

An Examination of the UK Treasury's Evidence Base for Cost and Time Overrun Data in UK Value-for-Money Policy and Appraisal

Allyson M. Pollock, David Price and Stewart Player

UK government procurement policy rests on Treasury claims that the private finance initiative (PFI) has reduced cost and time overruns. We review the five studies cited by the Treasury in support of this claim and find that only one purports to compare PFI with traditional procurement. The results of this single study are uninterpretable because of selection bias, small sample size (only 11 out of 451 PFI projects are included) and fundamental flaws in the analysis. There is thus no evidence to support the Treasury cost and time overrun claims of improved efficiency in PFI. We conclude that Treasury appraisal guidance, the 'Green Book' which compares PFI with other methods of procurement, is not evidence based but biased to favour PFI.

UK government procurement policy rests on Treasury claims that the private finance initiative (PFI) has reduced both the frequency and the magnitude of cost and time overruns. According to the Treasury's 2003 policy statement, *PFI: Meeting the Investment Challenge*:

PFI projects are being delivered on time and on budget. HM Treasury research into completed PFI projects showed 88% coming in on time or early, and with no cost overruns on construction borne by the public sector. Previous research has shown that 70% of non-PFI projects were delivered late and 73% ran over budget.

These data have been used by the government in response to criticisms of the policy, to inform the Treasury's guidance on PFI appraisal, and to support the whole of government public-private partnership (PPP) policy both in the UK and abroad. The strategy has been successful. Five years after their original publication the data continue to form the basis of newspaper articles (for example Timmins, 2006; Bentley, 2006).

Crucially, the data supports claims that, although private finance is more expensive than public finance, its use in public procurement saves money by introducing procurement efficiencies that are not available in the public sector. The claims are apparently based on hard evidence in line with a Treasury

statement (2003, p. 43) that the policy is by now sufficiently mature to allow proper assessment:

The government is determined to ensure that a sound evidence base informs the rigorous investigation of where PFI is delivering better facilities and value for money benefits in practice. The PFI programme has progressed to a point where, with 451 projects operational, sufficient evidence is available to assess many aspects of the early performance of the programme.

This article reports on the results of an evaluation of the evidence base which underpins the data used to revise guidance on appraisal and evaluation in central government (the Treasury's Green Book). We show that the Treasury claims about the benefits of the PFI in relation to comparative performance on time and cost overruns are not supported by their data and, as a result, Treasury guidance is biased in favour of the policy.

The Role of Cost and Time Overrun Data in PFI Appraisal

Public service investment in the past was financed through government grants or public loans funded either directly from taxation or from government borrowing. Since 1991 an increasing proportion of investment in the UK has been financed through private finance. By December 2004, Treasury figures show that

Allyson Pollock is head of the Centre for International Public Health Policy at the University of Edinburgh.

David Price is senior lecturer in the Centre for International Public Health Policy, University of Edinburgh.

Stewart Player is a research associate in the Centre for International Public Health Policy, University of Edinburgh.

Table 1. Recommended adjustment ranges for use by project appraisers for the public sector comparator during PFI appraisals.

Project type	Works duration		Capital expenditure	
	Optimism bias* (%)		Optimism bias* (%)	
	Upper end of range	Lower end of range	Upper end of range	Lower end of range
Standard buildings	4	1	24	2
Non-standard buildings	39	2	51	4
Standard civil engineering	20	1	44	3
Non-standard civil engineering	25	3	66	6
Equipment/development	54	10	200	10
Outsourcing	N/A**	N/A**	41	0

Source: HM Treasury, Supplementary Green Book Guidance (2003).

*Optimism bias is defined by the Treasury as the 'demonstrated, systematic, tendency for project appraisers to be overly optimistic' about the chances of schemes going over budget or being delivered late (HM Treasury, 2003, p. 1).

**'N/A' is not defined by the Treasury.

677 PFI projects had been signed with a capital value of £42.7 billion. Under PFI, the public service provider goes to a consortium of bankers, builders and service operators which raises finance on the government's behalf. The choice of finance route is determined according to the rules set out in the Green Book and the 'appraisal user guide', which require that public bodies compare the costs of public versus private financing methods. This determination is based on a formal appraisal process where the costs of the respective financing routes are compared. The comparison provides the basis for an assessment of which financing method, public or private, offers the best value for money.

