New approaches to estimating commissioning budgets for GP practices

Person-based Resource Allocation

Research summary
Martin Bardsley and Jennifer Dixon

December 2011
In April 2010 the Department of Health introduced a new approach to setting target allocations for hospital care to commissioning general practices – Person-based Resource Allocation (PBRA). This approach could be a significant component of the way that target allocations are calculated for the new clinical commissioning groups in the NHS in England. This research summary describes what lies behind the approach, developed by the Nuffield Trust and partners, and some of the possible implications.

Key points

• Up until recently it has not been possible to develop an accurate needs-based funding formula for calculating target allocations for commissioning practices. Recent improvements in access to computerised datasets and linking data at the level of the individual have made it possible to test and build statistical models that exploit and aggregate information about an individual’s predicted future costs.

• Using these innovative person-based methods, good-quality models for predicting future hospital costs were developed. The most important factors predicting future costs measured at person level were age and morbidity.

• The models performed well by international standards, predicting over 75% of the variation in the next year’s costs at practice level. Subsequent development has increased this predictive power to 85%. This is a real improvement on previous methods.

• Though the models performed well in explaining variation, there was still a relatively large proportion of practices where observed and expected costs varied by more than 10%.

• While the PBRA formula can be used to set target allocations, other factors outside the formula will have considerable impact on the actual budgets received by practices (and commissioning groups). These include the accuracy of GP lists of the population registered, the historic allocation and local policy of ‘pace of change’ to the target allocation, and arrangements for sharing financial risk.

• The PBRA approach is a sophisticated form of capitation payment. Its primary application is for commissioners – yet it could also be adapted to set capitated funding for providers spanning primary and secondary care, and so incentivise preventive care.

Find out more online at: [www.nuffieldtrust.org.uk/pbra](http://www.nuffieldtrust.org.uk/pbra)
Background

Tax funds for the NHS are distributed to NHS commissioning bodies, mainly on the basis of the number of people they commission care for and to a lesser extent on that population’s health needs. Resource allocation formulae have been used in the NHS since 1976 to distribute funds in a way that aims for equal opportunity of access to health care. Since 1999 a second aim has been to contribute to the reduction in avoidable health inequalities, not discussed further here.

For commissioning entities (primary care trusts (PCTs) and general practices) the resource allocation formulae produce a ‘target’ allocation for the next budget year, which usually differs from the previous year’s budget. The idea is that the commissioning body will move towards the target allocation over a period of time (usually years) at a ‘pace of change’ determined by the Department of Health (in the case of allocations to PCTs) or by PCTs (in the case of allocations to general practices for commissioning). This is usually arranged in such a way that no commissioning body will be given a lower budget in any given year, but there will be ‘differential growth’ in funds depending on their ‘distance from target’ and the policy on pace of change towards that target.

The way that funds are distributed may well change with the abolition of PCTs

The actual budgets received by a commissioning body will therefore depend upon several factors:

- the population size covered
- the health needs of the population
- the distance between the previous year’s actual budget and the target allocation
- the national or local ‘pace of change’ policy
- local factors such as the Market Forces Factor, which takes account of regional cost variations.

The way that funds are distributed may well change with the abolition of PCTs and with the advent of the new clinical commissioning groups from April 2013 onwards. There could be two main changes, as discussed below.

First, the way of assessing the number of people covered by a commissioning body might change. Funds may no longer be distributed first to commissioning groups and then to their constituent practices (based on the population size of the clinical commissioning group, defined by the number of persons on the registered list for each practice). Instead, it may be possible for funds to be allocated directly to practices; the total of these budgets will thus be the allocation to clinical commissioning groups.

In some scenarios this could mean that funds are shared out according to populations defined by GP registration data rather than area-based census data. This is important because, while GP registration data are becoming more accurate over time, there is a discrepancy between the sizes of populations defined by the census and those defined...
through GP registration data. In most areas of the country the population size defined with both methods is within 2%, but in some areas, notably London and other urban areas, the discrepancy is larger (ACRA, 2010) and so this approach will place a renewed emphasis on ensuring accurate GP registration lists in future. (This is not discussed further in this research summary.)

