

What's Measured Is Not Necessarily What Matters: A Cautionary Story from Public Health

Ce qui est évalué n'est pas nécessairement ce qui est le plus important : un récit instructif provenant de la santé publique



RAISA DEBER, PHD

*Professor, Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health, University of Toronto
Toronto, ON*

ROBERT SCHWARTZ, PHD, IHPME

*Professor, Institute of Health Policy, Management and Evaluation,
Dalla Lana School of Public Health
Executive Director and Principal Investigator, Ontario Tobacco Research Unit
University of Toronto
Toronto, ON*

Abstract

A systematic review of the introduction and use of outcome-based performance management systems for public health organizations found differences between their use as a management system (which requires rigorous definition and measurement to allow comparison across organizational units) versus for improvement (which may require more flexibility). What is included in performance measurement/management systems is influenced by ease of measurement, data quality, ability of organization to control outcomes, ability to measure success in terms of doing things (rather than preventing things) and what is already happening. To the extent that most providers wish to do a good job, the availability of good data to enable benchmarking and improvement is an important step forward. However, to the extent that the health of a population is dependent on multiple factors, many beyond the mandate of the health system, too extensive a reliance on performance measurement may risk unintended consequences of marginalizing critical activities.

Résumé

Une revue systématique sur l'introduction et l'utilisation de systèmes de gestion du rendement par les organismes de santé publique a relevé des différences entre leur utilisation en tant que systèmes de gestion (qui demande des définitions et des évaluations précises afin de permettre une comparaison des unités organisationnelles) et leur utilisation pour l'optimisation (qui exige plus de flexibilité). La sélection des paramètres qui seront utilisés dans les systèmes de gestion du rendement est influencée par : ce qui est facile à évaluer, la qualité des données, la capacité de l'organisation à contrôler les résultats et à évaluer le succès en fonction de ce qui se fait (plutôt qu'en fonction d'actions préventives). Dans la mesure où la plupart des intervenants souhaitent faire un bon travail, la disponibilité de données pertinentes pour permettre des évaluations comparatives et des améliorations est un pas important dans la bonne direction. Par contre, dans la mesure où la santé de la population dépend de plusieurs facteurs, qui sont souvent en dehors du mandat du système de santé, une trop grande dépendance sur la mesure du rendement risque d'avoir des conséquences inattendues, telles que la marginalisation d'activités critiques.

Introduction

The New Public Management has been associated with an increased emphasis on measuring performance, often summarized using the phrase "What's measured is what matters." A growing literature has found potential limitations to this view (Bevan and Hood 2006; Exworthy 2010; Kuhlmann 2010). This manuscript, which grew from a synthesis of the literature on performance measurement and management in public health, presents a conceptual framework for viewing performance measurement and suggests an additional set of risks inherent in over reliance on these approaches.

Materials and Methods

Literature search

We adapted Pawson et al.'s (Pawson et al. 2005) approach to literature review, which recognizes that much of the analysis will, of necessity, be thematic and interpretive (Dixon-Woods et al. 2005; Pawson 2002), including use of cross-case analysis (Mays et al. 2005; Pope et al. 2006). As the ESRC UK Centre for Evidence Based Policy has noted, social science reviews differ from the medical template in that they rely on a "more diverse pattern of knowledge production," including books and grey literature (Grayson and Gomersall 2003).

Our search strategy included multiple sources. We began with 213 references provided by our KT partner, the Public Health Practice Branch of the Ontario Ministry of Health and Long-Term Care. To capture published and grey literature, we searched such databases as PubMed, Web of Science and Google Scholar; these sites tend to capture different literatures, and thus helped ensure that key references were not missed, using such keywords as: indicators, accreditation, balanced scorecard, evidence-based public health, local public

health, performance measurement, performance standards and public health management, alone and in combination. We also searched relevant websites, both for the selected jurisdictions and for the papers and reports produced by the World Health Organization (WHO), Organisation for Economic Co-operation and Development (OECD) and the European Observatory on Health Systems and Policies. We then analyzed both backwards and forward citation chains from key articles – that is, checking the relevant articles cited by that paper (backwards) and the materials citing that article (forward). Other helpful sources were a US review of performance management in public health (Public Health Foundation 2009) funded by the Robert Wood Johnson Foundation, the materials on their website (available at http://www.phf.org/resourcestools/pages/turning_point_project_publications.aspx) and the proceedings of a WHO European Ministerial Conference on Health Systems, which focused on performance measurement for health system improvement (Smith et al. 2009).

The abstracts were then scanned for relevance by our team. The approach taken examined the general literature and then selected literature relevant to key case examples from Australia, New Zealand, the UK, the EU, the US and Canada. Case examples were chosen by looking at the jurisdictions selected, with a focus on those that matched, corresponded or contrasted with the Ontario Public Health Standards. This initial review yielded 970 references, which was subsequently augmented by new publications; we also deleted articles not relevant to this subject. The retained material on which this analysis is based was published between 1966 and 2015, with 13 references before 1990, 125 between 1990 and 1999 and 807 between 2000 and 2011, although we have subsequently examined additional more recent publications. Our analysis of the 55 public health measurement cases we selected has been published elsewhere (Schwartz and Deber 2016). This paper focuses on some key lessons for applying performance management and measurement approaches to public health.

