

## How to do (or not to do) . . .

---

# Evaluating payment mechanisms: how can we measure unnecessary care?

XINGZHU LIU<sup>1</sup> AND ANNE MILLS<sup>2</sup>

<sup>1</sup>*School of Public Health, Shandong Medical University, China, and* <sup>2</sup>*Health Economics and Financing Programme, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, UK*

There has been substantial concern that linking hospital or physician remuneration to the amount of care provided will encourage excessive provision of health care. Studies that seek to explore this relationship require methods to measure unnecessary care. This paper presents and discusses a method used to assess the magnitude of unnecessary care in the context of an evaluation of the impact of paying bonuses to physicians employed in public hospitals in Shandong Province, China.

### Introduction

In recent years, the choice of payment method for health care providers (hospitals and physicians) has attracted increased attention. There is concern that some forms of payment, particularly fee-for-service, encourage the provision of health care that is ‘unnecessary’. While there is a literature that addresses this question particularly in the US and Europe, it has been little studied in low- and middle-income countries, where one might expect a greater magnitude of unnecessary care because of weaker regulatory mechanisms for controlling provider behaviour, and a generally less well-informed patient population. There are also practical difficulties in doing such studies in these contexts, due to less well-developed information systems in hospitals.

Unnecessary care is generally defined as care provided but medically not needed. More precisely, care is considered unnecessary if on balance it does not benefit patients (Leape 1989). In operational terms, care is defined as unnecessary when, for an average group of patients presenting to an average physician, the expected health benefit of care provided does not exceed the expected negative consequences by a sufficiently wide margin, excluding consideration of monetary costs (Hopkins 1993). Unnecessary care includes unnecessary outpatient visits, inpatient admissions, days in hospital, diagnostic and treatment procedures, and drugs.

Economists have become interested in the issue of unnecessary care because of their interest in financial incentives and the relationship between the quantity of care given and the system used to pay providers. If the quantity of care given contributes to determining the income of providers, there is a risk that doctors and hospitals will be motivated to provide more care than is strictly necessary to achieve health benefit.

This will then result in an increase in health care costs, waste of resources, potentially damaging effects on patients due to inappropriate care, and reduction in the efficiency of medical care from a social perspective. Research efforts have focused on the relationship between the economic incentive that is built into provider payment systems and the degree of over-provision (Liu and Mills 1999), and on the effectiveness of different policies and approaches to controlling over-provision (Foreman 1990). One of the key tasks in this type of research is to measure the existence and magnitude of unnecessary care.

This short paper presents a case-study of the use of a method for measuring unnecessary care. Following the introductory section, it presents a detailed description of the approach used in a study in China, and then illustrates the sort of information that can be obtained by presenting the results which show the magnitude of unnecessary care. Final sections discuss the appropriate uses and limitations of this method, and draw conclusions on its value and relevance elsewhere.

To set the method used in context, the purpose of the study was to evaluate the effect of paying cash bonuses to hospital doctors. In order to improve the productivity and revenue generation of Chinese hospitals, a policy was introduced in the late 1970s whereby doctors were paid bonuses on top of their basic government salary. Over time, the bonus system developed from a flat rate bonus to all staff, to bonuses that were related to the quantity of services provided and the amount of revenue generated. As a result of this and other changes, Chinese public hospitals now generate the great bulk of their income from fees. The study’s concern was county hospitals (of around 250 beds) where the great majority of patients pay out of pocket for health care and are not covered by insurance. One of the main objectives of the study

was to explore the impact of bonus payment on the nature of health care provided: hence the need to measure unnecessary care. Both cross-sectional (comparing hospitals using different bonus systems) and longitudinal (studying how hospital performance changed when bonus systems changed) analyses were done, requiring data to be obtained over a period which encompassed the main changes in bonus system (around ten years).

