

Benchmarking and Learning in Public Healthcare Properties and Effects

Buckmaster, Natalie; Mouritsen, Jan

Document Version

Accepted author manuscript

Published in:

Australian Accounting Review

DOI:

[10.1111/auar.12134](https://doi.org/10.1111/auar.12134)

Publication date:

2017

License

Unspecified

Citation for published version (APA):

Buckmaster, N., & Mouritsen, J. (2017). Benchmarking and Learning in Public Healthcare: Properties and Effects. *Australian Accounting Review*, 27(3), 232-247. <https://doi.org/10.1111/auar.12134>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Jun. 2025



Benchmarking and Learning in Public Healthcare: Properties and Effects

Natalie Buckmaster and Jan Mouritsen

Journal article (Accepted manuscript)

This is the peer reviewed version of the following article: Benchmarking and Learning in Public Healthcare : Properties and Effects. / Buckmaster, Natalie; Mouritsen, Jan. In: *Australian Accounting Review*, Vol. 27, No. 3, 2017, p. 232-247, which has been published in final form at <https://doi.org/10.1111/auar.12134>.

This article may be used for non-commercial purposes in accordance with
[Wiley Terms and Conditions for Self-Archiving](#).

Uploaded to [CBS Research Portal](#): January 2019

Benchmarking and Learning in Public Healthcare:

Properties and Effects

Natalie Buckmaster, Australian School of Business, University of New South Wales
Jan Mouritsen, Department of Operations Management, Copenhagen Business School

Abstract

This research investigates the effects of learning-oriented benchmarking in public healthcare settings. Benchmarking is a widely adopted yet little explored accounting practice and which is part of the paradigm of New Public Management (NPM). Extant studies are directed towards mandated coercive benchmarking applications. The present study analyses a situation within public settings in which benchmarking is voluntary and oriented towards learning. The study contributes by showing how benchmarking can be mobilised for learning and offers evidence of the effects of such benchmarking for performance outcomes. It concludes that benchmarking can enable learning in public settings but that this requires actors to invest in ensuring that benchmark data are directed towards improvement.

Benchmarking and Learning in Public Healthcare:

Properties and Effects

Introduction

Accounting research about public settings is often premised on New Public Management (NPM) as a control practice for increasing efficiency, effectiveness and accountability (Cole and Cooper, 2006; Sharma and Lawrence, 2015). Such research uncovers a coercive form of accounting (Ahrens and Chapman, 2004). Less common are examples of situations in public settings where accounting is associated with learning and development, or the enabling form of accounting (Ahrens and Chapman, 2004). While existing studies suggest that in order for the benefits of benchmarking to be realised, a focus on ideas or learning is necessary (Van Helden and Tillema, 2005; Northcott and Llewellyn, 2005, Llewellyn, 2005; Tillema, 2010), they speculate more than demonstrate how this might function in practice. Existing studies posit that there is little evidence of positive effects of benchmarking for learning and performance (Tillema, 2010; Laine and Vinnari, 2014). This paper analyses a specific situation, a Consortium, where benchmarking in public healthcare made learning more probable; additionally, it offers descriptive data about performance effects. Such a situation is rarely explored in the literature.

Benchmarking is a practice which uses relative performance – the ranking of entities' performance – to identify entities with superior performance, and this entity is then taken to be one from which it makes sense to learn (Murby, 2008). Thereafter, as Hinton, Francis and Holloway (2000) suggest, it is possible to enhance performance by learning from the successful practices of others. Bowerman, Francis, Ball, and Fry (2002) explain that benchmarking involves a search outside the organization for, and the subsequent incorporation of, best practices into the organizations. Van Helden and Tillema (2005)

describe public sector benchmarking as being based on ideas about managing organisations and improving their performances, and that these ideas include emulating best practices and comparing results for improving performance. Braadbaart (2007) proposes that recent decades have seen a surge of interest in the transfer of private sector management practices to the public sector and that benchmarking constitutes one of these imports. Laine and Vinnari (2014) note that benchmarking in the public sector is on the increase and Wynn-Williams (2005) argues that benchmarking has a role in the public sector, perhaps even more so than in the private sector. Despite recognition of the increasing importance of benchmarking, there are few investigations of how this potential may be realised (Bowerman, Francis, Ball and Fry, 2002; Hong, Hong, Roh and Park, 2012; Laine and Vinnari, 2014).

To date, academic interest in benchmarking has been minimal (Alstete, 2000; Bowerman and Ball, 2000; Laine and Vinnari, 2014). In practice, however, Hinton, Francis and Holloway (2000) note that benchmarking increasingly occupies time and energy of public sector personnel, and thus requires considerable resource commitments. Tillema (2010) proposes that the promise that public sector benchmarking will lead to learning and performance gains needs to be tested. Northcott and Llewellyn (2005) conclude that whilst benchmarking is a policy instrument, its practical relevance to healthcare improvement remains in doubt. Tillema (2010) suggests that public setting benchmarking exercises are frequently concerned with measurement and not learning. Elnathan, Lin and Young (1996) argue that it is important that we identify the most effective ways of gathering and sharing information, and under what conditions benchmarking is most likely to prosper.

On this basis, the aim of this paper is to show how it may be possible for benchmarking to have enabling rather than coercive effects in public settings. Based on the discovery of a situation where benchmarking seemingly was able to become learning-oriented, and therefore different from situations typically described in the literature, the paper's research question is

a dual one. First, how and under what conditions can benchmarking enable learning? Second, does learning have performance effects? This dual research question helps explain how benchmarking can become a constructive practice.

This paper begins with a literature review about benchmarking in public settings. This is followed by a description of the research method and entity. The results of a field study about voluntary benchmarking for learning via an inter-organisational public healthcare collective facilitated by a non-profit third party are then presented. Finally, some conclusions and implications are offered.

Relevant literature about benchmarking in public settings

Benchmarking is a technique that has been advocated by NPM (Webster and Hoque, 2005). NPM is a term coined by Hood (1995) to reflect a general shift in public sector management styles incorporating a range of private sector initiatives to make performance management a more prominent process (Guthrie, Parker and English, 2003). Australia has been reported by Hood (1995) as having a high level of NPM and its healthcare reforms have mirrored other countries including the United States, United Kingdom and Canada (Lillis and Grafton, 2005). Through NPM, governments have experimented with different initiatives for more efficient, effective and accountable public services (Guthrie, Parker and English, 2003; Newberry, 2003; Townley, Cooper and Oakes, 2003; Sandhu, Baxter and Emsley, 2008). Much literature examining the implementation of management accounting initiatives in public settings within NPM paradigms explores its coercive dimensions (see Ahrens and Chapman, 2004).