A revised Treasury Green Book, published in 2003, requires that all estimates of construction costs in non-PFI schemes are inflated to take account of 'optimism bias': that is, the 'demonstrated, systematic, tendency for project appraisers to be overly optimistic about risks' of schemes going over budget or being delivered late. The Green Book guidance, which includes estimates of optimism bias in PFI and conventional procurement, requires cost estimates for standard buildings procured under the conventional route (i.e. the 'public sector comparator') to be increased by between 2% and 24% of the original construction cost estimates, and between 1% and 4% of the original estimate of works' duration. These revised estimates of the cost of the publicly financed option are then compared with the PFI cost estimates.

The revised estimates of public finance costs are higher than the original estimates and usually result in the decision to use PFI. According to the Treasury, the upper-range

percentages in table 1 are based on an estimate of the extent to which, on average, capital costs and construction periods have in the past been underestimated in conventional procurement. The data in this table underpin the decision to use private finance and are thus crucial to government policy.

Evaluating the Evidence Base

The UK Treasury cites five research studies as its authority for its claims about time and cost overruns in conventional and PFI procurement. These are a Treasury internal research project conducted in September 2002; two NAO reports, *Modernizing Construction* (2001) and *PFI Construction Performance* (2003); and two studies by the private sector, *Agile Construction Initiative: Benchmarking Stage Two Study* (1999) and the Mott MacDonald Report: *Review of Large Public Procurement in the UK* (June 2002). The Mott MacDonald report is cited as authority for the optimism bias adjustments listed in the revised Green Book (see table 1).

The reliability of empirical studies depends very largely on how they have been conducted. The ideal study design to evaluate the influence of alternative procurement methods on cost and time overrun risks would be a randomized control trial where new projects were randomly (and 'blindly') allocated to each procurement method. The experimental method can, of course, rarely be used in connection with public policy, but its principles provide the yardstick for evaluating any ex post analysis. We need to judge how far scientific principles which randomized trials embody have been appropriately applied, to the extent that they can be. The main sources of error which have

to be guarded against are as follows:

Comparing Like with Like

Bias can arise from overlooking significant differences between the subjects in the comparison, which could account for observed differences in their behaviour (in the present case, in the cost and time overruns), independently of the hypothesized determinant of such differences (in this case, the procurement method used). Such significant differences may be of many different kinds.

The types of scheme need to be comparable: equal proportions of standard and non-standard projects must be included in each sample, since non-standard projects are more complex and more liable to overruns.

The time periods in which projects were undertaken also need to be controlled for. Public procurement has been the subject of government scrutiny for at least a decade and numerous reforms have been implemented to make the process more efficient. Following publication of the 1994 *Constructing the Team* (Commissioned by the Secretary of State for the Environment) (Latham, 1994), the government reformed procurement regulations and undertook a series of further reviews and evaluations. The UK Office of Government Commerce (OGC), created in 1999, was itself the product of a procurement review (Gershon, 1999). Sixty one guidance notes (known as CUP or Central Unit of Procurement guidance) were transferred from the Treasury to the newly established OGC. About a third of the 61 notes were withdrawn and most of the remainder were subsequently superseded by new guidance. Only four of the original UK Treasury guidance notes are still current.

According to the National Audit Office (NAO), these extensive reforms have led to improvements in all methods of government procurement since 1999. Following a review of performance in 142 projects completed between April 2003 and December 2004, the NAO reported greater cost certainty and fewer delays in both PFI and conventional procurement, compared with results obtained in 2001 (NAO, 2005).

This means that if one of the two samples compared has a higher proportion of projects that pre-date the 1999 reforms, it will probably show higher time and cost overruns because projects in the sample did not benefit from the later improvements in procedure.

Representativeness

A further source of error is the failure to ensure that the sampled cases are representative of the population from which they are drawn. If, for example, public procurement were to be used more for non-standard projects, or for smaller projects, then both the public procurement and the PFI samples should include a corresponding proportion of non-standard and smaller cases if general conclusions about the relative merits of public procurement are to be drawn.

Sample Size

Sample size is also important. The representativeness of a sample depends partly on its size. There is no single formula for determining when a sample is large enough to be representative of the whole population. Factors that have to be taken into account include the amount of variation that exists in the population and the confidence required in the survey results. Statistical tests that show whether or not survey results are significant cannot be carried out when samples are too small. In such cases, no conclusions can be drawn from differences between the two samples.

