PCG/Ts: Does size matter?

Chris Bojke* **Hugh Gravelle*** **David Wilkin****

Abstract.

Primary Care Groups and Trusts hold unified budgets for hospital and community health services, prescribing and general practice staff infrastructure. At levels 1 to 3 they commission secondary and community health services and at level 4 can also provide community health services and can make contracts with practices for the delivery of Primary Medical Services. They devise practice level budgets and incentives and are responsible for the clinical governance of their practices.

The government suggested in its White Paper *The new NHS: Modern Dependable* that PCG/Ts would cover populations of around 100,000. The 481 PCG/Ts established in April 1999 had means of 107,586 population, 19 practices and 58 GPs, and a range of 43,618 population, 5 and 22 GPs to 332,360 population, 58 practices and 196 GPs. In response to pressure to move from PCGs at levels 1 and 2 to PCTs at level 3 and 4 many PCGs are attempting to merge. The explicit justification is one of economies of scale: larger PCG/Ts are argued to be more efficient.

In this paper we report the results of reviews of the conceptual and empirical literature on the relationship between performance and size in health care organisations. Neither type of literature provides much support for mergers as a means of improving performance. Optimal size varies with function. We argue that the key issue is not one of scale but one of designing organisational structures, contracts and incentives to allow the different PCG/T functions to be performed efficiently.

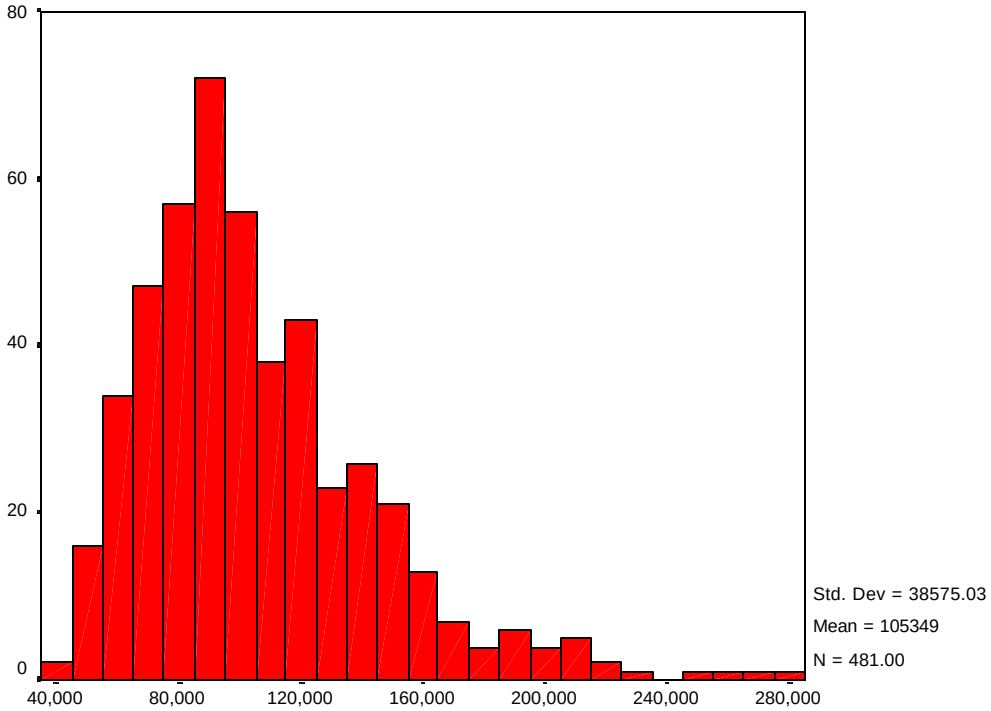
PRELIMINARY DRAFT: NOT FOR QUOTATION OR CITATION AT THIS STAGE

* National Primary Care Research and Development Centre, Centre for Health Economics, Alcuin Block D, University of York, York YO10 5DD; mad105@york.ac.uk; hg8@york.ac.uk. Funding from the Department of Health to NPCRDC is acknowledged. The views expressed are those of the authors and are not necessarily those of the DH. We are grateful for comments from Andy street, Brian Ferguson, Peter Smith and Mir Vulkan
** National Primary Care Research and Development Centre, University of Manchester.

Introduction

The 1997 NHS White Paper (Department of Health, 1997) indicated that Primary Care Groups (PCGs) would typically serve populations of about 100,000. When PCGs were established in April 1999, the average size was 105,000, but there was a wide range from less than 50,000 to more than 200,000.

Figure 1. Distribution of PCGs by population size



Even within their first year many PCGs were considering mergers with neighbouring PCGs, often in connection with applications for PCT status. Two thirds of PCGs in our National Tracker Survey of PCG/Ts (Wilkin et al, 2000) were considering mergers (28% definite and 39% possible) within six months of being established. If all of these mergers went ahead the average size of PCGs would rise to 180,000 people. Of those PCGs with a current population size of less than 75,000 less than one in ten would remain at this size.

It is important that decisions on whether PCG/Ts should merge and how big they should become are informed by an assessment of the available evidence on size of health care organisations, costs and performance. The purpose of this report is to review the available evidence, assess its relevance to determining optimal size for PCGs and summarise its implications. We examine two types of evidence: conceptual models of relationship between performance and size; and empirical findings. We identify the aims and functions of PCG/Ts and consider evidence on the relationship between size and performance for each function (economies of scale) and the implications for combining functions (economies of scope). We then examine some of the features of organisations other than size which need to be considered.

Conceptual Framework

Role of PCG/Ts

Descriptions of the role of PCG/Ts (for example in the White Paper *The new NHS* and in detailed administrative guidance such as HSC 1998) tend to run together the high level *aims* relating to the health of the populations they serve, the intermediate level *functions* which these aims imply and the managerial activities or *task*s which are required to perform the functions and achieve the aims. Separating out the levels helps to clarify the analysis of the arguments and evidence about the relationships between size and performance of PCG/T functions.

Aims

PCG/Ts have set of high level aims which are implicit or explicit in legislation and directives from the DH. A PCG/T is required to improve the health of its populations. This is not just a matter of how healthy the population is on average. PCG/Ts must also pay attention to equity: the distribution of health across the population.

PCG/Ts are not given explicit guidance as to the relative weights to place on equity and the overall health of their population. This will matter for detailed assessment of their performance, and therefore ultimately for judgements about the optimal scale and organisation of PCG/Ts. We ignore this difficult issue in what follows and take it that the relevant questions are what organisational structure and scale will maximise the health of the population subject to achieving some minimum level of equity given the resources made available to PCG/Ts. (Equivalently we could ask how equity could be maximised subject to achieving some minimum level of population health.) Although the approach will not identify the optimal structure and scale, it will permit an analysis in terms of efficiency. Structure and scale are efficient when no increase in population health can be achieved without either reducing equity or increasing resources.

Functions

To improve the health of their populations PCG/Ts must perform three principal functions.

Commissioning care.

PCG/Ts use their budget to purchase care from other organisations: community care from Community Trusts and secondary care from NHS Hospital Trusts. Level 4 PCTs will also be able to enter into Personal Medical Service agreements with providers of primary care, either practices or Community Trusts, effectively becoming commissioners of general practitioner services.

Providing care.

At level 4, PCTs will be able to hire staff and allocate resources to provide a wide range of community based services, including some specialist services provided in community settings. They will also be able to directly employ practitioners to provide primary care under PMS contracts.

Influencing care.

An important aspect of the work of PCG/Ts is to improve access to services, quality and the range of services available by influencing providers through a mixture of regulation, exhortation and budgetary incentives. Their role in developing clinical governance is an example of this. If projections (Department of Health, 2000) of the numbers of practices which are expected to change from GMS to PMS contracts are accurate, PCTs will have powerful levers to directly influence primary care services, rather than having to rely on more indirect methods through clinical governance and primary care investment plans

Tasks

In order to achieve their objectives PCG/Ts will need to undertake a wide range of management tasks or activities. The list in Table 1 is not exhaustive, but it does indicate the most important PCG/T management activities

Table 1. Conceptual Framework

PCG/T ROLES

Aims
Improving population health
Reducing inequities

Functions
Commissioning services: hospital, community and primary care.
Providing services: specialist, community and primary.
Influencing service quality, range, access via clinical governance, promotion of service development.

Managerial tasks include:
Need assessment.
Contract negotiation and bargaining.
Monitoring providers.
Co-ordination of providers.
Partnership working.
Managing risk.
Financial management.
Managing services.
Devising incentives and targets.
Regulation.
Human resource management.

+

ORGANISATIONAL FEATURES

Size: population, budgets, functions, practices, providers, partners

Governance: mechanisms for decision making and control (contractual, consensual, bureaucratic etc); structure (monolithic, hierarchical, divisional, networks, etc)

ENVIRONMENTAL FACTORS

Demographic mix
Socio-economic characteristics
Environment
Population density
Input prices
Initial set of practices and GPs

Þ

PERFORMANCE

Outcomes
Population health
Inequities in health
User satisfaction

Service delivery
Quantity
Access
Quality
Range of services
Inequity in access and use
Infrastructure stock

Costs
Service delivery cost
Transaction cost
Risk cost
Management cost

Human resources
Morale
Training
Skill development

Scope of Paper

The scope of this paper is limited to addressing the following key areas; negotiation and bargaining with providers; risk management and pooling; and empirical studies of cost and size. Given the previous discussion on the multi-task/function nature of PCG/Ts this is clearly not an exhaustive list, nevertheless it allows sufficient discussion to demonstrate the implications of size for PCG performance. A fuller analysis of size and performance, including other PCG/T functions such as clinical governance, is in the report from which this paper is drawn (Bojke, Gravelle and Wilkin, 2001).