Modell (2005) argues that NPM reforms around the globe imply an enhanced emphasis on mobilising performance measurement techniques because of the widespread “myth” of public sector wastefulness and inefficiency in comparison with the private sector (Malmrose, 2015). According to Modell (2005), however, the design and implementation of

these initiatives has been far from un-problematic. Benchmarking is an initiative often associated with the private sector that has nonetheless been applied to public healthcare (Webster and Hoque, 2005; Vagnoni and Maran, 2008) and usually with concerns directed to efficiency, formal accountability and control.

NPM has seen a range of initiatives applied to public settings including healthcare (Lillis and Grafton, 2005). For example, Northcott and France (2005) study the growing use of balanced scorecards for monitoring and managing hospital performance, noting that questions remain about their usefulness. In their study of New Zealand hospitals, the partial diffusion of balanced scorecards is said to have done little to illuminate performance management challenges facing hospitals in efforts to enhance efficiency and effectiveness. Instead of permeating practice and decision-making, Northcott and France (2005) find they may serve “an interim purpose as a symbolic shield against accusations of poor or outdated performance management” (Northcott and France, 2005, p.44). Nath and Sharma (2014) examine the coerced adoption of balanced scorecards in public housing, finding adoption to be a myth because of the widely held view that an institution’s performance can be improved by accounting-led initiatives (Modell, 2004). Such initiatives provide an image of rational management, whilst possibly having a negligible impact on operating-level action.

Studies of NPM in healthcare note a general tension between the accounting and healthcare disciplines (Kurunmaki, 2009). Malmrose (2015) writes that large parts of the management accounting discourse continue to mobilise NPM around performance measurement. Furthermore, some researchers even refer to NPM reforms as management accounting reforms (Pettersen, 2001). Malmrose (2015, p.146) suggests that accounting can translate governmental concerns about healthcare and service quality into quantitative performance measures. These measures can subsequently be enforced as they transform the

descriptive rhetoric into manifest facts, thereby creating numeric evidence for decision making.

Some literature about benchmarking in NPM settings identifies tensions between learning and control (Northcott and Llewellyn, 2005; Llewellyn and Northcott, 2005). Anderson, Henriksen and Spjelkavik (2008) describe benchmarking applications for control as flawed. Benchmarking is usually applied in healthcare for control by ranking organisations and comparing results through leagues of indicators (Northcott and Llewellyn, 2005). Some research finds that because the average cost is promoted as a norm to aspire to, “all an organisation can expect from benchmarking is to become a good average” (Llewellyn and Northcott, 2005, p.556). Studies of benchmarking voluntarily undertaken for learning have been overlooked as has their consequences for performance gains. Llewellyn (2005) suggests that learning in the public sector may be dependent on initiatives promoting inter-organisational cooperation (Borins, 2001). Llewellyn (2005, p.18) writes that cooperation is crucial in order to enable learning in healthcare where problems demand trans-disciplinary knowledge.

The remainder of this paper reports on an example of benchmarking which seemingly has more learning effects than control effects. It is an example of accounting used in ways that enable (Ahrens and Chapman, 2004), contrasting with the above mentioned studies which largely characterise NPM reforms and benchmarking as coercive uses of accounting. A focus on the enabling uses of accounting requires research to understand how accounting is improved by adding conditions such as the repairing of numbers (Adler and Borys, 1996; Ahrens and Chapman, 2004). Such a focus can also offer insights into how enabling and controlling uses can create unique organisational capabilities for improving performance (Mundy, 2010). Enabling uses can be mustered by developing conditions for flexibly mobilising and drawing on accounting (Ahrens and Chapman, 2004; Jordan and Messner,

2012; Neu, Rahaman and Everett, 2014; Wouters and Wilderom, 2008). Coercive use may be more likely when accounting is understood as the final arbiter of issues and problems; this is when benchmarking functions as coercive and controlling, rather than enabling for learning (Jordan and Messner, 2012). As noted by Wouters and Wilderon (2008), there is little empirical knowledge about what kind of process could enhance the enabling nature of accounting. The remainder of this paper provides a study relevant to a benchmarking application for learning, outlining its purpose, structure and effects.

Method

This accounting study consists of a qualitative research design employing case analysis (see Yin, 2003; Scapens 1990; Denzin and Lincoln, 2000). Cooper and Morgan (2008) argue that case research in accounting is important for understanding empirical situations of uniqueness such as the present study with its ambit of understanding how benchmarking might enable learning in public settings. As highlighted by Chapman (2012), accounting is not a natural phenomenon; rather, it is human made. Thus, a field study is appropriate to studying how public setting personnel could use benchmarking as a learning device. Evidence was collected from multiple sources including documentation, interviews and observation. A total of 42 semi-structured interviews were conducted with participants of a benchmarking Consortium formed for learning in public healthcare. Interviewees included personnel from a third party benchmarking Consortium and its participants, personnel from a large public hospital where the changes that resulted from benchmarking were implemented.

Documentary and interview evidence were a critical source for determining what benchmarking for learning in healthcare might look like because such evidence facilitated tracing the effects of benchmarking on performance outcomes. Documentary sources spanning almost two decades (1995-2014) included annual reports, media releases, government reports and confidential internal documentation from the Consortium's secure

website that provided examples of key innovations introduced to healthcare organisations. Additionally, selected reports made available by the Consortium to the general public were incorporated, as well as evidence provided by the New South Wales Independent Pricing and Regulatory Tribunal (IPART, 2003)¹ which independently prepared a report for the Government on benchmarking and performance measurement by public hospitals. Evidence sources also included the observation of measurement displays of benchmarked data. These displays contained summarised data as analysed by the Consortium and then redistributed to its participants.