Second, the way of assessing the health needs of the population covered will change slightly. At present there is a range of resource allocation formulae covering the main elements of expenditure, for example primary care, prescribing, mental health, and acute hospital and community health services. This summary covers developments in the way that needs are assessed in the formula for allocations for most acute hospital and community health services – the hospital share covers £22 billion, or approximately one third of the commissioning budget for future commissioning groups.

In the past, health needs have been assessed mainly using information on populations drawn from the census, such as socioeconomic deprivation. But because of advances in the availability of information and ability to link different sources of information in a way that protects individual confidentiality, much more information is now available at an individual level that can indicate health needs. This means two things. First, a formula can be built that more accurately reflects the needs of the individuals in the population served. Second, the health needs of practice populations can be calculated as the sum of the needs of each unique individual registered.
Scope of the formula

The PBRA formula covers allocations to general practices for commissioning NHS-funded hospital inpatient, outpatient and accident and emergency (A&E) care, and for community health services. It does not cover allocations for mental health care, maternity care, prescribing or general medical services, which are covered by separate formulae. It also does not cover expenditure for specialised services, since these are to be commissioned by the NHS Commissioning Board rather than by individual groups or practices. The formula covers allocations for the population of individuals registered with a general practice in England. Allocations for unregistered persons are made separately.

Approach to the formula

The broad approach to developing the PBRA formula is similar to that used in previous formulae (Bevan, 2009): a statistical model is built to try to explain the costs of hospital and community health services of a practice population, using information on the health needs of that population and the supply of care available to them. In calculating a target allocation for a population, only the health needs of the population are taken into account – supply factors (such as the number of hospital facilities available) do not directly influence the level of the target allocation. The model was developed to predict hospital costs, because of a lack of robust information on community services, but is used to help set target allocations for both hospital and community health services.

The formula attempts to explain the NHS hospital costs incurred by individuals in a particular year, by looking at a range of factors from the previous two years. These include the health needs of that individual, such as: their age, gender and recorded illnesses; area-based information associated with health needs (such as socioeconomic deprivation of the area); and information on the local supply of hospital or other NHS services that are then attributed to the individual. The precise needs and supply

Box 1: The scope of the PBRA model

<table>
<thead>
<tr>
<th>The model covers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>allocations for most acute hospital services (worth £22bn) and community health services</td>
</tr>
<tr>
<td>allocations for all people registered with a general practice in England.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The model does not cover:</th>
</tr>
</thead>
<tbody>
<tr>
<td>allocations for maternity services, mental health care, primary care</td>
</tr>
<tr>
<td>allocations for people who may be eligible for NHS care but are not registered with a GP, such as asylum seekers and military personnel – these funds are distributed using separate arrangements</td>
</tr>
<tr>
<td>private funding of NHS-provided care.</td>
</tr>
</tbody>
</table>
variables (and their weightings) that are included in the recommended formula are determined empirically by the extent of their influence on the individual’s costs, and their plausibility (the extent to which they have been shown in the published literature to have an impact on use or costs of care). Separating out what are health needs effects and supply effects is not straightforward, since the effects of each on costs may overlap.

Figure 1 illustrates how the model was developed to explain NHS hospital costs incurred in 2007/08 for each individual registered with a GP at the beginning of that year (on 1 April 2007), by using information about each individual taken from the previous two years (2005/06 and 2006/07). The prediction variable is the cost for each individual, the explanatory variables are the needs and supply variables that influence the cost. The resulting model was used to inform allocations for 2010/11. The formula has since been updated and developed further for allocations for 2012/13, using more recent data.