Negotiation and bargaining with providers

It is often suggested that larger PCG/Ts will do better when negotiating with providers over prices and service provision because they have more bargaining power. But the relationship between bargaining power and size is not straightforward. The bargaining power of a purchaser vis a vis a provider depends on the alternatives it has to making a bargain with that provider and the provider's alternative purchasers. The effect of size on a PCG/T's bargaining power will depend on the configuration of providers and other PCG/Ts.

We can illustrate this with a couple of very simple models which attempt to formalise the intuition that larger purchasers do better than smaller purchasers. Note that we are not here examining the potential gains from having one set of negotiations replace two or more sets with a resulting reduction in the time and other direct costs of bargaining. Rather we focus on the outcome: the terms of the transaction.

The first model is an adaptation of the price leader model usually formulated with a large seller facing price taking buyers and a fringe of price taking competitive sellers. Suppose that there are n identical buyers of a good who have utility functions $B(x) - px$ where B is their willingness to pay or the gross benefit from the good, and p is its price. If a buyer is a price taker then his demand D(p) is determined by $B'(x) - p = 0$. k of the n identical buyers form a buyers' cartel. (Or we can interpret k be a parameter describing the size of a single buyer.) The total consumption of the buyers' cartel is kx. All buyer face a competitive supply function $S(p)$, $S'(p) > 0.$

The market price is determined by the market clearing condition

$$(n-k)D(p) + kx - S(p) = 0$$

as $p = p(k, x)$ with $p_x = -k/[(n-k)D' - S'] > 0$ (since the denominator is negative to ensure a stable market). The cartel chooses its purchase per member x to maximise each members utility, taking account of the effect on the equilibrium price. Hence the optimal purchase per cartel member $x^*(k)$ is determined by the first order condition.

$$B'(x) - p(k,x) - xp_x(k,x) = 0$$

Since $p_x > 0$ each cartel member will purchase less than a price taker would at the same price: $x^*(k) < D(p(k, x^*(k)))$. Hence increasing the size of the cartel (k) reduces the price since $p_k = [D(p) - x]/[(n-k)D' - S'] < 0$, from the market clearing condition. The effect on the utility of a typical member of the cartel of increasing the size of the cartel is

$$\frac{d[B(x^*(k) - p(k, x^*(k))x^*(k)]}{dk} = [B' - p - xp_x]\frac{dx^*}{dk} - p_k x^* = -p_k x^* > 0$$

since x^* maximises the utility of each member of the cartel. Thus increasing the size of the non price taking purchasers' cartel makes all of them better off. Producers are worse off in aggregate. Since price has fallen the remaining non-cartelised buyers are also better off. Note that a buyer is better off outside the cartel, enjoying the benefits of the price reduction, without having to incur any of the cost in terms of reducing consumption. This may raise questions about the stability of the cartel but this is not our immediate concern which was to show that if two or more buyers agree to coordinate their purchasing decisions they can drive down the price from competing sellers and raise their consumer surplus.

Price taking behaviour is probably not what one would expect from NHS Trusts who operate in a market where distance is an important determinant of demand, so that they face a demand curve which is downward sloping with respect to price. As a contrast with the previous model consider a market where there are initially two buyers. Each buyer i wants to purchase a fixed number x_i of units of a good. Thus, if they were price takers their demand curves would be vertical up to some choke price. The two buyers face a single provider with cost function $C(x_0 + x_1 + x_2)$ where x_0 is the amount the sold to other purchasers whose decisions and bargains with the seller we take as given. The total revenue from sales of x_0 is R_0.

We assume that the terms of the bargain struck between the provider and each of the two purchasers are determined as the Nash bargaining solution to the bargaining game in which each purchaser negotiates separately with the seller. Both seller and purchaser i assume that bargain with between the seller and the other purchaser j is unaffected by the bargain between i and the seller, or even by whether they succeed in reaching agreement.

The threat or disagreement point of the seller when bargaining with buyer 1 is $R_0 + R_2 - C(x_0 + x_2)$ and the threat point of buyer 1 is zero: no other seller is available to provide the services required by buyer 1. The threat points are defined analogously for the bargain between buyer 2 and the seller. The seller's gain from trade with buyer i, as opposed to failing to reach an agreement and being left trading only with buyer j and the outside buyers, is

$$R_0 + R_1 + R_2 - C(x_0 + x_1 + x_2) - R_0 - R_j - C(x_0 + x_j) = R_i - [C(x_0 + x_1 + x_2) - C(x_0 + x_j)]$$

and buyer i's gain is $B_i(x_i) - R_i$. The total gain from trade between them is

$$B_i(x_i) - [C(x_0 + x_1 + x_2) - C(x_0 + x_j)]$$

and the Nash bargain price splits the gain equally between them.

Suppose that the two purchasers merge to form a single buyer with total willingness to pay $B_1 + B_2$. The newly merged purchasers' threat point is still zero. The seller's threat point in the event of a failure to agree is $R_0 - C(x_0)$ and the seller's gain from agreeing to supply the single purchaser in return for R is $R + R_0 - C(x_0 + x_1 + x_2) - [R_0 - C(x_0)] = R - [C(x_0 + x_1 + x_2) - C(x_0)]$.
The total gain from trade between the single buyer and the seller is

$$B_1(x_1) + B_2(x_2) - [C(x_0 + x_1 + x_2) - C(x_0)]$$

and the Nash bargain splits the difference to give them each half of the total gain.

Merging to bargain with seller makes the purchasers worse off if and only if

$$B_1(x_1) - [C(x_0 + x_1 + x_2) - C(x_0 + x_2)] + B_2(x_2) - [C(x_0 + x_1 + x_2) - C(x_0 + x_1)]$$
$$> B_1(x_1) + B_2(x_2) - [C(x_0 + x_1 + x_2) - C(x_0)$$

or

$$C(x_0 + x_2) + C(x_0 + x_1) - C(x_0) > C(x_0 + x_1 + x_2)$$

Hence whether the larger merged purchaser does better depends on the shape of the seller's cost function. In the simple case in which there are no other buyers ($x_0 = 0$) and the condition reduces to

$$C(x_1) + C(x_2) - > C(x_1 + x_2)$$

which implies that the cost function is subadditive. If there are economies of scale for the producer then the condition is satisfied and merger makes the combined purchaser worse off. Thus in this case the intuitively plausible notion of larger buyers doing better depends on the shape of the seller's cost function.

Threats by a PCG/T to take its business elsewhere also become less credible as it accounts for a larger proportion of the provider's revenue. It is unlikely to be allowed to drive a provider into bankruptcy. Even it was, it may find it difficult to find a replacement supplier with sufficient capacity. Further, a provider faced with fewer PCG/Ts may make more flexible but possibly less productive investment decisions to offer fewer hostages when bargaining once the investments have been made.

As a direct counter to the argument that larger purchasers have more bargaining power, it has been suggested that smaller are more flexible and better able to take advantage of spare capacity in providers. The evidence from fundholding suggests that providers can be influenced to provider better terms for purchasers who are very small in relation to HA populations (Goodwin, 1998).

Given the differential incidence of different conditions in the population, the required total PCG/T population to effectively commission for a service varies considerably (Killoran et al, 1998; Mays and Dixon, 1996) from around 50,000 or less (community nursing) to over 1,000,000 (organ tranplants). If the commissioning unit is the PCG/T such disparities imply difficult tradeoffs in choosing the optimal size of PCG/T. But the commissioning unit can be smaller than a PCG/T if budgets are delegate to locality groups of practices within the PCG/T. Or it can be larger than the PCG/T if several PCG/Ts form a commissioning cooperative to purchase care for treatment of rarer conditions. Economies of scale if they exist, may have implications for the organisation of commissioning within and amongst PCG/Ts, rather than for the size of individual PCG/Ts.

Risk management and risk pooling

An individual's health state is a random variable. Any organisation, such as a PCG/T, with a fixed budget to cover the costs of health care for a population faces uncertain expenditure. The introduction of fundholding in the NHS stimulated interest in the question of the relationship between the size of population covered and the risks of over or under spending by budget holders.

Crump et al (1991) used experience with admissions for fundholding procedures in the West Midlands RHA in 1989/90 to simulate the variability in the total cost of fundholding procedures. They found that there was a probability of 0.25 that for a pool of 9000 patients expenditure would be more than 5% above expected. With a pool of 24000 the probability fell to 0.05.

Bachmann and Bevan (1996) simulated the expenditure variations associated with 15 types of rare but costly referrals. Variation (defined as the difference between the 95th and 5th centile of the expenditure distribution divided by the its mean) fell by about 45% when the risk pool increased from 10000 patient years to 30000 and by a similar proportion when the risk pool increased from 30000 to 100000. They concluded that such rare costly referrals were unlikely to pose a threat to the solvency of risk pools of the size of a typical total purchasing pilot site (30000).

Martin, Rice and Smith (1998) considered the risks associated with expenditure on all types of HCHS admissions. Drawing on their work on resource allocation they were able to estimate the variability of annual HCHS expenditure for wards of 9648 individuals. They then calculated the 95% proportionate confidence intervals around mean expenditure and showed that with a population of about 10,000, the confidence interval is around 20% of the mean. With a population of 100,000 the confidence interval is 6.5%. Alternatively, the probability of a 10% deviation from the mean is 1in 3 for a population of 10,000 but only 1 in 400 for a population of 100,000

These and similar studies are informative in showing that there is sizeable per capita risk at populations around the size of a typical practice (6000) but that risk is greatly reduced when the population in the risk pool reaches the size of typical PCG/T (100,000).