Interviews were held over a 48-month period and were semi-structured. Snowball sampling was used by the researchers to identify potential subject interviewees (Goodman, 1961; Atkinson and Flint, 2004; Morgan, 2008). At the end of every interview, interviewees were asked to suggest other potential interviewees which further identified other participants for the purposes of this study. Each interview lasted between 60 to 120 minutes. Immediately upon leaving the interview site, notes were taken about the interview – what went well, what did not, the state of the physical space, etc. The conversation at interviews centred on how benchmarking undertaken for learning via the third party had unfolded and its perceived effects. The interviews took place in mostly formal surroundings, for example, office spaces of interviewees. The questions were asked in an open-ended manner in order to encourage interviewees to discuss key issues from their own perspective without their potential thought processes being influenced by the researchers. Suddaby (2006) suggests that an approach where careful attention is paid to the contrast between the daily realities of substantive areas and the interpretation of these realities by those who participate in them (interviewees) is suited to efforts to understand processes by which actors construct meaning out of intersubjective experience. This is apt for the present study which seeks to understand how

¹ A consulting report by a firm named Inspirit Management Services engaged to impartially assist IPART's efforts informs the findings also.

benchmarking was understood as a device for enabling learning. A summary of participants is below in Table 1.

Insert Table 1 here

Certain themes such as anonymity, the application of benchmarking for learning, roles of a non-profit third party, and sharing ideas via conversational techniques were identified from the responses. The data representing the key themes were clustered together and analysed and the research question was continually referenced in order to ensure data relevance (Miles and Huberman, 1994). The objective of the analysis was to understand in greater detail how benchmarking for learning could occur and what its effects might be - as told by the field study.

Research context

In Australia, as is the case in many other parts of the world, public healthcare is under enormous pressure to achieve more efficient and effective outcomes. This research focuses on the activities of an Australian and New Zealand Consortium, a third party, non-profit provider of services incorporating: benchmarking calculations, analysing data, and interactive methodologies for facilitating the application of good practice. Interviewees were Consortium personnel, as well as a Consortium member institution which was a large tertiary public hospital, where changes learned from benchmarking were implemented. The subject public hospital has a large and diverse population, and operates in Sydney, Australia.

As mentioned, the Consortium was a non-profit and voluntary organisation which organised the process of using the benchmarks to promote change. Initially, the aim was to promote networking amongst hospital chief executives but, over time, there was increased acceptance that other personnel from public hospitals would benefit from sharing issues and

innovations through the process of benchmarking. As a result, these personnel would participate in face-to-face gatherings where the processes behind good practice were identified. The Consortium had its own governance model without intervention in the day to day activities of participants. Apart from comparative benchmark data, the Consortium sought to provide opportunities to use these data in a trusted setting in order to enable change at local settings.

The case of the Consortium: Benchmarking as an enabling form of accounting

This section reports on the results of the study. The properties that facilitated benchmarking for learning in public healthcare are discussed and then selected effects on performance outcomes are described.

The study is a story about a benchmarking Consortium born in 1995 within the Australian and New Zealand healthcare context and still operating to this day. The Consortium was originally formed to disseminate the knowledge and implementation of best practices to hospitals, and it is therefore an example of how the ambition of learning from benchmarking can be realised. It was a *voluntary* initiative whereby official benchmarking data were collected and then analysed further and interpreted by participants of the Consortium's benchmarking workshops. These participants were from different public hospitals that together had to make sense of the benchmarking exercise. The Consortium was private, independent, and impartial, but the participants in its activities were from public hospitals that formed member institutions of the Consortium. This Consortium was of critical importance because it enabled participants to be able to deliberate under non-threatening conditions:

What the Consortium gives us is an independent, professional, analytical support service. There is great value in having an independent group facilitate this. Because it feels non-threatening it works a lot (Clinical Director Consortium Participant).

The Consortium facilitated meeting within peer-groups or inter-hospital networks, for sharing and interpreting performance information across public hospitals. The Consortium founders claimed in interview evidence that:

Back in 1995 hospitals had never talked with each other. Initially we got a leader from a hospital in each jurisdiction in Australia. But we have grown to 42 hospitals sharing data openly from States in Australia and New Zealand (The Consortium Founders).

The idea therefore was to make the Consortium a medium for undertaking comparisons between hospitals in a non-threatening environment where knowledge production was encouraged through peer communication. A participant describes what the Consortium offered:

The Consortium gives us a network and the flexibility to call up our peers when we have a problem. Or if we have a query about how they solved a problem somewhere in some other hospital, there is a group that we can trust (Director of Quality Consortium Participant).

The initial process was to collect public performance data and circulate benchmarks in order to initiate conversations which were played out at face-to-face meetings, and then followed up in virtual forums and inter-disciplinary teams. Teams were sent from large public (mainly tertiary) hospitals in Australia and New Zealand, as well as invited guests from overseas healthcare organisations. Sometimes, teams comprised both clinicians and administrators. Guven-Uslu and Conrad (2008) noted that cooperation between administrators and clinicians would constitute a necessary property in successful benchmarking in public healthcare. A participant describes this format below:

This is a unique opportunity to speak to peers across hospitals. We recently attended a two-day format on operating theatre efficiency. We went in an inter-disciplinary team comprising surgeon, anaesthetist, senior nurse, etc. There were some poster presentations and a fixed process. Our team gave a presentation as we do pretty well on this measure. A lot of discussion was encouraged around our practices (Manager of Surgery Consortium Participant).

The Consortium offered professional and analytical support by a two-step process. First, performance data were reported to the Consortium which then analysed and benchmarked the data, and then redistributed it to hospitals. Second, an opportunity was given to discuss the data in face-to-face workshop settings and in virtual forums. Hospitals shared two types of benchmark data, namely, in-patient episode data, such as reported to government, and more customised data collections which extended beyond case mix data to include quality, safety, and patient satisfaction.

Generally this illustrates Ahrens and Chapman's (2004) point that, in order for accounting to become enabling, it has to be associated with its active use, such as through recalculation which the Consortium mobilised to a large extent. It also shows that enabling uses may be more likely when people are motivated to add meaning and interpretation to the accounting numbers.

Benchmarking for learning

The approach employed by the Consortium was based on that of Everett Rogers². Adapted from Rogers, Table 2 below illustrates the Consortium's specific approach to benchmarking for learning and the diffusion of innovation in public healthcare. As described by the Consortium, this approach was rolled out in five sequential steps starting with computation, and ending with implementation and follow up.

Insert Table 2 here

The Consortium deliberately had no advocacy role with any level of government so as to be independent of government; it was designed to be non-coercive because it could not oblige things to happen (Kurunmaki and Miller, 2011). However, individual members could, if they

² Rogers, E. The Diffusion of Innovations. 4th Edition. The Free Press, 1995.

wished, use the knowledge gained via participation in Consortium activities to inform discussions with governments, provided that they abided by a code of ethics employed by the Consortium. As stated by the Consortium's General Manager:

We have stayed very quiet – do not get involved in any government policy and through us the hospitals have an agreement that all data are confidential (Consortium General Manager).

