CMPO Working Paper Series No. 01/37

Team-Based Incentives in the NHS:

An Economic Analysis

Marisa Ratto[1]

with

Simon Burgess[2], Bronwyn Croxson[1], Ian Jewitt[3] and Carol Propper[4]

[1]CMPO, University of Bristol
[2]CMPO, University of Bristol, and CEPR
[3]Nuffield College, University of Oxford and CEPR
[4]CMPO, University of Bristol, and CEPR

June 2001

Abstract

The NHS Plan welcomes the use of team rewards but does not specify how team based incentives are to be implemented or make clear what types of teams such incentives are to be given to. This paper looks for insights from economic theory on how to define teams and optimal incentive schemes within them. We give a brief description of the incentive mechanisms suggested in the NHS Plan and discuss the implications of the economic theory of team-incentives to the NHS. One implication is that the optimal compensation scheme depends on the type of team. Hence, the definition of teams in the NHS should precede the identification of the system of team rewards. We therefore examine which types of teams might be suitable for team rewards. We then consider issues in the use of financial incentives for such teams.

JEL Classification: D23, J41
Keywords: team production, NHS, financial incentives

Acknowledgements

We thank the Leverhulme Trust for funding CMPO and this project.

Address for Correspondence
Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
Tel: +44 (0)117 954 6944
m.l.ratto@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

INTRODUCTION

It is clear from the government's plans for its second term that its over-riding priority is a huge improvement in the delivery of public services. One of the flagship public services is health care, and the government has already begun the large task of reforming the delivery of health care through the publication of the NHS Plan. This plan outlines a performance measurement and assessment system, and proposes a key role for team-based bonuses for NHS staff. However, the Plan doesn't discuss either how such financial rewards should be implemented, or does it go in any depth into what constitutes a team.

In this paper we seek to address both issues using economic analyses of team based incentives. We begin by examining the classic arguments for the use of explicit financial team based rewards. When team bonuses can be used to motivate team members and when they alone are not sufficient. One implication of this analysis is that the optimal compensation scheme depends on the type of team. This suggests that the definition of teams in the NHS should precede the identification of the system of team rewards. We pursue this here and examine how teams might be defined if financial rewards are to be given at team level. Our analysis indicates the existence of several types of team in the NHS, some of which seem more suitable for the use of team based rewards than others. We then discuss issues that may arise in the use of explicit financial incentives at team level in the NHS.

1. THE NHS PLAN AND INCENTIVES

In this section we briefly discuss the NHS Plan as a whole. We then focus on the system for promoting performance and the role of incentives in improving performance.

1.1. The NHS Plan

The NHS plan, issued in July 2000, describes how the government intends to take forward the process of reform of the NHS. The reorganisation of the NHS started in 1997 with the White Paper, and aimed to move from the internal market to a new approach based on patient satisfaction. The need to adopt a new vision of the health care system, centred on the patient, emerged as a reaction to the logic of the internal market. According to the document, the previous system neglected the needs of patients in favour of market considerations, leading to poor standards and differences in patients' treatment.

The NHS Plan reaffirms the identification of patient satisfaction as *the* objective for any innovation to be introduced. The document emphasises two traditional points of strength of the health care sector that were undermined by the logic of the internal market, but which instead should be vigorously defended and promoted. These are teamwork between professionals and organisations and the dedication of staff working in the NHS.

The reforms described in the Plan encompass several aspects of health care: funding, investment in facilities and staff, the delivery system, roles, skills, rewards and contract conditions for all the professions involved in the provision of health care and attention to patients' concerns. We concentrate here on the recommendations for a new delivery system for health care, which are dealt with in Chapter 6 of the Plan.

The objective of the new system of delivery is to guarantee high standards of treatment for all patients, with "..*maximum devolution of power to local doctors and other health professionals*". The final stage of reform should therefore consist of devolving autonomy and responsibility for health care at a local level. But before giving complete autonomy to local organisations, however, the Government intends to raise and equalise performance across all NHS organisations. This is the objective of the (current) process of transition. To do this, the Plan calls for the creation of national and local structures to set priorities for improving the nation's health, develop standards of treatments at a national level and enhance performance among all NHS organisations.

In what follows we focus on the recommendations of the document on how to promote better performance in the NHS and focus in particular on the system of financial incentives.

1.2. Promoting better performance

According to Chapter 6 of the NHS Plan, the provision of better care and outcomes for patients is to be implemented in three steps. These are Phase 1: Measuring performance across NHS organizations, Phase 2: Classifying NHS organisations according to measured performance and Phase 3: Setting efficiency targets and delivering incentives to reach those targets. At present we are in Phase 1 and pilot schemes are to be implemented for phases 2 and 3.

Measurement of performance will be by the use of the Performance Assessment Framework (PAF), issued in 1999. The PAF is intended to help the NHS to improve the health of the population and provide better care. The framework identifies six areas in which activities will be measured. These are health improvement, fair access, effective delivery of appropriate healthcare, efficient use of resources, patient-carer experience and health outcomes of NHS care. For each of these areas a set of provisional performance indicators[1] are suggested as the basis for setting standards of performance and to assess how the NHS is doing. The current PAF has been designed for Health Authorities, but will be extended to all NHS organisations by April 2001[2]. Appendix 1 gives details.

Performance will be assessed on the basis of core national targets and the PAF measures. Health Authorities, NHS Trusts, Primary Care Groups, Primary Care Trusts and Health Action Zones will be conferred a *traffic light status* and be classified as *green*, *yellow* or *red* organisations. Red organisations will be those who fail to meet a number of core national targets. Those that meet all core targets and score in the top 25% of organisations

[1] We provide a list of the performance indicators in the appendix.

on the PAF will be labelled as green organisations. The yellow status will be given to those who meet all or most national core targets but are not in the top of 25% of PAF performance. It is worth emphasising that performance is measured both in absolute and relative terms: the core national targets will be used to measure performance for each single organisation, in absolute terms, whereas the PAF measures will be used to rank organisations and assess how they perform relative to the others.

The exact criteria for traffic light status are not completed yet. It is clear that they will take into consideration **joint production** across NHS organisations. In fact paragraph 6.28 of the document specifies" …*the criteria for traffic lights will explicitly include how well they* [individual NHS organisations] *work in partnership with others and how well the local health economy as a whole is performing on key shared objectives*."[3]

Finally, trusts will be set efficiency targets. These will again be based on PAF measures and will based on levels of service already being achieved by the best trusts around the country. The achievement of those targets will be promoted by use of an incentive scheme. The incentive scheme is designed as a system of rewards consisting of (a) greater autonomy (b) national recognition and (c) financial incentives.

Greater autonomy will be granted in the form of lighter monitoring by Regional Offices, less frequent monitoring by the Commission for Health Improvement, greater freedom to decide the local organisation of services, the possibility of taking over persistent 'red' organisations and even delegated Regional Office performance management functions.

National recognition will consist of being used as exemplars for the Modernisation Agency, which is an agency to be created for helping local clinicians and managers redesign local services**.** Greater autonomy and national recognition will be conferred to green light organisations.

---

[2] The Department of Health has run a consultation exercise for a new set of performance indicators for Health Authorities, Trusts and Primary Care Trusts (NHS Performance Indicators: A Consultation. Department of Health, May 2001).
[3] The NHS Plan, p. 63.

Financial incentives will be distributed out of the National Health Performance Fund, to be introduced in April 2001. The Fund, to be held and distributed regionally, will allow financial incentives worth on average £5M for each Health Authority. They will consist of a given share of the Fund, to be allocated conditional on traffic light status[4], rewards for staff and organisations that manage to succeed in particular tasks[5] and rewards for team production across and within NHS organisations.

Particular emphasis is put on rewards and team production. The document devotes three paragraphs to team production.

> *"6.36 For primary care groups and trusts one of the criteria for access to the funds will be the development and use of incentive schemes that ensure referrals to hospital are appropriate and will help achieve shorter waiting times. Incentives will also be developed for joint working between primary care groups, NHS Trusts and Social Services to achieve improvements in rehabilitation facilities for older people.*
>
> *6.37 The Performance Fund will enable NHS trusts and primary care trusts to offer greater incentives to staff in clinical teams and primary health care teams linked to their contribution to service objectives. The reward could take the form of:*
>
> - *Money to buy new equipment or upgrade facilities to improve patient care*
> - *Improved facilities and amenities for staff*
> - *Non-consolidated cash incentives for individuals and teams.*

[4] Green organisations will have access to their share of the Fund automatically, as of right. Yellow organisations will need to agree plans, signed off by the regional office, setting out how they will use their share of the fund. Red organisations' share will be held by the Modernisation Agency, which will also oversee their spending.

[5] Tasks listed in the document are: reducing waiting times and introducing booked admission, redesigning waiting out of the system and improving the quality of care, adopting local referral protocols based on national clinical guidelines.

6.38 As part of these new arrangements we intend to pilot the use of team bonuses in a number of NHS trusts from next year. The results of the pilots will inform our decisions on the further development of team rewards."

The measurement of performance will be dealt with at a national level, in that it is based on the core national targets and the PAF, issued by the Department of Health. The assessment of performance will be devolved to Regional Offices. The publication of the results of the PAF will be annual and under the responsibility of the Commission for Health Improvement (CHI), which is a separate inspectorate, created to guarantee an independent verification of the quality of the NHS. The Commission for Health Improvement will inspect every NHS organisation every four years and the organisations classified as red every two years.

It is worth noting that the functions of measuring, assessing and monitoring performance will not be managed by a single body (or principal). The Department of Health sets the procedure to measure performance, the regional authorities of the Department of Health make the assessment and an independent body, the CHI, verifies their activity and periodically inspects *every* NHS organisation.

We now turn to discuss the system of incentives in more detail.

1.3. The system of incentives to improve performance in the NHS

The incentive scheme suggested in the NHS Plan is based on both non-financial and financial incentives, although considerable emphasis is given to the latter. It offers incentives **for** and **within** organisations.