However, although the literature provides some indication of the effect of population size on the *magnitude* of risk, it does not address the issue of how the per capita *cost* of risk varies with population. The costs of risk depend on how the PCG/T reacts to when demand for treatment would imply an overspend or an underspend, as well as how likely such levels of demand are (Smith, 1999). Demand in excess of expected could be met reducing expenditure on other activities or the PCG/T could fail to meet such demand by not providing care, thus preserving expenditure on other activities. If fluctuations in demand arise solely for emergency cases and such cases are always treated, then fluctuations in emergency demand could be met by fluctuations in the number of elective cases treated in a period. The cost of excess demand would then be a lengthening of waiting times for elective care. Unexpectedly low demand for emergency care would benefit elective cases by reducing their waiting times.

A full treatment of the cost of risk would require a realistic specification of how PCG/Ts deal with demand fluctuations and the effects of their reactions on patient health. We do not attempt that here but instead present one simplified account which has the merit of generating strong but not obviously silly conclusions.

Assume that there are n PCG/Ts each with a population of s ex ante identical individuals. Each individual has probability p of being ill enough to require hospital treatment and all ill patients are treated at a cost of c per patient. Illness is independently distributed. Each PCG/T gets a total budget of $B = sy^o + spc$ which is sufficient to meet the expected demand for treatment sp and to provide per capita expenditure on other types of care of y^o. The realised random demand for hospital treatment is $D(s) = \sum_{i=1}^{s} d_i$ where d_i equals 1 if individual i is hospitalised and zero if not. Realised expenditure on hospital care is cD and realised per capita expenditure on other care is $y(s) = y^o + c\left(p - \frac{D(s)}{s}\right)$. Expenditure on other care is shared equally across all patients. A patient's utility depends on whether they are hospitalised or not and on the per capita level of other health care expenditure. Assume that utility is additively separable in other health care expenditure and let $u(y(s))$ be utility from other health care. Then the expected utility of a patient depends on the size of the PCG/T only via the random term $y(s)$.

Assuming that u is strictly concave so that patients are averse to risk in y. One measure of the cost of risk to a patient is the amount of y she would give up to have a certain level of expenditure on other health care. The cost of risk $r(s)$ for the individual patient is the difference between the mean level of expenditure on other health care y^o and the certain level of expenditure which yields the same expected utility as the uncertain prospect of $y(s)$. It can be shown (Pratt, 1964) that $r \approx A(y^o)s^2(s)$, where $A(y^o) = -u''(y^o)/u'(y^o)$ is the coefficient of absolute risk aversion and $s^2(s)$ is the variance of per capita expenditure on other health care $y(s)$.

The cost of risk to each patient is proportional to the variance of expenditure on other care. But

$$s^2(s) = Var\left(c\frac{D(s)}{s}\right) = \frac{c^2}{s^2}\sum_{i=1}^{s} Var(d_i) = c^2\frac{V}{s}$$

where V is the variance of d_i. Hence the per capita cost of risk is inversely proportional to the population of the PCG/T. With n identically sized PCG/Ts containing s patients each, the total cost of risk over the $N = ns$ patients is

$$nsr(s) \approx nsAs^2(s) = Ac^2nV = Ac^2N\frac{V}{s}$$

Thus there are economies of scale in pooling risks since the total and the per capita cost of risk are inversely proportional to the size of PCG/T.

We could extent the formula to allow for exposure to different conditions with different degrees of risk and cost but the same conclusions would hold. What is important for the cost of risk is the cost of the a typical episode of care and the variance of individual experience. Contrary to intuition, conditions which are rare events are not necessarily a sizeable source of risk. In this simple model the variance is $V = p(1-p)$ so that with rare events V is very small. Rare events are only a siz eable source of risk if they are also very costly to treat.

The conclusion that there are economies of scale from pooling risks is robust to other more plausible but more complicated specifications of PCG/T responses to fluctuations in demand for care. Any specification in which individuals gain but at a decreasing rate as the realised demand for care is smaller should yield the same kind of expression for the cost of risk as depending on the variance of the per capita residual expenditure. Note however that other methods of managing risk will affect the cost

of risk related gains from larger PCG/Ts. At the extreme, if the PCG/T had access to full insurance against demand fluctuations there would be no economies of scale for PCG/Ts in risk pooling. The need for a large population to pool risks is also reduced by allowing PCG/Ts to carry forward surpluses or defecits.

Empirical studies of cost and size

The sources of economies and diseconomies of scale are set out in the industrial economics literature (for example: Pratten, 1971; Milgrom and Roberts, 1992). For much of these (e.g. indivisibilities, specialisation, managerial diseconomies of scale) the theory is fairly straight forward and the emphasis is on empirical estimation. Literature relevant to the measurement of economies of scale in PCG/Ts was collected by a formal search of a number of bibliographic databases. The details of the databases used and the search strategies are given in Appendix A. In addition, a description of the study and a request for additional references was sent to an electronic mailing list for health economists and to academics thought to be active in the area. The authors also drew on their own knowledge of the health economics, industrial economics and health service literatures

The literature review revealed two main bodies of literature potentially relevant for the issue of cost and size in PCG/Ts: the TPP evaluation in the UK and the HMO managed care literature in the US. Much of the UK literature focuses directly on management and transaction costs, whilst the US literature tends to look at total/average costs in producing a unit with no special distinction being accorded to management costs. Appendix B contains a short account of the methods of estimating economies of scale and appendix C contains a summary of relevant UK and US papers. For the main body of the paper we concentrate on a limited subset of the most relevant papers and also discuss the limitations in using the TPP and US literature.

UK evidence

Although studies using data on UK health care organisations have fewer problems of context in using the experience of other organisations to predict the implications of size for performance of PCG/Ts, three difficulties are apparent:

- Unlike the US, where a much greater reliance on the market has generated a large body of data on activities and costs, UK data is poor in scope and quality. Consequently, there are few quantitative studies using UK data to estimate cost or production functions or to undertake DEA. Most studies tend to be small scale, comparing relatively few organisations, so that using multivariate techniques to allow for confounding is generally not possible.
- Because PCG/T represent the most fundamental organisational change in the NHS since its foundation, considerable care needs to be exercised in extrapolating from the experience of previous organisations to PCG/Ts. HAs commissioned HCHS services but did not have unified budgets and could not, until 1997, employ GPs or directly commission services from them. Fundholding practices were able to purchase only a subset of HCHS. TPPs had no clinical governance functions.
- HAs are much larger than PCG/Ts and fundholders much smaller so that even when there is evidence on the effect of size on performance it will not cover the size range of most PCG/Ts.

The main body of evidence relevant for PCG/Ts appears to be the experience of total purchasing pilots (TPPs). TPPs consisted mainly of a number of fundholding practices which were banded together and given a budget with which to purchase most types of hospital and community health care. Fifty three first wave TPPs were established in England and Scotland in April 1995, with a further 35 second wave TPPs in April 1996. Both waves had a preparatory year before going 'live'. The second wave of TPPs differed markedly from the first wave in that it included non-fundholders, had smaller median TPP populations (28,500 vs. 18,000), more two-practice TPPs, fewer 3, 4 or 5 practice TPPs and included two very large TPPs of 15 practices and 45 practices.

Table 2 summarises the differences between TPPs and PCG/Ts

TPPs	PCG/Ts
Average 30,000 population	Average 100,000 population
GP led	GP and nurse led with lay representation
Volunteer practices and time-limited (three years)	Compulsory for all practices and not time-limited
Rural and suburban	All types of area.
Many simple/informal organisations	Complex formal organisations
Ring-fenced TP and SFH budgets. GMS not included	Integrated budgets, including HCHS, prescribing and GMS-CL but not GMS
Selective purchaser of services able to block back to HA responsibility for some services	Total purchasing/provision of services. Lower level PCGs can block back.
Intended to be a purchasing organisation (although some practices developed provider roles)	Responsible for commissioning services plus health improvement and primary care development
No structure for clinical governance within TPPs	Arrangements for clinical governance aimed at improving quality and consistency of primary care delivery
No requirements to address public health or inequality issues	Required to address public health and inequalities
Average management cost allowance of £8 - £9 per capita per year	Average management cost allowance of £3 per capita per year

Table 2. Differences between TPPs and PCG/Ts.
(Source: adapted from Killoran et al, 1999)

The papers generated by the national evaluation of TPPs (Mays, et al, 1998) provide a great deal of detailed information about TPPs. However, the relatively small number of TPPs, the considerable disparity in the activities they undertook, and the differences in their environments means that the types of multivariate analysis common in the US literature was generally not possible, though some attempts were made to statistically examine the determinants of transaction costs (Posnett et al, 1998). The fact that TPPs consisted of self selected practices, reinforces the need for considerable caution in drawing conclusions for size and performance in PCG/Ts.

TPP evidence on management and transaction costs

Posnett et al (1998) examined TPP direct management costs, defined as those explicitly identified with the operation of the TPP, but excluding other staff time in HAs devoted to TPPs. They found no statistically significant relationship between TPP management costs per capita and size. Nor did they find any difference between costs in the preparatory year and the first 'live' year (first wave TPPs only). Second wave TPPs had lower preparatory year costs than first wave TPPs, in part due to the different characteristics of the TPPs in the two waves. There were wide ranges in per capita management costs (£0.02 to £7.49 for first wave TPPs and £0.48- £7.53 for second wave TPPs).

