion was taken up by DoH Leeds without any changes to the RARP 26 formulae that were adopted.

B10 It becomes clear that faith in subjective judgement is at the heart of the utilisation-based approach when one considers an admittedly unrealistic hypothetical scenario — the kind of extreme case that is widely used to test the logic of propositions in mathematics. Suppose that the RARP 26 investigation had discovered a smoothly varying function of the observables $\mathbf{P}$ and $\mathbf{S}$ (with a modest number of adjustable coefficients) that fitted well-nigh perfectly the values of the then current utilisation U in 8414 wards. Could it be confidently asserted that the function would serve as a funding formula to help meet real need? The RARP 26 investigators would be right to say "No!" to that, and to argue that judgement would have to be used to strip out the socially unacceptable content — not willing to echo Alexander Pope's "One truth is clear, Whatever is, is right." But what assurance would we have that such judgement would leave a formula that would express real need with the precision that any funding formula requires?

B11 Specifically, how strong is the statistical support for the RARP 26 judgements and recommendations relating to the two need indices, for acute & maternity services and for mental health services, that DoH Leeds took from RARP 26? Detailed and rather technical critiques of these are given in Refs 5 - 7, for which what follows offers a readable summary.

B12 So here is an outline of how the need index for acute and maternity services (I_1 in Ref. 5) was extracted from a huge data base of values of U, $\mathbf{P}$ & $\mathbf{S}$ for 8414 electoral wards and the knowledge of which health authority the ward had been in during the years to which the data refer.

B12.1 The "simple combination" introduced in B8 is taken to be a product of two expressions:

$$U \;=\; U_{national} \times M$$

where $U_{national}$ has no adjustable coefficients, being the calculable estimate of ward utilisation you get when you cost every individual by the national average for the individual's age-band alone (so that males quite reasonably share the cost of maternity services). Even before looking at the formulation of the ward-specific multiplier M that raises or lowers the ward's $U_{national}$, it is noteworthy that this particular product is what is dictated (by a simple mathematical argument I will not reproduce here) by two construction principles:

(i) For each ward, the formula for utilisation should be the sum of contributions from each individual that are the same for individuals in the same age-band.

(ii) The ratio of the contributions in any two age-bands should be the same for all 8414 wards in the country, regardless of their socio-economic characteristics.

B12.2 Principle (i) is quite acceptable, but a fine "judgement" might be needed before (ii) is accepted — as a principle. I can find no such "judgement" in RARP 26 or any statistical analysis to show that its implicit adoption might be over-ruling what the data would reveal. (The data on age-band costs at the ward level in 19 age-bands were at hand in the calculation of $U_{national}$.) Rather it seems that the RARP 26 investigators were reluctant to accept the product form (what RARP 26 calls a "two-step age-adjustment") as a way of handling the very different and cost-influential age-profiles of different wards. Part of their remit from DoH had been to investigate "alternative methodologies to replace the age-related components of the existing formulae" (which the DoH steering group may have seen as a form of "age-ism"). The alternatives were investigated but finally rejected in favour of "further investigation".

B12.3 The multiplier M is given a simple and statistically convenient form that has no prior justification:

M = Average level of $U/U_{national}$ for the health authority (HA for short)
+ (Coefficient of 1^{st} proxy $\times$ Deviation of 1^{st} proxy from its HA average) + **...**
+ (Coefficient of 1^{st} supply variable $\times$ Deviation of 1^{st} supply variable from its HA average) + **...**
+ Inexplicable residual.

The dots **...** save paper — they stand for additional contributions (each of them like the products within brackets) from the large number of additional proxy and supply variables involved in adjustment of the coefficients. Such simple expressions (technically known as "linear combinations of adjustable coefficients") were at the heart of the new theory of "combination of observations" that the mathematical genius of Karl Friedrich Gauss discovered over 200 years ago for use in geodetic surveys. Justification for their simple form depended entirely on the fact that, for Gauss, the quantities corresponding to the proxy deviations here were very small, so that a linear combination could be taken to be a satisfactory first-order approximation to the true relationship even if it were not expressible as a linear combination. Such an assumption cannot be made in the present context since proxies can vary considerably e.g. in some health authorities the variable "proportion of non-whites", that plays an important role for the indices, will have been close to 0% in some wards and 100% in others.

B12.4 The adjustment of coefficients ("least squares estimation") is made for each considered choice of variables from a much larger collection — to minimise the sum of the squares of the 8414 differences between M and the *per capita* ratio $U/U_{national}$. Using the term "estimation" to describe the adjustment is somewhat prejudicial — it suggests that there is some scientific entity waiting to be determined. It

should always be borne in mind that whatever reality there is in this construction of a funding formula is man-made.

B12.5 RARP 26 reveals that great attention was paid to the DoH remit that the study should propose " a methodology to adjust for unmet need". The DoH steering group knew, from what the University of York study[12] found for the previous formula, what the new study might be able to find for the relationship between utilisation and the proportion of non-whites in small areas (such as wards). With a different measure of ethnicity ("proportion *not* in black ethnic groups", taken negatively to avoid problems with the logarithm of zero!), the York study had found a statistically significant positive coefficient — utilisation goes down (in the formula) as black ethnic proportion goes up. RARP 26's intermediate "basic model" with ten proxies and eight supply variables duly found a statistically significant "wrong sign" for the coefficient of its measure of ethnicity (as well as for a measure of unemployment). This was interpreted as a "sign" (i.e. indication) of unmet need, especially when three more proxies (indices of morbidity) were added in the finally recommended "model" and the sign remained both "wrong" and statistically significant.