### Methods for measuring unnecessary care

The two main approaches to the measurement of unnecessary care are the simple 'implicit criteria' approach and the more sophisticated 'explicit criteria' approach. In the former, physicians review the entire patient record and make a summary judgement of whether the care provided is appropriate or necessary. In the review process, neither the information used in making the judgement nor the process for weighting the various aspects of the information are spelt out by the researchers (Morehead 1967). The validity of the method therefore depends entirely on the knowledge, skill, and judgement of the reviewers. Although simple and inexpensive, the method of implicit criteria may produce biases due to the difference in judgement of reviewers, as shown by low agreement rates across reviewers (Payne 1987).

In the explicit criteria method, a set of criteria for judging what care is unnecessary has to be developed either for each type of service (Gertman and Restuccia 1981) or for each specific diagnosis (Chassin et al. 1987). The major advantages of this method are that it is likely to have higher validity than the implicit criteria approach and that the patient record review can be performed by non-physicians (e.g. nurses). However, it lacks feasibility due to its high cost and the difficulties in developing the sets of criteria for all types of services or all types of diagnosis, and in capturing all necessary criterion indicators.

In the study in China, an intermediate method was used which was considered more feasible than the explicit criteria method and more reliable than the implicit criteria method. As with these methods, the first step was to decide on 'tracer' diseases, since unnecessary care can only be judged in relation to the nature of the cases being treated. The tracer diseases were appendicitis and child pneumonia, which were selected according to the following criteria:

- (1) They had to be common diseases so that there were enough cases for each hospital and for each year.
- (2) They had to have clear-cut diagnoses so that the sample would be homogeneous.
- (3) There had to be a common plan of treatment so that variation in treatment would not be due to treatment uncertainty.

To increase the homogeneity of the sample, the target appendicitis patients were defined as those with: acute condition; no perforation; no major accompanying diseases; and age ranging from 10 to 70 years. The target pneumonia patients were defined as those with: simple condition (e.g. no heart failure or other major diagnoses); age ranging from 0 to 7 years.

Two groups of doctors were recruited to do the record reviews, consisting of six surgeons (who were responsible for appendicitis cases) and six paediatricians (who were responsible for pneumonia cases). These doctors were taking post-graduate courses on research methods in Shandong Medical University and had experience of clinical work in county hospitals. They were judged to have the necessary knowledge of both good medical practice and the reality of medical care in county hospitals. The two groups of doctors worked together with the investigators to develop guidelines for appropriate management of the two tracer diseases and for identification of unnecessary care. These guidelines were positive lists of types and quantity of services and drugs that were thought by the doctors to be necessary for the improvement of the health outcomes of patients. Inpatient records were collected from six county general hospitals in Shandong Province and covered roughly a ten-year period, though the precise period depended on the availability of records. The inpatient records for each year, for each disease and in each hospital were drawn from the beginning of the year until the sample size reached 30 or until there were no more records available. Altogether, the study included 1161 appendicitis patients and 1142 pneumonia patients.

The patient files of each tracer disease and each year were randomly distributed to the relevant doctors who reviewed the inpatient files and recorded the types and quantities of services and drugs actually used, and the types and quantities of unnecessary services and drugs. If the removal of unnecessary items rendered care inadequate, substitute necessary services and drugs were added in by the panel doctors. Finally, to identify the financial magnitude of unnecessary care, actual expenditures and unnecessary expenditures were computed by accountants according to the 1997 provincial fee schedule. Expenditures and unnecessary expenditures included the total and its breakdown in terms of drugs, tests and examinations, surgical services, and routine services (including hotel services, regular nursing and physician services) which would be reduced proportionally with a reduction in length of stay. The 1997 fee schedule was used to control for the changes in medical prices over time.

Because of the burden of work involved, each record, randomly allocated, was reviewed by only one doctor. However, to check for consistency between doctors, 61 patient records for appendicitis and 57 for pneumonia were selected and randomly distributed to the panel doctors for re-reviews, without their knowledge. The means of the values recorded by different doctors for the same appendicitis and pneumonia patients were compared.