Table 3. TPP mean direct management costs per capita (1996/97 prices)

	Prep. year	Live year	N for prep. year (live if different)	Prep year	N for prep. Year *
Single Practice	£2.72	£2.70	18	£3.03	12
Two practices	£2.41	£2.96	3 (3)	£2.46	7
Three practices	£3.36	£3.19	9 (8)	£2.00	3
Four or more practices	£2.75	£2.73	20	£1.43	7
All TPPs	£2.83	£2.82	50 (49)	£2.40	29

*excludes Wakefield TPP (45 practices)
Source: Derived from Posnett et al (1998)

In a more detailed small-scale case study of seven TPPs, Place et al (1998) estimate transaction costs. They identify four types of transaction cost: information costs (acquiring information needed to transact); negotiation and contracting; monitoring and enforcement; co-ordination and organisation. Place et al. (1998) suggest that co-ordination and organisation costs arise from the need to get practices to agree on the appropriate package of services to commission. PCG/Ts with larger practices may therefore have smaller co-ordination costs for a given population. But whether increases in the number of practices increase co-ordination costs more than proportionately is unclear *a priori*.

Place et al (1998) attempted to measure the actual costs incurred in managing TPPs by interviews with TPP participants. Their transaction costs are therefore probably a more accurate measure than the management costs estimated in the less extensive study by Posnett et al (1998). Place et al (1998) find that transaction costs per capita are greatest for single practice groups, but that there is no relationship between size and costs for TPPs for groups with more than one practice. They conclude that any economies of scale in management are eroded by the coordination problems of larger commissioning groups. Street and Place (1998) also note that coordination problems were seen as the main barrier to effective working in PCG/Ts. Place et al (1998) are careful to point to the problems in this kind of exercise. Apart the obvious difference in functions between TPPs and PCG/Ts (see section 3.x), the study by Place et al (1998) compares the *additional* transaction costs in TPPs compared to those they incurred as general practice fundholders. Even if TPP and PCG/Ts transaction costs were similar, the difference between TPP and fundholding transaction costs does not provide any information about TPP costs.

Killoran et al (1998) estimated the appropriate level of management costs for PCGs with an average population of 100,000 patients may be in the region of £17 - £18 per capita. Place and Newbronner (1999) in a detailed study of the likely costs on managing East Yorkshire PCG (population approx. 146,000), estimate the likely management resources to be around £9.25 for a level 3 PCT and £11.20 for a level 4 PCT. The different definitions of management costs in these two studies do not permit sensible inference concerning the effect of size on per capital management costs. Note, however, that both estimates are well above the DoH allowance of £3 per head.

An earlier study of administrative costs in FHSAs (Giuffrida et al, 2000) found that there were some economies of scale at the average size of FHSA (around 500,000). There are considerable differences in the functions of PCG/Ts and FHSAs: FHSAs did not commission or provide care: their main tasks were administering payments to primary care practitioners. This is now carried out by Health

Authorities. Thus the finding that there are economies of scale in administration may be of more relevance to health Authority size than PCG/T size.

US managed care

Due to the greater availability of data over a number of years, there is a large empirical literature of the effects of scale in US health care organisations, much of it using multivariate statistical analysis. Lewis, Malbon and Gillam (2000) argue that US managed care organisations such as Health Maintenance Organisations share a number of key characteristics with PCG/Ts and that these shared characteristics provide scope for reasonable comparisons. Like PCG/Ts, HMO's have a single unified budget to provide universal health care for a defined population. As a consequence both organisations face the same incentives when managing risk, when deciding to commission or directly provide services, to allocate resources to health promotion versus treatment, and to introduce clinical governance to influence clinical activity via evidence-based guidelines, audit and utilisation reviews.

Robinson and Steiner (1998) in their review of US managed care organisations are more circumspect, pointing out that drawing lessons from abroad is a potentially hazardous activity. Not only may historical and cultural factors partially determine findings, it is difficult to determine the extent to which they do so. Drawing on their work we can identify a number of problems in using the findings from studies of managed care to draw conclusions about the existence and extent of economies of scale.

HMOs and PCG/Ts differ in a number of important respects:

- The basis for enrollment of populations leads to differences in the heterogeneity of populations and in their average characteristics. HMO enrollees choose their HMO on the basis of its premiums and benefit coverage. PCG/Ts populations do not pay for cover and will typically have no little choice of which PCG/T to join via their chosen practice.
- Objectives. Some HMOs are profit orientated.
- Competition. Patients in the US have a greater choice of health care organisation.
- greater variety of managed care organisations (see Robinson and Steiner, 1998). Consequently, differences in other organisational features may confound the effects of size.
- Input prices. Differences in the relative costs of different types of health care staff, pharmaceuticals, and plant can lead to differences in the relationship between size and costs.

It is in principle possible to allow for differences in these factors which may confound the effect of size. For example, a number of the papers in our review (Given,1996; McCue et al, 1999; Rosenman et al, 1997; Wholey et al, 1996, 1998), attempt to estimate the effects of competition and for-profit status.

A further difficulty is that many of the studies which include size as a determinant of performance were not primarily interested in the existence of economies of scale per se. Hence measures of size and estimation techniques may not be the most appropriate.

Our judgement is that the US studies are of some use in suggesting whether PCG/T size can affect performance but are not helpful in quantifying the extent of economies of scale in the UK context.

US HMO Studies

Due to the large expansion of HMOs and decline of the traditional FFS models in the US, a number of recent articles (including Given (1996), Wholey (1996) and Draper et al (2000)) concentrate not on measuring the differences in efficiency between HMOs and FFS, but on identifying the characteristics of HMOs which affect efficiency. Throughout this recent empirical literature a fairly standard approach has been adopted in that common key characteristics of interest are defined and identified (including HMO size) and their effects on efficiency measured, but with variation in the HMOs used, over different time periods and analysed using different econometric techniques.

Given the similarity in approaches, it is not surprising that the papers share common strengths and weaknesses. For example, no study includes a direct measure of the quality of care. Given the HMOs'

incentive to minimise cost, this is an important consideration – we recognise that HMOs can 'improve' efficiency by reducing the quality of their product, but have little technique or measurement to counter this. Other weaknesses are as a result of the particular econometric technique used. For example with Translog cost functions, it is assumed that output is exogenous and that factor input prices are unaffected by the actions of the HMOs. This assumption appears undesirable in light of the previous discussion on the possible increased bargaining power of large organsiations. However the consistency in approaches allow reasonable comparisons to be made between the studies.

In all the studies, HMOs provide a joint product; health insurance coverage and health care services (inpatient care, outpatient care and administrative services.) As a result, most studies define the unit of output as being a member month of coverage. HMO size is generally defined as the number of enrolees.

Economies of scope are also examined in some studies and are consistently defined as the economies gained from providing services for a number of enrolee markets, rather than providing different services. In the US there are two major enrolee market categories; private and public (e.g. Medicare and Medicaid.) This slightly unusual definition of scope has again been influenced by the direction in which the US HMO has taken since the early 1980s. Prior to 1980 HMOs served commercial enrolees almost exclusively, whereas by the mid-1990s the Group Health Association of America found that a sizeable 29% of all HMOs were serving both private and public markets with many more planning to do so in the near future.

Other factors which have been of particular interest to US researchers are HMO characteristics (for-profit status, federal qualification status, chain affiliation, model type and age) and environment characteristics (competition and population density) but not enrolee/population characteristics like morbidity which may equally affect the cost of providing cover.

Given (1996) studies a sample of 31Californain HMOs over 1986 – 1992 and uses a translog cost function. She finds that the average sized HMO (approx. 145,000 enrolees) operates in a region of constant returns to scale and that any economies of scale are exhausted by about 115,000 enrolees. She also finds that there exist diseconomies of scope, that is, it would be more efficient to provide for public and private enrolees in separate HMOs. Thus, she finds little evidence to support the current trend of HMO mergers and service provision across different markets. The level of competition was the only significant control variable. It was hypothesised that increasing levels of competition would have a negative influence on costs, but the regression result has a significant positive coefficient. Given attributes this finding to the possible endogeneity of location, with HMOs choosing to locate themselves in high cost areas.

Wholey et al (1996) apply a similar methodology to a larger sample of 599 national HMOs over 1988 to 1991. Their findings broadly agree with Given: diseconomies of scope in providing for both public and private enrolees and economies of scale which are exhausted relatively quickly, at around 50,000 enrolees. Unlike Given, Wholey at al find several significant control variables with increasing competition, for-profit status and an increasing population density all reducing costs in Group HMOs (but not IPAs)

Draper et al (2000) use a variable return to scale DEA on a national sample of 249 HMOs in 1995. Efficiency scores, ranging from 1 – perfectly efficient to 0 – perfectly inefficient, are calculated for all HMOs and these scores are tested against HMO characteristics. The efficiency scores demonstrate a wide range with 29 HMOs deemed efficient (at least relative to the sample of HMOs) and a sample mean efficiency score of just 0.427.

Using various different groupings of size, Draper et al find that in all cases a U-shaped relationship between size and efficiency exists. HMOs with mid-levels of enrolment (40,000 – 60,000 or 15,710 – 37,778 depending on the grouping of size used) are less efficient than small or very large HMOs, with the small HMOs being the most efficient. By using only the upper 50% of HMOs, they find a linear relationship between size and efficiency with increasing size implying increasing efficiency. These results do not agree with the Wholey and Given studies.

Draper also finds that HMOs which are not federally qualified are more efficient than those that are. HMOs that participate in Medicare-funded programmes are less efficient than those that don't, but

those HMOs that participate in both Medicare and Medicaid programmes are the most efficient (indicating economies of scope) and chain-affiliated HMOs were found to be less efficient. For-profit status and HMO model type had no significant effect. In short, the Draper paper produces results that appear inconsistent with other studies.

