A PCT funding formula for England based on faith may be wasting billions

**Mervyn Stone and Jane Galbraith**

In 2003/4 the 304 Primary Care Trusts that control England’s National Health Service were funded by slicing a Treasury cake of £44B. The division was guided by targets given by a weighted capitation formula, not yet fully implemented but heavily influenced by two statistically-derived indices based on socio-economic variables acting as proxies for actual health needs. This article maintains that these indices are in effect based on faith rather than science and implicitly expresses the hope that this particular faith/science divide will not block future attempts to devise a funding formula based on direct measurement of health needs rather than unjustifiable and politically manipulable proxies.

Mervyn Stone is an emeritus professor in the Department of Statistical Science, UCL.

Jane Galbraith is a senior teaching fellow in the Department of Statistics, LSE.

Two years ago “Beggars belief” (Hacking, 2003) was the engagingly ambiguous title of an article in Health Service Journal on the then new Primary Care Trust (PCT) funding formula that would have been used in 2003/4 to allocate £44B to the 304 PCTS of England if it had been fully implemented from year one. For its author, it beggared belief that it was going to take more than 20 years to reach the allocation targets dictated by the formula — and thereby to “close the north-south divide”. For others, however, it beggars belief that the allocation of 44 billions could be even influenced by a formula based on faith rather than science. The Beggars Belief article claimed that

> “Unlike the derivation of the formulae themselves which, after the decision to commission the work and the issuing of a broad remit, is a technical and statistical process, the setting of the pace of change inevitably constitutes a practical and political judgement [our italics].”

It went on to calculate how many billions some PCTs (“winners”) will gain from “losers” in the two decades until target allocations are reached. These billions are not wasted in the strict sense of the word, but there is potential waste in their allocation to PCTs that, according to formula, do not need them as much as “losers” — waste being a relative matter. The win/lose calculations and the apparent acceptance of the “technical and statistical process” that delivered the formula suggest that Hacking then thought

---

1Research Report No. 254, Department of Statistical Science, University College London. Date: July 2005.
there might be no such waste in the finally implemented formula i.e. that it was “about right”. For reasons that were set out in some detail in Stone and Galbraith, 2006, and that will be reinforced here, we see the “process” itself as seriously defective — and in no way providing a sound basis for thinking that the formula is “about right”.

**Appreciating the formula**

By “appreciation” of a formula we mean the informal mental processes (value judgements, intuition based on previous knowledge, considerations of plausibility, and the like) that help people to come to a view about its “rightness” without much attention to the question of its formal justification.

The current “weighted capitation formula” (Department of Health, 2003a; WCF for short) is what determines the target allocations for each of the 304 PCTs of England. Figure 1 shows how the “opening composite” target allocations for 2003/4 deviated from the (unacceptable) formula that would provide the same funding for every individual. For that formula, allocation would be proportional to population and all the PCTs would lie on the same line through the origin.

For 2003/4, over 93% of the total target allocation was for the so-called “unified” targets (Department of Health, 2003a). These were calculated for each PCT by the formula:

\[
\text{(Unified Target Population)} = 0.8276 \times \text{(HCHS)} + 0.0067 \times \text{(HIV/AIDS)} + 0.1407 \times \text{(Prescribing)} + 0.0250 \times \text{(GMSCL)}
\]

(1)

where (HCHS) is a “weighted population” — a notional fraction of the national population proportional to an estimate of what the PCT is supposed to need for its Hospital & Community Health Services.

Likewise for the other three components. (GMSCL stands for General Medical Services Cash-Limited.) Because the weights 0.8276, 0.0067, 0.1407 and 0.0250 sum to one, formula (1) is also a weighted population. Each of the four components has its own specialised provenance. The HCHS component is calculated as

\[
\text{(HCHS)} = \text{(PCT population)} \times I_{\text{age-profile}} \times I_{\text{need}} \times I_{\text{MFF}} \times I_{\text{EACA}}
\]

