



PERFORMANCE INDICATORS: GOOD, BAD and UGLY **{Embargoed until 10pm on 23 October 2003}**

Performance monitoring in the public services is poorly conducted, according to a report published today by the Royal Statistical Society.

Despite some good examples - such as enhanced size of British Crime Survey, Europe's informative BSE testing of cattle, and Scotland's reporting of confidence intervals - scientific standards, in particular statistical standards, have been largely ignored. This criticism applies not just in target setting, but also in the design, analysis and reporting of performance indicators. The RSS report offers practical solutions for resolving these critical issues, against which current and future performance monitoring of the public services should be judged.

A striking feature of UK public services over the past 10 years has been the rise of performance monitoring. Performance data can be used in establishing 'what works' among policy initiatives; to identify well-performing or under-performing institutions and public servants; and, equally important, to hold Ministers to account for their stewardship of the public services.

Hence, government is both monitoring the public services, and being monitored, by performance indicators. Because of government's dual role, performance monitoring must be done with integrity and shielded from undue political influence, in the way that National Statistics are shielded.

The Royal Statistical Society is calling for:

- * **performance monitoring protocols** - to ensure that statistical standards are met. Protocol is an orderly record not only of decisions made (from design to analysis and reporting) but also of the reasoning and calculations that led to those decisions;
- * **independent scrutiny** - to safeguard the wider-than-government public interest, the individuals and institutions being monitored, and methodological rigour;
- * **the reporting of measures of uncertainty** whenever performance data are published, including as league tables or star ratings – to avoid over-interpretation and the false labelling of performance;
- * **research on different strategies than 'name and shame'** for the public release of performance data, **and better designs** (including randomization) for evaluating policy initiatives – the first to allay ethical and effectiveness concerns, the second for robust evidence about 'what works';
- * **much wider consideration of the ethics and cost-efficiency of performance monitoring.**

Professor Sheila Bird, who chaired the Royal Statistical Society's Working Party on Performance Monitoring in the Public Services, appealed to journalists to champion better reporting:

"When league tables or star ratings are published, we'd like journalists to insist on access to (and reporting of) the measure of uncertainty that qualifies each ranking or rating. Without this qualifier, no-one can separate the chaff from the wheat, the good from the bad."

There is a precedent for this type of statistically-savvy journalism – though sometimes honoured in the breach. When reporting social or polling surveys, journalists know to cite the number surveyed, response rate, and a 'margin of error' (accounting for random variation) in any headlined percentage.

Policy initiatives often aim to change performance indicators. But, public money spent on inferior (usually non-randomized) study designs that result in poor-quality evidence about how well policies actually work is an evaluation-charade. Costly, inefficient by denying scientific method, and a loss in public accountability.

The Working Party looks to both the Treasury and Delivery Unit to review the resources spent on inferiorly-designed policy initiatives, and to hasten the implementation of better designs (including randomized) for cost-efficient and robust policy evaluation.

Professor Bird said: "I very much regret that randomized trials of policy initiatives are rare. Missed opportunities for proper evaluation were mandatory drugs testing of prisoners (rolled out from 1995) and drug treatment and testing orders (piloted from 1998) to name but two. Despite targets on recidivism and crime clear-up rates, UK judges prescribe sentences on lesser evidence (about 'what works') than doctors prescribe medicines."

Performance monitoring done well is broadly productive for those concerned. Done badly, it can be very costly and not merely ineffective but harmful and indeed destructive - of morale, reputations and the public services.

Professor Andy Grieve, President of the Royal Statistical Society, said: "The Royal Statistical Society wants to promote well-informed public debate on performance monitoring in the public services. Therefore, we'll engage with journalists in pursuit of better reporting standards, and with government and parliament to foster good practices in performance monitoring by implementing them across government. The Royal Statistical Society will start this by hosting a workshop on PM protocols at which a template can be worked through for a series of existing PM procedures, and then disseminated with these as exemplars."