Non-financial incentives may play a powerful role. The non-financial incentives in the Plan to be delivered in the form of earned autonomy and national recognition focus on independence and reputation respectively. Intuitively, earned autonomy acts as an incentive in that being constantly assessed and verified imposes some costs in terms of

time and other resources. The possibility of avoiding inspection by exhibiting outstanding performance may induce organisations to improve their performance. National recognition allows an organisation to build a reputation, which has been shown to be beneficial in terms of attracting good workers (known as sorting effects) and team motivation. Another aspect of non-financial incentives is the annual publication of the list of green, yellow and red organisations. Making the classification of NHS organisations open to the public can be a powerful incentive to induce better performance. The possibility of being top in the list of best organisations allows an organisation and those within it to build a reputation for high quality. This may both be beneficial in terms of career progression for individuals in the organisation and give the individuals intrinsic satisfaction. Such ideas lie behind the notion of Beacon status in schools and the use of league tables more generally in the public sector. Being at the bottom of the list is argued similarly to improve performance (the idea behind 'name and shame')[6].

The financial incentives in the Plan consist of a system of monetary rewards linked to performance. These have four components that can be formalised as follows:

$$y_i = a + b_i + c_i + d_i$$

Where y_i is the total amount of money given under the National Performance fund to organization *i*, and the other terms are components of this (discussed below).

The Government will introduce a National Performance Fund to be distributed regionally. Each Regional Office will have a fixed amount of money to devote to all the NHS organisations in its region. Each NHS organisation will receive y_i, on the basis of overall performance and the attainment of specific objectives.

An organisation *i* will receive a fixed share *a* of the Fund, which will be equal for all organizations in the region. However, organisations will have a different accessibility to it on the basis of their traffic-light status. Green organisations will have access to their

[6] The limits to the effects of league tables in health are discussed in Smith et al (see Grout et al 2000).

share of the fund automatically as of right. Yellow organisations will need to agree plans, signed off by the Regional Office, setting out how they will use their share of the Fund. Red organisations' share will be held by the Modernisation Agency, which will also oversee their spending. Hence the idea is to give the same amount of money to every organisation, so that they can have the same opportunities to make improvements. But the money will be administered differently according to overall absolute and relative performance[7]: those organisations that exhibit better performance will have more freedom on how to spend their share of the fund.

The second term in the overall compensation y_i includes an extra reward, b_i, which can be considered as performance related pay. It is to be awarded to those organisations which manage to reach specific targets, such as reducing waiting times and introducing booked admission, taking 'waiting out of the system' and improving the quality of care, adopting local referral protocols based on national clinical guidelines. The document does not set any precise criteria on how to allocate these extra rewards, but we can assume that the performance of each organisation in these specific aspects of their activity will be measured against some critical level, which identifies a target to be reached. Those organisations that reach the targets will be rewarded.

The final two components of y_i consist of extra rewards for team production. c_i represents extra rewards for team production *across* organisations, whereas d_i stands for extra rewards for team production occurring *within* the same organisation. This distinction reflects the distinction adopted in paragraphs *6.36* and *6.37* of Chapter 6 between joint work *across* and *within* organisations.

1.4. The definition of teams

Although team production is recognised as an important aspect of NHS activity, it is quite vaguely defined in the NHS Plan. The document identifies two specific forms of

[7] The traffic-light status is conferred on the basis of core national targets and the PAF measures. Performance is measured both in absolute terms and relative to the top 25% organisations.

joint production among different organisations, but these are not exhaustive of examples of joint production across the whole NHS.

Paragraph 6.36 suggests rewarding primary care groups and trusts that succeed in decreasing waiting times, and rewarding the joint effort of primary care groups, trusts and social services to improve rehabilitation facilities for older people. The second of these is a clear example of joint production across NHS organisations. The provision of rehabilitation facilities requires the collaboration of different units in different organisations and is therefore a clear example of joint production across primary care groups, trusts and social services. In the first case, referrals to hospitals are the productive inputs provided by primary care groups to trusts. The availability of beds is the input provided by the staff in the trusts, which, together with referrals, will determine the waiting lists. This list can then be considered the joint output of Primary Care Groups and Trusts. In both cases, the definition of a team seems to rely on complementarity in production: the activities of one group of staff determine the productivity of another group.

Paragraph 6.37 supports the award of extra rewards for joint production within the same organisation at the level of clinical teams and primary health care teams. In the Plan team production within the same organisation is defined in a very broad and general way: the existence of a team seems to be due to a common service objective that is shared by team members. Extra rewards will consist of monetary compensation and also improved facilities and amenities for staff. This implies that an important aspect of the production process within a NHS organisation is team motivation. Improved facilities and amenities for staff affect the quality of the working environment and are aimed at creating a team spirit among workers in order to facilitate their cooperation.

The two paragraphs seem therefore to distinguish between joint production among different organisations and joint production within the same organisation, even if a clear definition of a team and of the incentive scheme to promote team-work remains to be

identified. Paragraph 6.38 postpones the arrangement of a suitable reward system until the results of a trial of the use of team bonuses in a number of NHS trusts are available.

These definitions of teams are very general. As we shall consider in section 3, the set of common objectives shared among staff working in NHS organisations could be quite diverse. For example, individuals may be organized in teams to exploit complementarity in production or to pool risk or to monitor each other, or to communicate and learn from shared job experiences. In each of these cases the incentive mechanism has to act on a different motivation and hence needs to be designed in a different way.

In summary, the Plan outlines methods for measuring, assessing and rewarding NHS organisations and emphasised team production. The idea of using team-based incentives to promote joint production in the NHS is consistent with the awareness present in the public sector that production within government agencies depends on the co-operation and public spirit of staff, which may be undermined by individual incentives. However, the actual detail on the design of teams and the best way to reward them is sketchy at best. Understanding the rationale behind the existence of a team is a priority if an effective incentive scheme is to be designed. The reason for the existence of a team determines both the selection of the agents to take part in the team and the targets to be promoted by use of incentives. To this end, we now turn to the theoretical literature on team-based incentives for some insights into how to define team-based rewards.

2. THE THEORY OF TEAM INCENTIVES

In a team the value of final output is determined by the joint effort of the team members and may be measured with more or less accuracy. Similarly, the individual actions or efforts by team members may be measurable with greater or less accuracy. This raises the problem of inducing team members to exert proper levels of effort by using the proper combinations of team-based and individual-based incentives.

There are two main directions suggested by theoretical literature: the use of *explicit* incentives or exploitation of existing *implicit* incentives. Explicit incentives are delivered in the form of a compensation package that rewards (penalises) outstanding (poor) performance and induces individuals to behave in line with the team's objectives. Implicit incentives may arise through individuals' motivation to further their future careers, or, perhaps most significantly they may arise through intrinsic motivation (e.g. to care for sick people). In the case of implicit incentives devices other than performance related pay may be warranted, such as organising the production process in a way that is possible for team members to monitor each other, or motivating workers by giving them more job independence. The interactions between implicit and explicit incentives may be of paramount importance, for instance whether explicit incentives crowd-out intrinsic motivation for current workers and whether focussing on explicit incentives attracts workers with less intrinsic motivation.

The NHS Plan focuses on explicit incentives in that it calls for the use of financial incentives to promote team production. Given this, we typically also focus on explicit incentives in this paper, but the relationship between explicit and implicit incentives needs also to be borne in mind in a full analysis. We will make some comments in what follows but a fuller analysis is deferred.

2.1. The Holmström analysis of moral hazard in teams

Holmström (1982) is one of the seminal contributions to the theory of incentives in teams, and we review some of the lessons from this paper in order to further our discussion. Holmström examines how to design explicit rewards to solve moral hazard[8] issues in a team. A team is defined as " …a group of individuals who are organised so that their productive inputs are related"[9]. The paper distinguishes between two features of multi-agent organisations: *the free-rider problem* and the role of *peer performance evaluation*.

[8] *Moral hazard* refers to situations in which the actions taken by individuals are not directly observed. The term originated in the insurance industry to describe situations e.g. in which individual's took less care to avoid the insured event once insurance had been purchased.

The free-rider problem is due to the fact that workers in a team may have an incentive to shirk even when joint output is perfectly observable. In the case of a single agent this would not occur: in this case when output is perfectly observable it is an optimal indicator of the agent's inputs. In a team, final output is made up by all team members' contributions, and hence even if we observe final output perfectly, it is not a good indicator of each agent's inputs.

Peer performance evaluation can be used in teams to ameliorate moral hazard problems. When agents are subject to common uncertainty, even if their contribution to final output is separately observable, it can be useful to use relative performance evaluation, in that one agent's output provides information about the uncertainty facing another agent's.

2.1.1. The free-rider problem

Holmström applies his analysis to a setting where team members depend on each other to produce final output. In this case, when agents' individual inputs are imperfectly observed, if all the output of the team is shared among team members, team members are induced to free ride. We provide Holmström's formalisation of this argument in Appendix 2. The intuition behind this idea is that paying an amount equal to total output to team members (known as the balanced budget rule) creates a negative externality for the team. If output is fully shared among team members, when an agent decreases his/her contribution, the value of total output will decrease and the sum of all agents' shares will decrease. Hence the agent who cheats will not pay in full for the consequences of his/her act. The cost of one person's shirking (in terms of the share of lower joint output) will be passed onto the others. The private marginal cost of shirking will be less than the social marginal cost (borne by all members of the team) and the level of effort chosen by the individual, that maximises his/her payoff, will be lower than the Pareto efficient level. Intuitively, this free-rider problem becomes greater in large organisations. Rewarding everyone in the NHS for improvements to the NHS as a whole will lead to negligible

[9] Holmström (1982), p. 325.

incentives for individuals for any moderately sized incentive budget. This is the negative externality caused by the balanced budget-sharing rule, which induces team members to free ride. Note that this is in contrast to individual production – paying an individual the full value of her output will induce the efficient level of effort.

The solution that Holmström suggested to this incentive compatibility problem is to allow the aggregate compensation of the team to be less than the value of output, i.e. to break the budget balancing constraint, when the output level falls below a critical level[10]. In particular, when output falls below a critical level, agents should be penalised, in that they should receive only part of the value of joint production. The potential surpluses generated by not distributing the entire value of production to workers acts as an incentive to exert efficient levels of effort.

A recent contribution by Marino and Zábojník (2001) takes a different approach and argues that the free rider problem can be overcome by dividing the organisation into a group of sub-teams which compete against each other. Marino and Zábojník in fact show that close to optimal incentives can be restored by creating a system of teams which play a type of tournament against each other – the winning team receiving an optimally designed reward. Intuitively, the free-rider problem within each team is balanced by the fact that the winner-takes-all incentive is sufficiently high powered. Marino and Zábojník therefore find a role for teams which has little to do with intrinsic complementarities in production, but more with the balancing and creation of overall incentives within the organisation as a whole. Even where there do exist intrinsic complementarities in production within subsets of the organisation, the creation and reward of teams should clearly be designed with the overall impact on the organisation in mind.