Measurement Bias

Measurement bias occurs when different baselines are used to compare the two groups. Cost and time overrun data will be subject to error if insufficient account is taken of the procurement process in establishing the baseline or in establishing and adhering to a standard definition of costs.

The procurement process takes several years and consists of several stages which are set out in the OGC's Gateway Review guidance. This requires a series of business cases to be produced as negotiations take place. Costs are reported at three business case stages of the procurement process:

- The strategic outline case (SOC).
- The outline business case (OBC).
- The full business case (FBC).

In comparing cost escalation, it is essential that cost changes are measured from the same baseline or business case stage.

Tables 2 and 3 show that significant cost escalation occurs between SOB and OBC stages and between OBC and FBC stages, and therefore that the choice of baseline is crucial to the performance measurement. Table 2 shows that in five schemes reported to the Health

Table 2. PFI capital costs increases between strategic outline case and outline business case stages.

<i>NHS hospital project</i>	<i>Capital cost at SOC (£M)</i>	<i>Capital cost at OBC (£M)</i>	<i>Change (%)</i>
Bradford	116.0	191.0	64.7
Tameside and Glossop	41.0	84.2	105.4
Plymouth	101.0	274.4	171.7
Colchester	79.0	127.0	60.8
Sherwood Forest	66.0	125.0	89.4

Source: Health Select Committee, Public expenditure survey, session 2002–2003.

Table 3. PFI capital cost increases between outline business case and full business case, first wave NHS PFI schemes.

<i>Project</i>	<i>Capital cost at SOC (£M)</i>	<i>Capital cost at OBC (£M)</i>	<i>Change (%)</i>
Swindon	45	148	229
Worcester	49	116	137
South Manchester	40	89	123
Norfolk	90	200	122
Bishop Auckland	26	52	100
South Tees	65	106	63
North Durham	60	96	60
Bromley	80	120	50
Dartford	97	137	41
Calderdale	55	77	40
Wellhouse	30	40	33

Source: Gaffney and Pollock (1999).

Select Committee in 2003, PFI costs increased from SOC to OBC stages by between 64.7 and 171.7%. (These data were only collected for one year and are no longer requested by the Health Select Committee.) Table 3 shows that, in first wave of hospital PFI schemes, the costs increased from OBC to FBC stages by between 33 and 229%.

Evaluation of the Treasury's Time and Cost Overrun Evidence Base

Although five research studies are cited by the government when comparing time and cost overrun in traditional and PFI procurement, only one, the Mott MacDonald report, actually contains comparative data.

The first study was undertaken by the Treasury in 2002. It provides the empirical data for a chapter of its 2003 policy document, *Meeting the Investment Challenge*. In July 2003, it was announced that it would be published on the HM Treasury website in the following autumn. However, publication failed to materialize and in response to a subsequent request for the data made by the present authors to the Treasury on 18 April 2005 under the Freedom of Information Act, the Treasury

replied:

The information requested is held by HM Treasury and all fall within provisions of the Act which exempt it from disclosure. Disclosure of the information may be detrimental to the commercial interests of specific PFI contractors or the financial interests of procuring authorities and would therefore be exempt from disclosure under S43 and S29.

After further correspondence, the survey results, but not a report of the study, were eventually released to one of the present authors in July 2005 in the form of a spreadsheet. These data cannot be evaluated because the study methodology has not been published. The Treasury's data are therefore uninterpretable and do not provide support for the government's policy nor for the Treasury's guidance.

The Treasury also draws on data apparently contained in two NAO reports: *Modernizing Construction* (2001) and *PFI Construction Performance* (2003). But neither of these studies compare performance under different procurement routes. The first one is based on

interviews with the industry about the scope for improved construction performance. The second is a census of 38 PFI project managers. Neither study examines the relative performance of PFI and conventional procurement. Indeed, the authors of *PFI Construction Performance* conclude: 'it is not possible to judge whether these projects could have achieved these results using a different procurement route' (NAO, 2003).

A fourth source of comparative data is cited in the NAO's 2001 report and in *Meeting the Investment Challenge*. This is the 1999 *Agile Construction Initiative: Benchmarking Stage Two Study*. This study, however, was designed to develop a method for comparing performance, not to evaluate performance. Although it is cited by the Treasury and the NAO as the source for the claim that, historically, time overruns occur in 70% of conventionally procured projects, and cost overruns in 73%, no data bearing on these claims are provided in the report.