Results

Defining our terms

Increasing attention is being paid to the use of information to improve performance. Much of this dialogue is couched in terms of accountability (Smith et al. 2009). There is an extensive literature from management science and from new public management on the use of performance measurement and management in both the public and private sectors (Bouckaert 1993; Freeman 2002; Julnes 2009; Kuhlmann 2010; Poister and Streib 1999). These authors place heavy emphasis on the role of organizational culture and political support in being able to implement change.

Accountability is defined as having to be answerable to someone for meeting defined objectives (Emanuel and Emanuel 1996; Fooks and Maslove 2004; Marmor and Morone 1980). It has financial, performance and political/democratic dimensions (Brinkerhoff 2004) and can be *ex ante* or *ex post*. This may translate into fiscal accountability to payers, clinical accountability for quality of care (Dobrow et al. 2008) and/or accountability to the public.

The actors involved may include various combinations of providers (public and private), patients, payers (including insurers and the legislative and executive branches of government) and regulators (governmental, professional); these actors are connected in various ways (Shortt and Macdonald 2002; Zimmerman 2005). As noted in a series of sub-studies on approaches to accountability published as a special issue of *Healthcare Policy* (Deber 2014), the tools for establishing and enforcing accountability are similarly varied, and they require clarifying what is meant by accountability, including specifying for what, by whom, to whom and how. Performance management and measurement is frequently suggested as an important tool for improving systems of accountability. As our review clarified, there is some variation within the literature and the cases examined in how various terms are defined and in the purposes of the performance measurement exercise (Solberg et al. 1997). Underlying most of these examples is the sense that managing is difficult without measurement (Gibberd 2005).

Performance measurement has been defined by the US Government Accountability Office (GAO) as “the ongoing monitoring and reporting of program accomplishments, particularly progress toward pre-established goals” (US Government Accountability Office 2005). Their definition notes that such activities are typically conducted by the management of the program or agency responsible for them. The GAO contrasts this with program evaluation, which is often conducted by experts external to the program, and may be periodic or *ad hoc*, rather than ongoing. The GAO definitions, like many performance measurement systems in healthcare often use the framework of Donabedian, which focuses on various combinations of structures, processes, outputs and outcomes (Donabedian 1966, 1980, 1988).

A number of approaches to performance measurement can be found in the literature (Abernethy et al. 2005; Adair et al. 2003, 2006a, 2006b; Arah et al. 2003; Stoto 2014; Veillard 2012). The focus of performance measurement systems can also vary, but increasing attention has been paid to using performance management as a way of improving system performance. Goals may also vary but are often aligned with quality. Published reviews of performance measurement efforts include both examination of individual countries and comparisons among OECD countries, including Canada, the US, the UK and Australia (Baker et al. 1998, 2008; Hurst 2002; Hurst and Jee-Hughes 2001; Kelley and Hurst 2006; Mattke et al. 2006; Smith 2002). Much of the literature focuses on using performance measurement to improve clinical quality of care across a variety of settings, including primary care and emergency care (Barnsley et al. 1996; Linder et al. 2009; Lindsay et al. 2002; Phillips et al. 2008). Other projects focus on using performance measurement to improve governance, often using the language of accountability. For this to occur, ongoing data collection is important, so that management and stakeholders can use up-to-date information to monitor the quality of care being provided (Loeb 2004). One approach is to use performance indicators.

Performance management, by contrast, both paves the way for and requires a performance measurement system. Many measurement systems are developed with the goal of defining where improvements can be made, with the assumption that managers can use them once the

measurement results are examined (Lebas 1995). Performance management can be defined as the action of using performance measurement data to effect change within an organization to achieve predetermined goals (Folan and Browne 2005). There is now broad recognition that while public sector organizations are doing a great deal of performance measurement, they often do not use the data well in full-fledged performance management systems (Schwartz 2011). Nevertheless, there are a number of success stories in public management of using well-designed measurement systems to improve performance (Ammons 1995). Although measurement may be necessary for management, not all performance measurement systems assume that they will be used for management.

Implementing performance measurement: Goals and indicators

The first step to developing a successful performance measurement system is to clearly define what will be measured. McGlynn and Asch suggest that three considerations should be taken into account when choosing an area to measure: (1) how important the area of health-care being measured is, (2) the amount of potential this area holds for quality improvement and (3) the degree to which healthcare professionals are able to control quality improvement in this area of healthcare. They define importance in terms of mortality/morbidity, but also utilization of health services and cost to treat (McGlynn and Asch 1998). Again, there is likely to be variation, depending on whether one is focused on particular patient groups or on the health of the population. However, from the viewpoint of public health, these considerations point to the importance of surveillance systems to provide decision-makers with information about the prevalence of conditions, how they are being addressed and the outcomes of interventions.