2.1.2. The determinants of the optimal compensation scheme

The optimal compensation scheme for each individual (as distinct from the organizational form) varies according to whether joint final output is perfectly observable. In cases

[10] The definition of critical level varies according to whether joint output is perfectly observable.

where it is perfectly observable there is no uncertainty in production: output is solely determined by agent's actions and it is perfectly measurable. When final output is affected by random events, there is uncertainty in production and output will be only imperfectly measurable.

Holmström shows that the optimal compensation scheme will depend on the extent of certainty in production, the size of the team and the attitudes of team members towards risk. His analysis of the optimal compensation scheme is summarised in Chart 1 and the components are now discussed.

Certainty in production

When joint final output is perfectly measurable, a first best solution is attainable by imposing a system of group penalties whenever output falls below the Pareto optimal level. The compensation scheme can, for example, be such that final output is fully shared among team members if it is greater than or equal to the Pareto optimum, otherwise the principal will take some or all of the value of production away. In this case team members will receive a reward only if final output does not fall below the Pareto optimum.

When output is perfectly observable the compensation scheme is not affected by the team size or the individuals' preference towards risk.

Uncertainty in production

If joint output is not perfectly measurable the design of the optimal incentive scheme is more complex and depends on the size of the team and agents' attitude towards risk.

*Small teams***.** Holmström defines a small team as a team where each agent has a substantial impact on the probability distribution of final output. In this case group penalties can still be used to enforce a first best solution. The balanced budget rule should be broken whenever final output falls below a critical level. The compensation scheme is such that if observed output falls below the critical level an individual's share will be

decreased by a penalty[11]. A Pareto optimal solution can be obtained because the contribution of each member of the team has a substantial impact on the distribution of output and even if only one agent shirks final output may fall below the critical level, in which case penalties will be applied. Hence the probability of being punished is high, even when only a single individual in the team decides to cheat. In this case agents have a significant impact on the probability of receiving a penalty and the principal can set the critical level of output at such a level that the probability of punishment is very high.

Large teams. In a large team, the contribution of each worker to joint production is less clear to identify. Each agent's shirking will have only a mild impact on final output, and hence the probability of receiving a punishment if the individual shirks will be low. In this case, the penalties, which would be needed to enforce a critical level of output, have to be very high, because the probability of punishment is very low. Substantial losses of efficiency can be incurred if agents have wealth constraints or are risk averse[12]. In large teams, with risk averse individuals, a first best solution is not attainable. In other words, withholding payment of the total output when the efficient level is not delivered is not enough to bring about an efficient outcome. In this case it is necessary also to monitor the performance of agents. This is because the available measure of output does not contain sufficient information about the production process and a system of rewards based on that measure would not solve the moral hazard problem.

Monitoring should be aimed at extracting valuable information about the production process. Holmström investigates a particular type of monitoring that provides valuable information, in the sense that it improves welfare (better risk sharing, for the same effort by agents and no higher cost for the principal). The type of monitoring proposed by Holmström is based on the idea of extracting the best information out of the imperfectly observed actions of individual team members. This is known as a sufficient statistic. The argument is that if the principal uses an incentive scheme based on a sufficient statistic,

[11] The penalty may very across individuals.
[12] If, instead, agents are risk neutral and the principal is endowed with unbounded wealth, the author shows that it is possible to approximate a first best solution by using a system of bonuses for output above a critical level.

no other incentive scheme will do better[13]. Therefore the monitoring required in large teams with risk-averse individuals should be aimed at getting a measure of performance which is simple but also most informative – i.e. a sufficient statistic. Holmström shows that if the principal rewards the team members on the basis of such a measure of performance, she'll be able to improve welfare.

In conclusion, for large teams with risk-averse individuals, a system of group penalties or bonuses cannot deliver efficient production and some form of monitoring, on top a system of bonuses or penalties, is required. This monitoring is designed to produce a signal of agent's actions that is sufficiently informative about their actions. The value of the information that can be extracted then becomes an important determinant of the choice of the optimal incentive mechanism.

2.1.3. Relative performance evaluation

Holmström (1982) also provides an application of the rather abstract concept of sufficient statistic by analysing the rationale for relative performance evaluation. Relative performance evaluation can be used to improve the overall design of individual incentives in situations where individual actions are observed imperfectly because of some common shocks influencing the measurements[14].

In his exposition Holmström assumes that the information available is rich enough that the contribution of each individual can be identified, but there is common uncertainty which links agents' contributions. When agents' contributions to final output are observable, but are affected by a common uncertainty, peer performance is a source of valuable information and the optimal system of rewards should be based on peer performance.

[13] Hölmstrom proves that, given any collection of incentive schemes, there exists a set of schemes based on a sufficient statistics that weakly Pareto dominates all others.
[14] For instance, contamination of a theatre might make all surgeons' operations appear less proficient than otherwise.

This is still a "team" issue in that agents whose contributions are related by common uncertainties form a team of sorts – in that it is not optimal to reward them separately. but this is a different type of team than the one above where agents provide imperfectly observed inputs. The preceding analysis assumed that agents' productive inputs were related because of the organisation of the production process. In that case we had a "natural" production function defining the team: the only way of producing was by pulling together the (unobservable) efforts of different individuals. In this second case the team is still formed by those agents whose inputs are related but the reason for the correlation is not the production process, but rather the need to filter away common uncertainty and improve risk sharing.

Holmström shows that if total output can be itemised according to the contribution of each individual and there is a common source of uncertainty, a compensation scheme based on peer performance is optimal. This is because peer performance can give information on the common uncertainty. He considers the case of an output structure where each team member's contribution to final output is separately observable and is a linear function of the agent's effort, a common uncertainty, faced by all team members (systematic risk) and an independent risk, which differs across individuals (idiosyncratic risk).

In this situation the optimal action would be to filter away the common uncertainty and reward each worker for his/her personal contribution. A means of filtering away common uncertainties is linking individual *j*'s compensation to all other agents' performance. Peer performance provides information about the common uncertainty (systematic risk). Holmström demonstrates that in this case an optimal solution is to use a sharing rule based on the agent's performance and a weighted average measure of all other agents' performance:

$$s_j = s_j\left(x_j, \sum_i \alpha_i x_i\right)$$

where s_j is agent j's reward, x_j is agent j's (observed) contribution to final output, and the second term in brackets is the weighted average of all the team members' contributions.

So the agent's reward is a function of his/her observable contribution to final output and a weighted average of the other team members' contributions. This is an optimal reward scheme, in that no other sharing rule can achieve a higher effort and/or be implemented at a lower cost for the principal. The weights α_i should be such that the output of an individual whose idiosyncratic risk is low should be counted more in that the output of this individual gives more information about the common uncertainty. (A low idiosyncratic risk means that the variance of the idiosyncratic risk is low and hence the precision of measurement of the idiosyncratic risk is high). The contribution to final output from agents with low idiosyncratic risk is more informative of the common uncertainty and hence should be assigned a greater weight for the evaluation of aggregate performance to be used in the sharing rule.

Note that the rationale for peer performance is to extract valuable information, rather than to induce competition among agents. Competition among agents is the result of the attempt to exploit the information that can be gained from peer performance. If agents' contributions were not correlated there would be no reason to use peer performance: the optimal scheme would consist of rewarding each agent for his/her own contribution. In this case competition among agents would be valueless. The implication of the analysis that is worth stressing for our purposes is that when there exists common uncertainty the optimal compensation scheme should link agent's rewards, even if individual contributions can be separately observed. A team in this case is identified by the common uncertainty faced by its members.

Note that under the scheme of Marino and Zábojník (2001) teams are defined in such a way as to factor out common shocks to performance measurements, there will also be a degree of relative performance evaluation between teams. Under such a scheme, there will be an advantage of defining teams so that each team is hit by similar shocks to its performance, rather than defining teams so that the shocks hit one team or another but not

all. Unfortunately, there will typically be other compelling reasons to define teams which may run counter to this one, for instance physical proximity, mutual monitoring, etc. These reasons are discussed further in section 3.

In context of Holmström (1982) there is no role for teamwork per se, so it is important to extend the analysis to cases where teamwork does have a positive role. Itoh (1991) addresses the question when optimal incentive schemes should be designed to explicitly facilitate co-operation. His model, which is discussed further below, shows that in reasonably realistic formulations contracts should either be individualistic or foster substantial degrees of co-operation. That is, it gives support to either having team based incentives paramount or not having them at all.

A major concern of introducing individual based incentives in an organisation such as the NHS is that is that doing so may "crowd-out" co-operative effort. Holmström and Milgrom (1991) first emphasises the crowding out aspects of incentives in a formal model. Although the paper does not itself emphasise the team aspects, the conclusions can be extended to team contexts which may be especially pertinent in the NHS context. The general message is that if co-operation within a team is important for the overall organisational objectives, then rewarding individual performance can detract from team performance by raising the marginal cost of effort in co-operating. It is important to note that similar arguments will also hold with respect to 'small' and 'large' teams, or with respect to teams and the NHS as a whole. Rewarding individuals on the basis of Team A performance will very likely make them less willing to contribute to Team B performance even when team A and team B are not directly in competition through relative performance evaluation. Indeed, team B may even suffer if it contains team A as a sub-team or indeed if team B is the NHS as a whole. To put it baldly, just because individuals within the NHS are rewarded on the basis of teamwork, does not mean automatically that teamwork will be promoted in the wider sense.

Similar remarks can be made with respect to implicit incentives, if certain aspect of individual performance - even if based through teams - is emphasised to the public and

profession, then those aspects are likely to become the focus of individual career concern motivation. Working hard to achieve these aspects will typically increase the marginal cost of effort for other aspects which may be equally if not more important to the NHS overall.

In conclusion, to define the optimal compensation for a team, we first need to understand the nature of the team to which the incentive is to be applied, as the optimal scheme depends on the type of team. It will depend on whether output can be perfectly measured and on whether team members' contributions can be separately observed. The informational content that can be extracted from the available measure of performance is crucial to deliver appropriate team based incentives. It will be important to understand which, if any other, activity individuals may substitute their efforts away from and the consequences of this for the NHS. Hence in order to implement of team rewards in the NHS we first need to examine what types of teams currently operate in the NHS organisations.