All healthcare participants of the consortium signed off on a code of ethics which held that no information (incorporating benchmarking numbers and the internal work practices of any other hospital) be released outside of Consortium activities without permission. No releases of collated benchmark data were permitted without the consent of all Consortium members (participating hospitals) and this required their sign off. This is another sign of non-obligatory or non-coercive practices. All data were securely stored in the Consortium's information systems. The Consortium adjusted and transformed many benchmark numbers in order to make them meaningful to participants. Contextual differences unaccounted for in government benchmarking exercises for control, as associated with NPM, were acknowledged and accounted for. Thus, the public benchmark numbers were made more enabling by being recalculated to suit the situation and by making them more meaningful via articulated interpretation. Participants reported that this process of adjustment was collaborative with input often provided by clinicians:

Clinicians do have an active interest in making sure it is right if we are involved and you show us the data. We are happy to change and update the data until it is corrected. We senior clinicians need to have responsibility for the data, be involved in adjusting the data so it suits our local operating conditions and the flexibility to take action (Nurse Unit Manager Emergency Consortium Participant).

Analytical approaches, making benchmarking more enabling, drew on "data mining" by Consortium personnel to adjust the data. For instance, data were regrouped into Australian Refined Diagnostic Related Groupings (AR-DRGs) to achieve more appropriate measurement and adapted with the following variables: type of admission (planned or

unplanned with unplanned defined as less than 24 hours of expected arrival in hospital), source of admission (normal/other, to reflect other defined as a transfer from another institution or a statistical admission due to a change in care status), discharge destination (home/hospital transfer/died/other with other including transfers to nursing homes and statistical discharges due to a change in care status, age group (0-14, 15-54, 55-69, 70-79 and 80 years and above); and co-morbidity level (low, 0-2 diseases; high, 3 or more diseases). Related issues of data quality were discussed among participants (often clinicians) until adapted benchmarks were agreed upon.

A measure called a *relative stay index* took into account important differences in patient categories that were unaccounted for by government and was calculated by the Consortium to reflect the local operating conditions of each individual hospital more meaningfully. Over a million in-patient episodes in the 2000/2001 data base were utilised to set the different length of stay benchmarks in each category.³ To illustrate this, and as shown below in Table 3, the relative stay index for hip replacements of elderly patients (10.1 days) would differ substantially from that of middle-aged patients (7.2 days), and thus grouping them together had the effect of distorting results and rendering them meaningless. Therefore, the Consortium adapted the benchmark as reported by government to better reflect local setting operating conditions where, for instance, different patient demographics, such as a higher proportion of elderly patients than one's peer hospitals, could be incorporated into the result to better reflect the individual hospital's local setting.

Insert table 3 here

³ Source: Health Consortium, (2007). Unpublished document. Health Consortium Limited. Sydney.

Once the benchmarks were reconstructed and adapted to reflect local settings – hospital participants would meet in face-to-face settings not only to further refine the benchmarks but also to discuss them and share their work practices. Storytelling, narrative and dialogue were central to the Consortium's diffusion of innovation process. The overall group size of a face-to-face workshop was 40 persons so that approximately 10 to 12 hospitals could participate at any one time. Each organisation would send a multi-disciplinary team because this was considered more powerful for transferring innovations when these teams returned to their own organisational setting. Groups had to be small enough so that participants could speak freely, share ideas and innovative practices, be vulnerable through exposing their practices and listen to each other:

We employ a working group approach and the idea is to trust in the group. What the pioneers of the consortium, Rod and Bill did was to initiate the concept of benchmarking for learning. We start by identifying a problem, then set a general agenda and structure but have no idea what the answer will be (Consortium General Manager).

Participants shared their stories at face-to-face meetings while they exposed their practices. Networking and sharing stories played a critical role. Participants reported going back to the organisation after Consortium participation and revising their practices as a result of discussions. Storytelling brought people together and the focus was on the patient as opposed to the organisation which each participant represented. Thus, the patient was the common starting point for their stories:

We all came together and basically said how do we best run an Emergency Department. Then we shared our war stories. We may not always know what works but we certainly know what didn't work (Manager of Emergency Consortium Participant).

Storytelling and narrative, as described by the Director of Clinical Operations, offered an opportunity for clinicians to discuss processes, performance and results across hospitals. Of importance to the storytelling process was that the focus was on the patient and what might

be done to improve services to the patient. Important also was that innovations in practices were presented as suggestions, rather than mandates, with the understanding that conditions at each local setting could differ:

The nice thing about it is that the focus is on the patient rather than the organisation – and how do you improve services for the patient. And it is a success because while all participants are encouraged to share ideas it is not forced (Director of Clinical Operations Consortium Participant).

The majority of interviewees suggested that learning was possible partly because it felt like ownership of the process was held by participants (usually clinicians). The Consortium took the view that, ultimately, it is at local levels where its participants will decide what can and cannot work. The emphasis therefore was on practical everyday issues as confronted by healthcare workers. All participants were expected to share ideas and innovations but nobody was forced to adopt any particular idea or innovation. Thus, the solutions shared by participants were presented as offers and not orders or demands. Participants would evaluate different innovative practices as shared by their peers from other organisations and decide for themselves whether they were relevant to their local setting. Many innovations came about as a result of the Consortium. For example, innovations were implemented with respect to the improvement of bed management, such as “sitters” to reduce risk of falls by high risk patients. Examples of innovations also included the recognition and management of chronic patient “frequent attendees” and increasing presentations by patients on the day of surgery for operations leading to reductions in in-patient bed days and reductions in hospital acquired infections.

Effects of benchmarking for learning

Some effects of benchmarking for learning are traced in this section. The analysis presents summarised results of improvements achieved by member organisations.

The results of reductions in length of stay that were achieved are shown in Figure 1⁴ which reports the change in length of time of in-patient episodes of Consortium participants versus that of public hospitals nationally. Overnight in-patient episodes for participant organisations in the 2002/2003⁵ financial year took seven per cent less time than by the same group of organisations during the period 1994/1995. Conversely, only a three per cent reduction in length of stay was achieved by organisations which did not participate in Consortium activities.