Often implicit are what policy goals are being pursued. Different goals may imply different policies. Key goals are usually some combination of access, quality (including safety) (Baker et al. 2004), cost control/cost effectiveness and customer satisfaction (Monahan 2006; Myers and Lacey 1996). Behn suggests the objectives for accountability should be improved performance, fairness and financial stewardship (Behn 2001). This affects what organizations are accountable for. Often, policy goals may clash (Deber et al. 2004). An ongoing issue is the potential for unintended consequences if the measures selected do not reflect the full set of policy goals (Townley 2005). Indeed, one of the purposes of balanced scorecards is to make such potential conflicts between goals and measures more evident (Baker and Pink 1995; Kaplan and Norton 1996; Pink et al. 2001; Ten Asbroek et al. 2004; Weir et al. 2009).

Once an appropriate area has been identified for measurement, the next step in developing a performance measurement system is to identify potential indicators that will be used in the measurement system. Indicators have been defined as “a measurement tool used to monitor and evaluate the quality of important governance, management, clinical and support functions” (Klazinga et al. 2001). Indicators can be classified. For example, some authors assume that because performance must be measured against some specification, performance

indicators do infer quality. Others (who do not necessarily represent a common view) distinguish between “Activity Indicators,” which measure how frequently an event takes place; “Quality Indicators,” which measure the quality of care being provided; and “Performance Indicators,” which do not infer quality but measure other aspects of the performance of the system (for example, the use of resources) (Campbell et al. 2003).

The issue of measurement

Loeb (2004) argues that not everything in healthcare can or should be measured. Challenges may arise when outcomes are influenced by factors other than the interventions being assessed or beyond the control of those being held accountable. There are also issues associated with balancing the number of indicators needed to provide enough information, with usability and costs associated with having too many indicators. Developing and running a performance measurement system is often expensive, and the data produced needs to be useful and interpretable for its users.

Many indicators are developed through a rigorous process by which they are developed, defined and reviewed (Lindsay et al. 2002; McGlynn and Asch 1998). Data sources also need to be identified when developing and choosing a set of indicators, with the most common sources coming from healthcare enrolment, administrative data, clinical data and survey data. Clear definitions will ease implementation of the measurement system and its data collection processes across different organizations/users in a consistent fashion and help to ensure that the data collected within the measurement system will be comparable and reliable across different users of the system. As Black has noted, this is not always the case (Black 2015).

Considerable efforts have been made to develop comparable indicators to enable cross-jurisdictional comparisons. These include the OECD quality indicators project (Arah et al. 2006) and the reporting standards for public health indicators (Armstrong et al. 2008). An offsetting concern is the recognition that strategic scorecards also must include locally relevant indicators. Achieving the right mix between local relevance and the ability to compare across organizations is crucial.

Discussion

One ongoing issue is what sorts of indicators should be used. A promising development is the Canadian Institute of Health Information (CIHI) 2012 *Performance Measurement Framework for the Canadian Health System* (CIHI 2012), which attempts to link performance dimensions through expected causal relationships in four interrelated quadrants: Health System Outcomes, Social Determinants of Health, Health System Outputs and Health System Inputs and Characteristics. Proper application of this and similar frameworks may help to ensure a more balanced approach to what is measured and what matters.

However, our review suggests that the factors important to those individuals providing clinical services to clients often differ from those important to program managers, payers or health systems (Tregunno et al. 2004). One class of indicators focuses on adverse outcomes, either at

the individual level (e.g., adverse events) or at the system level (e.g., avoidable deaths). Klazinga et al. argued that “epidemiological research has shown the difficulties in validating [negative health outcomes] as indicators for the quality of care that was delivered” (Klazinga et al. 2001).

In selecting indicators, a key factor is the extent to which the elements affecting the measurement are under control of decision-makers. Chassin et al. emphasized that for an outcome indicator to be relevant, it must be closely related to the healthcare processes that have an effect on the outcome (Chassin et al. 1998). In addition, there may be differences in what would be done with information; although the information may be valuable, it is difficult to hold managers accountable for things they cannot control. One obvious example is geography, which will often affect travel costs or access. Another, which affects population health, is the extent to which the various determinants of health (e.g., income, housing, tobacco use, etc.) are under the control of public health organizations. Information may thus be helpful in affecting policy levers (e.g., pricing of alcohol, tobacco) that other actors control, but less useful if program managers will be rewarded (or punished) for variables they cannot affect.

Other factors include whether different indicators are correlated (which can lead to double counting), how easy they are to measure (transaction costs), extent to which they are subject to “gaming” and whether they cover the outcomes of interest (Bevan 2010; Exworthy 2010; Ham 2010; Hamblin 2008; Irwin 2010; Klazinga 2010; Provincial Auditor of Ontario 2003).

Likely impacts

Another set of issues involves what will be done with the performance measures, including how they will be applied. Frequently, performance measurement involves setting performance targets and assessing the extent to which these are being met. In turn, these may be used for funding (e.g., results-based budgeting) and/or to identify areas for in-depth evaluation. External bodies may use the information to ensure accountability. Managers may use them to monitor activities and make policies. Townley argued that “the use of performance measures reflects a belief in the efficacy of rational management systems in achieving improvements in performance” (Townley 2005). In the UK, use of fiscal levers is sometimes referred to as “targets and terror” (Propper et al. 2008).