The Mott MacDonald report, *Review of Large Public Procurement in the UK* (June 2002), is the only one of the five documents cited by the Treasury which compares PFI with conventional procurement. The study was commissioned by the Treasury in 2001 to gather evidence for a review of the Green Book.

The Mott MacDonald Report

The aim of the MacDonald study was:

...to gather a representative sample of projects procured traditionally and through the Private

Finance Initiative (PFI) and implemented over the last 20 years [in order] to assess the past delivery of major projects in the UK procured by the public sector over the last 20 years and from the lessons learned provide best practice guidance for reducing optimism in project estimates for current and future projects.

A main objective was to measure 'optimism bias' in a sample of both PFI and conventionally procured schemes.

The study results are summarized in table 4, which shows the numbers of projects included in the study by one of five categories (non-standard building, non-standard engineering, standard building, standard engineering, other), and the cost and time overrun data. The table shows the small number of studies and the absence of data on some schemes.

The MacDonald study does not include an account of the sampling methodology used, nor the representativeness of the samples studied. Although 80 projects were selected for inclusion in the study, 60 by the Treasury and 20 by Mott MacDonald, neither the populations nor the time periods involved are described. Furthermore, 29 projects had to be excluded from the sample because of insufficient data, but the characteristics of the excluded projects are not indicated. The PFI sample contained only 11 projects, although 451 PFI construction schemes were completed by April 2003, according to the Treasury. This compares with 39 schemes included in the non-PFI sample, although by 1999 there were very few non-PFI deals.

Table 4. Time and cost overruns as percentage of original estimates by type of procurement and project reported by Mott MacDonald.

Description of projects	Number of schemes		Percentage of total	Time overrun optimism bias (%)	Cost overrun optimism bias (%)
Non-standard buildings	PFI	0	-	-	-
	Traditional	7	(18)	39	51
Non-standard engineering	PFI	0	-	-	-
	Traditional	13	(33)	15	66
Standard buildings	PFI	3	(30)	16	2
	Traditional	14	(36)	4	24
Standard engineering	PFI	4	(40)	No information	No information
	Traditional	3	(8)	34	44
Other	PFI	4	(40)	28	No information
	Traditional	2	(5)	54	214
Total	PFI	11	(100)		
	Traditional	39	(100)		

There is clear evidence in the report of selection bias. The conventionally procured project sample includes projects commissioned under very different conditions, and very different policy guidance, from those governing PFI projects. Most conventional procurement projects predate the procurement reforms of 1999 and some predate the introduction of PFI by more than two decades. For example, of the schemes included in the public procurement sample, the Thames Barrier was conceived in the 1960s, commissioned in the 1970s and completed in 1982; the first lines of the Tyne and Wear Metro were opened in August 1980; and the Jubilee Line Extension was inaugurated in 1979. No conclusions can be drawn from a comparison of PFI with procurement performance in these eras because procurement guidance and government policy has since changed radically.

There is also evidence of selection bias through over-representation of atypical schemes in the conventional procurement sample and under-representation of them in the PFI sample. Seventy per cent of the PFI sample, seven projects in all, were standard projects, compared with only 44% (17 projects) in the public procurement sample. In contrast, the PFI sample included no non-standard projects in either the building or the engineering categories, compared with 40% (20) non-standard projects in the public procurement sample. The PFI sample also notably does not include any of the numerous failed PFI IT projects, such as those for National Insurance Recording System 2 (NIRS2) and the Passport Office.

As noted above, non-standard projects usually involve more cost increases because of their complexity. At least three of the 14 standard publicly-procured schemes included in the sample (Guy's Hospital, the Jubilee Line Extension, and the British Library) were referred to in the Public Administration Select Committee as examples of 'overruns remaining a serious problem in conventional public sector capital procurement' and have been regularly cited by ministers. But the alleged failings of at least two of these schemes should be treated with caution.

For example, although Guy's Hospital is frequently cited as an example of inefficient conventional procurement, the scheme was originally described by government as the first hospital public-private partnership, and the cost increases recorded at Guy's were partly attributable to the withdrawal of the private partner, leading to its eventual completion via

conventional procurement. The government attributes the Jubilee Line Extension time and cost overruns not to the method of procurement but to geological conditions. Lord Whitty told parliament in November 1998:

It is the case that even less than 100 years ago we were told that, geologically, south London could not have a tube line. The methods of construction therefore have had to be particularly careful...[That] was the major cause for delay on the jubilee line (Hansard, 16 November 1998, column 975).