3. APPLYING THE THEORY TO TEAM INCENTIVES IN THE NHS

3.1. The complexity of defining teams in the NHS

A general criterion that defines a team is joint production. However, while joint production may be widespread in the NHS, it is difficult to identify with precision. There are cases when the patient care pathway is well defined. If there are precise protocols and guidelines describing the procedure to be taken, the identification of the team members could be quite straightforward, in that the production process is clearly outlined. But these cases are few and far between. Even where joint production can be defined, final output may be difficult to observe and the identification of the input of individuals to these output even more difficult, in that the uncertainty of effects of treatment and the

intrinsic difficulty[15] of treatment can be quite substantial. Further, individuals may belong to a number of groupings, all of which might be considered more or less teams.

One suggestion made in the NHS Plan is that a team could be defined around each patient's care. This would mean following the patient from the moment she is referred from her GP to the moment when she is discharged from an NHS organisation. This approach, however, has significant drawbacks. One feature of medical care is that the diagnosis of a patient is intrinsically difficult to perform. Due to the imperfect observability of symptoms, the diagnosis is subject to uncertainty[16]. The more difficult the diagnosis the more difficult the treatment and hence the more uncertain the outcome of the production process. Another feature of medical care is that in some cases there may be more than one treatment. So patients with the same condition can follow different care paths, either because they get a different diagnosis or a different treatment. This implies that if we define the team around the patient we will get a definition of a team that is quite ad hoc. It should also be noted that the organisational structure is very heterogeneous among NHS organisations, so that the definition of a team around the patient would differ even among organisations with a similar function but a different structure, like for example hospitals.

While this may not be a sensible team, team production clearly exists in the NHS. And some of these teams may be suitable structures for the purposes of team rewards. While we cannot, without detailed empirical analyses, identify exactly which teams should have financial rewards and which not, we can consider those features of production which make some teams more suitable for team rewards. To do this we first briefly outline the features of team production which have benefits and disbenefits respectively. Then we consider, of the broad types of teams that currently exist in the NHS, which appear to

[15] A treatment could be intrinsically difficult, as in the case of a surgical treatment, or it could be made more complicated if the diagnosis of the patient's condition is particularly difficult.

[16] The skills and experience of a doctor or specialist do affect the diagnosis, but even if we were able to control for those aspects, we wouldn't be able to eliminate the uncertainty. This is due to the fact that the doctor has to make the diagnosis on the basis of the symptoms revealed by the patient. Patients and doctors have completely different sets of information, i.e. medical knowledge. The patient could underestimate the seriousness of a symptom when describing his/her condition to the doctor. As a result the doctor could not take into consideration a certain test, which is instead crucial to get the right diagnosis.

have the strongest case for organisation as a team and which may therefore also be more suitable if financial rewards are to be used in a team context.

3.2. Defining teams on the basis of the production process and the interactions among the NHS staff

3.2.1. Types of teams which exist within the NHS

Joint production in the NHS can take a number of different forms. In some cases patient care requires the intervention of specialists from different disciplines with different backgrounds. In other cases, the treatment for particular conditions has to be delivered by different organisations. For some procedures, joint production requires senior and junior staff. Sometimes each member's contribution to joint output is observable by the other members of the team and sometimes it is not.

On the basis of the types of joint production found in health care, we can identify five broad types of team:

- teams across disciplines
- teams across organisations
- hierarchical teams within an organisation
- small teams
- large teams

We define hierarchical teams as teams formed by senior and junior staff. The distinction between large and small teams depends on the observability of each member's contribution to final output: in small teams it is possible to observe each member's contribution. In reality teams are not so rigidly defined: there might be, for example, teams across disciplines, with members belonging to different organisations and they can be large or small. For the moment we ignore this issue to keep the analysis as simple as possible.

3.2.2. The nature of interactions among team members

Teamwork may have positive features. Those we may find in teams in the NHS include some or all of the following:

- diffusion of information and learning from shared job experience
- mutual monitoring of staff
- group cohesion
- risk pooling
- sharing of common resources
- division of labour

Negative features include:

- free riding
- exposure to greater risk because team members have different abilities or because of correlated output
- conflicts between professional values and team members priorities

Examples of these in NHS settings are given in Appendix 3.

Some of the positive externalities are negatively correlated with each other: for example, more complete diffusion of information is negatively correlated with division of labour. The more specialised the task to be performed, the less job experience can be shared. If team members specialise in particular tasks they cannot benefit from the use of informal ways of diffusion of information. Division of labour is also an obstacle for mutual monitoring: when the tasks of the team members do not overlap, it becomes more difficult to monitor each other. Division of labour also undermines group cohesion: the more team members specialise in particular tasks, the less they feel part of a team.

Some positive externalities may compensate for other negative externalities. For example, mutual monitoring compensates for different innate abilities, in that it may help identify the weakest member of the team. It could also make the free-rider problem less severe and decrease the uncertainty of the production process. More complete diffusion of information is another aspect that contributes to reduce the uncertainty of effects of treatment and the intrinsic difficulty of the diagnosis. Group cohesion could overcome the free-riding problem.

3.3. The likely importance of externalities in different types of teams

In Table 1 we consider the strength of these externalities in each of the type of teams that we identified above. The relevance of the externalities varies across the different types of teams: an externality may be more important in one type of team and less important or not important at all in others. In the table, a blank space means that the externality is not important for that typology of team, a cross means that it is important and two crosses distinguish the cases where the externality is particularly important.

We can first read the table by rows. This allows us to analyse how externalities vary across the different types of teams.

3.3.1. How externalities vary across types of team

More complete diffusion of information through communication and learning from shared job experience.

This positive externality is particularly important for teams across disciplines and small teams. In the former, uncertainties are particularly relevant when tasks to be accomplished require the intervention of specialists from different disciplines. The diffusion of information is particularly important in this case. If those specialists are organised in a team, the diffusion of information is easier, in that it can occur through formal and informal communication. In the latter teams the diffusion of information is easier and there are both formal and informal communication processes.

In teams across organisations, team members are located in different places. Hence the process of communication is less easy and while the diffusion of information is still important it plays less of a role. In hierarchical teams the diffusion of information is essential but there may be obstacles. The contribution of junior staff is limited by the fact that they are learning, whereas senior members may have an interest not to fully disclose information (for example to keep power over junior staff).

Mutual monitoring.

Mutual monitoring is important in teams across disciplines, in small teams and in hierarchical team. In teams across disciplines, monitoring may be limited by the different backgrounds of team members, though the ability to monitor will depend on the nature of tasks: for routine tasks or simple tasks, monitoring could be relatively easy. In small teams it is easy to observe the actions of others even if individuals have different competencies. In hierarchical teams, team members have the same background, hence it is easier to monitor than in teams across disciplines, though monitoring is likely to be only one-way as junior members of staff can monitor senior members only to a limited extent.

Group cohesion

Group cohesion is an important aspect of small teams where it is easier to build team spirit. In hierarchical teams, group cohesion exists but is probably weaker. Although team members have the same competencies the different status between senior and junior could make it more difficult to volunteer co-operation.

Risk pooling

Risk pooling is a relevant aspect of production in large teams and those that cross organisations. Working in a team could lower the costs imposed by external shocks. An example is A&E units of different hospitals working in contact with one another in case of major accidents: patients can be quickly moved to the hospital with available beds so that each hospital does not need to keep so many spare beds. Another example of risk

pooling across organisations is the Commissioning Consortia created by Primary Care Groups to pool the risk of high cost/low incidence treatments (e.g. spinal injuries).

Economies of scale

Economies of scale can be significant in teams across organisations (for example, sharing of high-tech equipment across organisations). In large teams, high fixed costs could be shared over a large number of units/staff. However, economies of scale could be present even if the team is small. An example is a GP practice where the administrative cost of having staff at the reception is shared among several doctors and nurses.

Division of labour

Division of labour is particularly important in teams across disciplines, where people have different competencies and it is possible to differentiate their tasks. In large teams, division of labour could also be an important aspect of production. In hierarchical teams, division of labour could be relevant in that senior members may delegate tasks to junior staff.

Free-riding

The free-riding problem is always an issue in teams. It is probably most serious in large teams. It is possibly mitigated in teams across disciplines and across organisations where each member's contribution is rather unique and therefore their contribution to joint output is more easily distinguishable. Small and hierarchical teams can benefit from mutual monitoring so mitigating the free-rider-problem.

Different innate abilities

Different innate abilities are particularly difficult to spot when people are from different disciplines and different backgrounds, so we can expect that in teams across disciplines and across organisations and in large teams this negative externality is particularly important. In small and hierarchical teams it is easier to spot low abilities because it is easier to observe one's peers, or because team members have the same background.

Correlated output

In large team and those across organisations, it is more difficult to communicate and monitor and this negative externality can be a big problem. In small teams and hierarchical teams mutual monitoring may mitigate the uncertainty of correlated output. In teams across disciplines the presence of different specialists with different backgrounds may increase the risk of getting the diagnosis or the treatment wrong, but the positive externality of more complete diffusion of information which may be found in this type of team could partly compensate for this.

Conflicts between professional values and team members' priorities

Conflicts may arise in many team settings. In teams across disciplines conflicts may arise in that people have different competencies and each one may feel "the" expert[17]. In teams across organisations, people may feel less part of a team, because they are located in different organisations and may be more attached to the organisation to which they belong, rather than identifying with their team. Different organisations may have different priorities or procedures, so that conflicts may be possible. In large teams the decision process in large teams is more costly and conflicts may arise. Finally, in hierarchical teams conflicts may arise because senior and junior members may have different priorities[18].

3.3.2. Which types of NHS team exhibit more benefits from team production?

We can read the table by columns to establish how the positive and negative externalities interact. In so doing we can assess which types of team are more likely to deliver better production. Note that a strong case for team production does not necessarily mean a strong case for financial rewards, as such teams may be motivated by factors other than financial. On the other hand, if the case for team production is not strong then there is

[17] In some cases language may raise co-ordination problems due to the different epistemology used by the different disciplines This may be the case when, for example, psychologists work together with clinician, or public health specialists or social workers collaborate with clinicians.

only a weak case for financial rewards of such a team. We examine financial rewards (and the issue of implicit rewards) in Section 4. Here we focus on the types of team which are best defined in terms of production.