The way in which measures are likely to affect behaviour varies. Clearly, measurement is simplest if organizations produce a small number of services, have a limited number of goals, understand the relationship between inputs and results and can control their own outcomes. As Townley notes, “A failure to ground performance measures in the everyday activity of the workforce is likely to see them dismissed for being irrelevant, unwieldy, arbitrary, or divisive.” Other potential downsides are that “the time and resources taken to collect measures may outweigh the benefits of their use” (Townley 2005).

A related set of factors relates to the organizational infrastructure (Alexander et al. 2006). The workplace culture, including differences between the explicit goals and what some have called the “implicit theories” or “theories in use,” which affect day-to-day functioning, may affect the extent to which change initiatives are embraced and performance changes

(Aitken 1994). This is in turn related to concepts of “street level bureaucracy,” which deals with the extent to which it is simple to manage and observe the activities of those responsible for providing the given services (Lipsky 1980). Other less desirable organizational responses to performance measurement may include decoupling, a term used to refer to situations where specialist units are responsible for performance measurement, but where the measures have little impact on day-to-day activities and may lead to a sense that the measurement approach is “ritualistic” and “bureaucratic” rather than integral to improvement (Townley 2005). Even more alarmingly, measurement can lead to dysfunctional consequences, including focusing on measures rather than actual performance, impairment of innovation, gaming and creative accounting, potentially making performance worse (Hamblin 2008; Leggat et al. 1998). Other effects can be subtle; one example is placing less emphasis on prevention than on treating existing problems. The extent to which these positive or negative effects are realized may be heavily dependent upon context.

Conclusions

Selecting indicators

We found considerable differences in what sorts of performance measurement and management are actually being done, not just by jurisdiction (which we expected) but also by type of service. We found heavy emphasis on surveillance and far less on explicitly using the indicator data for management. Additionally, there is more focus on processes of how services are provided than on outcomes.

A number of rationales are provided for this state of affairs. An excellent synthesis can be found in the proceedings of a WHO symposium, which stresses the importance of clarifying causality and the difficulty in holding providers accountable for outcomes that they cannot control. As one example, “physicians working in socio-economically disadvantaged localities may be wrongly blamed for securing poor outcomes beyond the control of the health system” (Smith et al. 2009: 12). Risk adjustment methodologies can control for some, but not all, of this variation. Composite indicators can be useful, but only if transparent and valid. Similarly, it may be necessary to deal with random fluctuations before determining when intervention is needed to improve performance.

One striking finding that emerged from our review of how performance measurement and management are used in public health was the extent to which they focused on clinical services addressed to individuals (Smith et al. 2009). Activities directed towards improving the health of populations, particularly those with a preventive orientation, tend not to be included. As one example, the chapter in the report of the WHO symposium purportedly devoted to population health focuses almost exclusively on clinical treatment, including heavy focus on tracer conditions. One rationale given by these authors is that the performance measurement/management experiments they reported on wished to focus on the healthcare system. Their reaction to the fact that “it is often difficult to assess the extent to which variations in health outcome can be attributed to the health system” (Nolte et al. 2009) was accordingly to omit such measures.

One concern arising from our review is that performance measurement approaches, by focusing so heavily upon the healthcare system, may skew attention away from important initiatives directed at improving the health of the population. Indeed, another chapter in the WHO symposium volume on “measuring clinical quality and appropriateness” explicitly states (pp 88–89): “A number of potential actions to improve population health do not operate through the health-care system (e.g., ensuring adequate sanitation, safe food, clean environments) and some areas do not have health services that are effective in changing an outcome. Neither of these areas is fruitful for developing clinical process measures” (McGlynn 2009). Omitting such areas from measurement systems, however, may falsely imply that they do not matter.

Our review stresses the importance of being aware of unintended consequences. For example, in the UK pay-for-performance (P4P), success tended to be measured as doing more of particular things (e.g., screening tests, medication, some immunization) for particular populations (e.g., people with chronic diseases); prevention and population health risk being lost in the shuffle.

Some key variables that appear to influence what is being included in performance measurement/management systems include:

- + Ease of measurement.
- + Data quality. Jurisdictions vary considerably in how good the data are. For example, Canada does not yet have good data about immunization at the national level.
- + Ability of organization to control outcomes.
- + Ability to measure success in terms of doing things (rather than preventing things).
- + What is already happening. One example is the UK P4P for physicians, which is generally considered to have been highly successful. However, there was some suggestion that what was being rewarded was better recording rather than changes in practice. The indicator systems appear to, in part, reward providers for things they were already doing, which in turn raises questions about who gets to set the indicators.

One important caveat for any performance measurement/performance management system is that it does not, and cannot, capture all activities. In that connection, as Black (2015) has noted, it is important to recognize that most providers are professionals who want to do a good job. Performance measurement/management is only one component, but can give tools to allow all stakeholders to know how they are doing and enable the use of benchmarking to improve performance. A second caveat is that we focused on published information; this may or may not reflect current activities in those jurisdictions. Successful interventions are also more likely to have been published.

To the extent that the health of a population is dependent on multiple factors, many beyond the mandate of the healthcare system (both personal health and public health), however, our review suggests that too extensive a reliance on performance measurement may risk unintended consequences of marginalizing critical activities. As ever, balance is key.