Membership of Working Party

Sheila M. Bird (chair) MRC Biostatistics Unit, Cambridge & Strathclyde University
[phone: 01223 330368; mobile: 07800 639 269; sheila.bird@mrc-bsu.cam.ac.uk]

Sir David Cox FRS Nuffield College, Oxford
[phone: 01865 278690; email david.cox@nuffield.oxford.ac.uk]

Vern T. Farewell MRC Biostatistics Unit, Cambridge
[phone: 01223 330371; email vern.farewell@mrc-bsu.cam.ac.uk]

Harvey Goldstein FBA Institute of Education, University of London
[phone: 0207 612 6652; mobile 07866 735 299; h.goldstein@ioe.ac.uk]

Tim Holt CB Department of Social Statistics, Southampton University
[phone: 023 8059 2531; email tim.holt2@virgin.net]

Peter C. Smith Department of Economics & Related Studies, York University
[phone: 01904 321443; email pcs1@york.ac.uk]

BACKGROUND NOTES

1. The Royal Statistical Society was founded in 1834. It is a learned and professional society, with some 7200 members based in the United Kingdom and overseas. The purpose of the Royal Statistical Society is to develop, foster and disseminate statistical knowledge, methodology and good practice for the benefit of all society.

By long tradition, therefore, from 'passionate statistician' Florence Nightingale in the 19th century to recent presidential addresses, such as *'Mad cows and ecstasy: change and choice in an evidence-based society'* by Professor Adrian FM Smith FRS, the Royal Statistical Society supports the application of scientific method and insightful analysis to improve the condition of society, and to evaluate objectively government policies.

2. Press conferences are convened exceptionally by the Royal Statistical Society. Previous reports - similarly launched - were on *'Official Statistics: Counting with Confidence'* (in 1990) and *'The Measurement of Unemployment in the UK'* (in 1995).

3. The remit of the Royal Statistical Society's Working Party on Performance Monitoring was to deal with scientific, and in particular statistical, aspects of many issues in PM which the Public Administration Select Committee highlighted; and to make recommendations.

The Working Party has a breadth of experience with statistical applications in education, health, economics, official statistics, criminal justice, agriculture and industry. It is also grateful to the Fellows and others who commented on its consultation report.

4. DIGESTS

On PM PROTOCOL:

Before introducing performance monitoring [PM] in any public service, a PM protocol should be written. This is an orderly record not only of decisions made but also of the reasoning or calculations that led to those decisions. And there should be calculations. A PM protocol should cover: objectives; design considerations and the definition of PIs; sampling versus complete enumeration; the information to be collected about context; the likely perverse behaviours or side-effects that might be induced as a reaction to the monitoring process, and also, very importantly, the practicalities of implementation.

Procedures for data collection, analysis, presentation of uncertainty and adjustment for context, together with dissemination rules, should be explicitly defined and reflect good statistical practice. PIs should be seen as 'screening devices' and not over-interpreted.

On TARGETS:

If quantitative performance targets are to be set, they need to have a sound basis, take account of prior (and emerging) knowledge about key sources of variation, and be integral to the PM design.

Aspirational targets have a distinctive role, but one which is largely irrelevant in the design of a PM procedure; motivational targets which are not rationally based may demoralize and distort. Anticipated and actual side-effects of performance monitoring, including on individuals' behaviour and priorities, may need to be monitored as part of the PM process.

On SCRUTINY:

Independent scrutiny of PM schemes for the public services should be set up and must report publicly. The extent and nature of this scrutiny should be related to the assessed drawbacks and benefits, reflect ethical concerns, and conform with good statistical practice.

On ETHICS:

PM lacks statistical integrity if it fails to identify, and design out, major corruptions of measured indicators; if its chosen PIs are not measured with sufficient precision to reveal whether targets have been met; if targets are set irrationally; if its design is cost-inefficient; or if analysis lacks objectivity, or is superficial.