(2)

where MFF stands for Market Forces Factor and EACA for Emergency Ambulance Cost Adjustment. The first index, \(I_{\text{age-profile}}\), is the ratio of (i) the estimated average cost of the PCT’s HCHS, when the individuals in its population are assigned the per capita national HCHS cost for the age-band (one of
Figure 1: Target allocation for 2003/4 against population for the 304 PCTs. The straight lines are lines along which per capita allocation would be constant. The main reason for their divergence is the allowance the formula makes for inequalities in socio-economic variables and regional costs over and above the allowance for differences between PCT age-profiles.
seven) in which they fall, to (ii) the national *per capita* HCHS cost. The variation of $I_{age-profile}$ from PCT to PCT therefore depends only on differences in the age-profile (factor 1 in Table 1). The second index $I_{need}$ is a combination of two indices concocted from socio-economic factors— $I_1$ for acute & maternity services and $I_2$ for mental health services. $I_1$ is a function of no less than nine factors (2-10 in Table 1) but $I_2$ requires only four (11-14). Factors 15-19 complete the list of factors that need to be estimated for each PCT as a whole in order to calculate (HCHS) in (2) and (Unified Target Population) in (1).

**Table 1. The factors or variables directly involved in calculating the Unified Target Population.**

1. Age-profile
2. Standardised mortality ratio under 75 years
3. Proportion of low birthweight babies
4. Standardised birth ratio
5. ID2000 education domain score
6. Proportion of over 75s living alone
7. ID2000 income domain score
8. Nervous system morbidity index
9. Circulatory morbidity index
10. Musculo-skeletal morbidity index
11. Comparative mortality under 65 years
12. Proportion of over 60s claiming income support
13. ID2000 housing domain score
14. Psycho-social morbidity index (PSMI)
15. MFF
16. EACA
17. HIV/AIDS
18. Prescribing
19. GMSCL

The four morbidity indices are themselves combinations of socio-economic factors.

For comparison of PCT unified targets, the quantity of interest has to be the *per capita* ratio of the Unified
Figure 2: Comparison of four PCTs in the sequential build-up of the partial index for equation (1), as more and more factors are “incorporated” (i.e. allowed to express their variation from PCT to PCT). The first point plotted is for the null model of equal per capita allocation. The last point is the target index for the recommended formula i.e. the partial index when all 19 factors have been incorporated (i.e. the *per capita* ratio of Unified Target Population to Population).

Target Population (1) to the PCT population — which we will call the *target index*. We have devised a computer-based graphical exposition that provides insight into the influence that particular factors have on the target index of any particular PCT. This is done by calculating a *partial index* — defined as what the target index would have been if a number of factors were given their PCT values and the remaining factors were set at their national population-weighted average values (the same for all PCTs). The 19 factors are incorporated sequentially as in Table 1 (which is how they entered the calculations of WCF’s Appendix 12). Figure 2 uses the data for 2003/4 from the relevant Resource Allocation Exposition Book (Department of Health, 2003b) to compare Hillingdon (the first author’s PCT) to three PCTs with
contrasting socio-economic profiles. The target index is the last value plotted in each graph — 1.04 for Hillingdon, 0.94 for East Devon, 1.55 for Tower Hamlets and 0.74 for Wokingham. The first point in Figure 2 corresponds to the “null” equality model in which everyone has the same per capita allocation so that there is no weighting of PCT populations. The second point has the partial index in which all factors in the formula except factor 1 (age-profile) are set at their national average values. The third point then adds the education variable to the age-profile factor to show its additional influence on the target index, while factors 3-19 stay at their national values. And so on, until we get to the target index itself.

We can see, in Figure 2, the initial dip for Tower Hamlet’s youthful population in contrast to the rise for East Devon’s aged. Variations in age-profile can have a major influence — but it is one that is usually outweighed by the cumulative effect of the multiplicity of factors intended to “replace the age-related components of the existing formulae” (as AREA was pressed to do in its remit). The other major single influence is Market Forces (factor 15) where only East Devon does not benefit, in contrast to the three PCTs in the London area.

