

Performance Measurement in Healthcare: Part I – Concepts and Trends from a State of the Science Review

La mesure du rendement dans les soins de santé :
Partie I – Concepts et tendances issus d'un examen
de l'état de la science



by CAROL E. ADAIR, MSC, PHD
*Associate Professor, Departments of Community Health Sciences and Psychiatry
University of Calgary, Calgary, AB*

ELIZABETH SIMPSON, MSC
*Health Research Consultant
Red Deer, AB*

ANN L. CASEBEER, MPA, PHD
*Associate Professor, Department of Community Health Sciences
Associate Director, Centre for Health and Policy Studies
University of Calgary, Calgary, AB*

JUDITH M. BIRDSELL, MSC, PHD
*Principal Consultant, ON Management Ltd.
Adjunct Associate Professor, Haskayne School of Business
University of Calgary, Calgary, AB*

KATHARINE A. HAYDEN, MLIS, MSC, PHD
Associate Librarian, Information Resources
University of Calgary, Calgary, AB

STEVEN LEWIS, MA
President, Access Consulting Ltd.
Adjunct Professor, Department of Community Health Sciences
University of Calgary, Calgary, AB

Abstract

Objective: Performance measurement is touted as an important mechanism for organizational accountability in industrialized countries. This paper describes a systematic review of business and health performance measurement literature to inform a research agenda on healthcare performance measurement.

Methods: A search of the peer-reviewed business and healthcare literature for articles about organizational performance measurement yielded 1,307 abstracts. Multi-rater relevancy ratings, citation checks, expert nominations and quality ratings resulted in 664 articles for review. Key themes were extracted from the papers, followed by multi-reader validation. Information was supplemented with grey literature.

Results: The performance literature was diverse and fragmented, and relevant evidence was difficult to locate. Most literature is non-empirical and originates from the United States and the United Kingdom. No agreement on definitions or concepts is evident within or across disciplines. Study quality is not high in either field. Performance measurement arose in public services and business at about the same time. The evolution of thought on performance measurement ranges from unfettered enthusiasm to sober reassessment.

Conclusions: The research base on performance measurement is in its infancy, and evidence to guide practice is sparse. A coherent multidisciplinary research agenda on the topic is needed.

Résumé

Objectif : La mesure du rendement est considérée comme étant un important mécanisme de responsabilisation organisationnelle dans les pays industrialisés. Ce document décrit un examen systématique de la littérature sur la mesure du rendement dans les domaines des affaires et des soins de santé afin d'éclairer l'élaboration de programmes de recherche sur la mesure du rendement dans les soins de santé.

Méthodes : Une recherche d'articles sur le rendement organisationnel dans des publications sur les affaires et les soins de santé évaluées par les pairs a permis de repérer

1 307 résumés. Des évaluateurs multiples ont déterminé la pertinence de chaque article; on a vérifié les citations, désigné des experts et établi des cotes de qualité. Ces critères ont permis de réduire à 664 le nombre d'articles à examiner. On a dégagé des thèmes clés des documents, puis on les a fait valider par des lecteurs multiples. On a aussi eu recours à la littérature grise pour compléter les données.

Résultats : La littérature sur le rendement était diversifiée et fragmentée, et les preuves pertinentes difficiles à repérer. La majeure partie de la littérature est de nature non empirique et provient des États-Unis et du Royaume-Uni. On ne décèle aucun consensus quant aux définitions ou aux concepts, ni entre les disciplines ni à l'intérieur de celles-ci. La qualité des études laisse à désirer, et ce, dans les deux domaines. La mesure du rendement a fait son apparition dans la fonction publique et dans le monde des affaires à peu près en même temps. L'évolution de la pensée sur la mesure du rendement varie d'un enthousiasme sans réserve à une réévaluation sobre.

Conclusions : La recherche sur la mesure du rendement en est encore à ses balbutiements, et les preuves pouvant guider la pratique sont rares. Un programme de recherche multidisciplinaire et cohérent sur le sujet s'avère nécessaire.

—

“... no one knows the extent to which some of the expenditures have brought good or ill to the recipients, and certainly no one knows whether better service might not have been achieved for smaller outlay more intelligently applied.”

– Pennsylvania Children's Commission, 1920s
(cited in Beck et al. 1998: 164)

“Collectively (and perhaps belatedly) we have recognized the most important issue facing the health service is not how it should be organized or financed, but whether the care it offers actually works.”

– Walshe and Ham, 1997 (Perkins 2001: 9)

IN THE CONTEXT OF RISING HEALTHCARE EXPENDITURES, PERFORMANCE MEASUREMENT (PM) is increasingly integral to accountability. Healthcare accountability mechanisms have traditionally included business planning, annual reporting and contracting (Alberta Government 1998). In recent years a richer sense of accountability has emphasized the achievement of goals effectively and efficiently and has stimulated PM. PM has been described most simply as “the use of statistical evidence to determine progress toward specific defined organizational objectives” (State of California 2003), although more comprehensive definitions exist. The PM process

typically involves four stages – conceptualization and strategy; measures selection and development; data collection and analysis; reporting and use – and can occur at multiple levels of organizations and systems.

The literature includes reports on performance measurement initiatives across the healthcare spectrum from primary (e.g., Proctor and Campbell 1999) through tertiary care (e.g., Rowan and Angus 2000), public health (e.g., Corso 2000) and the voluntary sector (e.g., Dunn and Matthews 2001) that have been mounted in response to demands from governments, other payers, consumers, proponents of evidence-based practice and accreditation organizations. Substantial resources have been invested in PM system development from the policy level through front-line care delivery (Goddard et al. 1999). Given the scale of investment, a commensurate level of research on PM would be expected. However, because scientific and experiential information about PM spans multiple sources and disciplines, there is no easy way to identify and summarize relevant evidence.

We conducted a systematic review of the business and health PM literature to inform a research agenda on healthcare PM. Our intent was to draw broad issues and themes from diverse sources. This paper outlines methods, summarizes general PM concepts and trends from business/management and healthcare sources and offers recommendations for a research agenda to produce evidence for practice. Part II (which will be published next issue) details findings from the review according to the four stages of PM and outlines implications for practice, including common problems and solutions suggested by PM authors. It also provides a brief update on new developments at the policy level in PM in Canada.