Clement(1995) uses survival analysis to measure optimal size. The survival analysis approach looks at changes in the numbers of HMOs in group sizes over time, 1977 – 1986 in this case. Those group sizes that grow in membership over time are deemed to be the most efficient. This simple approach does not allow for the measurement of the effects of other covariates. Clement finds that a minimum enrolment of 25,000 members should be an immediate short-term objective for HMOs and an enrolment of 40-60,000 is required for long-term survival.

The Wholey, Given and Clement papers broadly agree that economies of scale are exhausted at a relatively low level, this despite using different data, and in the case of Clement, a different analytical technique. Draper finds the opposite result, that mid-range sized HMOs are the most inefficient and that there are increasing returns to scale over the larger sizes of HMOs. Draper uses DEA, an operations research method of scoring relative efficiency. The DEA technique itself is not without its criticisms and having obtained the efficiency scores, Draper then estimates the effects of HMO characteristics in isolation, using a simple one-way ANOVA.

A further problem with the Draper paper is that 54% of the original HMOs were omitted from the final sample due to item non-response, although Draper does argue that the sample is broadly representative of the US HMO population. However, the definitions of size categories look slightly skewed, with more smaller HMOs than expected.

UK and US literature

Our conclusions about the usefulness of the US and UK literature for estimating economies of scale in PCG/Ts are:

- the better quality of US data and the greater reliance on multivariate methods is offset by the institutional differences between US managed care and UK PCG/Ts
- US studies can identify the existence of economies of scale in health care organisations but not their importance in the UK context.
- there are major differences between the functions of PCG/Ts and earlier NHS organisations
- UK studies can provide some limited evidence on the costs of particular functions now undertaken by PCG/Ts but not on the relationship between overall performance and size when the functions are combined within PCG/Ts.

Discussion and Conclusions

The government's decision in 1997 that PCGs should typically serve populations of around 100,000 did not appear to be based on a systematic review of evidence about the optimum size of health care organisations. However, it probably did reflect a mixture of experience (eg in relation to various models of commissioning) and consultation with managers and health professions involved in primary care. Whatever the reasons for this decision, when PCGs came to be established in April 1999, they did indeed conform broadly with the policy, having average populations of about 100,000. However, it has rapidly become apparent that many Health Authorities and PCGs themselves consider 100,000 population to be an absolute minimum size, particularly for transition to Trust status. Two thirds were considering mergers within the space of their first six months, seemingly reflecting a widely held view that optimum size was probably closer to 200,000 than 100,000. PCGs seem to have been caught up in the same 'merger mania' which has been a phenomenon in other parts of the health care systems of both the UK and USA. In our review of the evidence concerning optimum size of primary care organisations we have set our interpretation and conclusions in the context of these trends. We have asked the question 'Is bigger better?' because this appears to be the assumption underlying the current trend towards larger PCGs and PCTs. Whatever the evidence for or against the current size of these organisations, we argue that a significant change from the status quo should be based upon evidence that such a change will yield benefits for patients and for the NHS. We have not focused on whether a move to smaller organisations would be beneficial because this is not an option. We know of no PCG/T which is actively considering dividing itself into two or more organisations.

In considering the evidence relevant to optimal size of PCG/Ts we argue that size is but one organisational characteristic and that optimal size will vary for different functions, with different organisational and governance structures and in different environments. The apparently simple question posed in our title becomes much more complex and capable of many different answers when broader contextual framework. Despite the importance of the issue of organisational size, the evidence available from published research is limited both in quantity and relevance to the problem of determining optimal size for PCG/Ts. We have relied mainly on evidence from the UK, principally relating to commissioning organisations, and the USA where managed care organisations offer parallels with PCG/Ts. As is the case with most organisational change prompted by changes in policy, there is no direct evidence relating to the newly implemented changes precisely because they are different from previous models. In these circumstances we must rely on evidence derived from other organisations where there are sufficient parallels to warrant comparisons and on theoretical models supported by empirical evidence.

Despite the limitations of the evidence currently available, we believe that four general conclusions can be drawn from our review which are pertinent to the decisions facing the NHS in relation to PCG/T mergers:

- There is no evidence that increases in the size of PCG/Ts beyond 100,000 will generate significant improvements in overall performance.
- Optimal size will vary substantially for different functions of PCG/Ts.
- Organisational structures and organisational alliances can be used to achieve differing levels of aggregation for different functions.
- The relevant question is the optimal *organisation* of PCG/Ts not just their *size*.

The issues facing PCGs, Health Authorities and NHS Executive concerning mergers are immediate and pressing. Many mergers have already been approved and many more are at an advanced stage of negotiation. It is probably too late to reconsider mergers that have already been agreed. Indeed many PCGs are already well advanced in making preparations for the completion of mergers in April 2001. However, many more potential mergers are still at the discussion stage and have not yet been finalised. We think the evidence presented here should cause managers and health professionals to consider very carefully whether mergers will deliver the anticipated benefits and at what costs. All PCG/Ts, regardless of population size, should look very carefully at ways of creating organisational structures and alliances which are capable of delivering the benefits of larger size where these exist and smaller size where this is beneficial to particular functions. For PCG/Ts which remain relatively small (below 100,000) this may mean working in alliance with neighbouring PCG/Ts for certain aspects of commissioning and strategic planning. For larger PCG/Ts it may mean creating devolved budgets and responsibilities for locality groups in areas such as primary care development and commissioning.

Anecdotal evidence suggests that many PCGs have felt under considerable pressure from their Health Authorities, and in some cases even instruction, to merge with their neighbours. While this pressure is often justified in terms of the perceived benefits to the PCG/T of a larger population, it also seems to reflect a lack of financial and managerial resources at Health Authority and PCG/T level. Health Authorities have found themselves severely stretched in providing support and performance managing the numbers of PCGs established in April 1999. Smaller PCGs have found themselves severely constrained by the allocation of a management budget on a per capita basis (usually around £3 per head of population) and have thus seen mergers as the only way of increasing their managerial capacity. Both Health Authorities and PCG/Ts have found themselves drawing on the same limited pool of managers to staff a larger number of organisations. All of these factors have fuelled the trend towards mergers, but they have little to do with the issue of whether or not mergers will deliver improved services at lower cost. We believe that the NHS Executive should look carefully at the constraints placed on PCG/T management budgets and try to ensure that these do not become the major factor in decisions to merge.

Where PCG/T mergers have already taken place or will do so over the coming year it will be important to take account of existing knowledge of the factors which influence successful outcome. In particular, this indicates that successful mergers require substantial investment in managerial capacity, careful attention to professional morale and sufficient time to realise potential benefits.

References

Astley WG (1985) Organisational size and bureaucratic structure, Organisation Studies, 6, 201-228

Bachmann, M. O. and Bevan, G. (1996). "Determining the size of total purchasing site to manage the financial risks of cost referrals: computer simulation model", *British Medical Journal*, 313, 1054-1057.

Bailey, S. J. (1999). *Local Government Economics: Principles and Practice*, Macmillan, London.

Baker, L.C. & Brown, M.L. (1999) Managed care, consolidation among health care providers, and health care: evidence from mammography. *RAND Journal of Economics,* **30,** 351-374.

Barr DA (1995) The effects of organisational structure on primary care outcomes under managed care, Annals of Internal Medicine, 122, 5, 353-359

Baumol, W. J., Panzar, J. C. and Willig, R. D. (1982). *Contestable Markets and the Theory of Industrial Structure*. Harcourt Brace Jovanovich, New York.

Baxter, K., Bachmann, M. & Bevan, G. (2000) Primary care groups: Trade-offs in managing budgets and risk. *Public Money & Management,* **20,** 53-62.

Bazzoli DL and Burns LR (1998) Assessing the extent of integration achieved through physician-hospital arrangements, Journal of Healthcare Management, 43, 3, 242-261

Bazzoli, G. J., Shortell, S.M., Dubbs, N., Chan, C. and Kralovec, P. (1999). "A taxonomy of health networks and systems: Bringing order out of chaos", *Health Services Research*, 33, 6, 1683 – 1717

Bevan, G. (1998) Resource allocation within health authorities : lessons from total purchasing pilots In *National Evaluation of Total Purchasing Pilot Projects report*.

Bevan, G., Baxter, K. & Bachmann, M. (1998) Survey of budgetary and risk management of Total Purchasing Pilot Projects, 1996-97.

Blau PM (1970) A formal theory of differentiation in organisations, American Sociological Review, 35, 201-218

Boyne, G. A. (1995). "Population size and economies of scale in local government", *Policy and Politics*, 23, 213-222.

Carr-Hill, R., Rice, N. & Smith, P.-C. (1998) Towards locally based resource allocation in the NHS.

Clement, D.G. (1995) HMO survival: determination of optimal size. *Health Serv Manage Res,* **8,** 10-22.

Crump, B. J., Cubbon, J. E., Drummond, M. F., Hawkes, R. A. and Marchment, M. D. (1991). "Fundholding in general practice and financial risk", *British Medical Journal*, 302, 1582-1584.

Dawson, D. & Street, A. (1998) Reference costs and the pursuit of efficiency in the 'new' NHS.

Debrock, L. & Arnould, R.J. (1992) Utilization control in HMOs. *Quarterly Review of Economics and Finance,* **32,** 31-53.

Department of Health (1997). *The new NHS: Modern, Dependable*

Department of Health. (2000). *The NHS Plan*, Cm4818-I, HMSO, London.

Department of Health (1998). HSC 1998/xxx Annex with functions.