The technical quicksand of faith

Where does the current formula come from? Equations (1) and (2) have a multifaceted provenance but the responsibility for the components I_1 and I_2 of I_{need} in (2) is more specific. Their construction and the recommendations that they be used are documented in a key report entitled “Allocation to English Areas” (Sutton et al, 2002; “AREA” for short). This study (by a team led by the University of Glasgow and ISD Scotland) was commissioned by the Department of Health (of England) with the broad remit to produce a “fairer formula for the new NHS [National Health Service]”. The indices I_1 and I_2 were then taken from AREA without updating the numerical values of the coefficients that had been estimated from data going back to the 1991 Census. Twenty-one pages of WCF are devoted to step-by-step arithmetic illustrations of how the targets would be calculated for four hypothetical PCTs. These pages have helped us to derive the simple algebraic formula (Stone and Galbraith, 2006) that allows one to see how I_1 and I_2 influence the value of (HCHS) in (1), over and above an allowance by standardization for differences in the age-profile of PCTs.

The Department of Health documents provide the reassuring arithmetic detail of the calculation of the target allocation for each PCT — but do not offer any theoretical insight. The AREA thinking was set out with commendable clarity as follows:
“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. The basic assumption underlying the resource allocation procedure is that use of health services is determined by patient need and by supply. Statistical modelling of the relationship between utilisation, socio-economic variables (including measures of health) and supply factors can identify which socio-economic variables are indicators of need because of their effect on utilisation. Information on the level of such socio-economic variables across areas can then be used to allocate resources in a way that reflects need.” (AREA, §1.1)

The separate “statistical modelling” of existing levels of utilization of acute & maternity and mental health services (but with different age-profile standardisations of utilization) turned out to be little more than the well-known, if often misused, technique of least-squares multiple regression. The quotation from Hacking 2003 implicitly assumes that the formula modelling the relationship will provide a measure of the true need for the corresponding health services. But it will not do that if it has serious logical and statistical deficiencies. The modelling (based on data at the level of electoral wards) eschewed simple usage data (as part of the formula itself) as well as any prospect of employing direct measures of actual health need. Instead, it employed over 80 socio-economic factors (described as “need” variables) to which were added a large number of measures (described as “supply” variables) of how wards differ in their access to services — in all, many more than the thirteen (9+4) factors that break surface in \( I_1 \) and \( I_2 \). In this section we look at just one feature of the logic involved in the construction of \( I_2 \), the simpler of the two indices — and find disturbing deficiencies that complement those already noted in Stone & Galbraith (2006) which concentrated on \( I_1 \).

Of the £34B (44 × 0.93 × 0.8276) for HCHS about £5B are for mental health. AREA got its numbers from something of a witches’ cauldron — a brew into which went the very large assortment of “need” and “supply” variables. The idea was that an automatic statistical procedure based on repeated multiple regression would boil these down to a smaller number that would explain the differences, between electoral wards, in estimates of utilization of mental health resources standardized for differences in ward age-profile.
The “explain” here was in the sense that, lurking in the data for 7982 wards with their suspected but unknown unfair allocations of resources, is a relationship between utilization and selected “need” and “supply” variables that expresses the true need for mental health care — a relationship that can be pulled out of the statistical brew and used to formulate a fairer division of the five billions. If a “need” variable in the estimated relationship has a positive coefficient, that is to be interpreted as an unbiased expression of true need — but, if negative, as an indicator of unmet need in wards that have an above average level of the variable. “Supply” variables such as the distance of the ward from mental health hospitals are expected to have negative coefficients. Their inclusion in the modelling is so that their variation can be excluded, in the form of a reduced formula that will impose equality between wards in those variables, at least conceptually.

The reduced mental health index $I_2$ that AREA finally recommended was:

$$0.385 + 0.358 \times \text{Comparative mortality factor under 65} + 0.338 \times \text{Proportion over 60 claiming income support} + 0.034 \times \text{Deprivation index for housing} + 0.636 \times \text{Psycho-social morbidity index}$$

where italics indicate that a variable has been standardised to have a population-weighted average over PCTs of zero and a standard deviation given by the coefficient of variation of the variable. This formula does not include the “need” variable to which the AREA report pays great attention — the proportion of non-white ethnic minorities (PNW) — for a reason that merits further explanation. PNW was included in the pool of variables used by AREA for the statistical brew — but it was not among the nine statistically significant variables. AREA’s logic was that this non-appearance might be the expression of a mutual cancellation — an unmet need among non-white ethnic minorities (for a fixed level of need, utilization goes down as PNW goes up) cancelled by the effect of a correlation between PNW and some morbidity measure not yet thought of (PNW would act as a proxy for the omitted morbidity measure and, according to the formula, utilization would go up with PNW).