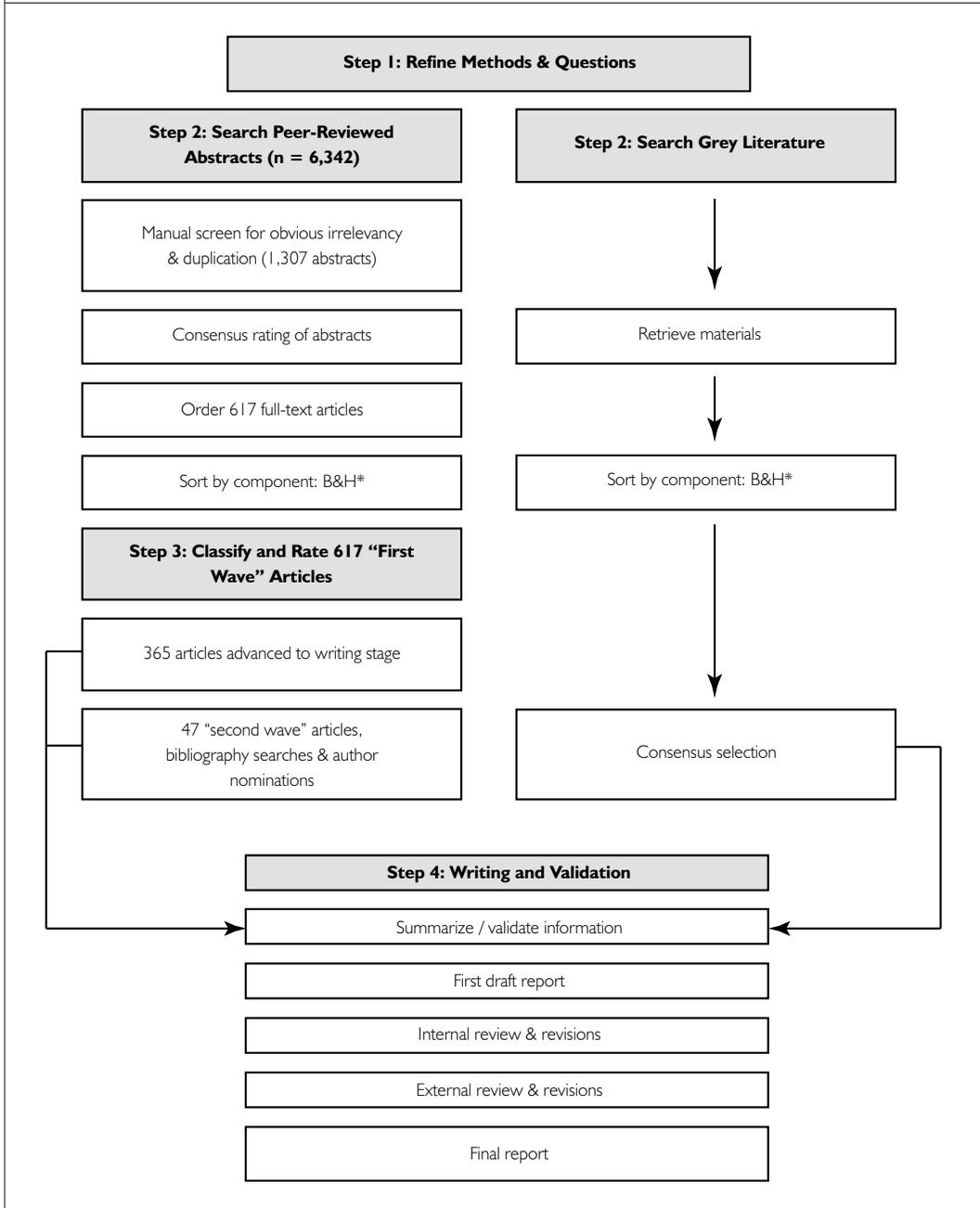
Methods

Methods for systematic reviews of clinical effectiveness studies (ScHARR 1997; Clarke and Oxman 2001) guided our review process, but were not entirely suitable for a broad policy subject. We based our approach on the principle of replicability; in the end, the review was a hybrid of systematic and narrative review methods. There were four steps in the process: (a) refining the review questions, (b) searching for and selecting articles, (c) classifying and rating the articles for quality and (d) writing and validation.

First, to refine the review questions, we received feedback on initial drafts from 22 (54%) of 41 healthcare decision-makers surveyed from across Canada, and in a focus group with four of them. Participants strongly endorsed the need for the review and provided suggestions for scope and content. Specifically, they called for inclusion of the business/management literature based on an interest in cross-sector comparisons and business innovations that might apply to healthcare.¹ The final questions focused on predominant models and frameworks, evidence for impact and recommendations for research.

Performance Measurement in Healthcare:
Part I – Concepts and Trends from a State of the Science Review

FIGURE 1. The Article Selection and Review Process



*B = Business, H = Health

Second, a professional librarian (KH) designed and ran searches of the business and healthcare peer-reviewed literature and the grey literature. The peer-reviewed literature searches included the databases ABI Inform, Business Index ASAP, PsycInfo, ABI Inform, Medline, Embase, Cinahl and Web of Science. Key words varied by database, but were close approximations to *performance measurement*, *performance measurement system* and *health system performance*. Searches were limited to English-language materials for 1992–2002 for health and 1997–2002 for business (to focus on more recent developments). Grey literature provided an appropriate policy and practice context; sources included PM-related Internet sites, conference proceedings, newsletters, press clippings and periodical indexes. The article selection and review process is shown in Figure 1.

The initial 6,342 abstracts from the peer-reviewed literature were screened manually by CA for obvious irrelevancy, since the search terms yielded abstracts about unrelated topics such as occupational and athletic performance. After removal of duplicates, teams of three investigators for each of the two fields independently rated the 1,307 remaining abstracts from the two fields for relevancy based on a criterion statement focusing on organizational-level PM concepts or evidence and research recommendations that had been pre-tested on 50 articles. Full-text papers that rated a “yes” by two of three raters were retained for the review, and the percentage of agreement for this criterion was high (85.3% $K = .77$; $p = .000$). This “first wave” of articles numbered 617.