Department of Health (1998) Developing Primary Care Groups, HSC 1998/139:17

Draper, D.A., Solti, I. & Ozcan, Y. A. (2000) Characteristics of health maintenance organizations and their influence on efficiency. *Health Serv Manage Res,* **13,** 40-56.

Earwicker, S. (1998) GP Commissioning Groups - The Nottingham Experience. OHE.

Encinosa, W.E., III & Sappington, D.E.M. (1997) Competition among health maintenance organizations. *Journal of Economics and Management Strategy,* **6,** 129-150.

Faulhaber, G. R. (1975). "Cross-subsidization: pricing in public enterprises", *American Economic Review*, 65, 966-977.

Feldman, R., Wholey, D. & Christianson, J. (1996) Economic and organizational determinants of HMO mergers and failures. *Inquiry,* **33,** 118-132.

Ferguson, B., Posnett, J. and Sheldon, T. (eds.). (1997). *Concentration and Choice in Healthcare*, Royal Society of Medicine Press, Edinburgh.

Garside P (1999) Evidence based mergers? Brit. Med. Journal, 318, 345-346

Garside P and Rice J (1994) Merger mania, Health Service Journal, 21, July, 22-24

Gask, L., Lee, J., Donnan, S. & Roland, M. (1998) Total purchasing and extended fundholding of mental health services.

Giuffrida, A., Gravelle, H., and Roland, M. "Measuring quality of care with routine data: performance indicators should not be confused with health outcomes". *British Medical Journal*, 319, 10 July 1999, 94-98.

Giuffrida, A. and Gravelle, H. "Measuring performance in primary care: econometric analysis and DEA", *Applied Economics*, 2000, to appear. [See http:// www.users. york.ac.uk /~econ32 /9936.pdf]

Giuffrida, A., Gravelle, H. and Sutton, M. "Efficiency and administrative costs in primary care", *Journal of Health Economics*, 19, 6, 2000; 983-1006.

Given, R.S. (1996) Economies of scale and scope as an explanation of merger and output diversification activities in the Health Maintenance Organization industry.
Journal of Health Economics, **15,** 685-713.

Goddard M and Ferguson B (1997) Mergers in the NHS: Made in heaven or marriages of convenience, London: Nuffield

Goddard, M., Mannion, R., Ferguson, B. (1997) Contracting in the UK NHS: Purpose, Process and Policy. Centre for Health Economics, York.

Goldsmith JC (1994) The illusive logic of integration, Healthcare Forum, 37, 5, 26-31

Goodwin, N. (1998). "GP fundholding", in Le Grand, J., Mays, N. and Mulligan, J. (eds.) , *Learning from the Internal Market*, Kings Fund, London

Goodwin, N., Mays, N., McLeod, H., Malbon, G. & Raftery, J. (1998) Evaluation of total purchasing pilots in England and Scotland and implications for primary care groups in England: personal interviews and analysis of routine data. The Total Purchasing National Evaluation Team. *BMJ,* **317,** 256-259.

Hargraves, J.L., Palmer, R.H., Orav, E.J. & Wright, E.A. (1996) Practice characteristics and performance of primary care practitioners. *Medical Care,* **34,** SS67-SS76.

Harris, A. (ed.) (1997). *Needs to Know: A Guide to Needs Assessment in Primary Care*. Churchill Livingstone, London.

Holmstrom, B. and Milgrom, P. (1991), "Multitask principal-agent analysis: incentive contracts, asset ownership and job design", *Journal of Law, Economics and Organization*, 7, Special Issue, 24-52.

Kaluzny AD Zuckerman HS and Rabiner DJ (1998) Interorganisational factors affecting the delivery of primary care to older Americans. Health Services Research, 33, 2, 381-401.

Kaluzny AD, Zuckerman HS and Ricketts TC eds (1995) Partners for the Dance: Forming Strategic Alliances in Health Care. Chicago: Health Administration Press.

Killoran, A., Abbott, S., Malbon, G., Mays, N., Wyke, S. & Goodwin, N. (1999) The transition from TPPs to PCGs : lessons for PCG development In *National Evaluation of Total Purchasing Pilot Projects report*. King's Fund, London.

Killoran, A., Griffiths, J., Posnett, J. & Mays, N. (1998a) What can we learn from the total purchasing pilots about the management costs of Primary Care Groups?: a briefing paper for health authorities In *National Evaluation of Total Purchasing Pilot Projects working paper.*, pp. 17. King's Fund, London.

Killoran, A., Mays, N., Griffiths, J. & Posnett, J. (1998b) Growing pains. *Health Service Journal*, **108,** 32-34.

King, D. N. (1996). "A model of optimum local authority size", in Pola, G., France, G. and Levaggi, R. (eds.), *Developments in Local Authority Finance: Theory and Practice*, Edward Elgar, Cheltenham, 55-76.

King, D. N. and Yue Ma. (2000). "Local authority size in theory and practice", *Environment and Planning C: Government and Policy*, 18, 3, 255-270

Lake, T. (1999) Do HMOs make a difference? Consumer assessment of health care. *Inquiry-the Journal of Health Care Organization Provision and Financing*, **36,** 411-418.

Leutz, W., Malone, J., Kistner, M., O'Bar, T., Ripley, J.M. & Sandhaus, M. (1990) Financial performance in the social health maintenance organization, 1985-88. *Health Care Financ Rev*, **12,** 9-18.

Light, D.W. (1998) Effective Commissioning: Lessons from purchasing in American managed care. OHE.

Long, M.J. (1985) Minimum optimum scale of federally qualified health maintenance organizations. *Am J Prev Med*, **1,** 41-45.

Mahon, A., Leese, B., Baxter, K., Goodwin, N. & Scott, J. (1998) Determining success criteria for Total Purchasing Pilot Projects.

Malbon, G., Mays, N., Killoran, A. & Goodwin, N. (1998) Profile of second wave Total Purchasing Pilots : lessons learned from the first wave.

Martin, S., Rice, N. and Smith, P. C. (1998). "Risk and the general practitioner budget holder", *Social Science and Medicine*, 47, 1547-1554.

Mays, N. & et al. (1998) Total purchasing : a profile of national pilot projects.

Mays, N., Goodwin, N., Killoran, A. & Malbon, G. (1998) Total purchasing: a step towards primary care groups : national evaluation of total purchasing pilots In *National Evaluation of Total Purchasing Pilot Projects main report*. King's Fund, London.

Mays, N. and Dixon, J. (1996). *Purchaser Plurality in UK Health Care*. Kings Fund, London

McClenahan, J. (1999) Emerging problems with merger policy. *Health care UK 1999/2000 : the King's Fund review of health policy*„ 94-97.

McCue, M.J., Clement, J.P. & Luke, R.D. (1999) Strategic hospital alliances - Do the type and market structure of strategic hospital alliances matter? *Medical Care,* **37,** 1013-1022.

McKelvey B (1975) Guidelines for the empirical classification of organisations. Administrative Science Quarterly, 20, 509-25.

Milgrom, P. and Roberts, J. (1992). *Economics , Organization and Management*, Prentice Hall, New Jersey.

Miller D and Friesen PH (1984) Organisations: A Quantum View. Englewood Cliffs NJ: Prentice Hall Inc.

Mueller DC (1972) A life cycle theory of the firm, Journal of Industrial Economics, 20, 199-219

Myles, S., Wyke, S., Popay, J., Scott, J., Campbell, A. & Girling, J. (1998) Total purchasing and community and continuing care : lessons for future policy developments in the NHS.

Newhouse RP and Mills ME (1999) Vertical systems integration, Journal of Nursing Administration, 29, 10, 2-29

Nutting PA, Burkhalter BR, Dietrick DL and Helmick EF (1982) Relationship of size and payment mechanism to system performance in 11 medical care systems, Medical Care, XX, 7, 676-690

Place, M. & Newbronner, E. (1999) Roles, Functions and Costs of Primary Care Trusts: the Expected Costs of Managing a Primary Care Trust. York Health Economics Consortium.

Place, M., Posnett, J., Street, A. (1998) An Analysis of the Transaction Costs of Total Purchasing Pilots. Final Report. King's Fund, London.

Posnett, J., Goodwin, N., Griffiths, J., Killoran, A., Malbon, G., Mays, N., Place, M. and Street, A. (1998). *The Transactions Costs of Total Purchasing*, Kings Fund, London

Pratt, J. W. (1964). "Risk aversion in the small and in the large", *Econometrica*, 32, 122-136.

Pratten, C. F. (1971). *Economies of Scale in Manufacturing Industry.* Cambridge University Press, Cambridge.

Pringle, M. (2000). "Participating in clinical governance", *British Medical Journal*¸321, 737-740.

Reschovsky, J.D. & Kemper, P. (1999) Do HMOs make a difference? Introduction. *Inquiry-the Journal of Health Care Organization Provision and Financing,* **36,** 374-377.

Rich, E.C., Kralewski, J., Feldman, R., Dowd, B. & Bernhardt, T.S. (1998) Variations in the management of primary care - Effect on cost in an HMO network. *Archives of Internal Medicine,* **158,** 2363-2371.

Robinson, J. C. (1996). "The dynamics and limits of corporate growth in health care", *Health Affairs*, 15, 2, 155-169

Robinson, J.C. & Casalino, L.P. (1996) Vertical integration and organizational networks in health care. *Health Affairs,* **15,** 7-22.

Robinson, R., Robison, J. & Raftery, J. (1998) Contracting by Total Purchasing Pilot Projects, 1996-97.

Robinson, R. & Steiner, A. (1998) *Managed Health Care,* Open University Press.