Insert Figure 1 here

During the period 2002/2003, the variation for length of stay between benchmarking Consortium members and the average of Australian public hospitals, as reported in national benchmarks (as calculated by the national government), was four percentage points, as shown in Figure 1; that is, length of stay as per the national length of stay benchmark was 97 per cent. This compares with a result for participants of the benchmarking Consortium which was 93 per cent. It should be noted that the length of stay for Consortium benchmarking participants decreased from 100 per cent in 1994/1995 to 93 per cent in 2002/2003. This means that Consortium participants took seven per cent less time for overnight inpatient episodes than did non-Consortium members. This compares with a reduction of only three per cent for the period 1994/1995 to 2002/2003 by non-members. While a difference between Consortium and non-Consortium participants of four percentage points might not appear dramatic, it translates into a substantial number of bed days because the Consortium members are typically amongst the largest public hospitals in Australia.

⁴ Health Consortium, (2007). Unpublished document. Health Consortium Limited. Sydney.

⁵ These are the last data available tracing the effects of benchmarking for learning by the Consortium because of the time and costs required to do so therefore directing efforts away from their ambit of disusing good practices.

The number of bed days saved by participating Consortium hospitals (that were Consortium members since its inception), compared to the number of bed days that would have been required if the hospitals only reduced their length of stay in line with the Australian national average (national benchmark as calculated by government) in each year, is shown in Figure 2. For example, in 2002/2003, the gap mentioned above previously was 97 per cent minus 93 per cent equals 4 per cent. The Consortium hospitals actually used 1,433,200 bed days for overnight patients during the year. The figure of 55,000 bed days in Figure 2 below represents the additional four per cent of bed days that were saved compared to the number that would have been needed by the hospitals if they had not reduced their length of stay beyond the national average. The cumulative savings between 1995 and 2003 by seven major teaching hospitals that were members since its inception, these being the earliest participants of the Consortium, was over 279,000 bed days. The marginal cost of providing food, fresh linens and minimal nursing support to patients in these beds is well over \$80 per day, providing a notional value of over \$22,000,000 for these bed day savings. The Consortium estimated that even if their role had contributed only 10 per cent to this improved efficiency, the result far exceeded the overall costs for membership to benchmarking Consortium participants.

However, it could also be argued that, while such group-wide comparisons are interesting, they tend to overlook the differences between participant hospitals themselves. Each hospital had different expertise and service offerings according to Consortium personnel that makes high-level comparisons problematic. As discussed earlier in this paper, the Consortium had adapted the benchmarks highlighting specific differences in operational performance by using a “relative stay index. Figure 2 reveals that the Consortium organisations used 1,433,200 bed days for overnight patients during the year. Again, as mentioned above, a figure of 55,000 days shown in Figure 2 represents the additional four per

cent of bed days that were saved. The cumulative savings between 1995 and 2003 by the major participants was over 279,000 bed days. This was a significant result.

Insert Figure 2 here

As discussed earlier, the actual length of stay was adapted by the Consortium to develop what they considered a relative stay index, considered to be a more appropriate measure because benchmarks published by government distorted key differences in length of stay by putting different types of patients into the same category. For example, patients who fractured their hips and arrived in hospital as emergency patients were in the same category as those who were having elective hip surgery after weeks of pre-planning. Therefore, the Consortium chose to adapt the DRG measure to incorporate differing conditions, demographics and situations across hospitals.⁶ A value of 100 on the Index suggested that the length of stay as measured by government matches benchmarking Consortium participant hospital average as it stood in 2000/2001. A lower value by benchmarking may have resulted from participating in Consortium activities and therefore the subsequent diffusion of innovations. Changes in the relative length of stay are shown in Figure 3, representing the results for 2000/2001, and Figure 4 representing results for 2003/2004. In the year, 2000/2001, when the measure was first developed, only one benchmarking Consortium member had a relative index value of 90 per cent or below. However, by the year 2003/2004, at least nine of the 30 Consortium members achieved a relative index of 90 per cent or below, suggesting the implicated effects of benchmarking for knowledge production.

Insert Figure 3 here

⁶ Over a million episodes in the 2000/2001 database were used to set different length of stay benchmarks in each subcategory.

Insert Figure 4 here

Other examples of effects are discussed below. Participants of a Consortium meeting had inter-disciplinary discussions which identified a series of explanations about what might constitute good practice for patient service levels in major teaching hospitals. One of these was the expectation that at least 70 per cent of surgery patients who were to have major elective surgery should be able to have their operation on the day of admission, rather than the day after arriving for surgery. The day of surgery admission (DOSA) measure was viewed as a way to reduce patient stress, but also provide opportunities for Consortium participants to process elective surgery patients in a more cost-effective way. One participant hospital championed this concept and then shared the work it had done to design pre-admission and peri-operative processes to enable an improved DOSA rate. The champion hospital which shared its good practices had already achieved a rate of over 60 per cent, while most of the peer hospitals were well below 40 per cent.

Consequently, the benchmarking Consortium began monitoring the DOSA rate for all member hospitals in 1996, reporting the results back on a six-monthly basis. The Consortium then conducted a Study Tour of the organisation with the most superior results and then ran a series of workshops to diffuse innovations to project leaders from participant Consortium hospitals that were working on improving their DOSA rates. The overall average DOSA rate rose from 44 per cent in 1996/1997 to 74 per cent in 2003/2004, as shown in Figure 5. The data for individual participating organisations as shown in Figure 5 highlights that the lowest DOSA rate had risen from 27 per cent to 65 per cent. The rates then continued to rise above the target level, with some Consortium hospitals even achieving results of nearly 90 per cent. This would appear to be a significant effect.

Insert Figure 5 here

Moving to ‘day of surgery admissions’ was reported by the Consortium to be especially challenging for certain specialties, such as Neurosurgery, Orthopaedics, and Cardio-Thoracic surgery but improvements were nevertheless reported. For example, there was evidence that several benchmarking Consortium hospitals were able to increase their overall DOSA for Neurosurgery to 70 per cent, as shown in Figure 6.

Insert Figure 6 here

At a major consortium meeting in March 1999 in Christchurch, New Zealand, ways were examined to improve the treatment of “complex medical patients.” An operational definition of these patients was developed to aid in comparing practices across hospitals. A complex medical patient was defined by the Consortium as a patient with multi-system disease affecting at least three processes such as: 1. Arrives as an emergency patient, 2. Stays at least one night, and 3. Does not receive a major surgical procedure. Using in-patient episode data collected over a three-year period, the Consortium discovered that patients who fitted this definition were likely to stay twice as long in hospital per episode, and were also likely to have multiple in-patient episodes per year.