In the third step, six readers in two teams read, classified and rated the papers for quality. Articles were classified by publication year, country, organizational level, patient population (health only) and type of research (non-empirical or empirical). Reading teams also flagged frameworks, definitions, innovations and recommendations for research. Quality rating scales were developed, pre-tested on 20 papers for each article type and revised. Non-empirical and empirical papers were rated on 10- and 15-point scales, respectively. Inter-rater agreement (unblinded) was high (intra-class correlation coefficients were .88 (95% CI .82–.92, $p = .000$) for empirical articles and .92 (95% CI .88–.94, $p = .000$) for non-empirical articles. Papers with ratings above the mean were advanced to the writing stage ($n = 365$), although the cut-off was not strict and some subjective judgments were made at that stage about inclusion based on the pertinence of material to the major themes that were identified to that point. Reference lists for the highest-rated papers were searched, and 21 authors of the highest-rated papers (of 42 contacted) nominated an additional 39 “best” papers on the topic. Forty-seven more papers were included as a result of these citation searches and author nominations (“second wave” activities). All papers were tracked using Access and EndNote databases.

In step four, one reader for each field prepared a written summary, which was validated by a second reader. The principal investigator then wrote a comprehensive draft report after reading all first- and second-wave papers and readers’ summaries. Grey

literature was incorporated at this stage if it provided important context or additional insight on emerging themes. The draft review was then read and revised by team members. Feedback from two external reviewers was also incorporated.

Results

The “first wave” group of selected peer-reviewed articles included 81 (22%) from the business literature and 284 (78%) from health. First-author country of residence was the United States (69%), United Kingdom (15.3%), Australia (3.6%), Canada (3.6%) and all others (8.5%). The ratio of non-empirical to empirical papers was about two to one (69.3% vs. 30.7%). The business literature had more empirical papers (40.7%) than the health literature (28.0%), which might be attributable to the difference in search timeframes. Mean quality ratings were moderate and variable overall: 9.1 out of 15 (SD 2.93) for empirical papers and 6.6 out of 10 (SD 1.81) for non-empirical papers, even after selection for relevancy and quality. Mean quality ratings were not significantly different for business vs. health papers for non-empirical papers (6.1 vs. 6.0; $p = .66$) or empirical papers (8.8 vs. 9.2; $p = .58$).

Concepts

The concept *performance measurement* has no agreed-upon definition in or across the literature reviewed. It sits within a dizzying array of related theoretical ideas, research- and practice-based tools, initiatives and rhetoric. Nevertheless, the generally implied purpose of PM across all materials was the pursuit of excellence in organized human enterprise. *Performance measurement systems* are also conceptualized in multiple ways with many synonyms, including “organizational performance assessment system” (Leggat et al. 1998), “outcomes management system” (Jones and Brown 2001) and even the more general “continuous measurement process” (Nadzam and Nelson 1997; Rosenblatt et al. 1998). Definitions for all concepts were extracted to an omnibus glossary in the main report (Adair et al. 2003). For each concept, the principal investigator (with concurrence from the review team) selected the definition that seemed to fit best with the most frequently implied meaning across the body of literature (Table 1).

With respect to its function in organizations, some authors characterize PM as only one part of a larger health information framework (Schneider et al. 1999). Others view it as an overarching activity that encompasses a variety of other related initiatives (Neely et al. 1995). PM is described by some as an internal organizational activity only, while others consider its primary purpose to be external, e.g., “PM has become a preferred means of ensuring external accountability” (McCorry et al. 2000: 636). Still others recognize that performance measures have utility for both internal and external reporting (e.g., Trabin et al. 1997).

TABLE 1. Selected definitions for healthcare performance measurement terms

TERM	DEFINITION
Performance	What is done and how well it is done to provide healthcare (JCAHO 2002)
Performance Measurement*	The use of both outcomes and process measures to understand organizational performance and effect positive change to improve care (Nadzam and Nelson 1997)
Performance Indicator**	Markers or signs of things you want to measure but which may not be directly, fully or easily measured (Alberta Government 1998)
Performance Measure	A quantitative tool, such as rate, ratio or percentage, that provides an indication of an organization's performance in relation to a specified process or outcome (JCAHO 2002)
Process Measure	A measure focusing on a process that leads to a certain outcome, meaning that a scientific basis exists for believing that the process, when executed well, will increase the probability of achieving a desired outcome (JCAHO 2002)
Outcome Measure	Not simply a measure of health, well-being or any other state; rather, it is a change in status confidently attributable to antecedent care (intervention) (Donabedian 1968)

*Like this one, many performance measurement definitions included the use of measurement results for organizational improvement that implies performance *management* – and resulted in these two terms being used interchangeably in the literature.

**Despite these distinctions, the terms performance measure and performance indicator were usually used interchangeably in most general discussions about PM because either or both are used in PM.

The relationship between PM and related activities, such as total quality management (TQM), is an area of great conceptual confusion. For example, contemporary quality improvement processes are seen to rely on PM, but PM exercises are also described as having quality improvement activities as a key component. These endeavours seem to be increasingly blurred with time. Accreditation processes now emphasize the use of performance measures, benchmarking activities now often link directly to quality award schemes and TQM is touted as a mechanism to take action on PM results. The lack of conceptual clarity is not surprising, given the breadth of organizational types (e.g., for profit/not for profit) and levels (individual to system) at which PM is applied, and the diversity of disciplines producing PM-related research (at least 17 sub-disciplines were identified). Studies across disciplines were rare.

Performance measures for health have been developed for the three classical components of care defined by Donabedian (1988) (structure, process and outcomes) and at

all care levels, from patient to population (Tansella and Thornicroft 1998; McIntyre et al. 2001). At the patient level, PM manifests primarily as measurement of the process or outcomes of treatment; at the service or program level, as program evaluation (measurement of program attributes for planning); at the system level, as a mechanism for organizational control and ensuring efficient use of resources; and at the population level, as the collection and reporting of high-level information for broad planning, policy making and accountability (Thompson and Harris 2001; Coyne 2002; Studnicki et al. 2002). While there are some commonalities in PM across levels, and data for one level are often rolled up for higher-level use, stakeholder priorities for and uses of PM data can be very different by level (Goddard et al. 1998; Legnini et al. 2000).