To develop the model, data on each individual registered with a general practice in England as of 1 April 2007 were identified from GP registration data. If an individual had incurred an inpatient admission, outpatient or A&E attendance in a hospital in the previous two years, this information was obtained, mainly from Hospital Episode Statistics (known as HES data) linked using a unique, but anonymous, patient identifier. The information was used to calculate the hospital costs incurred to the NHS for each individual, and also contained information on individual-level health needs – diagnoses recorded during inpatient admission (ICD-10 codes). The GP registration data showed the area (at Census Lower Super Output level 1) in which the patient resided, and the practice of registration. From this, area-based information available from the census (related to health needs) and the NHS (supply factors) could be attributed to the individual. Thus a longitudinal record for each unique individual registered with a general practitioner at 1 April 2007 could be constructed over three years, including explanatory and prediction variables.

---

1 Super Output Areas (SOAs) are aggregations of adjacent Census Output Areas with similar social characteristics. Lower Layer Super Output Areas (LSOAs) typically contain four to six output areas, and have a total population of around 1,500.
We were able to develop robust models and achieved good levels of prediction

Use of care by each individual was costed using the national tariff, or in a minority of cases, where a tariff was not available, national reference costs. All data were anonymised and linked in a way that preserved patient confidentiality, in line with advice from the National Information Governance Board.²

Developing the formula
Aside from age and gender, the main source of information on individual-level needs was the ICD-10 diagnoses recorded in the two previous years 2005/06 and 2006/07. In any one year, 14% of individuals had at least one inpatient admission in a year and 30% had at least one outpatient appointment. For these, more information was available relating to health need, such as the diagnoses recorded on admission.

While a number of different ICD-10 groupings were tested, their overall performance in the statistical modelling was similar. We therefore opted for simplicity and used an established scheme of 152 diagnostic groups used by the NHS Information Centre. In subsequent analyses to update the formula to set target allocations in 2012/13 more sophisticated ways of incorporating diagnostic information were tested and recommended, for example taking into account the possible impact on costs of having certain combinations of diagnoses. A few other variables denoting individual-level needs were also tested. These included whether or not the individual had had a privately funded episode of care in a NHS facility in 2005/06 or 2006/07 (which was associated with lower NHS costs for the individual). In subsequent analyses to develop the formula further for use in 2012/13, a variable denoting whether or not the person had newly registered with the practice was also tested.

In line with best international practice, we did not include variables denoting the number of times an individual had either used outpatient or inpatient care, since it is recognised that these variables are highly associated with supply factors (the available level of local health services).

The majority of individuals registered with a GP had no inpatient admission or outpatient attendance in any one year. For these individuals all that was known about them was the basic information in the member file – age, gender, area of residence and general practice registered with. For these individuals the individual level data on health need were their age and gender, and health need was attributed by linking a variety of area-level information from census and other sources (as outlined above) according to their area of residence.

Over 300 attributed (area-level) needs variables were tested in the models. These included the extent of socioeconomic deprivation in an area, or the proportion of the population from black or ethnic minority groups. Supply variables included the

² See www.nigh.nhs.uk for more information.
number of hospital beds locally available, hospital staffing levels, the specific PCT area the person resided in (a PCT ‘dummy’ variable) and the main hospital treated at (a provider dummy variable).

Testing alternative models and results

Using these datasets we were able to build statistical models that tried to predict future hospital costs (in 2007/08) at individual level, and at the level of the practice population. The models were developed on a 10% (five million persons) random sample of the total population, and the performance of the models was assessed using two different 10% samples (a second random sample of individuals, and a random sample of practices).

All data were anonymised and linked in a way that preserved patient confidentiality

With over 300 explanatory variables to choose from, many different models were tested and we went through a process to create a series of parsimonious models – that is, those with relatively few variables. The process of selecting the best variables from the 300 was mainly statistical (looking at the effects on variance explained) tempered by an evidence-based view on the variables that were plausible. The final models recommended to the Department of Health were relatively simple. These included age and gender, the diagnostic morbidity variables, seven attributed needs variables, three supply variables and the PCT the person resided in (the PCT dummy variable). After age, the most powerful predictors of costs were the diagnostic groups that described individual-level morbidity and the PCT dummy variables. The latter, being a supply variable, would not influence the calculation of the target allocations in the fair shares toolkit.