Roland, M. and Baker, R. (1999). *Clinical Governance: A Practical Guide for Primary Care Teams*. NPCRDC, Manchester

Rosenman, R., Siddharthan, K. & Ahern, M. (1997) Output efficiency of health maintenance organizations in Florida. *Health Economics*, **6,** 295-302.

Royce, R. (2000) *Primary Care and the NHS Reforms: A Manager's View*, OHE.
Salinas-Jimenez, J. & Smith, P. (1994) Data envelopment analysis applied to quality in primary health care.

Schlesinger, M., Blumenthal, D. & Schlesinger, E. (1986) Profits under pressure. The economic performance of investor-owned and nonprofit health maintenance organizations. *Med Care,* **24,** 615-627.

Shortell SM (1988) The evolution of hospital systems: unfulfilled promises and self-fulfilling prophecies, Health Services Research, 45, 2, 177-214

Smith, J. & Barnes, M. (2000) Developing primary care groups in the new NHS : towards diversity or uniformity? *Public Money and Management,* **20,** 45-52.

Smith, J., Knight, T. & Wilsin, F. (1999a) Supra troupers. *Health Service Journal,* **109,** 26-28.

Smith, J., Regen, E. & Shapiro, J. (1999b) Lending a hand. *Health Service Journal*, **109,** 24-26.

Smith, P. (1993) Large scale models and large scale problems : the case of the health services.

Smith. P. C. (1995). "On the unintended consequences of publishing performance data in the public sector", *International Journal of Public Administration*, 18, 277-310

Smith, P. C. (1999). "Setting budgets for general practice in the new NHS", *British Medical Journal*, 318, 776-779.

Smith, P.-C. (1997) Devolved purchasing in health care : a review of the issues.

Street, A. & Place, M. (1998) The management challenges for primary care groups.

Stump, R.L. (1995) Antecedents of purchasing concentration - a transaction cost explanation. *Journal of Business Research*, **34,** 145-157.

Sullivan, K. (2000) On the 'efficiency' of managed care plans. *Health Affairs,* **19,** 139-148.

Tu, H.T., Kemper, P. & Wong, H.J. (1999) Do HMOs make a difference? Use of health services. *Inquiry - the Journal of Health Care Organization Provision and Financing,* **36,** 400-410.

Walston SS, Kimberly JR and Burns LR (1996) Owned vertical integration and health care: Promise and performance, Health Care Management Review, 21, 1, 83-92

Weill TP (2000) Horizontal mergers in the United States: some practical realities, Health Services Management Research, 13, 137-151

Wholey, D., Feldman, R., Christianson, J.B. & Engberg, J. (1996) Scale and scope economies among health maintenance organizations. *Journal of Health Economics,* **15,** 657-684.

Wholey, D.R., Burns, L.R. & Lavizzo Mourey, R. (1998) Managed care and the delivery of primary care to the elderly and the chronically ill. *Health Services Research,* **33,** 322-353.

Williamson OE (1981) The economics of organisation: The transaction cost approach, American Journal of Sociology, 87, 3, 548-577

Wyke, S., Hewison, J., Piercy, J., Posnett, J., Macleod, L., Page, L. & Young, G. (1998) National evaluation of general practice-based purchasing of maternity care : preliminary findings.

Appendix A: *Literature Review Search Details*

Databases used

MEDLINE (1966-8/2000)
EMBASE (1980-6/2000)
EconLit (1969-6/2000)
Kings Fund Database (1979-8/2000)
Health Management Information Service database (1984-1998)
DH-Data (1983-8/2000)
Cochrane Library (2000 Issue 2)
Database of Abstracts of Reviews of Effectiveness (to date)
NHS Economic Evaluation Database (to date)
Health Technology Assessment database (to date)
Science Citation Index (1981-8/2000)
Social Science Citation Index (1981-8/2000)
Index to Scientific and Technical Proceedings (1981-8/2000)

Search Terms

Organisation	**Size**	**Performance**	**Economic**
Primary care group*	Population	Performance	Economies of scale
Primary care trust*	Size	Quality	Economies of scope
Health maintenance organi*ation*	Scope	Efficiency	Optimal size
Hm	Grouping	Outcomes	Optimal efficiency
Hmos	Number near (practice* or gp* or physician* or service*)	Output	Efficiency measure*
Managed care organi*ation*	Size near (practice* or gp* or physician* or service*)	Risk management	Transaction cost*
Total purchasing	Large near small	Productivity	Management cost*
Locality purchasing	Geographical coverage	Budget manag*	Data envelopment analysis
Internal market*	Area covered		DEA
			Organi*ational performance

Appendix B : *estimating economies of scale.*

"Engineering" estimates
Experts familiar with the current best practice technology estimate the inputs required to produce different levels of output. The method has the advantage that like is being compared with like, avoiding the difficulty with empirical studies that plants of different size may be of different vintage or have been built when input prices were different, so that the effects of size are confounded. The simulation studies of the relationship between size and risk (Bachmann and Bevan, 1996; Crump et al., 1991; Martin, Smith and Rice, 1998) and the attempts to estimate the required level of PCG/T management costs (Place and Newbronner, 1999) could be thought of as simple examples of engineering estimates.

Survivor techniques.
In competitive markets only organisations which minimise the cost of production are profitable and survive. By examining the change in the size distribution of firms over time it is possible to identify the size which minimises the cost of production.

The US literature has a number of studies of the survival probability of for profit health service firms and its relationship with size. Inference about the average cost minimising size of firm based on such studies needs to be treated with caution since it is by no means clear that such firms operate in a competitive environment. If their products are close but imperfect substitutes, perhaps because of location, then the most profitable size of firm depends on demand conditions as well as the relationship between cost and size.

We found no examples of survivor techniques applied to UK health care organisations. Such studies would in any case be dubious. Changes in the size distribution of public sector or other not for profit health service organisations which do not face binding bankruptcy constraints cannot provide information about relationship between cost and size. The method could in principle be applied to general practices which are privately owned and do face binding financial constraints. However, practices are imperfect substitutes for each other because of the effect of location on patient choice of practice and government policies have been directed at increasing practice size. It is doubtful whether changes in the practice size distribution provide much information about the existence of economies of scale at practice level.

Data envelopment analysis
Data envelopment analysis (DEA) compares information on inputs, outputs and the environment of a set of organisations to calculate efficiency scores for each organisation. The scores are derived by weighting outputs and inputs to construct a productivity ratio as an efficiency score. DEA has the advantage that it requires very few prior assumptions about the underlying technololgy. In some UK health service contexts efficiency scores have been shown to be highly sensitive to the number of inputs and outputs and to their definitions (Giuffrida and Gravelle, 2000). Further, DEA also does not lend itself readily to investigation of economies of scale because calculation of the efficiency scores requires an assumption about whether there are constant or variable returns to scale. It is possible to assume that there are constant returns and then to investigate whether the resulting efficiency scores are correlated with size but the methodology seems overly ad hoc.

Regression methods
Most of the quantitative studies found by the literature review used regression methods to estimate the relationship between cost and output. Such methods require assumptions about the underlying technology and environment but the assumptions are in principle testable. Their results are more readily interpretable than DEA, especially for the relationship between size and costs.

Literature relevant to economies of scale in PCG/Ts was collected by a formal search of a number of bibliographic databases. The details of the databases used and the search strategies are given in Appendix A. In addition, a description of the study and a request for additional references was sent to an electronic mailing list for health economists and to academics thought to be active in the area. The authors also drew on their own knowledge of the health economics, industrial economics and health service literatures.

Appendix C:

Size and performance of health care organisations: US Literature

Paper	Data	Dependent variable	Measure of size	Other variables	Method	Results	Comments
Barr (1995)	N/A	Patient and physician satisfaction	Practice size – number of physicians	N/A	Literature review	Patients and physicians tend to be less satisfied with larger practice settings.	Barr concludes that the optimal size for a primary care organisation remains unknown.
Clement (1995)	985 National HMOs 1977-86	Numbers of HMOs in size categories	Number of enrollees	none	BV. Survival analysis	A minimum enrollment of at least 25,000 members should be achieved as quickly as possible. A higher enrollment of 40,000 to 60,000 appears necessary for longer-term survival.	
Debrock and Arnould (1992)	40 Illinois HMOs 1985-87	Hospital visits per capita	Number of enrollees	Physician compensation arrangements, hospital incentives, managerial control, for-profit status	MR. Generalised Least Squares	The numbers of enrollees had a positive and significant effect on admissions per capita, lending support to the contention that size increases the cost of monitoring.	Physician behaviour influenced by financial incentives. Bonus systems most effectively minimise free rider problem when based on individual rather than group behaviour
Draper et al (2000)	249 National HMOs	Cost	Number of enrolees	HMO age and model type (Group or IPA), for-profit status, federal qualification, national affiliation	DEA and ANOVA	Small and very large HMOs are the most efficient. Looking at only the top 50% of the largest HMOs reveals a linear relationship between size and efficiency	
Encinosa and Sappington (1997)	N/A	Welfare; insurance contracts	Number of enrollees	HMO competition	Theoretical model of HMO competition - asymmetric information	Market power and scale economies can sometimes admit socially preferred outcomes when they would not otherwise arise	
Feldman et al (1996)	All HMOs 1986-93	HMO outcomes: merge and fail or survive	Number of enrollees	HMO age and model type (Group or IPA), for-profit status, federal qualification, national affiliation	MR. Multinomial logit	Large and profitable HMOs are more likely to merge and survive, less likely to merge and fail and less likely to fail.	Mergers more likely in markets with more competing HMOs. There was no evidence that mergers increased cost
Given (1996)	California. 31 HMOs 1986-92	Total cost	Member months (Medical, Medicare, commercial)	Input prices. HMO age. Type. Competition	MR. Translog with share eqns	EoS exhausted at 115,000 enrolees. Diseconomies of scope	Compet positively correlated with cost.