Current practice in healthcare

Table 2 provides a summary of the relationships among PM-related initiatives and tools as they are described in the healthcare literature according to the predominant level of activity and how the various initiatives seem to have evolved over time.

The business PM literature

In business, PM was traditionally a top-down activity for organizational control (Lebas 1995; Smith and Goddard 2002). In a paper considered profoundly influential, Eccles (1991) predicted widescale redesign of business PM. The idea has now evolved into a more holistic, organizationwide, strategic management approach (Smith and Goddard 2002). Neely (1999) describes an exponential increase in the number of PM publications in the 1990s and the diffusion of PM language in corporate annual

TABLE 2. Healthcare improvement activities and trends		
HEALTH SYSTEM LEVEL*	TRADITIONAL ACTIVITIES	CONTEMPORARY ACTIVITIES
Inter-system, international	Comparisons of basic statistics (e.g., life expectancy) (R)	PM with aggregate indicators compared across nations (R)
System (total health system), e.g., provincial health department, Regional Health Authority, Health Maintenance Organization	Global financial measures/ budgeting (P) Administrative data-based health services research (R)	PM across services and programs (P) Outcome-oriented health services research including economic analysis (R)

TABLE 2. CONTINUED

HEALTH SYSTEM LEVEL*	TRADITIONAL ACTIVITIES	CONTEMPORARY ACTIVITIES
Program/Service unit organization	Single program evaluation studies (P/R) Accreditation (P) Quality Assurance (P) Simple comparative benchmarking (P)	PM within programs (P) Organizational development and leadership enhancement (P/R) Multi-service evaluations (P/R) Total Quality Management (P) Accreditation (including quality improvement and PM) (P) Provider profiling (organizational) (P) External audits (a variety in US & UK) (P) Portfolio management measurement tools, e.g., Balanced Scorecard (P/R) Benchmarking (P/R) Management quality awards programs, e.g., Baldrige (P)
Individual client/patient or provider	Clinical judgment and subjective impression of improvement based (usually) on physiologic measures (P) Client satisfaction measurement (P)	PM, including routine process and outcome measurement (P) Evidence-based practice (clinical practice guidelines, care pathways/algorithms, systematic reviews) (R/P) Clinical governance (P) Provider profiling (individual) (P) Individual physician accreditation (US; AMA) (P) Professional development and recertification (CME, audit and feedback, etc.) (P) Outcome effectiveness research, including patient-level longitudinal studies using functioning/quality of life (R)

Developed from the ideas of Bartlett 1997, Grol et al. 2002 and others.

*Where an activity or tool is applicable at more than one level, we have classified it according to the level where it originated or is used most predominantly: R = predominantly a research activity; P = predominantly a practice activity

reports. This explosion of interest is attributed to a need for organizational change spurred by globalization and increasingly competitive markets, as well as an increase in active marketing by PM system and service vendors (Smith et al. 1997; Neely 1999; Malmi 2001).

The non-empirical business literature is characterized by the promotion of largely proprietary systems and the reporting of practical problems (Holloway 2001). Holloway (2001: 170) charges that much of the literature “merely proselytizes for particular models or approaches,” contains “a litany of failed or abandoned PM systems” and provides little in the way of analysis of the problematic side of PM. Busby and Williamson (2000: 336) describe the attitude towards PM in business as “simply an unquestioned belief that it leads to positive improvement.” Very few organizations have actually quantified change in performance associated with the implementation of PM (Mooraj et al. 1999), and cost–benefit studies are non-existent (Neely et al. 1995).

Few examples of empirical studies of PM system effectiveness were found in the business literature, and most were case series or surveys. In a New Zealand study, Upton (1998) investigated the association between performance measures and organizational performance. Overall, organizations using non-financial measures of the type found in newer PM schemes performed better. Unfortunately, the finding was based on the self-report of one respondent from each of 85 firms, rather than more rigorous evidence. In a longitudinal survey, Fawcett and Cooper (1998) found that between 1989 and 1994, higher-performing companies reported increasing use of measures. Key methodological challenges noted in the business literature included the tendency to attribute any and all organizational improvement to the PM system when other, uncontrolled factors may be present, and the problem of endogeneity (the same measures that make up the PM system are used to evaluate it) (Holloway 2001; McAdam and Bannister 2001). High-quality articles (both non-empirical and empirical) about PM in business that emerged from the review include Neely et al. 1995; Neely 1999, 2000; Bourne et al. 2000; and Holloway 2001.

The health PM literature

PM arose in the public service and private sectors at virtually the same time, unlike other management innovations that typically originate in the private sector and then subsequently move into the public realm (Smith and Goddard 2002). This situation may result, at least in part, from an urgent need for PM in organizational structures where no natural market exists. Business models and tools for PM (e.g., the Balanced Scorecard) have been applied in public services including health, but the public sector has also developed its own systems, instruments and analysis techniques (e.g., Khim and Hian 2001; Smith and Goddard 2002). Although there is no doubt that PM has been implemented more extensively in business (Malmi 2001), the chorus of calls in

policy and research documents for its application in public and health services has been pronounced since the late 1980s (e.g., Relman 1988).

Several stages of the evolution of thought about PM emerged from the healthcare literature. During the 1980s and early 1990s, the calls for PM were abundant, along with unfettered enthusiasm for its promise for improving healthcare. This stage could be characterized as the *performance measurement imperative* (Relman 1988; Williams et al. 1992; Hall 1996). The mid-1990s brought rapid, uncoordinated proliferation of measures and systems and, in the United States, a burgeoning support industry. This *proliferation and fragmentation* stage has been colourfully described by Hermann et al. (2000: 149) as a time of “letting many flowers bloom.” The later 1990s heralded a distinct *sober reassessment/reflection* period. This stage was likely stimulated by practice experiences that revealed the great cost and complexity of system implementation, multiple failures and lack of standardization that impaired comparability. At this stage, some authors questioned whether PM, or some aspects of it, were even useful or feasible (Epstein 1995; Baumgarten 1998; Jencks 2000). More recently, the literature reflects a trend towards *consensus and initial solutions*, including acknowledgment of the complexities of PM, a redirection of energies towards more thoughtful, problem-solving approaches in practice (Viccars 1998; Jarvi et al. 2002; Mannion and Davies 2002; Smee 2002) and calls for broader consensus about PM at all levels (e.g., Bishop and Pelletier 2001; McLoughlin et al. 2001).