We were able to develop robust models and achieved, by international standards, good levels of prediction – the best models could explain almost 12% of the variation in costs between individuals, and 77% at practice level. In the more recent analysis to develop the PBRA formula for 2012/13, the models recommended to the Department of Health could predict cost variations of 85%. This level of prediction is at least as high as models used internationally.

In the models we also found that the individual-level needs variables, in particular those drawing on prior ICD-10 diagnoses recorded at inpatient admission, were especially powerful predictors of future costs. Once these, and PCT dummy variables, had been added to the model, other attributed needs and supply variables added little. Though many possible explanatory variables were tested, the final set was relatively small (see Figure 2).
Comparing predictability of the model at practice level

Using the formula, it was possible to calculate a predicted cost based on the needs elements of the formula for every practice in the country in 2007/08. This was compared to the ‘observed’ costs for each practice in 2007/08. The ‘observed’ costs were simply the inpatient and outpatient activity of each individual in the practice in 2007/08 multiplied by a national tariff or reference cost as outlined above. Thus ‘observed cost’ is not necessarily the actual price paid for this activity, since in some areas PCTs may have negotiated a different local price to that in the tariff, or due to elements such as the Market Forces Factor.

*It was possible to calculate a predicted cost based on the needs elements of the formula for every practice in the country*

The ratio of observed to predicted costs was then calculated for each practice in England for 2007/08. The variability in value of this ratio by practice gives some indication of the financial risk in the system in terms of the number of practices for...
which observed exceeded expected costs by a significant margin. At this stage we did not include risk pooling arrangements (as is planned by the NHS Commissioning Board). However it is only an indication, because clearly many if not most practices are not given the target allocation as the budget – the budget allocated is some way between the historic budget and the target allocation, depending upon local policy and negotiation.

The result suggested that approximately two thirds of practices (65.4%) were within 10% of those expected, but one third exceeded 10%. Even though the models were performing well, it was clear that many practices were still some way from their expected allocations. Further examination showed that many practices beyond a 10% variance between observed and expected costs covered a small population size, as shown in Figure 3. A rule of thumb was the smaller the practice, the larger the variance.

The figure also shows that some practices appear as extreme outliers – these were atypical practices, for example ones that have been subject to large changes in practice populations. Our conclusion was that much more analysis of the variance was needed, and that if practices were to hold hard budgets for commissioning in future, adequate risk pooling arrangements would need to be designed that brought the variance between observed and expected costs of the commissioning body to below 5%. In developing the model further to set target allocations for shadow commissioning groups and practices in 2012/13, much further analysis was carried out, which will be reported separately.
Applying the model results in practice

Though the model can be used to predict cost at an individual and practice level, there are challenges when using it in this way to calculate a target allocation. The first major challenge is the time lag in available data, in particular from HES, which may be a year out of date at the time of budget setting. The second major challenge is that there can also be changes to the practice-registered population over the lag period: births and deaths, people moving practices, and practices forming, merging or splitting. Tracking a person’s use of hospital care (used to quantify person-level health needs) around the NHS wherever it occurred in England was theoretically possible but impractical, given the way populations move and the time needed to set and agree budgets.

To overcome these challenges, we tested a pragmatic solution whereby we assigned ‘needs weights’, not to each unique individual in each unique practice but to people according to their age/gender group in each practice. This meant calculating specific weights for 38 individual age/gender groups in each of the 8,200 practices in England. We found that over time, the practice-specific health needs weights of these age and gender bands varied very little compared to the health needs weights for each unique individual.

We assigned ‘needs weights’ to people according to their age/gender group

This solution meant that the model was less susceptible to lags in HES data, and also the most recent registered population in the practice before ‘target allocation’ setting day could be used. The health needs weight for each age/gender group in each practice could be multiplied by the latest information on the number of people on the practice list in the age/gender group. The PCT could make in-year or end-of-year adjustments to the target allocation if the practice population changed radically.