Acknowledgements

This review has been drawn from a Canadian Institutes for Health Research (CIHR)-funded Expedited Synthesis, in partnership with the Ontario Ministry of Health and Long-Term Care, Public Health Practice Branch. The authors appreciate the contributions of their research partners and of the research team: Professors Ross Baker, Jan Barnsley, Andrea Baumann, Whitney Berta, Brenda Gamble, Audrey Laporte, Fiona Miller, Tina Smith and Walter Wodchis; students Kathleen Gamble, Corrine Davies-Schinkel, Tim Walker; Project Manager Kanecy Onate; and Administrative Support Christine Day.

Correspondence may be directed to: Raisa Deber, PhD, Institute of Health Policy, Management and Evaluation, Health Sciences Building, 155 College Street, Suite 425, Toronto, ON M5T 3M6; tel.: 416-978-8366; e-mail: raisa.deber@utoronto.ca.

References

- Abernethy, M.A., M. Horne, A.M. Lillis, M.A. Malina and F.H. Selto. 2005. "A Multi-Method Approach to Building Causal Performance Maps from Expert Knowledge." *Management Accounting Research* 16(2): 135–55.
- Adair, C.E., L. Simpson, J.M. Birdsell, K. Omelchuk, A.L., Casebeer, H.P. Gardiner et al. 2003 (January 17). *Performance Measurement Systems in Health and Mental Health Services: Models, Practices and Effectiveness. A State of the Science Review*. Retrieved October 31, 2016. <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.195.2219&rep=rep1&type=pdf>>.
- Adair, C.E., E. Simpson, A.L. Casebeer, J.M. Birdsell, K.A. Hayden and S. Lewis. 2006a. "Performance Measurement in Healthcare: Part I – Concepts and Trends from a State of the Science Review." *Healthcare Policy* 1(4): 85–104. doi:10.12927/hcpol.2006.18248.
- Adair, C.E., E. Simpson, A.L. Casebeer, J.M. Birdsell, K.A. Hayden and S. Lewis. 2006b. "Performance Measurement in Healthcare: Part II – State of the Science Findings by Stage of the Performance." *Healthcare Policy* 2(1): 56–78. doi:10.12927/hcpol.2006.18338.
- Aitken, J.-M. 1994. "Voices from the Inside: Managing District Health Services in Nepal." *International Journal of Health Planning and Management* 9(4): 309–40.
- Alexander, J.A., B.J. Weiner, S.M. Shortell, L.C. Baker and M.P. Becker. 2006. "The Role of Organizational Infrastructure in Implementation of Hospitals' Quality Improvement." *Hospital Topics* 84(1): 11–20.
- Ammons, D.N. 1995. "Overcoming the Inadequacies of Performance Measurement in Local Government: The Case of Libraries and Leisure Services." *Public Administration Review* 55(1): 37–47.
- Arah, O.A., N.S. Klazinga, D.M.J. Delnoij, A.H.A. Ten Asbroek and T. Custers. 2003. "Conceptual Frameworks for Health Systems Performance: A Quest for Effectiveness, Quality, and Improvement." *International Journal for Quality in Health Care* 15(5): 377–98. doi:10.1093/intqhc/mzg049.
- Arah, O.A., G.P. Westert, J. Hurst and N.S. Klazinga. 2006. "A Conceptual Framework for the OECD Health Care Quality Indicators Project." *International Journal for Quality in Health Care* 18(Suppl. 1): 5–13.
- Armstrong, R., E. Waters, L. Moore, E. Riggs, L.G. Cuervo, P. Lumbiganon and P. Hawe. 2008. "Improving the Reporting of Public Health Intervention Research: Advancing Trend and Consort." *Journal of Public Health* 30(1): 103–09.
- Baker, G.R., N. Brooks, G. Anderson, A. Brown, I. McKillop, M. Murray and G. Pink. 1998. "Healthcare Performance Measurement in Canada: Who's Doing What?" *Healthcare Quarterly* 2(2): 22–26. doi:10.12927/hcq.16555.
- Baker, G.R., A. MacIntosh-Murray, C. Porcellato, L. Dionne, K. Stelmacovich and K. Born. 2008. *High Performing Healthcare Systems: Delivering Quality by Design*. Toronto, ON: Longwoods Publishing.
- Baker, G.R., P.G. Norton, V. Flintoft, R. Blais, A.D. Brown, J. Cox et al. 2004. "The Canadian Adverse Events Study: The Incidence of Adverse Events among Hospital Patients in Canada." *Canadian Medical Association Journal* 170(11): 1678–86. doi:10.1503/cmaj.1040498.