The literature on the effectiveness of healthcare PM is also sparse, but the general view is that there is little evidence that PM schemes have had either indirect impact on behaviour or direct impact on the quality of patient care or health gain *at the system level*. While many isolated examples of successful quality improvement initiatives can be found, few evaluations of organizationwide PM exist. Among the notable reported successes on a large scale include the transformation of the Veterans Health Administration system in the United States in the 1990s (Kizer et al. 2000) and improvements in access to primary healthcare in the United Kingdom (Pickin et al. 2004), published since our review. The extent to which these system improvements are attributable to the PM initiatives that made it possible to chart progress is not reported.

Many initial claims about PM effectiveness are purely anecdotal. For example, in the context of accreditation in Australia, Collopy (1998: 175) reports that the Australian Council on Healthcare Standards “has evidence of numerous alterations in practice and improvement in patient care induced by indicator monitoring,” and in the context of an application of the Balanced Scorecard (BSC) at Duke Children’s Hospital, Voelker et al. (2001) provide an example of reduced cost per case, profitability, patient satisfaction, reduced length of stay and reduced readmissions. Grol et al. (2002) contend that there is more evidence about the effectiveness of traditional clinical-level interventions, such as audit/feedback and continuing medical education, than about newer strategies for health services improvement, such as portfolio learn-

ing and organizational development. In a review of PM models, Leggat et al. (1998: 8) conclude that most were “in the early stages of development with no hard evidence as to their long-term impact.”

Some weak evidence of positive impact comes from PM-system user reports. These range from general satisfaction with the PM system to user opinions about specific changes, such as in administrative or care processes (Lemieux-Charles et al. 2000), to reports of improvement resulting from PM-based action (Turpin et al. 1996). Smith (1993) also provides some examples of reported potential for adverse effects of PM derived from case study interviews. However, uncontrolled case studies and opinions about systems have significant limitations in supporting causal inferences about the impact of PM on health services improvement itself (Hulley et al. 2001). These study designs, as their authors frequently acknowledge, cannot rule out other explanations for their findings, including historical trends and reporting or selection biases (Turpin et al. 1996; Smith 1993).

Our review uncovered only a handful of empirical studies that examined PM effectiveness more directly by including actual measures of health services outcomes. Studies supporting and refuting attribution of reductions in coronary bypass surgery mortality to publication of performance data are reviewed by Jencks (2000). Longo et al. (1997) report positive changes in response to indicator information in the context of obstetric care, and Kazandjian and Lied (1998) found positive impact of participation in a multi-hospital PM project on caesarean-section rates. Finally, in a time-series study, Petitti et al. (2000) demonstrated that even in the absence of incentives, physician-specific profiling of process and outcomes positively influenced diabetes care organizationwide.

Overall, the clarity of ideas in the non-empirical literature and the quality of empirical work is stronger in more recent health literature. However, empirical studies remain varied with respect to questions and methods, and no coherent research direction is evident. Multidisciplinary papers were rare. Of particular note are the variety of perspectives and approaches in the US literature, the depth and general theoretical quality of the UK literature and the relative paucity of Canadian literature. At least two dozen high-quality papers on healthcare PM were identified and listed in an appendix to the full report (Adair et al. 2003). These are recommended as key readings in the field for researchers and decision-makers (Adair et al. 2003).

Discussion

Our review was challenged by the diverse sources and locations of the knowledge and practice base on PM. It required a breadth of search and synthesis far greater than the scope of a systematic review on a clinical question. This situation limited the depth of the review on any single aspect of PM, but allowed us to identify broader themes and

issues that were generalizable across sources. More in-depth reviews of specific topics and the many strands of relevant empirical literature are warranted. The diversity of terms and discipline-specific perspectives also necessitated that we cast a large search net (favouring sensitivity over specificity), resulting in a labour-intensive review process. Of particular note, there was almost no convergence on author-nominated papers and only 33% overlap between nominated and selected papers, further evidence that a bounded and identifiable PM literature does not yet exist. Despite these limitations, the review identified a rich set of PM research and practice-related issues, which themselves raised two fundamental questions for our team.

First, there seems to be a simple unifying concept of performance, defined as how an entity does in relation to articulated goals and/or other similar entities. Performance in this context is entirely relative; its meaning is rooted in the gap between the “is” and the “ought.” There can be measures that do not connote performance; indeed, the healthcare system is full of measures that merely describe quantities. A very frequent sentiment in both the business and health literatures is that PM must become more about *management* than just measurement, more about action rather than just awareness.

Second, the issue of causality is a difficult one to disentangle, that is, the relative roles of PM systems and data, leadership and management factors in producing change. PM systems are essential to *documenting* improvement; it is not clear how and whether they are supposed to cause it, or whether a culture of improvement and focused achievement of performance, united by well-understood goals, can lead to good performance even in the absence of much performance measurement. The causal assumptions of authors whose articles were reviewed were rarely made explicit. A unifying theory for PM that encompasses this complexity for both fields – and perhaps across fields – does not exist and is greatly needed. Such questions require a greater level of research sophistication in both theory and methods.

Conclusion

An unequivocal finding of this review was that the science of PM is in its infancy, lagging far behind practice in both healthcare and business. The necessity of PM and its potential benefits are widely supported, but rhetoric and good intentions appear to outweigh demonstrations of successful implementation. Where PM is implemented, there is little substantive evidence of positive impact on decision-making, improvement in health services delivery or health outcomes. Many authors advocate generally for more PM research; others call for a specific research program or agenda (Neely et al. 1995; McLoughlin et al. 2001). For example, Kaplan and Norton (2001: 160) suggest that “researchers can now begin a systematic research program, using multiple research methods to explore the key factors in implementing more effective measurement and

management systems.” Johnson (1998: 253) advocates an international research effort: “The shape of a research agenda, so badly needed in this field, should not depend on one vision of future healthcare delivery. The utility of research will be enhanced if it is independent of any particular form of healthcare delivery. Furthermore, there may be important, practical reasons to plan and orchestrate the research internationally.”