The calculation of target allocations by practice, and how this compared with their target allocations in previous years, was beyond the remit of the research team. Traditionally the Department of Health commissions the development of an empirical formula from independent researchers who are ‘blind’ to the impact in specific geographies. The empirical formula recommended by the research team, as outlined in this briefing, was accepted by the Advisory Committee on Resource Allocation (ACRA) and has been in use since 2010/11 in setting target allocations for commissioning practices.

Issues to consider in calculating target allocations using PBRA

List inflation

If the target allocations of commissioners are to be set by summing the target allocations of their constituent practices, then they will be influenced more than anything else by the size of population registered with those practices. Within England there is the recurrent problem that the practice list may not be accurate and usually overestimates the number of people registered. This is inevitable to some extent: as people move between areas their GP registration details will lag behind. The problem arises because list inflation is not distributed evenly – it is generally much higher in
inner-city areas. While inflation of practice-registered lists is not a major problem for the modelling within this project, or the construction of age/gender needs indices for each practice, it is more a challenge for commissioners (currently PCTs) in the subsequent allocations to practices. The accuracy of GPs’ registered lists is a wider issue than the PBRA formula – it is now up to PCTs, and in future the clinical commissioning groups and the NHS Commissioning Board, to take steps to assure accuracy of list size. Recent analysis by the Audit Commission’s National Duplicate Record Initiative shows that the accuracy of registered lists is improving and in most parts of the country (Audit Commission, 2006). Challenges still remain, in particular in cities, especially London, where a robust external audit of practice lists is needed.

Perverse incentives

One of the deliberate choices made in the PBRA work was to exclude information on the number of prior inpatient admissions or outpatient attendances of the individual. These variables may indicate health needs and would be powerful predictors of future costs of the individual. However including them in the model risks reinforcing historical patterns of supply rather than current health needs (Asthana and Gibson, 2011). Concern about incentives that might reward hospital activity is also evident in similar work internationally, where these variables have generally been kept out of formulae that set per capita risk-adjusted payments. Worries about perverse incentives internationally are exacerbated where there are few policy instruments for controlling the costs of care.

It could be argued that the ICD-10-based morbidity variables used to denote individual health need in the model depend upon a person being admitted; thus hospitals might have a perverse incentive to code more to attract a higher target allocation to their local commissioners. This incentive is present but indirect, and small. In fact in the modelling we tested whether the hospital the person was treated in had any influence on costs over and above other factors, and it did not. This suggests that differences in coding are not systematic or large enough to have a significant influence on costs. This result has been confirmed by later analysis.

It could also be argued that practices that have worked hard to reduce the number of avoidable admissions and outpatient attendances in their registered population may be unfairly penalised if these resulted in lower levels of recorded morbidity (resulting from fewer hospital admissions) as this would lead to a lower allocation over the medium to longer term. The PBRA formula focuses on allocations for hospital care, not allocations for community services or primary care. It relies upon hospital-recorded diagnoses to indicate individual-level health need, because of a current absence of individual-level data from primary care or community health services. However it does include attributed variables to indicate need, for example local prevalence of some conditions (as measured through the Quality and Outcomes Framework or QOF) and socioeconomic deprivation. But as noted earlier, these variables added little to the model over and above individual-level variables.
Application to practices bordering Scotland and Wales, and new practices

For those practices bordering Scotland and Wales, special arrangements will have to be made by the Department of Health to account for practices with patients using hospitals that were not captured in our datasets. Similarly other arrangements will have to be made for new practices forming. There are also likely to be other locally important issues which will need to be resolved by the Department of Health and by PCTs.

Improving and updating the formula

While PBRA is one step forward from using area-based data to indicate health needs, in future formulae should develop further as and when better data become available. In particular formulae should incorporate information on individual-level needs recorded in primary care and community health services. 