### Results

A substantial amount of unnecessary expenditure was identified by panel doctors (Table 1). The average expenditure for an appendicitis patient was 774 yuan (US\$95), of which 143 yuan was expenditure for services and drugs that were thought by doctors to be unnecessary, or 18.47%. The average expenditure for a pneumonia patient was 559 yuan (US\$68), of which 107 yuan (19.14%) was judged unnecessary. Although the monetary values were meaningless in

**Table 1.** The extent of unnecessary expenditures for appendicitis and pneumonia patients (all expenditure in 1997 yuan)

Variables	Appendicitis (n = 1161)		Pneumonia (n = 1142)	
	Mean	Std Dev	Mean	Std Dev
<b>Length of stay:</b>				
Length of stay (days)	8.68	3.30	8.81	4.30
Unnecessary hospital stay (days)	1.37	1.90	0.90	1.74
<b>Unnecessary expenditures:</b>				
for examinations	1.86	7.62	1.07	8.35
for lab tests	2.38	6.78	1.02	4.56
for drugs	69.78	190.12	77.77	197.42
for hotel service	6.86	9.50	4.52	8.68
for doctor and nurse services	61.71	85.51	22.59	43.39
Total unnecessary expenditure	142.58	216.87	106.98	202.64
<b>Expenditures:</b>				
for examinations	3.61	10.87	16.23	23.95
for lab tests	14.55	13.78	11.53	11.92
for drugs	181.27	213.66	227.29	317.82
for hotel services	43.42	16.49	44.05	21.50
for doctor and nurse services	460.78	148.37	260.27	107.51
for operation	70.00	0.00	–	–
Total expenditure	773.63	282.96	559.38	368.48
<b>Itemized unnecessary expenditure as % of itemized expenditure (%):</b>				
for examinations	51.52	–	6.59	–
for lab tests	16.36	–	8.85	–
for drugs	38.49	–	34.21	–
for hotel services	15.80	–	10.26	–
for doctor and nurse services	13.39	–	8.68	–
<b>Itemized unnecessary expenditure as % of total unnecessary expenditure (%):</b>				
for examinations	1.30	–	1.00	–
for lab tests	1.67	–	0.95	–
for drugs	48.94	–	72.70	–
for hotel services	4.81	–	4.23	–
for doctor and nurse services	43.28	–	21.12	–

terms of actual expenditure because the cases covered a long time period, the proportion of unnecessary expenditure indicated that the extent of the problem must be substantial.

Further analyses showed that more than one-third of the expenditure for drugs was judged by panel doctors, on the basis of the guidelines, to be unnecessary for both appendicitis (38.49%) and pneumonia (34.21%) patients. Unnecessary expenditure for drugs made up a dominant proportion of total unnecessary expenditure, and as a percentage of total unnecessary expenditure for appendicitis patients was 48.94%, and for pneumonia patients 72.70%.

Unnecessary expenditure for doctors' and nurses' services, which was almost proportional to unnecessary days in hospital, made up the second largest proportion of unnecessary expenditure. It can be seen from the table that although it accounted for only 13.39% (for appendicitis) and 8.68% (for pneumonia) of the expenditure for doctors' and nurses' services, it made up 43 and 21% of the total unnecessary expenditure for appendicitis and pneumonia patients respectively.

Although more than 50% of the examination expenditure for appendicitis patients was unnecessary, it made up only about

1.3% of total unnecessary expenditure since little was spent on examinations. Unnecessary expenditure for laboratory tests was small in terms of both its percentage of expenditure for tests (6.59%) and its share of total unnecessary expenditure (1%).

The results suggest that for the given level of health outcome at the time of patient discharge, the period of hospital stay could be cut by 16% (1.37/8.68) for appendicitis patients and 10% (0.9/8.81) for pneumonia patients; the total expenditures for appendicitis patients and pneumonia patients could be reduced by 18 and 19% respectively; and efforts related to the reduction of unnecessary days in hospital as a measure to decrease doctors' and nurses' unnecessary services are important. Considerable savings could be generated by cutting unnecessary drug prescriptions, because they dominated unnecessary expenditure.

## Discussion

The measurement of unnecessary care is important for: the estimation of the magnitude of over-provision and the level of social efficiency of health care resource use; the study of the effects of changing incentives (such as changes in the

payment mechanism for providers) on provider behaviour; the analysis of the determinants of over-provision; and the development of policies to reduce over-provision.