The good news is that more rigorous studies of system-level PM are starting to appear. A stellar example is the recent publication by Beck et al. (2005), published after our review was complete. While this study found no evidence that a PM-related intervention (report card feedback on acute myocardial infarction care) changed practice, more research of this calibre will no doubt identify specific PM approaches and mechanisms that can achieve health gains. In our view, effective research in this field needs to go beyond the emergence of quality research by individual research teams. Advancement of the field depends on a comprehensive and coordinated research agenda, including programmatic funding, standard nomenclature, theory development, international and transdisciplinary projects, innovative research–practice partnerships and mechanisms to optimize knowledge transfer and exchange.

NOTES

1. The mental health literature was separated as a special case study of the health literature. Findings specific to that review are not presented in this paper.

Correspondence may be directed to: Carol E. Adair, MSc., PhD, Associate Professor, Depts. of Psychiatry and Community Health Sciences, Room 124, Heritage Medical Research Building, 3330 Hospital Dr. NW, Calgary, Alberta T2N 4N1, Tel: 403-210-8805, Fax 403-944-3144

ACKNOWLEDGMENTS

The State of the Science Review was funded by the Alberta Heritage Foundation for Medical Research, and significant in-kind support was received from the Alberta Mental Health Board. Thanks are due to K. Omelchuk, H. Gardiner, S. Newman, S. Clelland, A. Beckie, K. Lewis-Ng, I. Frank, J. Osborne, D. Ma, X. Kostaras and O. Berze for their assistance on parts of the broader review. T. Sheldon and C. Baker provided methodologic consultation, and E. Goldner and S. Lewis reviewed the main report. Findings have been presented in part at Academy Health, Nashville, Tennessee, June 2003; World Psychiatric Association, Paris, France, July 2003; International Conference on the Scientific Basis of Health Services, Washington, DC, September 2003; and American Evaluation Association, Reno, Nevada, November 2003.

REFERENCES

Adair, C., L. Simpson, J.M. Birdsell, K. Omelchuk, A. Casebeer, H.P. Gardiner, S. Newman, A. Beckie, S. Clelland, K.A. Hayden and P. Beausejour. 2003. *Performance Measurement Systems in*

Health and Mental Health Services: Models, Practices and Effectiveness. A State of the Science Review. Calgary: University of Calgary.

Aidemark, L. 2001. "The Meaning of Balanced Scorecards in the Health Care Organisation." *Financial Accountability and Management* 17(1): 23–40.

Alberta Government. 1998. *Results-Oriented Government: A Guide to Strategic Planning and Performance Measurement in the Public Sector.* Edmonton: Author.

Bartlett, J. 1997. "Treatment Outcomes: The Psychiatrist's and Health Care Executive's Perspectives." *Psychiatric Annals* 27(2): 100–3.

Baumgarten, A. 1998. "Performance Measurement: A Good Idea in Theory." *Minnesota Medicine* 81(1): 6–7.

Beck, C., R. Hugues, J.V. Tu and L. Pilotte. 2005. "Administrative Data Feedback for Effective Cardiac Treatment. AFFECT, a Cluster Randomized Trial." *Journal of the American Medical Association* 294(3): 309–17.

Beck, S., P. Meadowcroft, M. Mason and E.S. Kiely. 1998. "Multiagency Outcome Evaluation of Children's Services: A Case Study." *Journal of Behavioral Health Services and Research* 25(2): 163–76.

Bishop, W. and L. Pelletier. 2001. "Interview with a Quality Leader: Janet Corrigan on the Institute of Medicine and Healthcare Quality." *Journal for Healthcare Quality* 23(5): 21–24.

Bourne, M., J. Mills, M. Wilcox, A. Neely and K. Platts. 2000. "Designing, Implementing and Updating Performance Measurement Systems." *International Journal of Operations and Production Management* 20(7): 754–71.

Busby, J. and A. Williamson. 2000. "The Appropriate Use of Performance Measurement in Non-Production Activity: The Case of Engineering Design." *International Journal of Operations and Production Management* 20(3): 336.

Clarke, M. and A. Oxman. 2001. *Cochrane Reviewers' Handbook, The Cochrane Collaboration.* Chichester, UK: John Wiley and Sons/The Cochrane Collaboration.

Collopy, B. 1998. "Health-Care Performance Measurement Systems and the ACHS Care Evaluation Program." *Journal of Quality in Clinical Practice* 18(3): 171–76.

Corso, L., P. Wiesner, P. Halverson and K. Brown. 2000. "Using the Essential Services as a Foundation for Performance Measurement and Assessment of Local Public Health Systems." *Journal of Public Health Management and Performance* 6(5): 1–18.

Coyne, J. 2002. "The World Health Report 2000: Can Health Care Systems Be Compared Using a Single Measure of Performance?" *American Journal of Public Health* 92(1): 30–33.

Donabedian, A. 1968. "Promoting Quality through Evaluating the Process of Patient Care." *Medical Care* 6(6): 181–202.

Donabedian, A. 1988. "The Quality of Care: How Can It Be Assessed?" *Journal of the American Medical Association* 260(12): 1743–48.

Dunn, B. and S. Matthews. 2001. "The Pursuit of Excellence Is Not Optional in the Voluntary Sector, It Is Essential." *International Journal of Health Care Quality Assurance* 14(2–3): 121–25.

Eccles, R. 1991. "The Performance Measurement Manifesto." *Harvard Business Review* 69(1): 131–37.