The search for the missing measure led AREA to Health Survey for England (Erens et al, 2000) samples of individuals that show a higher incidence of long-standing mental illness in communities with higher values of PNW. It was a logistic regression of the binary variable “whether or not an individual has such illness” on a multiplicity of socio-economic variables that delivered the psycho-social morbidity index (PSMI) used in $I_2$ as a proxy for the incidence of mental illness (despite its having a very low pseudo-$R^2$). As expected, PNW was one of the selected socio-economic variables with a major influence.
on the value of PSMI (as AREA’s Appendix 7 reveals). When PSMI was included in the basic model regression brew, AREA got the numbers it was looking for. The unreduced formula attributed statistical significance to PNW and the ID2000 deprivation index for education — as the two “need” variables with negative coefficients that AREA interpreted as evidence of unmet need. The part of the formula that involved PNW and PSMI was $0.636 \times \text{PSMI} - 0.034 \times \text{PNW}$, in which predicted utilization goes down as PNW increases (other things, including PSMI, being held constant). This was taken as evidence of unmet need because the alternative explanation — negative correlation between PNW and a true measure of mental health morbidity inadequately represented by PSMI — was ruled out by the finding in the Health Survey for England data that PNW is positively correlated with mental illness. To correct (in the allocation formula) for the variation in this putative unmet need, AREA logic replaces PNW (in the term $0.034 \times \text{PNW}$) by its national average and thereby gets a mental health index that will benefit PCTs with large values of PNW at the expense of PCTs with low values. By dealing with PNW in this way, AREA appears to have satisfied the remit to propose a “methodology to adjust for specific unmet need”.

We make the following comments on the general and statistical logic of all that:

(i) The authors of the formula displaced by AREA conceded that their underlying model was “almost certainly a simplification of reality” (Carr-Hill et al., 1994, Sec.3.2.1) and that they had “no incontrovertible reason a priori for excluding any [socio-economic variable] from possible consideration — only those socio-economic variables that proved statistically important were eventually retained” (Sec.5.4). They emphasized the purely descriptive character of their model thus:

“Because of the high level of correlation between many of the variables, the variables were to some extent arbitrary, in the sense that a different combination of variables may have been able to yield roughly similar explanatory power ... not too much meaning should be attached to the actual variables selected, as there may be other combinations of variables which capture the same needs characteristics. That is, the variables selected are proxies for unmeasurable social factors which could be equally successfully captured by other variables, so the precise variable selected is less important than the social factor it partially captures.” (Sec.5.16)

In other words, you have been warned!
(ii) AREA did not dispute the grounds for the warning — rather there is a suggestive transition from the hardness of the term “measure” to the softness of “indicators” in our quotation of AREA thinking. It is the ad hoc character of the AREA “methodology” for PNW that emphasizes the lack of scientific objectivity in what was done. Such objectivity would offer some justification for AREA’s faith that their statistical modelling has some causal rather than purely descriptive relationship to true need — and that money reallocated according to some coefficient in the favoured relationship would actually correct an inequality. Without such justification, concepts such as “unmet need” and the equalization of supply have no theoretical base. The multiple correlation coefficient between the fitted and observed values of age-profile standardized utilization was only 0.60. This leaves plenty of room for other better or equally badly specified models, including ones that might have been just as responsive to other interpretations of the Department of Health remit. (A correlation of 0.60 explains only 36% of the conventional statistical measure of variation, and the statistical model of which $I_2$ is a portion actually failed by $P < 0.0001$ to pass the only model-specification test used in the AREA document!) The “methodology” is assumed to make sense in the framework of the highly specific model presented, in which utilization is a product of a factor based on age-profile and a sum of non-interacting terms for the selected explanatory variables. The AREA account does not adequately acknowledge the enormous influence that selection of apparently statistically significant variables, by automatic procedures such as backward elimination from a large number of “explanatory” variables, has on the outcome. Such methods are misleading enough in open-ended scientific research where it is possible to reject — or follow up by further research — any clues they may generate. They risk misapplication when the outcome is to be used, as here, in large-scale financial management.