Because of the problems associated with the implicit criteria method, and the unacceptably high cost of the explicit criteria method, a combination approach was designed which proved feasible. It is termed a combination approach because the guidelines developed were based on normal medical practice and doctors' opinions, in contrast to the explicit criteria approach where criteria are developed through the ratings of panel doctors based on the results of a thorough literature review of the scientific evidence on the effectiveness of care. In addition, the guidelines used in this study were much simpler than those used in the explicit criteria approach.

The re-analysis of a small sample of records showed that there was considerable consistency between doctors in terms of their judgements of unnecessary care: there were no statistically significant differences between any of the pairs of the means of the variables in the first and the second reviews. However, the value of the method depends on how it is applied. Firstly, the establishment of the treatment guidelines is of key importance. It should be made clear to the panel doctors who make the guidelines that unnecessary care refers to those services, drugs and other medical goods that can be removed without reduction in patient health outcomes. Without this clarification, doctors may tend to develop guidelines based on usual and customary practices that may not be medically necessary, and the level of unnecessary care would be under-estimated. Secondly, after the identification of unnecessary services and drugs, substitutes are often required: unnecessary expenditure should be the expenditure for unnecessary care minus the expenditure for substitutes. Thus, in the process of record review doctors should be required not only to identify unnecessary care but also to prescribe substitutes when necessary. For example, if CT examination for a pneumonia case was judged unnecessary, the doctor might recommend a chest X-ray as a substitute. Failure to do this will result in over-estimation of unnecessary expenditure. Thirdly, even though doctors are required to follow uniform guidelines in the process of record review, inconsistency in judgement is unavoidable, firstly because the guidelines may not include all possible variations in disease diagnosis and treatment, and secondly because individual judgements are sometimes necessary. To deal with this problem, two measures can be taken, as done in this study. One is to distribute patient records to each of the panel doctors randomly and evenly; and another is to require panel doctors to discuss and make a collective decision if the guidelines do not allow for the specific circumstances of the case.

Interpretation of results needs to be carefully done. This point is illustrated by reference to particular features of this study. Firstly, the concept of unnecessary expenditure in this paper was unnecessary expenditure for medical care payers, not cost to providers, since unnecessary care was valued using the fee schedule. The level of unnecessary cost depends on the cost/price ratio. If the unit cost of care is on average higher than its price (as in China), the additional social cost due to unnecessary care would be greater than the estimated

unnecessary expenditure. If the unit cost of care is lower than its price (as would be the case for for-profit hospitals), the actual social cost of unnecessary care would be less than the estimated unnecessary expenditure, because the difference between the unit cost and price would reflect the transfer of money from payers to providers, not the consumption of resources. The actual savings if unnecessary care could be decreased would depend on the extent to which hospital capacity could be reduced or diverted to other, more valuable uses.

Secondly, the interpretation of results depends on the scope and representativeness of the selected tracer conditions. In this study, the magnitude of unnecessary provision of hospital care was under-estimated, because the two tracer diseases allowed for only a limited degree of over-provision. For example, neither encouraged the use of high technology diagnostic services such as a CT scanner. However, hospitals in China have been able to charge for such services at higher-than-cost levels, thus greatly encouraging their use and making it likely that a substantial amount of unnecessary care results from this. Moreover, it was deliberately decided in this study not to seek to judge whether or not the admissions were necessary, and also whether surgical operation was necessary in the case of appendicitis. Panel doctors believed that a predominant proportion, if not all, of children with pneumonia should be hospitalized, and that it was a clinically correct option to operate for acute appendicitis, although conservative treatment was also possible. This is another reason for the under-estimation of unnecessary care in this study. In order to measure the overall magnitude of over-provision, the tracer sample should be representative of all hospital patients (including diseases that allow for different degrees of over-provision), and the scope of care should cover both inpatient and outpatient care.

The development of treatment guidelines can take both positive and negative approaches. The former refers to specifying what services are necessary, and the latter to what are unnecessary. In this study, the positive approach was used and found to be feasible. Further research could explore the negative approach, and whether the two different approaches yield similar results.