Additionally, patients with chronic conditions, such as Chronic Obstructive Airways Disease (COAD), frequently had multiple disease conditions which rendered them complex to treat. They also had multiple admissions to hospital each year, as the following chart, Figure 7 from 1999, highlights. Following a face-to-face Consortium meeting, many of the member hospitals began a year-long project to address the issue of better treatment for

complex medical patients, particularly the “frequent attenders.” Staff at Middlemore Hospital in New Zealand, for instance, highlighted an intervention program they had pioneered to provide influenza immunisations to COAD patients, prior to the onset of winter, which reduced the readmission rate. Participating benchmarking hospitals visited Middlemore Hospital and developed their own action plans to begin better ways to address ways of improving services to these patients.

Insert Figure 7 here

After the above Consortium initiative began, the New South Wales and Victorian governments also launched major initiatives to improve the care of complex medical patients. In 2001, it was flagged by the New South Wales Government Action Plan - “Improving Health Care for People with Chronic Illness” and, in 2002, Victoria followed suit launching the Hospital Demand Management Strategy and the Hospital Admission Risk Program (Health Consortium, 2007). This was to promote improved management of complex medical patients with chronic diseases. Both of these government initiatives were informed by the work of the Consortium and its participant institutions that shared insights with governments without compromising the code of ethics. In-patient episode data were collected by the Consortium from the time of its original meetings on complex medical patients, revealing that the average emergency length of stay was reduced by about seven per cent since 2000/2001; this is shown in Figure 8.

Insert Figure 8 here

The analysis of total bed days per patient with COAD suggests that the average days per episode were reduced from seven days to six days (nearly 20 per cent reduction).

Discussion of how benchmarking can enable learning and effects

Llewellyn (2005, p.21) argues that leveraging the performance of public healthcare organisations is a pressing policy issue. NPM policy initiatives have become embedded in Australian and international public settings (Guthrie, Parker and English, 2003; Llewellyn, 2005; Lawrence, 2005; Sandhu, Baxter and Emsley, 2008), including healthcare (Malmlose, 2015). Benchmarking is one such initiative (Webster and Hoque, 2005). The literature reveals a well-researched tension between the implementation of NPM initiatives that draw from management accounting initiatives (such as benchmarking) and healthcare organisations (Grafton and Lillis, 2005; Malmlose, 2015). As mentioned earlier in the paper, this literature frequently explores the coercive (see Ahrens and Chapman, 2004) adoption of accounting. One persistent issue is that institutions “must” respond through the adoption of accounting initiatives that emphasise efficiency, accountability and control (Bogt, 2008), but ones nonetheless that offer few learning effects (Lye, 2006).

Benchmarking in the present study moved from a coercive tool to an enabling one. The case of the Consortium is, in a sense, the opposite of Jordan and Messner’s (2012) account of the transformation of an enabling type of accounting into a coercive one. The primary difference is that, in the case of the Consortium, the movement is from coercive to enabling control. Jordan and Messner explain the movement from enabling to coercive control by suggesting that accounting has become increasingly a tool for managing at such a distance that complexities are disregarded. In contrast, it is possible to understand the Consortium as a tool for enabling uses of accounting because it becomes a tool for managing *in situ* where complexity is embraced.

Embracing complexity is also the theme of Wouters and Wilderom's (2008) investigation of the development (design and implementation) of performance measurement which identifies enabling effects whereby people are able to influence accounting to be meaningful for their job situations. The Consortium does this. It assumes that mandated benchmark data could be useful, but the usefulness is not directly apparent as they are tools for action at a distance. When shifted through the processes of re-calculation and re-interpretation by the Consortium, accounting becomes more meaningful because it embraces the type of complexity that actors feel they face. The Consortium creates accounting *in situ* and people add interpretations *in situ* which are relevant to their own experiences. This addition of complexity does not make accounting less relevant and less transparent; on the contrary, it adds to relevance and transparency. This is therefore an iterative process which leads to learning from existing experiences and experimentation with new practices (see also Wouters and Wilderom, 2008).

Learning from benchmarking requires both meaningful calculations and a context for interpretation. The Consortium offers both. This created a situation where no-one was prematurely held to be accountable for his or her propositions; it also allowed different solutions to be relevant for different hospitals. Such a 'contingent' view on solutions is an effect of a process wherein sharing of ideas is more important than agreeing on ideas, and where a particular solution might be adapted and used differently in different hospitals.

Analysis and understanding were collective, but application was individual. Participants were therefore willing to be 'vulnerable' and not only reveal their practices but also to experiment. This perhaps is because change was a choice and not an obligation. In the present study, part of the enabling process was repairing accounting numbers used in a different coercive reporting system and making them meaningful by taking into account differing local conditions.

This study shows how benchmarking can lead to performance gains; this is a situation often neglected in the literature and where empirical evidence is needed (Lye, 2006; Tillema, 2010). This situation seems to offset some of the control dilemmas identified in the literature (Van Helden and Tillema, 2005; Laine and Vinnari, 2014). Braadbaart (2007) and indicates that benchmarking for control might not always inhibit learning in the way that some literature suggests (Bowerman, Francis, Ball and Fry, 2002; Llewellyn and Northcott, 2005, Tillema, 2010). An instance of this latter is seen in the proposition that “designing a successful benchmarking system might not be feasible” (Laine and Vinnari, 2014, p. 303).

The Consortium offers a space where benchmarking can enable learning. Such a space is important and needs to be understood. As suggested by Adler and Borys (1996) generally, and Ahrens and Chapman (2004) in relation to accounting, enabling uses of control require a series of additions and repairs to happen. These additions and repairs add context. They add context by providing a larger perspective from which a control issue is to be understood, such as the ability to understand the hospital in its context, and repair of calculation to make benchmarks more meaningful. Benchmarking is hardly meaningful without context; and it is not easy to avoid premature conclusions on mandated benchmarking. A more considered process appears to take more time, requires shielding participants and developing new understandings that build on mandated benchmarking, which is then transformed to become more meaningful. It is by this transformation that otherwise coercive benchmarking turns into tools that enable learning. This study suggests that there are two main axes of an enabling space which can turn benchmarking associated with coercive uses into a learning orientated benchmarking. First, the space is a safe place where it is possible to debate ideas and practices that will not prematurely be taken as blue prints for all situations. Second, the space will help participants to develop more intelligent benchmark calculations that fit local situations and make it possible for people to develop understandings of the causes behind the

benchmark numbers, as well as develop conditions in order for good practices not only to be identified but also to be relevant. The idea is that one size does not fit all.