(iii) Of the 8414 electoral wards in England that were available for analysis, 432 had to be excluded just because their populations had made no use of mental hospital services. (For such wards, the particular “supply” measures in the highly specific statistical modelling were, like $0/0$, indeterminate. This exclusion puts one in mind of the statistical howler that resulted in disaster on the final mission of the Challenger shuttle: missions were excluded from the safety analysis if, at their typically higher temperatures, there had been no problems with freezing O-rings.)

(iv) The negative coefficient finally uncovered for PNW may have been a combined result of (a) inadequate modelling of the effect of age by the adopted age-standardization of utilization that was used and (b) the relationship between PNW and age-profile — PNW would try to correct the inadequacy. The
strong dependence of PSMI on PNW as one of its inner constituents may also have had a similar artefactual consequence.

(v) If PNW required special consideration on grounds of omitted variables, the same consideration might have been equally justifiable for other potential “need” factors such as “Proportion without access to a car or van”.

Faith rather than science is needed to see evidence here of unmet need — and even greater faith to believe that $I_2$ delivers a fair measure of true need for mental health services. Similar faith is needed for the second formula that determines the “acute and maternity” inequality index, $I_1$, and for the questionable way (Stone & Galbraith, 2006) in which a weighted combination of the two indices is then used to multiply an age-profile standardized utilization without separating mental health from acute & maternity, to give the target allocations for PCTs.

**Opinion, authority, or argument**

Dividing the Treasury cake for the NHS to take account of the different health needs of its autonomous administrative areas has a contentious history (Department of Health, 1989a). It goes back to the simple “Crossman formula” of the 1970s, when the areas were Regional Health Boards and the formula used only rudimentary information about usage (bed and case numbers). The following opinions reveal a deep and widely held scepticism about the use of increasingly complex funding formulae:

(i) “...modelling techniques cannot deal adequately with a system where demand, utilization and supply are so inextricably intertwined...The use of the formulae gives too much importance to variables which only account for some of the variation...The fundamentally political nature of the construction of an allocation policy and the choice of indicators of need must be acknowledged.” (Sheldon & Carr-Hill, 1992)

(ii) “...the search for an empirically based resource allocation formula of high precision in the name of promotion of equity is largely fruitless...Attention should increasingly be focused on how resources are used — the effectiveness and appropriateness of health service interventions...weighting in the dark [is] an industry which is abstracted from research into, and the management and delivery of, health services.” (Sheldon *et al*, 1993)

(iii) “We have become besotted with the production of ever more refined empirically based formulas.”
(Sheldon, 1997)

(iv) “...a professional consensus should be developed regarding the major methodological issues in this field. If this cannot be provided, it implies either an arrogance among statisticians or the absence of a robust methodological core to the work of the profession, neither of which I believe to be the case...I wonder whether there is a role for the [Royal Statistical] Society in encouraging and endorsing transparency in this area.” (Derbyshire in Smith et al, 2001)

(v) “My third concern is with the existence of the large amounts of unexplained variation in the author’s models.” (Goldstein in Smith et al, 2001)

(vi) “…however sophisticated the methods, estimating needs weights is essentially contested, that is to say a concept whose application is inherently a matter of dispute.” (Bevan in Smith et al, 2001)