Performance Measurement in Healthcare:
Part I – Concepts and Trends from a State of the Science Review

- Epstein, A. 1995. "Performance Reports on Quality – Prototypes, Problems and Prospects." *New England Journal of Medicine* 333(1): 57–61.
- Fawcett, S. and M. Cooper. 1998. "Logistics Performance Measurement and Customer Success." *Industrial Marketing Management* 27(4): 341–57.
- Goddard, M., R. Mannion and P. Smith. 1998. "Performance Indicators. All Quiet on the Front Line." *Health Service Journal* 108: 24–26.
- Goddard, M., R. Mannion and P. Smith. 1999. "Assessing the Performance of NHS Hospital Trusts: The Role of 'Hard' and 'Soft' Information." *Health Policy* 48(2): 119–34.
- Grol, R., R. Baker and F. Moss. 2002. "Quality Improvement Research: Understanding the Science of Change in Health Care." *Quality and Safety in Health Care* 11: 110–11.
- Hall, J. 1996. "The Challenge of Health Outcomes." *Journal of Quality in Clinical Practice* 16(1): 5–15.
- Hermann, R., H. Leff, R.H. Palmer, D. Yang, T. Teller, S. Provost, C. Jakubiak and J. Chan. 2000. "Quality Measures for Mental Health Care: Results from a National Inventory." *Medical Care Research and Review* 57(Suppl. 2): 136–54.
- Holloway, J. 2001. "Investigating the Impact of Performance Measurement." *International Journal of Business Performance Management* 3(2–4): 167–80.
- Hulley, S.B., S.R. Cummings, W.S. Browner, D. Grady, N. Hearst and T.B. Newman. 2001. *Designing Clinical Research* (2nd ed.). Philadelphia: Lippincott Williams and Wilkins.
- Jarvi, K., R. Sultan, A. Lee, F. Lussing and R. Bhat. 2002. "Multi-Professional Mortality Review: Supporting a Culture of Teamwork in the Absence of Error Finding and Blame-Placing." *Hospital Quarterly* 5(4) (Summer): 58–61.
- Jencks, S. 2000. "Clinical Performance Measurement – A Hard Sell." *Journal of the American Medical Association* 283(15): 2015–16.
- Johnson, K. 1998. "Fundamentals of Performance Measurement." *Health Affairs* 17(6): 253.
- Joint Commission on Accreditation of Healthcare Organizations (JCAHO). 2002. *Glossary of Terms for Performance Measurement*. Oakbrook Terrace, IL: Author.
- Jones, E. and G. Brown. 2001. "Outcomes: Why Isn't Everyone Doing This?" *Behavioral Healthcare Tomorrow* 10(5): 22–23, 42.
- Kaplan, R. and D. Norton. 2001. "Transforming the Balanced Scorecard from Performance Measurement to Strategic Management: Part II." *Accounting Horizons* 15(2): 147–60.
- Kazandjian, V. and T. Lied. 1998. "Cesarean Section Rates: Effects of Participation in a Performance Measurement Project." *Journal on Quality Improvement* 24(4): 187–96.
- Khan, K.S., G. Riet, J. Glanville, A.J. Sowden and J. Kleijnen. 2001. *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Those Carrying Out or Commissioning Reviews*. York, UK: University of York/NHS Centre for Reviews and Dissemination.
- Khim, L. and C. Hian. 2001. "Balanced Scorecard: A Rising Trend in Strategic Performance Measurement." *Measuring Business Excellence* 5(2): 18–26.
- Kizer, K., J.G. Demakis and J.R. Feussner. 2000. "Reinventing VA Health Care: Systematizing Quality Improvement and Quality Innovation." *Medical Care* 38(6) (Suppl. 1): I7–16.

- Lebas, M. 1995. "Performance Measurement and Performance Management." *International Journal of Production Economics* 41(1-3): 23-35.
- Leggat, S., L. Narine, L. Lemieux-Charles, J. Barnsley, G.R. Baker, C. Sicotte, F. Champagne and H. Bilodeau. 1998. "A Review of Organizational Performance Assessment in Health Care." *Health Services Management Research* 11: 3-23.
- Legnini, M., L. Rosenberg, M.J. Perry and N.J. Robertson. 2000. "Where Does Performance Measurement Go from Here?" *Health Affairs* 19(3): 173-77.
- Lemieux-Charles, L., N. Gault, F. Champagne, J. Barnsley, I. Trabut, C. Sicotte and D. Zitner. 2000. "Use of Mid-Level Indicators in Determining Organizational Performance." *Hospital Quarterly* 3(4): 48-52.
- Longo, D., G. Land, W. Schramm, J. Fraas, B. Hoskins and V. Howell. 1997. "Consumer Reports in Health Care." *Journal of the American Medical Association* 278(19): 1579-84.
- Malmi, T. 2001. "Balanced Scorecards in Finnish Companies: A Research Note." *Management Accounting Research* 12(2): 207-20.
- Mannion, R. and H. Davies. 2002. "Reporting Health Care Performance: Learning from the Past, Prospects for the Future." *Journal of Evaluation in Clinical Practice* 8(2): 215-28.
- McAdam, R. and A. Bannister. 2001. "Business Performance Measurement and Change Management within a TQM Framework." *International Journal of Operations and Production Management* 21(1-2): 88-107.
- McCorry, F., D. Garnick, J. Bartlett, F. Cotter and M. Chalk. 2000. "Developing Performance Measures for Alcohol and Other Drug Services in Managed Care Plans." *Journal on Quality Improvement* 26(11): 633-42.
- McIntyre, D., L. Rogers and E.J. Heier. 2001. "Overview, History and Objectives of Performance Measurement." *Health Care Financing Review* 22(3): 7-21.
- McLoughlin, V., S. Leatherman, M. Fletcher and J.W. Owen 2001. "Improving Performance Using Indicators. Recent Experiences in the United States, the United Kingdom, and Australia." *International Journal for Quality in Health Care* 13(6): 455-62.
- Mooraj, S., D. Oyon and D. Hostettler. 1999. "The Balanced Scorecard: A Necessary Good or an Unnecessary Evil?" *European Management Journal* 17(3): 481-91.
- Nadzam, D. and M. Nelson. 1997. "The Benefits of Continuous Performance Measurement." *Nursing Clinics of North America* 32(3): 543-59.
- Neely, A. 1999. "The Performance Measurement Revolution: Why Now and What Next?" *International Journal of Operations and Production Management* 19(2): 205-28.
- Neely, A. 2000. *Performance Measurement – Past, Present and Future*. Bedford, UK: Centre for Business Performance, Cranfield School of Management, Cranfield University.
- Neely, A., M. Gregory and K. Platts. 1995. "Performance Measurement System Design – A Literature Review and Research Agenda." *International Journal of Operations and Production Management* 15(4): 80-116.
- Perkins, R. 2001. "What Constitutes Success? The Relative Priority of Service Users' and Clinicians' Views of Mental Health Services." *British Journal of Psychiatry* 179: 9-10.
- Petitti, D., R. Contreras, F.H. Zeil, J. Dudl, E.S. Domurat and J.A. Hyatt. 2000. "Evaluation of the Effect of Performance Monitoring and Feedback on Care Process, Utilization and Outcome."