PM often requires access to data about people (patients, school-children, prisoners, employees). Generally, research on humans needs 3rd-party approval of its ethics and methodology. While staff who deliver public services may well have agreed implicitly to the monitoring of their performance, it is far from clear that this extends to the publication of analyses in a form in which individuals can be identified. Legal requirements on the protection of confidentiality govern official statistics.

5. Outline of the Working Party's Report in seven sections:

1. Introduction gives the background to performance monitoring in the public services, refers to Public Administration Select Committee's recent report "On target? Government by Measurement", and sets out the remit of the Royal Statistical Society's Working Party as being to deal with scientific, and in particular statistical, aspects of many issues in PM which the Public Administration Select Committee has highlighted.

2. PM Design, Target Setting and Protocol includes choosing what to measure, definition of indicators, how to set targets, how NOT to set targets, and observes that getting a PM protocol right is no more, or less, onerous than applying scientific method to any other field. Sampling approaches are discussed because, although they may add slightly to the complexity of analysis, their potential cost-effectiveness and safeguarding against perverse behaviours commend them as sometimes both more accurate and less costly than complete enumeration.

3. Analysis of data starts with general principles which, for PM data, are again no different from those for other types of data and include that analysis should be sharply focused on the appropriate objectives and as simple as possible, subject to avoiding misleading oversimplification.

Analyses need to examine not just overall average values of PIs but to look at variability - because variability is an intrinsic part of the real world and often of interest and importance in its own right; but also because variability is one of the determinants of uncertainty in the primary conclusions. Frequency of analysis and adjustment for context to achieve comparability are also dealt with, together with the important issue of multiple indicators of performance.

Amalgamation of indicators into a single summary figure should be resisted as far as possible.

4. Presentation of PIs. When a new PI is introduced, initial analyses should be designed to provide insights to its characteristics, particularly major components of variation, which should inform decisions about its routine presentation. League tables are frequently produced on an annual basis when a better basis for ranking or estimation may be achievable through use of a more extended period of observation.

In all cases, the uncertainty of ranking should be indicated through the use of plausible ranges of rank for each institution. Banding, as in 'star' banding of hospitals, does NOT circumvent the need to represent the uncertainty of an institution's banding.

There is discussion of extremes and that 'naming' is NOT a pre-requisite for public accountability and may have dis-benefits, that are bad or ugly, besides its apparent attractiveness in promoting public choice.

5. The impact of PM on the public services recalls the salutary lesson from New York that the publishing of named cardiac surgeons' unadjusted patient death-rates led to increased cardiac mortality.

There is caution against targets which require no breaches of a waiting time guarantee for NHS patients, and warning that PM risks over-emphasis on the easily measurable. It is therefore essential that when a new PM protocol is proposed, a broadly-based assessment should be made of its likely costs and consequences.

6. Evaluating PM initiatives recognises difficulties of evaluation - how soon to start, constraints imposed when the roll-out of a new policy has been pre-determined by financial or political imperatives, and the confounding of policy effects when a raft of new initiatives is launched more or less contemporaneously. Difficulties apart, the report is clear that experiments & randomization should have a stronger role in policy evaluation; and that if a new PM initiative is itself costly to implement in relation to the likely gain from it in terms of new insights or performance enhancement, then it should be discontinued for exceeding reasonable 'value-for-money' thresholds. After all, we limit NHS treatments on this basis.

7. Integrity, confidentiality and ethics of PM. PIs for public accountability need wider consensus than just from within government. They also need independent safeguards for their design and monitoring, as for National Statistics; and to have statistical integrity.

PM often requires access to data about people (patients, school-children, prisoners, employees). Generally, research on humans needs 3rd-party approval of its ethics and methodology. While staff who deliver public services may well have agreed implicitly to the monitoring of their performance, it is far from clear that this extends to the publication of analyses in a form in which individuals can be identified. Legal requirements on the protection of confidentiality govern official statistics.

Royal Statistical Society, 12 Errol Street, LONDON EC1Y 8LX

[phone: 0207 638 8998; email rss@rss.org.uk; nearest tubes: Barbican, Liverpool Street; website: www.rss.org.uk]