- Baker, G.R. and G.H. Pink. 1995. "A Balanced Scorecard for Canadian Hospitals." *Healthcare Management Forum* 8(4): 7–13.
- Barnsley, J., L. Lemieux-Charles and R. Baker. 1996. "Selecting Clinical Outcome Indicators for Monitoring Quality of Care." *Healthcare Management Forum* 9(1): 5–21.
- Behn, R. 2001. *Rethinking Democratic Accountability*. Washington DC: Brookings Institution Press.
- Bevan, G. 2010. "If Neither Altruism Nor Markets Have Improved NHS Performance, What Might?" *Eurohealth* 16(3): 20–22.
- Bevan, G. and C. Hood. 2006. "What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System." *Public Administration* 84(3): 517–38.
- Black, N. 2015. "To Do the Service No Harm: The Dangers of Quality Assessment." *Journal of Health Services Research and Policy* 20(2): 65–66. doi:10.1177/1355819615570922.
- Bouckaert, G. 1993. "Measurement and Meaningful Management." *Public Productivity and Management Review* 17(1): 31–43.
- Brinkerhoff, D.W. 2004. "Accountability and Health Systems: Toward Conceptual Clarity and Policy Relevance." *Health Policy and Planning* 19(6): 371–79. doi:10.1093/heapol/czh052.
- Campbell, S.M., J. Braspenning, A. Hutchinson and M. Marshall. 2003. "Research Methods Used in Developing and Applying Quality Indicators in Primary Care." *BMJ* 326: 816–19.
- Canadian Institute for Health Information (CIHI). 2012. *A Performance Measurement Framework for the Canadian Health System*. Ottawa, ON: Author. <https://secure.cihi.ca/free_products/HSP-Framework-ENweb.pdf>.
- Chassin, M.R., R.W. Galvin and National Roundtable on Health Care Quality. 1998. "The Urgent Need to Improve Health Care Quality: Institute of Medicine National Roundtable on Health Care Quality." *JAMA* 280(11): 1000–05. doi:10.1001/jama.280.11.1000.
- Deber, R., A. Topp and D. Zakus. 2004. *Private Delivery and Public Goals: Mechanisms for Ensuring That Hospitals Meet Public Objectives*. Washington, DC: World Bank. <<http://siteresources.worldbank.org/INTHSD/Resources/376278-1202320704235/GuidingPrivHospitalsDeberetal.pdf>>.
- Deber, R.B. 2014. "Thinking About Accountability." *Healthcare Policy* 10(Sp): 12–24. doi:10.12927/hcpol.2014.23932.
- Dixon-Woods, M., S. Agarwal, D. Jones, B. Young and A. Sutton. 2005. "Synthesizing Qualitative and Quantitative Evidence: A Review of Possible Methods." *Journal of Health Services Research and Policy* 10(1): 45–53.
- Dobrow, M.J., T. Sullivan and C. Sawka. 2008. "Shifting Clinical Accountability and the Pursuit of Quality: Aligning Clinical and Administrative Approaches." *Healthcare Management Forum* 21(3): 6–12. doi:10.1016/S0840-4704(10)60269-4.
- Donabedian, A. 1966. "Evaluating the Quality of Medical Care." *Milbank Quarterly* 44(3, Part 2): 166–203.
- Donabedian, A. 1980. *The Definition of Quality and Approaches to Assessment*. Ann Arbor, MI: Health Administration Press.
- Donabedian, A. 1988. "The Quality of Care: How Can It Be Assessed?" *JAMA* 260(12): 1743–48.
- Emanuel, E.J. and L.L. Emanuel. 1996. "What Is Accountability in Health Care?" *Annals of Internal Medicine* 124(2): 229–39. doi:10.7326/0003-4819-124-2-199601150-00007.
- Exworthy, M. 2010. "The Performance Paradigm in the English NHS: Potential, Pitfalls, and Prospects." *Eurohealth* 16(3): 16–19.
- Folan, P. and J. Browne. 2005. "A Review of Performance Measurement: Towards Performance Management." *Computers in Industry* 56(7): 663–80.
- Fooks, C. and L. Maslove. 2004. *Rhetoric, Fallacy or Dream? Examining the Accountability of Canadian Health Care to Citizens*. Ottawa, ON: Canadian Policy Research Networks. <www.cprn.org/documents/27403_en.pdf>.
- Freeman, T. 2002. "Using Performance Indicators to Improve Health Care Quality in the Public Sector: A Review of the Literature." *Health Services Management Research* 15(2): 126–37. doi:10.1258/0951484021912897.
- Gibberd, R. 2005. "Performance Measurement: Is It Now More Scientific?" *International Journal for Quality in Health Care* 17(3): 185–86.
- Grayson, L. and A. Gomersall. 2003. *A Difficult Business: Finding the Evidence for Social Science Reviews*. Working Paper 19. London, UK: ESRC UK Centre for Evidence Based Policy and Practice, University of London. <www.evidencenetwork.org/Documents/wp19.pdf>.
- Ham, C. 2010. "Improving Performance in the English National Health Service." *Eurohealth* 16(3): 23–25.
- Hamblin, R. 2008. "Regulation, Measurements and Incentives. The Experience in the US and the UK: Does Context Matter?" *Journal of the Royal Society for the Promotion of Health* 128(6): 291–98.