Hargreaves et al (1996)	630 practitioners in 16 primary care group practices 5 years	% of review criteria for each patient care guideline met	Number of practitioners in the group	Full-time status, hospital or health care centre,	Mixed model ANOVA	A statistically significant negative association was found between practice size and performance was found for practices not receiving the intervention. For intervention practices this relationship was reversed.	Randomised controlled trial with quality controlled assurance intervention.
Leutz et al (1990)	4 Social HMOs 1985-88	Costs and revenues; break-even enrollment	Number of enrollees	Type of HMO, age	Qualitative case study	Two of the HMO-based plans could break even at around 2,000 members or less, but the two new SHMO models may need 3,000 to 5,000 members each.	3 out of the 4 HMOs showed losses over 1985 to mid 1987. The primary source of losses was slow enrollment **Comments**
Long (1985)	56 federally qualified HMOs	Average costs per unit of output (hospital days and ambulatory visits)	Number of enrollees, number of enrollees squared	Age of HMO, input prices, HMO type	MR. OLS regression	For staff HMOs and IPAs, size is negatively and significantly related to average costs. Size sqd. is positive and significant (indicating diminishing returns) . Size was not significant in the group model and when all types were combined. Minimum optimal scales are calculated as being 32,700 for staff, 42,300 for group and 21,900 for IPA.	Mean sizes for HMO types: Staff : 20,227 Group : 132,930 IPA : 13,705
Nutting et al (1982)	6 Govt owned health service units, 3 FFS practices, 2 HMOs 1974-76	% of consumers in need receiving adequate standards	Log of the number of direct care providers (physicians, nurses and physician extenders)	System Type	ANCOVA	Size was found to have a strong negative relationship to the quality of treatment and follow-up care. Size was also inversely related to performance of screening for hypertension.	Payment mechanism had no effect on quality of treatment or follow-up care. HMOs performed significantly better in screening for prenatal anemia.
Rich et al (1998)	153,000 enrollees -112 primary care practices in Minnesota 1995	Non-hospital cost per member per year	Number of physicians in the group	Enrollee characteristics (age, sex, etc) referral management, hospital care management primary care services.	MR. OLS regression	Group practices with larger numbers of physicians are associated with lower non-hospital costs. An increase in 10 for the size of the group would be associated with a 1% decrease in costs per member per year.	Efforts to have patients identify a primary care physician may be more important in larger multispeciality groups.
Rosenman et al (1997)	28 Florida HMOs 1994	Total costs	Number of enrollees	For-profit status, input prices, tax status, HMO type	DEA	Efficiency is positively related to both total enrollment and homogenous enrollment. I.e. there are economies of scale but diseconomies of scope	

Schlesinger et al (1986)	173 HMOs 1983	Costs and revenue per enrollee	Number of enrollees	For-profit status, federal qualification, input costs	MR. OLS regression	The number of enrollees is significantly and negatively associated with total costs	For-profit status increases costs (via higher ambulatory costs per enrollee) Average total enrollees is 46,400 (non-profit) and 30,700 (profit)
Wholey et al (1996)	US. 559 HMOs. 1988-91	Total cost	Member months (Medicare, non-Medicare)	Physicians. Group or IPA structure. Input prices. Population density. Competition	MR Translog with share eqns	EoS for non-Medicare and Medicare enrollment but small after 50000. Diseconomies of scope for non-Medicare and Medicare enrollees.	Compet positively, pop density negatively correlated with cost.
Wholey et al (1998)	HMOs census data 1991-95	Primary care staffing levels	Number of enrollees	HMO type, for-profit status, HMO competition, national based, federally qualified	MR. OLS regression	Larger HMOs have fewer physicians per 100,000 enrollees than do smaller ones.	For-profit status has a negative effect on the primary care physicians but a positive effect on the number of specialists.

MR: multiple regression. BV: bivariate analysis. OLS: ordinary least squares. DEA: Data envelopment analysis

Size and performance of health care organisations: UK Literature

Paper	Data	Outcome Measure	Measurement of size	Co-variates	Technique	Results	Comments
Bachmann and Bevan (1996)	Simulation	Costs of 15 categories of rare and costly referrals	Population	None	Simulation of 100 fund years assuming Poisson distribution of referrals	Risk of overspend increases rapidly with risk pools under 30,000 and decreases marginally with larger pools. Results robust to assumptions about prices and referral rates.	Rare costly referrals seem unlikely to bankrupt TP sites. The management of risk is not in itself a justification for merging organisations to generate large populations
Baxter et al (2000)	49 1st wave TPPs 1997	Budget management and success; decision making techniques	Multi- or single practice; small (populations < 30,000) vs large	None	BV	Small and single-practice pilots were more likely to; underspend their budgets, involve more GPs in decision making, integrate clinical and financial roles.	TPPs were questioned after being live 1 year. More complex multi-practice groups may catch up with single-practice groups over time.
Bevan (1998)	52 1st wave TPPs 1995 - 1997	Risk and TPP budget setting	Population	None	Descriptive	Small area variations in utilisation have only been partly explained Reliable estimates of target spending can be achieved provided populations are over 50,000	

Crump et al (1991)	Simulation based on historical data from Birmingham Health Authority GPs 1989 - 90	Simulated expenditure on 113 surgical procedures	Population of 179,400 divided into practice list sizes (listsizes 9,000 12,000 15,000 18,000 and 24,000)	none	Simulation of 100 fund years assuming Poisson distribution of referrals	Practices of 9000 would exceed their surgical budget by more than 5% 25% of the time; those of 24,000 would exceed budget 5% of the time.	
Earwicker (1998)	Nottingham Total Commissioning Pilot 1998	Management costs; contract setting with trusts; budgets, GP incentive setting	TCP compared to Nottingham HA average	None	BV Mostly descriptive but some comparisons with HA averages	Prescribing costs lower for TCP than HA average. Higher rate of generic prescribing. Emergency and elective referral rates below HA average	Nottingham TCP (with 455,000 popn.) is bigger than largest PCG. Fundholders and non-fundholders formed group.
Giuffrida et al (2000)	90 FHSAs. Health Service Indicator data 1989/90 – 94/95	Administration costs	Numbers of GPs, nurses and other medical staff	Input prices, population density, test rates, % GPs fundholders	MR Generalised Translog function	Random effects model suggests that there are EoS to be exploited at the mean 1994/1995 size of FHSAs. EoS exhausted at a level 20% larger than the average FHSA	
Goodwin et al (1998)	52 1st wave TPP 1996-97	Questionnaire response on the ability of TPs to achieve their own objectives (especially those functions beyond fundholding)	Single- and multi-practice; small (fewer practices and population) and large	None	BV	Single-practice and small multi-practice TPPs were more likely to report achieving their objectives.	Wide variation in achievements reported. Higher achievement associated with higher direct management costs.
Killoran et al (1998)	Hypothetical PCG/Ts	Management costs; commissioning		None	Estimates based on evidence from TPPs	Appropriate management costs may be £17 - £18 per capita due to high costs of co-ordinating large numbers of GPs. Current PCG sizes not large enough to commission metal health, emergency care; or to reconfigure acute and mental health services.	

Killoran et al (1999)	1st and 2nd wave TPPs. 1998	Objectives achieved	Practices; population	None	Postal survey of lead GPs. Case study of 12 TPPs & Wakefield whole district TPP.	Smaller TPPs were able to achieve objectives with relatively little investment; larger TPPs required substantial time and investment before progress was made. By the second live year, many larger projects caught up.	
Place and Newbronner (1999)	East Yorkshire PCG. 1999	Expected costs of managing a PCT at level 3 and level 4.	East Yorkshire population (approx 146,000)	PCT structure, management activities	Case study	Estimates adequate management resources to be £9.25 per head of population for level 3 PCT, £11.20 for a level 4 PCT	
Place et al (1998)	7 1st and 2nd wave TPPs	Management and transaction costs	Practices; populations		Descriptive detailed cost analysis	Cannot establish optimal size of TPP. EoS in management eroded by coordination problems in larger commissioning groups	Estimates are of difference between TPP and FH.
Posnett et al (1998)	52 1st wave TPPs, 35 2nd wave TPPs. 1995 -1997	Management and Transaction costs	Practices, population	None	BV	No statistically significant relationship between direct management costs per capita and size. For transaction costs (uses only 6 TPPs), total transaction costs per practice are higher for single-practice TPPs	
Smith and Knight (1999)	40 Birmingham GPs, HA and Trust managers 1998	Support for a supra-PCG	Supra-PCG of 12 Birmingham PCGs	None	Survey. Descriptive	Mixed reaction to supra-PCG. General support for a set of umbrella functions (clinical governance, strategic overview and management support)	
Smith and Regen (1999)	40 PCG Pilots 1998	IT arrangements, prescribing, participation	Groups of stakeholders	None	Descriptive	Involving different groups of stakeholders (GPs, nurses, lay) appears to enjoy a significant degree of support in principle but this enthusiasm may soon be tested by the practical considerations.	

Street and Place (1998)	81 stakeholders including 16 GPs, 6 nurses, 14 Practice managers and 30 HA staff 1998	Perceived advantages and weaknesses of working in PCGs	Number of workers (GPs, nurses, etc.)	None	Workshop to elicit views. Descriptive	The requirement that practices co-operate was seen as the main barrier to effective working (accounting for 31% of all barriers) Internal differences may be solved by creating (town-based) sub-groups.