The present study of benchmarking via the Consortium suggests that some calculation creates willingness to experiment (Wouters and Wilderom, 2008). Meaningful calculations enable learning and this meaning can be developed in addition to mandated benchmark numbers. The Consortium also says that meaning has to be enabled by construction of a social process which helps experimentation. The case of the Consortium illustrates that, given space, it is possible to render benchmark-data useful for learning, even though it emanates from processes tied to coercive reporting. This is not by using them directly, but by working to make them more meaningful. The Consortium makes no claim to a one-size-fits-all solution and consequently there is no obligation to use its findings. Additionally, it appears from the results of this study that when benchmarking is for learning, many different types of innovations and solutions might be preferable, given different situations. This view of benchmarking is not strong in the literature perhaps owing to NPM's primary focus on institutional arrangements, whereby governments delegate and control by means of accounting. Criticism, even confusion results (Lye, 2006) instead of learning from differences that instead becomes a ranking process (Van Helden and Tillema, 2005).

Conclusions

This study has important implications for researchers, practitioners and policymakers as it speaks to the public sector accounting literature and NPM, extending the boundaries of this literature. The findings suggest that benchmarking can enable learning effects and how this may happen. There is a need to pay greater attention to more local understandings of what is sensible and appropriate, at particular organisational and local settings (Grafton and Lillis, 2005; Modell, 2005; Lye, 2006; Sharma and Lawrence, 2015); as Nath and Sharma (2014).

p.15) say, dissemination does not guarantee diffusion (see also Modell, 2004; Grafton and Lillis, 2005; Cole and Cooper, 2006; Nath and Sharma, 2014). It is shown in the field study material that people decide at a local level what can and cannot work. Webster and Hoque (2005) note that the implementation of accounting initiatives in public organisations at local settings has often been slow to be devolved.

There is a lack of empirical evidence about how benchmarking can enable learning in public settings (Tillema, 2010; Laine and Vinnari, 2014). Additionally, the extant literature offers limited evidence of benchmarking processes that work, and especially ones with potential to impact on performance outcomes (Tillema, 2010). A contribution of this study therefore is in offering evidence of how benchmarking can be enabled for learning with insights into how this can happen and the flow on effects for performance outcomes. Northcott and Llewellyn (2005) suggest that the usefulness of public benchmarking in disseminating best practice appears to be limited and, instead, benchmarking exercises turn into cumbersome accountability situations with little possibility for learning. This is when benchmarking is exclusively employed for control purposes and its more desirable attributes, they argue, will therefore be difficult to achieve.

The present study suggests that enabling benchmarking for learning incorporates adjusting data for local settings, and at the same time taking into account that what might be considered an appropriate innovation or practice in one setting does not necessarily hold for another. As shown by this study, however, the way in which benchmarking is applied is the key to enhancing effects. Under the conditions pertaining to the Consortium, it is possible that benchmarking can enable learning which, in turn, produces improved performance results. Importantly, the study suggests that these effects were made possible given certain conditions. Such conditions incorporated the transformation of case mix data (reported for control purposes) into meaningful benchmarks via recalculation to fit local settings. Other

conditions are conversational methodologies, learning as an offer not an obligation, and understanding that the appropriateness of a practice innovation depends on the circumstances of local settings. The fact that the Consortium makes an offer and does not oblige its participants to apply innovations is important. It also appears that its role as a non-profit third party without a consulting or policy role that transforms data for hospitals while proposing an approach for learning is important. So, the conditions for benchmarking to be enabling are: that benchmarking data are made more useful; that many ideas are shared about the things that produce better performance; and, that there is an absence of premature answerability. In this regard, accounting is not only used in a more enabling way but also to a certain extent produced by those involved.

As with all research, this study has some limitations. The main limitation is that, while documentary and interview evidence inform the study, and some information systems were observed, the authors were unable to observe the face-to-face gatherings wherein learning from benchmarking happened; this was owing to confidentiality issues. This limitation could flag a basis for future research. Such research could be aimed at observing settings where benchmarking for learning happens. This could include face-to-face and virtual gatherings where conversational methodologies are employed to share practices using intelligent benchmarks that provide triggers for conversations, and enable learning about good practice. The willingness of participants to truly be vulnerable and expose their practices, as well as how this really pans out in practice, could elicit observational research methods, such as the study of body language, in order for this to be done objectively and more effectively.

We also need greater understanding of the characteristics of spaces that enable more learning-oriented than control-oriented performance measurement in public settings, including in healthcare (Kellogg, 2009). More detailed insight into the conditions under which collective inter-organisational cooperation and learning can happen is needed, as well

as the consequences of such cooperation for performance gains. Finally, how does all this best take place in complex settings which are already constrained owing to requirements to direct time and other resources to NPM initiatives?

References

- Adler, P and Borys, B., (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly*, 41:1, 241-258.
- Ahrens, T and Chapman., C., (2004). Accounting for flexibility and efficiency: A field study of management control systems in a restaurant chain. *Contemporary Accounting Research*. 21:2. 271-301.
- Alstete, J., (2000). Association-sponsored benchmarking programs. *Benchmarking an International Journal*. 7:3, 200-205.
- Anderson, B., Henriksen and Spjelkavik, I., (2008). Benchmarking applications in public sector principal-agent relationships. *Benchmarking: An International Journal*. 15:6, 723-741.
- Atkinson, R and Flint, J., (2004). *Encyclopaedia of Social Science Research Methods*. Sage Publications, Inc. 1044–1045. ISBN 9780761923633.
- Biernacki, P and Waldorf, D., (1981). Snowball sampling: problems and techniques of chain referral sampling. *Sociological Methods and Research*. 10:2, 141-163.
- Bogt. H.J.T., (2008). Management accounting change and New Public Management in Local Government: A reassessment of ambitions and results – an institutionalist approach to accounting changes in the Dutch public sector. *Financial Accountability and Management*. 24:3, 209-241.
- Borins, S., (2001). Encouraging innovation in the public sector. *Journal of Intellectual Capital*. 2:2, 310-319.
- Bowerman , M and Ball, A., (2000). The modernisation and importance of Government and Public Services: Great Expectations: Benchmarking for best value. *Public Money and Management*. 20:2, 21-26.
- Bowerman, M., Francis, G., Ball, A., and Fry, J., (2002). The evolution of benchmarking in UK local authorities. *Benchmarking: An International Journal*. 9:5, 429-449.
- Braadbaart, O., (2007). Collaborative benchmarking, transparency and performance: Evidence from the Netherlands water supply industry. *Benchmarking: An International Journal*. 14:6, 677-692.
- Cole, B and Cooper, C., (2006). Deskillling in the 21st century: The case of rail privatisation. *Critical perspectives on Accounting*. 17. 601-625.
- Chapman, C., (2012). Framing the issue of research quality in a context of research diversity. *Accounting Horizons*. 26:4, 821-831.
- Cooper, D and Morgan., (2008). Case study research in accounting. *Accounting Horizons*. 22:2.
- Denzin, N.K., and Lincoln, Y.S. (2000). The discipline and practice of qualitative research. In N.K. Denzin and Y.S. Lincoln, Y.S. (Eds.), *Handbook of Qualitative Research*, Thousand Oaks: CA, Sage Publications, 2nd Edition, pp. 1-28.