What's Measured Is Not Necessarily What Matters: A Cautionary Story from Public Health

- Hurst, J. 2002. "Performance Measurement and Improvement in Health Systems: Overview of Issues and Challenges." In P. Smith (Ed.), *Measuring Up: Improving Health System Performance in OECD Countries* (pp. 35–54). Paris, FR: Organisation for Economic Co-operation and Development.
- Hurst, J. and M. Jee-Hughes. 2001. *Performance Measurement and Performance Management in OECD Health Systems*. Paris, FR: Organisation for Economic Co-operation and Development. <[http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DEELSA/ELSA/WD\(2000\)8&docLanguage=En](http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DEELSA/ELSA/WD(2000)8&docLanguage=En)>.
- Irwin, R. 2010. "Managing Performance: An Introduction." *Eurohealth* 16(3): 15–16.
- Julnes, P.D.L. 2009. *Performance-Based Management Systems: Effective Implementation and Maintenance*. Boca Raton, FL: CRC Press.
- Kaplan, R.S. and D.P. Norton. 1996. "Using the Balanced Scorecard as a Strategic Management System." *Harvard Business Review* 74(1): 75–85.
- Kelley, E. and J. Hurst. 2006. "Health Care Quality Indicators Project: Conceptual Framework Paper." OECD Health Working Papers No. 23. Paris, FR: Organisation for Economic Co-operation and Development. <www.oecd.org/dataoecd/1/36/36262363.pdf>.
- Klazinga, N. 2010. "Health System Performance Management." *Eurohealth* 16(3): 26–28.
- Klazinga, N., K. Stronks, D. Delnoij and A. Verhoeff. 2001. "Indicators Without a Cause: Reflections on the Development and Use of Indicators in Health Care from a Public Health Perspective." *International Journal for Quality in Health Care* 13(6): 433–38.
- Kuhlmann, S. 2010. "Performance Measurement in European Local Governments: A Comparative Analysis of Reform Experiences in Great Britain, France, Sweden and Germany." *International Review of Administrative Sciences* 76(2): 331–45.
- Lebas, M.J. 1995. "Performance Measurement and Performance Management." *International Journal of Production Economics* 41(1/3): 23–35.
- Leggat, S.G., L. Narine, L. Lemieux-Charles, J. Barnsley, G.R. Baker, C. Sicotte et al. 1998. "A Review of Organizational Performance Assessment in Health Care." *Health Services Management Research* 11(1): 3–18.
- Linder, J.A., E.O. Kaleba and K.S. Kmetik. 2009. "Using Electronic Health Records to Measure Physician Performance for Acute Conditions in Primary Care: Empirical Evaluation of the Community-Acquired Pneumonia Clinical Quality Measure Set." *Medical Care* 47(2): 208–16.
- Lindsay, P., M. Schull, S. Bronskill and G. Anderson. 2002. "The Development of Indicators to Measure the Quality of Clinical Care in Emergency Departments Following a Modified-Delphi Approach." *Academic Emergency Medicine* 9(11): 1131–39.
- Lipsky, M. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York, NY: Russell-Sage Foundation Publications.
- Loeb, J.M. 2004. "The Current State of Performance Measurement in Health Care." *International Journal for Quality in Health Care* 16(Suppl. 1), i5–i9. doi:10.1093/intqhc/mzh007.
- Marmor, T.R. and J.A. Morone. 1980. "Representing Consumer Interests: Imbalanced Markets, Health Planning and the HSAs." *Milbank Memorial Fund Quarterly, Health and Society* 58(1): 125–65. doi:10.1111/j.1468-0009.2005.00431.x.
- Mattke, S., E. Kelley, P. Scherer, J. Hurst, M.L.G. Lapetra and HCQI Expert Group Members. 2006. *Health Care Quality Indicators Project: Initial Indicators Report*. Paris, FR: Organisation for Economic Co-operation and Development. <www.oecd.org/dataoecd/1/34/36262514.pdf>.
- Mays, N., C. Pope and J. Popay. 2005. "Systematically Reviewing Quantitative and Qualitative Evidence to Inform Management and Policy-Making in the Health Field." *Journal of Health Services Research and Policy* 10(1): 6–20.
- McGlynn, E.A. 2009. "Measuring Clinical Quality and Appropriateness." In P.C. Smith, E. Mossialos, I. Papanicolas and S. Leatherman (Eds.), *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects* (pp. 87–113). Cambridge, MA: Cambridge University Press.
- McGlynn, E.A. and S.M. Asch. 1998. "Developing a Clinical Performance Measure." *American Journal of Preventive Medicine* 14(Suppl. 3): 14–21.
- Monahan, P.J. 2006. *Chaoulli V Quebec and the Future of Canadian Healthcare: Patient Accountability as the "Sixth Principle" of the Canada Health Act*. Toronto, ON: C.D. Howe Institute, ISPCO Inc. <www.cdhowe.org/pdf/benefactors_lecture_2006.pdf>.
- Myers, R. and R. Lacey. 1996. "Consumer Satisfaction, Performance and Accountability in the Public Sector." *International Review of Administrative Sciences* 62(3): 331–50.