In the application of PBRA models, the most up-to-date information available on health needs should be exploited and will place a more exacting requirement upon the Department of Health and the NHS Information Centre. As well as HES data and registration information, it is also important to collect accurate and timely data on services provided such as critical care. Recently there have been improvements in accessing more up-to-date HES data and progress will need to continue. Yet data on critical care (through the critical care minimum dataset) are patchy – the quality and coverage of these data need to be addressed as a matter of urgency.

Risk sharing and high-cost cases

The PBRA formula outlined in this briefing can predict approximately 77% of next year’s costs at practice level; subsequent work has increased this to 85%. However there will always be some variation in costs that is inherently unpredictable. This will be especially important when the random effects of a small number of high-cost cases mean that target allocations and expenditure do not match. Those practices with smaller populations are more likely to vary from the target allocation – the impact of fluctuations in the frequency of high-cost cases is greater in smaller populations. Imbalances between allocation and expenditure can broadly be linked to either chance or systemic differences in local clinical practice.

“With the NHS Commissioning Board being charged with developing appropriate risk pooling across commissioners, further analysis in this area is critical.”

At present, commissioning practices are cushioned from significant under- or overspending against a target allocation in a number of ways. First, as noted above, the budgets practices are assigned are usually not the ‘target allocation’ calculated through a PBRA formula, but are somewhere between the historic budget and the target allocation, in part determined by the ‘pace of change’ policy as outlined above. Second, commissioning practices are generally given ‘notional’ not hard budgets, and there are locally flexible rules on how to manage ‘overspends’ and ‘underspends’. Third, much high-cost (and relatively unpredictable) care is commissioned by specialised commissioning groups, which cover a larger population than a PCT.
The PBRA formula (for allocations in 2010/11 as presented above) included the costs for specialised services and other high-cost patients – these were included therefore in budget allocations for commissioning practices. Thus, the variance in observed to expected costs by practice, as shown in Figure 3, is likely to be an overestimate, especially when the proposed NHS Commissioning Board takes over responsibility for commissioning this type of care.

There are at least three ways of spreading the financial risk within the commissioning of unpredictable and high-cost cases. The first is to increase the population size covered by the commissioning entity (so that low-cost cases might balance out high-cost cases). The second is to allow commissioning bodies to spread financial risk over more than one financial year. The third is to implement a ‘stop loss’ arrangement, whereby the commissioning entity is at risk for costs up to a certain ceiling (say £20,000 per person per year), with costs above this level borne by a commissioning body covering a larger population size. Clearly with the proposed new NHS Commissioning Board being charged with developing appropriate risk pooling across commissioners, further analysis in this area is critical. Further analysis of the impact of these approaches will be published by the Nuffield Trust early in 2012.

Conclusion

The PBRA approach is a step forward in the development of resource allocation formulae to distribute resources for health care on the basis of need. The advantages are that it can be used to calculate needs-based target allocations for small populations, and for populations not bounded by a census-defined area, such as practices and clinical commissioning groups. By exploiting more information on individual-level health needs, the PBRA formula thus can be more predictive of the future costs incurred by a practice. The PBRA formula in its latest development can predict 85% of next year’s costs by the practice population.

Application of these models to set target allocations for practices requires a number of related policy considerations.

Application of these models to set target allocations for practices requires a number of related policy considerations. Of these the most critical in the shorter term is to make sure there are robust systems in place to audit the accuracy of GP registered lists, particularly for practices in cities, where population mobility is higher. The next key step will be to develop systems for the management of financial risk. At practice level the models show how for small populations, costs can be both much higher and much lower than expected. In some cases the excess costs may be due to relatively few atypical or specialist cases. Using the PBRA formula, and including specialised services, we estimate that one third of all practices in England will undershoot or exceed their target allocation by 10% or more without appropriate risk-sharing arrangements. More analysis in this area is important in order for appropriate risk pooling across commissioning entities to be developed. We will report further on this issue in 2012.
References


Acknowledgements

This document summarises the approach developed in a series of analyses over time and is based on work by a range of organisations including: Nuffield Trust, University of York, University of Manchester, New York University, Health Dialog and the London School of Hygiene and Tropical Medicine.