Teams across disciplines

When the delivery of patient care requires the joint work of specialists across disciplines, if these specialists are organised in a team they can benefit from informal channels of communication and take advantage of better diffusion of information about the diagnosis and the treatment of the patient. Being in the same team can allow observation of the effort, or the outcome, of other teams members' actions and so allow detection of possible mistakes or careless behaviour. These positive externalities tend to offset the free-rider problem and the other two negative externalities due to different innate abilities and correlated output. In fact, the uncertainty of the production process due to different innate abilities and correlated output may become less relevant if team members are able to monitor each other and communicate easily.

In some cases working in a team across disciplines can avoid unnecessary duplication of effort by allowing division of labour. This is particularly useful for tasks that are less influenced by uncertainty and for which mutual monitoring is not important (e.g. routine or simple tasks). However, in this case the diffusion of information, mutual monitoring and group cohesion will be less effective.

On the negative side, conflicts between professional values and team members' priorities may be quite substantial in a team across disciplines. This undermines group cohesion and makes it more difficult to build team spirit. However, in teams across disciplines there might be less direct competition, in that the different professionals are not directly competing with each other and hence we could expect more tolerance and respect.

[18] For example, senior and junior staff may disagree on the importance given to research versus clinical work. Or junior staff may have career concerns which clash with the desires of their seniors.

Teams across organisations

In teams defined across organisations the organisational structure can be quite complicated as, by definition, team members are located in different organisations. This is an obstacle for mutual monitoring, in that it is not possible to observe peers' effort or actions[19], and also for the diffusion of information. In many cases formal procedures of communication do not work and they have to be substituted with informal channels of communication[20]. The free-rider problem and the uncertainty due to different innate abilities and correlated output are more difficult to tackle than in teams across disciplines. In addition, possible conflicts between professional values and team member priorities may emerge and this makes it more difficult for team members to identify themselves as part of a team. However, it might still be worth promoting team production across organisations if risk pooling and/or economies of scale are substantial.

Large teams

Large teams appear to be problematic for team production. Negative externalities are substantial and difficult to offset. In some cases, however, different innate abilities, correlated output, and conflicts between professionals' values and team members' priorities may not be relevant and the production process could be still organised in large teams. This is more likely where the tasks are routine (for example, in some of the service functions in hospitals e.g. portering, administration, maintenance, the provision of ambulance services).

Small teams

In a small team individuals are better able to observe their peers and share information. Mutual monitoring and more complete diffusion of information contribute to decrease the uncertainty which characterise the production process and make the free-riding problem

[19] In some cases it could be possible to observe the outcome from an organisation. It is the case, for example, of a patient leaving an Acute Hospital to go to a Community Hospital: if the patient comes to the Community Hospital with an infection or malnourished, the outcome of the Acute Hospital is observable, and hence mutual monitoring can be, to some extent, effective.

[20] For example, for the cure of infectious diseases, which involves professionals based in many organisations, the complexity of the organisational structure requires the adoption of informal procedures of communication such as professional networks.

less severe. Sharing job experience can help to tackle the intrinsic difficulty of diagnosis and treatment and, to some extent, to control for different innate abilities.

Moreover, team members' contributions are more easily distinguishable, and hence each individual may be more aware of her role in joint production and recognise the importance of teamwork. As a result conflicts between team member priorities may be less difficult to manage and group cohesion may be stronger.

Hierarchical teams

Hierarchical teams benefit from mutual monitoring and more complete diffusion of information. If uncertainty in the delivery of health care is an important aspect for patient care, teamwork between senior and junior staff may reduce this uncertainty and so lead to a better outcome. The positive externalities of a team tend to offset the negative aspects, although there may be conflicts of interest among the team members arising from unequal status and career concerns.

3.3.3. When to promote teamwork

This analysis helps us to distinguish situations when there is a strong case for teamwork from situations where teamwork is less advisable. Given the pattern of crosses in Table 1, teamwork is likely to be particularly advantageous when organised in small teams, hierarchical teams and teams across disciplines within a single organisation. Those types of teams seem to perform better in delivering health care: the positive externalities, created by team members' interactions tend to outweigh the negative externalities. In particular, small teams defined across disciplines, where senior staff work together with junior staff, create substantial positive externalities, which are able to offset the negative externalities. Good grounds for teamwork also exist for small and hierarchical teams and small teams across organisations.

Team production has less in its favour in large teams which cross organisations and large teams across disciplines. Here team members' interactions create strong negative impacts

on the outcome of the production process. More generally, for large and cross organisation teams, a more careful analysis of the tasks to be accomplished and the organisational structure of the team is required before deciding if there is a case for teamwork.

Given this, there appears to be a case for team based rewards in teams *within* organisations, where the teams are based on those that already exist. Examples of such teams could include the 'firms' that operate within specialties, a ward or group of wards, possibly clinical specialties within hospital trusts, primary care groups and possibly whole primary care Trusts. Types of teams which do not appear suitable are those based on patient care pathways, or teams formed by a number of organisations which are also large – for example, a team defined as all health and social care personnel in one geographic area.

But a case for teamwork is not necessarily a case for financial rewards. Two examples may illustrate this. First, take the case of teams defined across disciplines. If we accept the argument advanced in Section 3 that conflicts between team members' priorities have a greater impact on team cohesion than the absence of direct competition, we should probably give priority to financial incentives. Investing in trust or sympathy among team members could be quite difficult and so explicit incentives may be more effective in teams across disciplines. So in this case, there may be a weaker case for teamwork than in a team within disciplines, but when we have such teams, financial rewards may be important. Second, take the case of teams across organisations. Such a team may be desirable if small, but in this case getting the organisation of the production process correct (e.g. setting up shared databases or shared facilities) may be more important than increasing effort through a system of financial rewards (or other implicit incentives).

4. SHOULD WE USE FINANCIAL REWARDS?

The NHS Plan, and other analyses of the public sector by policy makers (Makinson 2000), are clearly of the opinion that the NHS is under-incentivised. Our analysis above

has identified teams that may be suitable for team based rewards, but as we noted, a case for teams is not necessarily a case for financial incentives. In this section we survey a number of issues raised in the economics literature that are relevant to the use of explicit incentives. These are the role of implicit rewards, the applicability of the types of penalties suggested by Holmström to the NHS context, measuring and rewarding performance and dynamic issues in the use of financial rewards for teams.

4.1. The role of implicit rewards

At present, the teams that exist in the NHS are not rewarded financially as teams. But this does not necessarily mean that individuals working in teams within the NHS do not get rewards from team production: they may get implicit rewards. Individuals working in a team may develop a team motivation and tend to free-ride less than what would be predicted from a model which focuses only on explicit incentives. Motivation to act cooperatively may rely, for example, on peer pressure[21], trust between managers and employees[22], job independence[23]. The policy response is to create a working environment in which members of the team will behave in the interest of the team i.e. to reinforce or develop the strength of implicit incentives. Which policy response is appropriate depends which type of implicit incentive is most likely to be in operation.

For NHS organisations, peer pressure may be particularly relevant. Kandel and Lazear analyse ways in which peer pressure may induce team members not to shirk and which type of policy may enhance this[24]. They identify three forms of peer pressure: an additional source of utility from not shirking, mutual monitoring and/or group norms. In the case of the first form, it is important to distinguish between internal and external peer

[21] As we shall briefly consider, this argument is developed in Kandel-Lazear (1992) and Encinosa-Gaynor-Rebitzer (1997).
[22] See Auriol-Peschlivanos-Friebel (1999), Bruce-Waldman (1990), Chami-Fullenkamp (1999), La Porta et al. (1997) and Lorenz (1999).
[23] Mitusch (2000)
[24] Note that, in general peer pressure can be effective if one individual's effort affects the well being of the rest of the team (otherwise the rest of the team does not have any incentive to exert pressure on her). This

pressure. Internal peer pressure is present when an individual gets disutility from cheating, even if others cannot identify the offender. In this case the worker feels guilty if she shirks. It requires substantial past investment in team loyalty, but it is ideal when individuals work on their own, so that monitoring would be quite expensive, but total output depends on the team. An example of a strategy that can be used by a firm to instil guilt is bringing the worker's family into the organisation through childcare centres, organisations for spouses, etc., so that if the worker shirks she imposes a cost on her family. In this way the externality from shirking can be internalised. In general, investments in loyalty are most important in environments where it is not possible to observe the worker's effort and there is complementarity in production. Both forms of pressure require empathy, in that the individual must care for the team. Team members are able to affect each other's choices only if they empathise with the group. Therefore it becomes very important to invest in activities that create empathy among workers (e.g. team building exercises).

In case of mutual monitoring it is assumed that workers are able to monitor each other at a cost and workers who are caught shirking are penalised by means of a non-pecuniary penalty, such as mental harassment. A worker will be willing to engage in peer monitoring because she believes that other workers will increase their effort as a response. If there is some form of profit sharing it is in the interest of an individual to monitor her peers, in that a higher effort by other individuals will affect her own reward from effort. Mutual monitoring tends to be more effective in small teams and if workers' tasks overlap. An increase in size means more people to do the monitoring and so the total amount of monitoring increases. However, monitoring loses effectiveness as size increases[25].

---

requires some form of profit sharing, in that the choice of a worker's effort affects her peers. The team members must also have the ability to affect each other's choice.

[25] The existence of mutual monitoring changes the inverse relationship between firm size and individual effort that is the free-rider effect. With no peer monitoring, as the firm size increases, the free-rider problem becomes more serious. With peer monitoring it is possible to observe an increase in individual effort if firm size increases. If the number of workers increases, more workers will engage in peer monitoring and peer pressure will exert a greater impact on individuals' behaviour, in that sanctions imposed by the group will be greater. However, there is a limit to the increase in the firm size if incentives are to be effective: adding more workers, after a certain point, may make relationships more impersonal or sanctions more difficult to enforce.

Where peer pressure is active in the form of group norms, there will be an established group norm (for example, a certain level of effort) such that, if individuals deviate from that norm, they are punished and suffer a loss in utility. Group norms have to be established and are thus more likely to occur in teams where individuals have repeated interactions with each other, or where there are strong professional codes of practice.

It is clear that such peer pressure exists in the NHS, more strongly in some cases than others. More generally, implicit incentives are clearly important. The unanswered question is whether they are sufficient, or are explicit financial rewards needed as well? The NHS Plan clearly is of the view that such implicit incentives are not sufficient. Without empirical evidence we can draw no firm conclusions. However, from our analyses above, we can raise some issues in the use of financial incentives in teams. Some of these arise from the applicability of Holmström to the NHS context. Some arise from the tools currently at the disposal of those trying to measure performance. And some arise from the possible responses of individuals to the introduction of team rewards.