Performance Measurement in Healthcare:
Part I – Concepts and Trends from a State of the Science Review

Diabetes Care 23(2): 192–97.

Pickin, M., A. O’Cathain, F.C. Sampson and S. Dixon. 2004. “Evaluation of Advanced Access in the National Primary Care Collaborative.” *British Journal of General Practice* 54(502): 334–40.

Proctor, S. and C. Campbell. 1999. “A Developmental Performance Framework for Primary Care.” *International Journal of Health Care Quality Assurance* 12(7): 279–86.

Relman, A. 1988. “Assessment and Accountability: The Third Revolution in Medical Care.” *New England Journal of Medicine* 319(18): 1220–22.

Rosenblatt, A., N. Wyman, D. Kingdon and C. Ichinose. 1998. “Managing What You Measure: Creating Outcome-Driven Systems of Care for Youth with Serious Emotional Disturbances.” *Journal of Behavioral Health Services and Research* 25(2): 177–93.

Rowan, K. and D. Angus. 2000. “Don’t Let Perfection Be the Enemy of the Good: It’s Time for Optimism Over the Role of Severity Scoring Systems in Intensive Care Unit Performance Measurement.” *Current Opinion in Critical Care* 6(30): 153–54.

SchHARR. 1997. *The SchHARR Guide to Evidence-Based Practice*. Sheffield, UK: School of Health and Related Research, University of Sheffield.

Schneider, E., V. Riehl, S. Courte-Wienecke, D. Eddy and C. Sennett. 1999. “Enhancing Performance Measurement: NCQA’s Road Map for a Health Information Framework.” *Journal of the American Medical Association* 282(12): 1184–90.

Smee, C. 2002. “Improving Value for Money in the United Kingdom National Health Service: Performance Measurement and Improvement in a Centralized System.” In Health Canada, *Measuring Up: Improving Health System Performance in OECD Countries*. Paris: Organisation for Economic Co-operation and Development.

Smith, G.J., E.P. Fischer, C.R. Norquist, C.L. Mosley and N.S. Ledbetter. 1997. “Implementing Outcomes Management Systems in Mental Health Settings.” *Psychiatric Services* 48(3): 364–68.

Smith, P. 1993. “Outcome-Related Performance Indicators and Organizational Control in the Public Sector.” *British Journal of Clinical Governance* 4(3): 135–51.

Smith, P. and M. Goddard. 2002. “Performance Management and Operational Research: A Marriage Made in Heaven?” *Journal of the Operational Research Society* 53(3): 247–55.

State of California, Division of Transportation. 2003. *Transportation System Performance Measures*. Retrieved March 25, 2006. <<http://www.dot.ca.gov/hq/tsip/tspm/faqs.htm>>.

Studnicki, J., F. Murphy, D. Malvey, R.A. Costello, S.L. Luther and D.C. Werner. 2002. “Toward a Population Health Delivery System: First Steps in Performance Measurement.” *Health Care Management Review* 27(1): 76–95.

Tansella, M. and G. Thornicroft. 1998. “A Conceptual Framework for Mental Health Services: The Matrix Model.” *Psychological Medicine* 28: 503–8.

Thompson, B. and J. Harris. 2001. “Performance Measures: Are We Measuring What Matters?” *American Journal of Preventative Medicine* 20(4): 291–93.

Trabin, T., T. Kramer, A. Daniels and M.A. Freeman. 1997. “Cost and Value of Performance Indicators.” *Behavioral Healthcare Tomorrow* 6(5): 37–39.

Turpin, R., L. Darcy, R. Koss, C. McMahill, K. Meyne, D. Morton, J. Rodriguez, S. Schmaltz, P. Schyve and P. Smith. 1996. “A Model to Assess the Usefulness of Performance Indicators.” *International Journal for Quality in Health Care* 8(4): 321–29.

- Upton, D. 1998. "Just-in-Time and Performance Measurement Systems." *International Journal of Operations and Production Management* 18(11): 1101–10.
- Viccars, A. 1998. "Clinical Governance: Just Another Buzzword of the 90's?" *MIDIRS Midwifery Digest* 8(4): 409–12.
- Voelker, K., J. Rakich and G.R. French. 2001. "The Balanced Scorecard in Healthcare Organizations: A Performance Measurement and Strategic Planning Method." *Hospital Topics* 79(3): 13–24.
- Williams, I., D. Naylor, M. Cohen, V. Goel, A. Basinski, L. Ferris and H. Llewellyn-Thomas. 1992. "Outcomes and the Management of Health Care." *Canadian Medical Association Journal* 147(12): 1775–80.

For an examination of lessons from this systematic review for research and practice see Volume 2 Issue 1 for:

Performance Measurement in Healthcare: Part II – State of the Science Findings by Stage of the Performance Measurement Process

A systematic review of the peer-reviewed and grey literature suggests the need for an approach to performance measurement that addresses issues of organizational culture, stakeholder needs, the identification of meaningful and strategic performance measures, the potential for unintended effects and the application of findings to achieve healthcare improvement.

Pour un examen des leçons tirées de cet examen systématique pour la recherche et la pratique, voir l'article intitulé **La mesure du rendement dans les soins de santé : Partie II – Résultats de l'examen de l'état de la science, par étape du processus de mesure du rendement** dans le Volume 2, Numéro 1.

Un examen systématique d'articles revus par les pairs et de la littérature grise suggère qu'on a besoin d'élaborer une méthode de mesure du rendement qui aborde la culture organisationnelle, les besoins des intervenants, la détermination de mesures du rendement significatives et stratégiques, le potentiel d'effets non prévus et l'application des constatations afin d'améliorer les soins de santé.