B12.6 The steps to get from the fitted formula M in B12.3 to the recommended I_1 were:
(i) replacement of the HA averages (in the first line of M) by their national average to make a 'level playing field' with respect to any supply effect at the HA level that applied equally to all wards in the health authority;
(ii) setting all the supply deviations to zero for the same reason;
(iii) removal of the terms for the two proxies with the "wrong" signs on the grounds that such removal is equivalent to a correction for "unmet need" (favouring populations with higher non-white proportions or higher unemployment).

B13 The analogous derivation of the mental health index I_2 has two notable differences.

B13.1 For I_1, the *per capita* ratio $U/U_{national}$ to which the formula M was fitted was recognised in RARP 26 as the "indirectly standardised" utilisation . As noted in B12.1, this ratio is what is needed for consistency with two principles for the construction of any utilisation model. The consistency must have been either unknown to, or rejected by, the RARP 26 team because the variable to which the analogous mental health formula was fitted was the "directly standardised" utilisation whose use can be shown to violate the first (reasonable) principle.

B13.2 Unlike the finding in B12.5 for the basic acute & maternity model , the basic mental health model fitted to utilisation in 7982 wards did not deliver a statistically significant coefficient for the 'proportion of non-whites' proxy. So, according to the rules of the statistical game being played, there was at that

stage no “indication” of possible unmet need. RARP 26 nevertheless introduced a fresh variable (the psycho-social morbidity index) and was able to get the statistically significant “wrong sign” that justified correcting for the supposed unmet need, as was done for I_1. Although the resulting very different model was “poorly specified”, it was said to “provide interesting and plausible results”. Such subjective judgements provide the theoretical basis of the mental health index I_2 that DoH built into the current formula.

B14 RARP 26 has to appeal to judgements because its statistical methods are in themselves so questionable. Before getting to the stage of least-squares estimation with a specific choice of variables, that choice has been subject to the screening-out of many variables on the grounds of mutual correlation, even before a residual set is subject to the hazards of “forward and backward stepwise selection”. Except when a remit-inspired objective such as the identification of “unmet need” supervenes, variables survive in the fitted formula only if their coefficients are statistically significant at the conventional two standard-deviations level. Models are considered to be well-fitted and not “poorly specified” if they survive the RESET statistic test against a specific class of alternatives.

B15 RARP 26 does not even tell us what fraction of the variance of the standardised utilisation across all 8414 wards for I_1 or 7982 for I_2 is “explained” by the proxy variables once the supply variables have had their say in the fitting of the formulae. The two values of R^2 (the statistic that is some measure of the combined explanatory power of both proxy and supply variables) — 0.76 for I_1 and 0.36 for I_2 — are in any case too low to exclude other equally plausible (and not simply linear) formulae, that might do much better and maybe (who knows?) get much closer to the measurement of real need.The conclusion has to be that the statistical basis for confidence in both I_1 and I_2 is very weak. I am afraid that, once acquainted with the details of what has been done, most statisticians would conclude that RARP 26 exhibits a naive belief that its fitted formulae — simple linear combinations of whatever the combination of mechanical variable selection techniques and “judgement” ultimately delivered — can be trusted even as guides in the cutting of a large financial cake.

Annex C: Complementary tabled publications

1. “How not to fund hospital and community health services in England.”[5]

2. “Eye of newt and toe of frog: a good formula for health?”[7]

3. “Fathoming the PCT funding formula with Excel graphics.”[8]

References

1. Department of Health (2005) *NHS organisations annual accounts surplus and deficits 2004-05,* www.dh.gov.uk.
2. Department of Health (2005) *NHS organisations forecast surplus and deficits 2005-06,* www.dh.gov.uk.
3. Department of Health (2005) *Resource Allocation: Weighted Capitation Formula.* DoH, London, www.doh.gov.uk/allocations/capitation.htm.
4. Sutton, M., Gravelle, H., Morris, S., Leyland, A., Windmeijer, F., Dibben, C. and Muirhead, M. (2002) *Allocation of Resources to English Areas: Individual and Small Area Determinants of Morbidity and Use of Healthcare Resources*. Report to the Department of Health: Research Allocation Research Paper RARP 26. Information and Statistics Division, Edinburgh, www.homepages.ucl.ac.uk/ũcakmst.
5. Stone, M. and Galbraith, Jane (2006) How not to fund hospital and community health services in England. *J. Roy. Statist. Soc. A (Statistics in Society)* **169** 143-164. Research Report 242, www.ucl.ac.uk/stats/research.
6. Stone, M. and Galbraith, Jane (2005) A PCT funding formula for England based on faith may be wasting billions. Research Report 254, www.ucl.ac.uk/stats/research.
7. Stone, M. (2006) Eye of Newt and Toe of Frog: A Good Formula for Health? Civitas's Statistics Corner, www.civitas.org.uk.
8. Stone, M. (2006) Fathoming the PCT funding formula with Excel graphics. Research Report 267, www.ucl.ac.uk/stats/research.
9. Stone, M. (2006) Corrections to Eye of Newt and to PCTgrapher.xls. Civitas's Statistics Corner, www.civitas.org.uk.
10. Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C. and Muirhead, M. (2003) Modelling supply and demand influences on the use of health care: implications for deriving a needs-based capitation formula. *Health Economics* **12** 985-1004.
11. Department of Health (2003) *Resource allocation exposition books,* www.doh.gov.uk/allocations/2003-2006.index.htm.
12. Carr-Hill, R., Hardman, G., Martin, S., Peacock, S., Sheldon, T. A., and Smith, P. C. (1994) *A Formula for Distributing NHS Revenues Based on Small Area Use of Hospital Beds.* Centre for Health Economics, University of York.