4.2. The applicability of the types of penalties suggested by Holmström in a not-for-profit context

Holmström suggests certain reward structures when teams are 'natural teams', in terms of agents requiring the input of others to complete their own tasks. In some cases there are NHS teams in this sense. For example, if there is a well defined care path way, involving professionals with different skills or from different specialties, and the care path is formal in that it is governed by protocols or guidelines, those protocols and guidelines will be the equivalent of a natural production function and will define the team. A 'firm' within a specialty is also clearly a natural team, as are individuals who work together on a ward.

It is reasonable to assume that professionals involved in those teams provide imperfectly observed inputs: what we can observe is the final output, i.e. the condition of the patient

after receiving treatments. Given the intrinsic difficulty of measuring patients' condition, we can assume that output in those teams is only imperfectly measurable.

The implications from Holmström for defining the optimal compensation scheme when agent's productive inputs are not observable presented is helpful in these contexts. In Chart 1, the results of Holmström's analysis that are relevant for our purposes are identified by boxes outlined with dotted lines. We can rule out the case of risk neutral agents in the NHS. Teams in the NHS could be both small, in that the contribution of each member has a substantial impact in determining final output, and large. The Holmström analysis indicates in the case of small teams, a system of group penalities could be useful, while in the case of large teams, monitoring is important to achieve optimal output.

In other cases, teams may not be natural teams, in that agents do not require the inputs of others to do their task, but they may be subject to common uncertainty. An example is a sudden increase in demand that can affect the work of individuals belonging to different organisations. Another example comes from the intrinsic difficulty of the diagnosis of patients. In this case there is a systematic risk for the staff involved in delivering health care to the same patient: irrespective of the abilities of the individuals, if someone gets it wrong the whole process of delivering patient care is compromised. In these cases, individuals' contributions are related by a common uncertainty and the Holmström analysis suggests the need for relative performance evaluation.

So from Holmström we can take the fact that the value of information about team members' contributions and the production process should play an important role in the design of an optimal compensation scheme in NHS teams. This will mean the need for measurement and comparison of agents' actions. But there is an issue as to whether the Holmström analysis can be applied to the case where there may be no residual claimant and the principal is more than the residual claimant.

We can distinguish three main differences between the NHS context and Holmström's setting. First, Holmström considers private for-profit partnerships, while teams in the NHS operate in a not-for-profit organisation where output is not in monetary units. Second, in the NHS the identification of the targets to be achieved and the assessment of performance are made by the DoH and the Regional Offices. These institutions carry on a regulatory function for the NHS organisations, so that they indirectly determine the output of those they regulate. Third, in the Holmström analysis, the funds to be allocated for team rewards come from the revenues from final output i.e. they are internally generated by the team, whereas in the NHS case rewards for team production come from extra funds to be distributed by the government, i.e. they are externally generated.

The first two points imply that the principal (the DoH/Regional Offices) may be involved in generating net revenues. Hence, the condition that the principal should not be involved in the production process is not satisfied and so the figure of the residual claimant – needed to give group penalties or bonuses - cannot be identified as the principal.

The third point may limit the applicability of the system of rewards suggested by Holmström. In the NHS, members of a team do not have surpluses to distribute among themselves gained from joint production: instead it is the government that provides the funds. The system of incentives consists of extra rewards for teams that achieve targets. How these extra rewards are managed is an essential feature of the incentive mechanism and able to affect its efficacy. Courty and Marshke (2001) have shown that while team rewards can be given in such a setting, the need for the government agency (e.g. the DoH) to balance its budget on an annual basis limits the power of such team rewards.

There may, in the case of relatively small team bonuses, be a way round this. It is not made clear in Chapter 6 of the NHS Plan how the annual funds to promote team production will be managed. If it is the case that when an organisation does not meet the requirements to get the extra funds those extra funds will be allocated to other organisations, the system of incentives as suggested by Holmström becomes feasible. In this case the residual claimant would be those organisations who meet the requirements

for extra rewards. This solution would also overcome the problem of not being able to identify the principal as the residual claimant (as discussed above)[26]. On the other hand, if the monies earmarked for team bonuses are not allocated but instead are saved for the following year, there is no residual claimant and the system of incentives may not be effective. This is because if the organisation thinks there is some chance of getting the funds next year they will not put effort into meeting targets this year. If the funds are given to other organisations then there is no chance of an organisation that does not put in effort of getting them next year.

4.3. Measuring and rewarding performance

As emphasised by Holmström, while teams may exist, the decision whether to reward the team by use of financial incentives should rely on the characteristics of the technology and the information on the uncertainty of the production process that can be extracted from the available measures of performance. The production of health care is subject to substantial sources of risk: uncertainty of the diagnosis, uncertainty of the effects of treatment, the availability of different treatments. If we do not have an appropriate measure of performance of the team and use a system of team bonuses or penalties based on that measure, we run the risk of penalising team members for events that are not under their control.

Measures of performance have been developed in the NHS over a number of years. However, none of them have yet been designed with team rewards in mind. This suggests that a direction for future analysis is to define a measure of team performance and establish to what extent it can be informative of each agent's contribution and of the underlying uncertainty of the production process. The measure of performance is related to the technology of production, so this exercise should identify what sort of technology

[26] The idea of using individual teams to act as each other 'budget breakers' has been analysed by Marino-Zábojník (2001). This paper suggests the use of inter-team tournament to solve the free-rider problem in large teams. Under the assumption that the output in a large organisation can be decomposed in two teams output, they demonstrate that if a firm can organise a tournament between these two teams and transfer output from the team with inferior aggregate performance to the team with superior aggregate performance,

leads to a good measure of team performance. If good measures are available then financial incentives in teams are more likely to be useful.

The design of financial incentive in teams in the NHS is made more difficult by another aspects of teamwork in the NHS: overlapping teams. The same individual may be member of different teams. Even if the identification of the different teams the individual belongs to may be straightforward, it could be the case that tasks performed in one team are more easily measured than others. It could also be the case that there are available good measures of performance for the different tasks but some of them are inversely correlated. Financial incentives in the first case might encourage individuals to devote effort to the team activities that are measured. More generally, if teams conflict, introducing financial incentives is less desirable and is another issue to be further investigated. The issue of overlapping teams has parallels with the issues of multi-tasking for agents. Theoretical analysis is provided in Dixit (2000) who notes that these issues weaken the case for the use of high powered incentives.

Where both team output and individual outputs are both imperfectly measured, we have seen that the classical analysis stresses the role of relative performance evaluation. Roughly speaking, for any given worker, the worse every one else performs the better. Teamwork is hardly likely to be promoted by relative performance evaluation therefore.

More positively, the issue of measuring team performance is closely related to the use of benchmarking as a way of improving performance in the NHS. There is currently considerable effort being put into the development of performance criteria that can be used to benchmark NHS organisations. In a review of the use of financial rewards for benchmark performance Grout *et al* (2000) argue that there are strong grounds for linking benchmarked performance to financial rewards. Without explicit incentives, there are dangers that organisations that are above the bottom of the distribution but below the top will take little action to improve their performance. But the difficulty of measuring output

---

a first best solution can be achieved. One effect of such a mechanism is that the individual teams serve as each other's budget breakers.

and the fact that agents undertake several tasks limits the extent of high-powered incentives. Grout *et al* (2000) also argue that any measure of performance that is initially adopted is likely to be refined over time, and that this should not deter the Department of Health from introducing measures. The experience from regulation of utilities suggests that waiting for an ideal system is pointless: any measure of will have to be refined over time as agents react to the new incentives embodied in the performance measure and the rewards associated with good and bad performance.

On the downside, the use of team rewards as envisaged in the current NHS Plan could lead to collusion between organisations. The Performance Fund will be introduced by the Government. This means the reward of team performance is not managed by the same authorities responsible for the monitoring of performance (which is to be shared between the Department of Health and CHI). The result is a decentralised supervision structure with no direct control on the funds to be used for promoting better performance. This may encourage collusion among assessed organisations. As already observed, targets to be promoted by use of incentives will be set on the basis of "…*levels of service already being achieved by the best trusts around the country.*" NHS organisations could collude and conform to a standard of service in order to lower the targets and hence increase the likelihood of getting extra rewards. With a decentralised supervision, those collusive activities would be difficult to be detected. And given that the monitoring authorities are not responsible for managing the funds, they would not have any direct incentive to prevent them. This problem could be aggravated when funds are allocated in a discretionary way, as it seems to be the case for the rewards for accomplishing particular tasks: the Plan only mentions the availability of extra funds to promote particular targets, without setting any precise criteria on how to allocate them.

On the other hand, a flexible approach for managing the funds to promote better performance is to be welcomed, in that a flexible structure for financial incentives allows recognition of actual outstanding performance and hence guarantees the success of

incentive schemes. The Makinson report[27] on improving services in the public sector emphasises the fact that if the money to devote to performance improvements is agreed in advance and there are rigid funding constraints it is not possible to reward outstanding performance and penalise unpredicted failures, and incentives are less effective[28].

Regardless of whether collusion emerges, there is a problem with the system envisaged in the NHS Plan - the lack of a unique and independent authority charged with both the measurement and assessment of performance and the management of the funds to be awarded. Before team based financial incentives are introduced this issue needs to be addressed.

4.4. Dynamic considerations

The introduction of financial rewards for team members will invoke responses by agents. These may include increasing effort, but are also likely to lead to other actions, some of which are less desirable[29]. Financial rewards for team performance might results in changes in team membership. This is particularly likely when the production function is endogenous, in that it is the compensation mechanism that defines the team rather than the production process. For example, individuals could pretend to be team members in order to get an extra compensation. Teams could start competing with each other. For the NHS this could have serious consequences for patient care: rivalry could emerge within and among organisations or teams could try to steal or dump patients on other teams. This would be a particular problem where there are common resources across teams.

Individuals might also try to change membership of teams. We might expect a process converging to equilibrium where high-ability workers tend to be matched together. This would lead to disparities in the productivity of different teams. On the other hand, more positively, if changing teams is limited or prohibited the team members may invest in

[27] John Makinson, "Incentives for change. Rewarding performance in national government networks. " Public Service Productivity Panel, 2000.
[28] For example, in terms of bonuses to be awarded for reaching certain targets, non-consolidated bonuses are to be preferred to consolidated bonuses.

activities that are beneficial for the team. It may also happen that, if the wrong team is designed, team members can react and change the production process.