These examples show that time and cost overruns cannot automatically be attributed to the method of procurement.

Finally, the sample sizes are too small. There were only 11 projects in the PFI arm. Three were 'standard' buildings and two were standard engineering. The numbers of standard schemes in both the PFI and non-PFI samples were too small to allow statistical tests to be conducted (see table 4). The authors acknowledge that this is a weakness: 'Statistically, the sample of projects in the Mott MacDonald study is necessarily small because, in the time period studied, large public sector procurement was restricted to a relatively limited number of projects' (Mott MacDonald, 2002, p. 8).

In short, the study samples were not representative of projects procured either traditionally or under PFI, were non-comparable, and too small to be significant.

There were also two important sources of measurement bias. In the first place, there are vagaries in definitions employed. Mott MacDonald researchers found variation in the definition of capital costs among the schemes studied:

Often when developing a business case, a contingency allowance is added to the estimate of capital expenditure [cost]. In some cases Mott MacDonald experienced difficulties determining whether the figures quoted in the reference material used included contingencies (Mott MacDonald, 2002, p. 18).

It is impossible to say, on the basis of the report, when real cost overruns are being measured, or whether apparent increases reflect simply the inclusion of provision for contingencies.

The study also measures changes in works duration, not late delivery, and it is therefore not possible to distinguish projects which were delivered late from those projects which were

delivered on time, even though the time taken increased. Mott MacDonald state: 'The measured optimism bias does not give any indication of whether the project was delivered on time, but only reflects the extent to which the works duration had increased' (Mott MacDonald, 2002, p. 12). So we do not know whether the reported increases in the duration of the work were expected or not.

In addition to these shortcomings, a major source of measurement bias arises from the fact that cost changes were measured from different baselines under PFI and under conventional procurement. Cost change in PFI projects was measured from the full business case stage to completion, whereas cost change in conventional procurement was measured from either the strategic outline case or the outline business case stages to completion (Mott MacDonald, 2002). Thus cost escalations included in conventional procured projects were excluded from PFI procured projects. The potential scale of these exclusions from measures of cost changes under PFI is illustrated in tables 2 and 3, and is very large. The effect was to inflate the cost changes under conventional procurement relative to those under PFI.

Mott MacDonald state that the use of different baselines for measuring cost changes is an accident of data availability, acknowledging but not addressing a major source of measurement bias.

Conclusion

Sampling and measurement flaws make uninterpretable the results of the only study cited by the government that actually compares conventional and PFI procurement performance, and all claims based on them

misleading. The study samples were not representative of projects procured either traditionally or under PFI. For all types of projects, the numbers in the PFI sample were so few as to provide no meaningful data. Measurement bias introduced additional sources of error. PFI performance, relative to public procurement, was not properly evaluated in the Mott McDonald study. The Treasury's claims about the superiority of the PFI is based on time and cost overrun arguments for which there is no evidence. Treasury appraisal guidance, which is based on these claims, is therefore biased. ■

References

- Bentley, N. (2006), Britain's health depends on private finance. *The Guardian* (3 February).
- Gaffney, D. and Pollock, A. (1999). Pump-priming the PFI: why are privately financed hospital schemes being subsidized? *Public Money & Management*, 19, 1.
- Gershon, P. (1999), *Review of Civil Procurement in Central Government* (HM Treasury, London).
- HM Treasury (2003), *PFI: Meeting the Investment Challenge* (HM Treasury, London).
- Latham, M. (1994), *Constructing the Team. Final Report of the Government/Industry Review of Procurement and Contractual Arrangements in the UK Construction Industry* (HMSO, London).
- Mott MacDonald (2002), *Review of Large Public Procurement in the UK* (Treasury, London).
- NAO (2003), *PFI Construction Performance* (HC371), Session 2002-3 (NAO, London).
- NAO (2005), *Improving Public Services Through Better Construction* (HC 364-II), Session 2004-05 (NAO, London).
- Timmins, N. (2006), Uncertain destination: will private finance for public projects soon run out of road. *Financial Times* (8 February).