## Conclusions

This study has demonstrated an approach to measuring unnecessary care that worked well in Shandong Province, China. The method involved selecting tracer diseases, using a panel of doctors who subsequently did the record reviews to establish treatment guidelines, identifying unnecessary care from patient records, prescribing necessary substitutes, and calculating unnecessary expenditures based on the fee schedule. It is recommended that to increase the accuracy of measurement, treatment guidelines should be carefully set and agreed based on the best knowledge of doctors and taking into account the minimization of cost for the required outcome; necessary substitutes need to be prescribed after identification of unnecessary care; and patient records should be distributed randomly and evenly among panel doctors. If the study is done to evaluate payment methods, the types of tracer diseases selected need particularly careful judgement.

The relevance elsewhere of the methods used in this study requires careful consideration. Studies based on record review commonly face problems, even in developed countries, because records often do not provide a complete description of the care provided to the patient. A further problem in low- and middle-income countries is that record storage is often poor. In the case of China, the recording of information on the record was extraordinarily full, and writing very clear: perhaps a legacy of a very controlled society. The quality of the medical record requires detailed assessment before this method is used in other settings. Obtaining old records was a problem in China, though was more of an issue in this study, given the longitudinal component, than it may be in others where only recent information is required.

Public hospitals in low-income countries rarely have computerized medical records or even a computerized patient information system that allows patients with a particular diagnosis to be quickly identified. Tracer disease studies thus usually involve the painstaking search for relevant cases through shelves and shelves of patient records. In the case of this study, four fieldworkers worked for two full days for each of the six hospitals to collect the patient records. Substantial persistence, as well as person power, is needed to conduct studies of this type especially where large sample sizes are required for statistical analysis.

Finally, it is worth commenting that the identification of policy implications from such a study needs to be carefully done. This paper has focused on the method, and has not discussed the issue of relating the findings on unnecessary care to the bonus payment system, which will be addressed in other papers. In practice, the pattern of health care provided will be influenced by a variety of factors, and some aspects may be more amenable to change than others. Thus it is important that any study of unnecessary care should seek to explore also the determinants of the patterns observed, and which determinants may be most readily influenced.

## References

- Chassin MR, Koseoff J, Solomon DH, Brook RH. 1987. Does inappropriate use explain geographic variations in the use of health care services? *Journal of the American Medical Association* **258**: 2533–7.
- Forman J. 1999. Practice parameters designed to reduce unnecessary medical care. *Archives of Ophthalmology* June, **108**(6): 789.
- Gertman P, Restuccia J. 1981. The appropriateness evaluation protocol: a technique for assessing unnecessary days of hospital care. *Medical Care* **19**(8): 855–71.
- Hopkins A. 1993. What do we mean by appropriate health care? *Quality in Health Care* 1993(2): 415–22.
- Leape L. 1989. Unnecessary surgery. *Health Service Research* **24**(3): 351–407.
- Liu X, Mills A. 1999. Supplier-induced demand: is it measurable? Paper presented to International Conference on Health Economics, Taiwan, March 1999.
- Morehead MA. 1967. The medical audit as an operational tool. *American Journal of Public Health* **57**: 1643–56.
- Payne SM. 1987. Identifying and managing inappropriate hospital utilisation: a policy synthesis. *Health Service Research* **22**(5): 709–69.

## Acknowledgements

This study was funded by support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (for PhD training), and from the UK Department for International Development (for fieldwork).

## Biographies

Xingzhu Liu is a medical doctor and professor of social medicine in Shandong Medical University School of Public Health (China) and has been doing his PhD in Health Economics at the London School of Hygiene and Tropical Medicine (UK) over the last three years. Now he is working at the Global Programme on Evidence for Policy in the World Health Organization.

Anne Mills is Professor of Health Economics and Policy at the London School of Hygiene and Tropical Medicine. She has over 20 years' experience in health-economics related research in less developed countries, and has published widely. She founded, and is Head of, the Health Economics and Financing Programme which receives a programme grant from the UK Department for International Development.

*Correspondence:* Dr Xingzhu Liu, Organising Health Systems, the Global Programme on Evidence for Policy, the World Health Organisation, 1211 Geneva 27, Switzerland.