When a team is rewarded financially, the nature of production may change. One example of this is how much individuals are prepared to help each other. As discussed above, Itoh (1991) analyses the relationship between financial incentives and 'helping' effort. An agent can choose between two types of effort: own effort, which improves the outcome of the task for which the agent is mainly responsible, and helping effort, which improves the outcome of the other agent's task. Itoh analyses whether it is always the case that, in moving from an individual based contract (i.e. one where individuals are paid only for their own output) towards one where rewards are based on teamwork, agents are induced to increase the level of helping effort.

He finds that strategic interactions among agents and agents' attitudes towards performing multiple tasks determine agents' response to a change in the contract. In particular, agents can be induced to provide help, even for a small change in the wage schedule, if they get positive benefit from both types of effort. If, instead, tasks are similar and agents only care about the total amount of effort, they are reluctant to provide even a small amount of help. In this case a large perturbation of the individual based-contract is required to induce any helping effort from the individual.

But even if co-operation can be induced relatively easily through financial rewards, it might not be in the interest of the principal to promote it. This is because there may be undesirable effects of an increase in help by one agent on the efforts of other agents. This happens when own effort and helping effort are not complementary and so an agent receiving help may respond by decreasing his own effort.

What we want to emphasise here is that the incentive effects of introducing interdependence among workers by rewarding team performance may not necessarily be positive. Interactions among team members and their preferences for own and helping

[29] Examples of responses to cost benchmarks are discussed in Grout et al (2000).

effort have to be considered, otherwise rewarding individuals for helping each other may be counter productive. More generally, there remain unresolved issues over how explicit rewards interact with implicit rewards.

5. CONCLUSIONS

The NHS Plan welcomes the use of team rewards but neither specifies how team based incentives are to be implemented nor makes it clear what types of teams such incentives are to be given to. This paper has examined relevant economic theory for insights into optimal incentive schemes within teams. We have focused on the seminal paper on explicit incentives in teams (Holmström 1982). The implications from this analysis are that the optimal compensation scheme will depends on the type of team. Teams that are small can be treated differently from those that are large. In the former, a system of group rewards and penalties is sufficient to obtain efficient levels of effort, while in the latter monitoring of performance will also be required. The aim of such monitoring is to derive measures of the actions of individuals. This will also be necessary in cases where individuals do not work in natural teams (where their actions are complementary) but can be treated as a team because they are subject to common uncertainty.

Hence the definition of teams in the NHS should precede the identification of the system of team rewards. The paper therefore seeks to identify which the attributes of production associated with a greater number of benefits from team production. Our analysis suggests teams that are small and hierarchical are likely to be more effective than those that are large and cross-organisation. Measuring the actions of single individuals, required for monitoring purposes for optimal team rewards when teams are large and/or not defined by a natural production function, may also be easier when teams are not too large. Teams defined at the level of a large organisation, or formed across organisations, are likely to be less effective. And, in contrast with some of the discussion in the NHS Plan, we do not favour the teams rewards where teams are defined by the care received by the patient.

The Holmström analysis also indicates that, in many cases, effective financial rewards will require more than a system of group bonuses and penalties at the team level. Monitoring of individual's actions will be required. Some of this could be of a comparative nature (the idea being similar to that used under the yardstick competition used in the utility industries). This may require comparison within and between teams.

While we can define those teams in which team working is more likely to be beneficial, there remains a set of complex issues in the use of financial incentives. These include appropriate balance between financial and non-financial rewards, the design of financial rewards to motivate team members, and the dynamic responses to rewarding of teams with financial incentives. Many of these are unresolved issues in the theoretical literature, and have not been systematically examined in the empirical literature either. Hence their importance cannot be quantified. Given the desire of the current administration to introduce teams in the NHS, what is now needed is empirical testing of team based rewards. Such tests should be established in a manner that allows inferences to be drawn as to whether and in what cases team based rewards can increase desired actions. This means testing team-based rewards against no team-based rewards, and testing them within different environments. The lessons from the theoretical literature is that the outcome will differ according to the nature of production, the ability of the principle to extract information on the actions of individuals within teams and the importance of implicit rewards for existing team members.

6. REFERENCES

L. Aiken et al. “Hospital organisation and outcomes”, Quality in Health Care, 7, 1998, 222-226.

E. Auriol – G. Friebel- L. Pechlivanos “Teamwork management in an era of diminishing commitment”, CEPR Working Paper 2281, 1999.

N. Bruce- M. Waldman , “The rotten kid theory meets the Samaritan’s Dilemma”, Quarterly Journal of Economics, 105, 1990, 155-165.

R. Chami- C. Fullenkamp “Trust and Efficiency”, MIMEO,1999 (subject to permission).

Department of Health “The NHS Plan”, Command Paper Cm 4818-I, July 2000. London:HMSO.

Department of Health “ The NHS Performance Assessment Framework”, April 1999.

Department of Health “NHS Performance Indicators: a consultation”, May 2001.

A. Dixit “Incentives and organisations in the public sector: an interpretative review” MIMEO, 2000.

W. Encinosa III- M. Gaynor- J. Rebitzer “The sociology of groups and the economics of incentives: theory and evidence on compensation systems. NBER Working Paper 5953, 1997.

M. Freeman et al. “The impact of individual philosophies of teamwork on multi-professional practice and the implications for education”, Journal of Interprofessional Care, 14, (3), 2000.

P. Grout, Jenkins, A and Propper, C, Benchmarking and Incentives in the NHS, London: Office of Health Economics 2000.

B. Holmström “Moral hazard in teams”, Bell Journal of Economics, 13, 1982, 324-340.

B. Holmström and P. Milgrom “Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership and Job Design” Journal of Law, Economics and Organization, 7, 24-52.

H. Itoh “Incentives to help in multi-agent situations”, Econometrica, 59, (3),1991, 611-636.

E. Kandel- E. Lazear “Peer Pressure and Partnerships”, Journal of Political Economy, 100 (4), 1992, 801-817.

R. La Porta et al. “Trust in large organisations”, American Economic Review, 87 (2), 1997, 333-338.

E. Lorenz, “Trust, contract and economic cooperation”, Cambridge Journal of Economics, 23, 1999, 301-315.

J. Makinson “Incentives for change. Rewarding performance in national government networks”, Public Service Productivity Panel, 2000.

A. Marino and J. Zábojník “ Profit Centers and Incentives in Teams”, MIMEO, 2001.

K. Mitusch “Job independence as an incentive device”, Economica, 67, 2000, 245-263.

C. Phelps “Diffusion of information in medical care”, Journal of Economic Perspectives, 6, (3), 1992, 23-42.

H. Varian “Monitoring agents with other agents”, Journal of Institutional and Theoretical Economics, 146, 1990, 153-174.

P. Wilcock and L. Headrick “Interprofessional learning for the improvement of health care: why bother?”, Journal of Interprofessional Care, 14, (2), 2000.

7. APPENDIX 1

The Performance Indicators identified in the PAF consist of 43 *High Level* indicators *(HLPI)*, available for each Health Authority, and 6 **Clinical** Outcomes indicators, available for each Health Authority and Trusts grouped by trust clusters **CI (**Small/medium acute, Large acute, Very large acute, Acute specialist, Acute teaching, Multi-service, Specialised community, Priority single service).

Area	Performance Indicator
Health Improvement	*(i)* *Deaths from all causes (ages 15-64)* *(ii)* *Deaths from all causes (ages 65-74)* *(iii)* *Deaths from cancer* *(iv)* *Deaths from circulatory diseases* *(v)* *Suicide rates* *(vi)* *Deaths from accidents* (vii) *Serious injury accident*
Fair Access	*(i)* *Inpatient waiting list* *(ii)* *Adult dental registration* *(iii)* *Early detection of cancer* *(iv)* *Cancer waiting times* *(v)* *Number of GPs* *(vi)* *GP practice availability* *(vii)* *Elective surgery rates* (viii) *Surgery rates – Coronary heart diseases*
Effective Delivery of Appropriate Healthcare	*(i)* *Childhood immunizations* *(ii)* *Inappropriately used surgery* *(iii)* *Acute care management* *(iv)* *Chronic care management* *(v)* *Mental health in primary care* (vi) *Cost effective prescribing* **(vii) CI 5 Rates of discharge to usual place of residence within 56 days of emergency admission from there with a stroke, for patients aged 50 and over** **(viii) CI 6 Rates of discharge to usual place of residence within 28 days of emergency admission from there with a hip fracture, for patients aged 65 and over**
Efficiency	*(i)* *Day case rate* *(ii)* *Length of Stay* *(iii)* *Maternity unit costs* *(iv)* *Mental health unit costs*

	(v)	*Generic prescribing*
Patient-Carer Experience	*(i)*	*Patients who wait less than 2 hours for emergency admission (through A&E)*
	(ii)	*Cancelled operations*
	(iii)	*Delayed discharges*
	(iv)	*First outpatient appointments for which patient did not attend*
	(v)	*Outpatient seen within 13 weeks of GP referral*
	(vi)	*Number of those on waiting list 18 months or more*
	(vii)	*Patients satisfaction*
Health Outcomes of NHS Care	*(i)*	*Conceptions below age 18*
	(ii)	*Decayed, missing or filled teeth in five year old children*
	(iii)	*Readmission to hospital following discharge*
	(iv)	*Emergency admission of older people*
	(v)	*Emergency psychiatric re-admission*
	(vi)	*Stillbirths and infant deaths*
	(vii)	*Breast cancer survival*
	(viii)	*Cervical cancer survival*
	(ix)	*Lung cancer survival*
	(x)	*Colon cancer survival*
	(xi)	**CI 1a Rates of deaths in hospital within 30 days of surgery (emergency admissions)**
	(xii)	**CI 1b Rates of deaths in hospital within 30 days of surgery (non-emergency admissions)**
	(xiii)	**CI 3 Rates of deaths in hospital within 30 days of emergency admission with a heart attack (myocardial infarction) for patients aged 35-74**
	(xiv)	**CI 2 Rates of deaths in hospital within 30 days of emergency admission with a hip fracture for patients aged 65 and over**

Source: The NHS Performance Assessment Framework, Department of Health, March 1999.