Do these opinions alone provide authority for saying that the formula the Department of Health has taken from AREA are damaging the NHS? We do not think they do — or that they should do. Recall that, when Galileo made his stand against Earth as the centre of the World, the Pope was the greater authority! The health economists who devised the latest formula may well claim that they are more up-to-date and therefore greater authorities than the experts who have gone before. As well as being unscientific, appeal to authority is risky! Two of the critics just quoted went on to help devise the formulae (Carr-Hill et al, 1994) that AREA now displace — what do they say now? In the end, any dispute has to be resolved by the quality of evidence-based and logically coherent argument. Comparison of the AREA formulae with those displaced reveals how academic recommendations that lack a driving logic can be changed by a slight reformulation of the political remit. This has happened even though the two teams of health economists involved appear to have a similar approach — if we can ignore the oddity that there is only one item in common between the 51 literature references of one team and the 49 of the other. It is not that one can prove (or that critics should have the onus of proof) that the present formulae are wrong. It is enough to establish that we have no way of knowing whether the target allocations they result in are making reasonable allowances for the inequalities in actual health need between PCTs. AREA is not just a simulacrum of statistical science — it is also a wasteful diversion from the kind of broader and harder statistical thinking (not excluding value judgements) that will be needed if we are to replace faith by science.
What can be done?
The problem remains of how to divide those billions. The first step is to lose respect for the idea that “Whitehall knows best” or, more to the point, that it knows who knows best. Under ministerial pressures, civil servants should be praised when they are able to get reasonably close to a scientific formulation of the issues in any complex technical problem — but they should not be required to decide which of a number of potentially disputing groups of outside “experts” responding to an Invitation to Tender would provide the best solution. Disputes between experts are not easily handled, but it should be a matter of concern that in many cases they have not been tackled head on in an effort to establish a widely-accepted academic consensus. One of us has suggested (Stone, 2005) that one of the obstacles to getting quality into government science may well be the restrictive contractual arrangements that inhibit the informative expression of a wide range of truly independent external opinions in the crucial early stages of any enterprise. As far as PCT allocations are concerned, we ask:

“Should we be deterred by the potentially high cost of finding, developing and testing a nationally acceptable way of directly measuring the aggregate health needs of individual PCTs? The acquisition of the necessary data, and the hard thinking and value judgements needed to guide and exploit it, would not come cheaply — but just one percent of the PCT budget for one year would give £450M. Perhaps we should ask ourselves how confident we can be that the current waste from the continued use of proxy formulae is not well in excess of that one percent.” (Stone & Galbraith, 2006)

Postscript
Some observers of the current scene believe that, given the improvement in the quality of data in General Practice we are promised over the next few years, it is perfectly reasonable to consider using direct measurements to estimate the aggregate cost of a PCT’s health needs, rather than proxy variables such as PNW or the proportion of over-75s living alone. For example, Hillingdon PCT appears to be using, as much as it can, direct measurement of health needs to allocate resources — for instance, in the location of diabetes clinics. An interesting feature of its activity concerns mental health, for which Hillingdon is the only London PCT providing its own services. The mental health inpatient costs used by AREA in devising its formula were for the year April 1st 2000 to March 31st 2001. During that period, “dump
codes” were used by Hillingdon for the unknown or non-existent postcodes of 12% of admitted patients, including those brought under Place of Safety orders from Heathrow and other public transport destinations in Hillingdon. (AREA gives its national figure as 13%.) In estimating the ward-by-ward utilization of mental health services for its multiple regression brew, AREA assigned the statistics for this “dump code” activity to electoral wards in the proportions that those wards were represented at the Hillingdon hospital by patients with genuine codes — thereby incorrectly associating their costs with the socio-economic characteristics of the artificially assigned wards. How seriously that may have influenced the form of AREA’s mental health index is a vexing question that would not have arisen with direct measurement of health needs.

References


Sheldon, T.A. (1997), Formula fever: allocating resources in the NHS (editorial). British Medical


Sutton, M., Gravelle, H., Morris, S., Leyland, A., Windmeijer, F., Dibben C. and Muirhead, M. (2002), Allocation of Resources to English Areas (AREA): Individual and small area determinants of morbidity and use of healthcare resources. Department of Health, Scotland: Information and Statistics Division (ISD), Edinburgh. [This paper, containing the only full theoretical defence of the formulae that will guide PCT fund allocation until 2006, has been withdrawn from official websites but is available for personal use subject to copyright by clicking the first author at www.ucl.ac.uk/stats/department/staff.html.]