- Nolte, E., C. Bain and M. McKee. 2009. "Population Health." In P.C. Smith, E. Mossialos, I. Papanicolas and S. Leatherman (Eds.), *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects* (pp. 27–62). Cambridge, MA: Cambridge University Press.
- Pawson, R. 2002. "Evidence-Based Policy: The Promise of 'Realist Synthesis.'" *Evaluation* 8(3): 340–58.
- Pawson, R., T. Greenhalgh, G. Harvey and K. Walshe. 2005. "Realist Review – A New Method of Systematic Review Designed for Complex Policy Interventions." *Journal of Health Services Research and Policy* 10(Suppl. 1): 21–34. doi:10.1258/1355819054308530.
- Phillips, C.D., M. Chen and M. Sherman. 2008. "To What Degree Does Provider Performance Affect a Quality Indicator? The Case of Nursing Homes and ADL Change." *Gerontologist* 48(3): 330–37.
- Pink, G.H., I. McKillop, E.G. Schraa, C. Preyra, C. Montgomery and G.R. Baker. 2001. "Creating a Balanced Scorecard for a Hospital System." *Journal of Health Care Finance* 27(3): 1–20.
- Poister, T.H. and G. Streib. 1999. "Performance Measurement in Municipal Government: Assessing the State of the Practice." *Public Administration Review* 59(4): 325–35.
- Pope, C., N. Mays and J. Popay. 2006. "Informing Policy Making and Management in Healthcare: The Place for Synthesis." *Healthcare Policy* 1(2): 43–48.
- Propper, C., M. Sutton, C. Whitnall and F. Windmeijer. 2008. "Did 'Targets and Terror' Reduce Waiting Times in England for Hospital Care?" *B.E. Journal of Economic Analysis & Policy* 8(2): 1935–1682. doi:10.2202/1935-1682.1863.
- Provincial Auditor of Ontario. 2003. *Annual Report of the Office of the Provincial Auditor of Ontario*. Toronto, ON: Office of the Provincial Auditor of Ontario. <www.auditor.on.ca/en/reports_2003_en.htm>.
- Public Health Foundation. 2009. *Performance Management in Public Health: A Literature Review*. Seattle, WA: Turning Point. <www.phf.org/resourcestools/Documents/PMcliteraturereview.pdf>.
- Schwartz, R. 2011. "Bridging the Performance Measurement-Management Divide? Editor's Introduction." *Public Performance & Management Review* 35(1): 103–107. doi:10.2753/PMR1530-9576350105.
- Schwartz, R. and R. Deber. 2016. "The Performance Measurement – Management Divide in Public Health." *Health Policy* 120(3): 273–80. doi:org/10.1016/j.healthpol.2016.02.003.
- Shortt, S.E.D. and J.K. Macdonald. 2002. "Toward an Accountability Framework for Canadian Healthcare." *Healthcare Management Forum* 15(4): 24–32.
- Smith, P.C. 2002. "Performance Management in British Health Care: Will It Deliver?" *Health Affairs* 21(3): 103–15. doi:10.1377/hlthaff.21.3.103.
- Smith, P.C., E. Mossialos, I. Papanicolas and S. Leatherman (Eds.). 2009. *Performance Measurement for Health System Improvement: Experiences, Challenges and Prospects*. Cambridge, MA: Cambridge University Press.
- Solberg, L.I., G. Mosser and S. McDonald. 1997. "The Three Faces of Performance Measurement: Improvement, Accountability, and Research." *International Journal for Quality in Health Care* 23(3): 135–47.
- Stoto, M.A. 2014. "Population Health Measurement: Applying Performance Measurement Concepts in Population Health Settings." *eGEMS* 2(4): 1132. doi:10.13063/2327-9214.1132.
- Ten Asbroek, A.H., O.A. Arah, J. Geelhoed, T. Custers, D.M. Delnoij and N.S. Klazinga. 2004. "Developing a National Performance Indicator Framework for the Dutch Health System." *International Journal for Quality in Health Care* 16(Suppl. 1): i65–i75.
- Townley, B. 2005. "Critical Views of Performance Measurement." In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (Vol. 1, pp. 565–71). Amsterdam, The Netherlands: Elsevier Academic Press.
- Tregunno, D., R. Baker, J. Barnsley and M. Murray. 2004. "Competing Values of Emergency Department Performance: Balancing Multiple Stakeholder Perspectives." *Health Services Research* 39(4): 771–92.
- US Government Accountability Office. 2005. *Performance Measurement and Evaluation: Definitions and Relationships*. Washington, DC: Author.
- Veillard, J.H.M. 2012. "Performance Management in Health Systems and Services: Studies on Its Development and Use at International, National/Jurisdictional, and Hospital Levels." (PhD), University of Amsterdam, Amsterdam, Netherlands. Retrieved October 31, 2016. <http://jeremyveillardresearch.com/thesis/Veillard_PhD_Thesis.pdf>.
- Weir, E., N. d'Entremont, S. Stalker, K. Kurji and V. Robinson. 2009. "Applying the Balanced Scorecard to Local Public Health Performance Measurement: Deliberations and Decisions." *BMC Public Health* 9(127). doi:10.1186/1471-2458-9-127.
- Zimmerman, S.V. 2005. *Mapping Legislative Accountabilities. Health Care Accountability Papers – No.5, Health Network*. Ottawa, ON: Canadian Policy Research Networks. Ottawa, ON: Canadian Policy Research Networks. <www.cprn.org/documents/35190_en.pdf>.