APPENDIX 2

According to Holmström, when joint output is fully shared among the team members, i.e. the budget balancing rule is adopted, team members tend to free-ride on their contribution to final output.
Holmström formalises this argument as follows.
A budget-balancing rule is defined as:

$$\sum s_i(x) = x. \qquad (1)$$

$s_i(x)$ is agent i's share of the outcome x, which is assumed to be perfectly observable, as in the case of a monetary outcome. According to this rule, total output should be fully distributed among workers, regardless of the level of effort they exert.
Joint output x is represented as a function of agents' actions:

$$x = x(a_1,...,a_n) \qquad (2)$$

A representative individual i exerts a level of effort a_i , which is not observable, and incurs a cost $v_i(a_i)$. His/her payoff is:

$$s_i(x(a)) - v_i(a_i) \qquad (3)$$

Individual i will choose the level of effort that guarantees the maximum payoff:

$$\frac{\delta s_i}{\delta x}\frac{\delta x}{\delta a_i} - \frac{\delta v_i}{\delta a_i} = 0 \quad i = 1,...,n \qquad (4)$$

The level of effort satisfying (4) will be Pareto optimal, if the marginal revenue of an extra unit of effort equals the marginal cost of that unit. This is the condition for efficiency in production:

$$\frac{\delta x}{\delta a_i} - \frac{\delta v_i}{\delta a_i} = 0 \quad i = 1,...,n \qquad (5)$$

The level of effort chosen by individual i *is Pareto optimal if conditions (4) and (5) are simultaneously satisfied. This requires:*

$$\frac{\delta s_i}{\delta x} = 1 \qquad (6)$$

But with a balancing sharing scheme, (1), we can only have

$$\sum \frac{\delta s_i}{\delta x} = 1 \qquad (7)$$

Hence the level of effort chosen by individual i *that maximises his/her payoff, as in equation (4), will be lower than the Pareto efficient level, resulting from equation (5).*

APPENDIX 3

Examples of advantageous and disadvantages features of production in a team in the NHS.

Positive externalities and example of these in the NHS include:

More complete diffusion of information and learning from shared job experience

The importance of communication amongst health professionals is widely recognised. Lack of communication can lead to poor outcomes (e.g. L. Aiken et al 1998). The difficulty of acquiring information can determine variation in medical care use (e.g. Phelps 1992). Interprofessional learning is important for the improvement of healthcare and the importance of exchange of skills and knowledge in teamwork (e.g. Freeman et al 2000, Wilcock and Headrick 2000).

More generally, the idea is that teamwork may facilitate a process of communication and sharing of job experience. This is a positive externality in that it allows for better risk management within a team. The delivery of health care is subject to different sources of risks: the intrinsic difficulty of the diagnosis and of the treatment, the uncertainty of effects of treatment and the availability of more than one treatment. If the production process is organised in a team, the team members can communicate with each other, formally and informally, and learn from directly observing the others' job experience. Hence they gain more information, which they can use to identify the diagnosis and the treatment more accurately, reducing thereby the uncertainty underlying the production process.

Mutual monitoring

An argument in favour of team production is that team members are better able to observe the others' effort and this could help in enforcing proper levels of effort (Varian 1990). As already noted, in the NHS the free-rider problem is mitigated by ethical

commitment to patient care. Even so, there is a role for mutual monitoring: for example, it could lead to team members spotting mistakes. As an example, in teams dealing with complex procedures, e.g. surgical teams, the team members can directly observe their peers and may intervene during the production process in order to prevent mistakes.

Sometimes it is not possible to observe the other team members' effort in that the production process requires the intervention of different specialists at different times, so that team members are not always in the same place. Teamwork could still create a positive externality in that team members could be able to observe final output and infer mistakes or inattention by their peers. This could motivate the team members to be more careful and responsible. In surgical teams, for example, a physician cannot observe a surgeon in that he's not in the operating theatre at the time the operation is carried out, but can observe the outcome of the surgical operation.

Group cohesion

One result of organising the production process in a team could be that agents taking part in joint production become more cohesive. This might improve motivation.

Risk pooling

Teamwork may create a positive externality in case of external shocks, in that team members are able to share the negative impact of adverse events**.** An example of an external shock could be a sudden epidemics, such as a flu epidemic. Teamwork across organisations could alleviate the problems of sudden increase in the demand for particular services, such as hospital beds.

Economies of scale

In some cases the organisation of production in teams is motivated by the presence of economies of scale. Organisations could, for example, share very expensive high-tech equipment, which otherwise would be not affordable for a single organisation. In this case joint production consists of making a technology available for more than one organisation. Or it could be the case that the production process involves high fixed costs

and teamwork could allow to increase output at a decreasing average cost. An example could be GP practices, where doctors working in the same surgery can save administration costs. Economies of scale are related to the underlying technology rather than to the type of team.

Division of labour

When the tasks to be accomplished by team members are simple and mutual monitoring and communication are not so relevant, division of labour within the team could be a positive externality that avoids unnecessary duplication of effort.

Negative externalities could include some or all of the following:

Free-riding

When team members' contributions to final output cannot be easily distinguished, the free-rider problem may become significant. We can assume that in the NHS organisations ethical constraints and mutual monitoring, when this latter is feasible, mitigate against this negative externality.

The impact of different innate abilities

Given different innate abilities, each member in a team is exposed to more risk than in an individualistic workplace. This is a source of risk for the team.

Correlated output

When the delivery of patient care requires the intervention of different specialists, working in teams, final output is subject to greater risk. Due to the intrinsic difficulty of the diagnosis and/or of administering treatment (for given abilities of team members), if somebody gets it wrong, the whole process of delivering patient care is compromised.

Conflicts between professional values and team members' priorities

When members of a team have unequal status and/or power, the working environment could be quite difficult and conflicts may emerge.

Chart 1: How to define the optimal compensation scheme when agent's inputs are not observable.

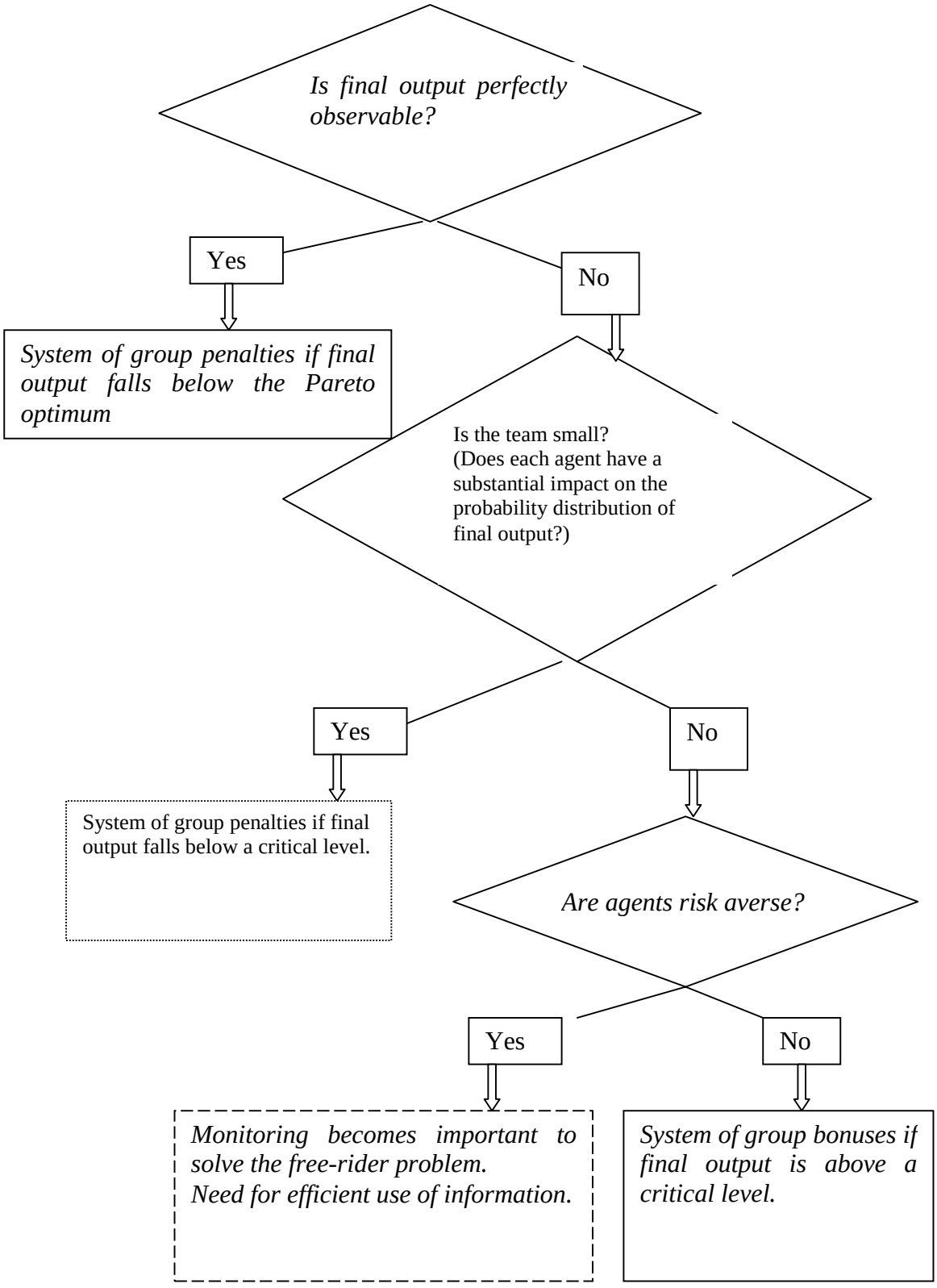

Table 1: Externalities in teams

	Typology of teams				
	Across disciplines	Across organisations	Large Teams	Small teams	Hierarchical teams (senior+junior staff)
Positive externalities					
More complete diffusion of information and learning from shared job experience	xx	X		xx	x
Mutual monitoring	x			xx	xx
Group cohesion				xx	x
Risk Pooling		X	xx		
Economies of scale		X	x	x	
Division of labour	xx		x		x
Negative externalities					
Free-riding	x	X	xx	x	x
Different innate abilities	xx	Xx	xx	x	x
Correlated output	x	Xx	xx	x	x
Conflicts between professional values and team member priorities	xx	Xx	x		xx