- Elnathan, D., Lin, T and Young, S (1996), Benchmarking and management accounting: a framework for research. *Journal of Management Accounting Research*, 8, 37-54.
- Goodman, L.A. (1961). "Snowball sampling". *Annals of Mathematical Statistics* 32 (1): 148–170. doi:10.1214/aoms/1177705148.
- Grafton J, Lillis A. (2005) "The Role of Performance Management and Control Systems in Public-Sector Reform: Victorian Health Care Networks", *Australian Accounting Review*, vol.15 (3), pp.25 - 33.
- Guthrie, J; Parker, L and English, L., (2003). A Review of New Public Financial Management change in Australia. *Australian Accounting Review*. 13:2, 3-9.
- Guven-Uslu, P and Conrad, L (2008). Uses of management accounting information for benchmarking in the NHS trusts. *Public Money and Management*. 28:4, 239-246.
- Health Consortium," Unpublished Document, Health Consortium Limited 2004, Sydney
- Hinton, M., Francis, G and Holloway, J., (2000). Best practice benchmarking in the UK. *Benchmarking: An International Journal*. 7:1, 52-61.
- Hong, P., Hong, S., Roh, J.J and Park, K., (2012). Evolving benchmarking practices: A review for research perspectives. *Benchmarking: An International Journal*. 19:4/5, 444-462.
- Hood, C., (1995). The "New Public Management": In the 1980's: Variations on a theme. *Accounting, Organizations and Society*. 20:2/3, 93-109.
- Independent Pricing and Regulatory Tribunal of New South Wales [IPART]. (2003). NSW Health – Focusing on Patient Care. Sydney.
- Jordan, S and Messner, M., (2012). Enabling control and the problem of incomplete performance indicators. *Accounting, Organisations and Society*. 37, 544-564.
- Kellogg, K. (2009). Operating room: relational spaces and microinstitutional change in surgery, *Journal of Sociology*. 115:3, 657-711.
- Kurunmaki. L., (2009). Management accounting, economic reasoning and New Public Management, In *Handbooks of Management Accounting Research Volume 3*. Chapman, C., Hopwood, A and Shields, M. (eds), Elsevier.
- Kurunmäki, L. and Miller, P. (2011) 'Regulatory hybrids: Partnerships, budgeting and modernising government', *Management Accounting Research*, 22 (4), 220-241.
- Laine, M and Vinnari, E., (2014). The dynamics of voluntary benchmarking in the water sector. *Public Money & Management*, 34:4. 297-304.
- Lawrence, S., (2005). Performance measures and cost containment in the New Zealand health sector: A case of Iatrogenic Disorder. *Australian Accounting Review*. 15:3, 4-14.
- Lillis, A and Grafton, J., (2005). Health sector performance management. *Australian Accounting Review*. 15:3. 3.

- Llewellyn, S., (2005). Performance, productivity and innovation in healthcare organisations: Comparing England and Scotland. *Australian Accounting Review*. 15:3, 15-24.
- Llewellyn, S. and Northcott, D. (2005) 'The average hospital', *Accounting, Organizations and Society*, 30:6, 555-583.
- Lye, J., (2006). Performance measurement in the public sector: A clarification and agenda for research. *Australian Accounting Review*. 16:2, 25-33.
- Malmrose, M. (2015), Management accounting versus medical profession discourse: Hegemony in a public healthcare debate – A case from Denmark. *Critical Perspectives on Accounting*. 27, 144-159.
- Miles, M., and Huberman A. (1994). *An expanded sourcebook: qualitative data analysis*. Second Edition, Sage Publications Inc, Beverly Hills.
- Modell, S., (2004). Performance measurement myths in the public sector: A research note. *Financial Accountability and Management*. 20, 39-56.
- Modell, S., (2005). Performance management in the public sector: Past experiences, current practices and future challenges. *Australian Accounting Review*. 15:3, 56-66.
- Morgan, D., (2008). *The SAGE Encyclopaedia of Qualitative Research Methods*. SAGE Publications, Inc. pp. 816–817. ISBN 9781412941631.
- Mundy, J., (2010). Creating dynamic tensions through a balanced use of management control systems. *Accounting, Organizations and Society*. 35, 499-523.
- Murby, L (with Technical Information Service: The Chartered Institute of Management Accountants)., (2008). *Benchmarking: Topic Gateway Series No. 11*. London.
- Nath, N. and Sharma, U., (2014). Performance management systems in the public housing sector: Dissemination to diffusion. *Australian Accounting Review*. 24:1, 2-20.
- Neu D, Rahaman AS, Everett J. (2014). Accounting and Sweatshops: Enabling Coordination and Control in Low-Price Apparel Production Chains. *Contemporary Accounting Research* 31:2, 322-346.
- Newberry, S., (2003). Sector neutrality and NPM 'incentives': Their use in eroding the public sector. *Australian Accounting Review*. 13:2, 28-34.
- Northcott, D. and France, N., (2005). The Balanced Scorecard in New Zealand health sector Performance Management: Dissemination to diffusion. *Australian Accounting Review*. 15:3. 34-46.
- Northcott, D and Llewellyn, S., (2005). Benchmarking in UK health: A Gap between policy and practice? *Benchmarking: An International Journal*. 12:5, 419-435.
- Pettersen, I.J., (2001). Implementing management accounting reforms in the public sector: The difficult journey from intentions to effects. *European Accounting Review*. 10, 561-581.
- Rogers, E., (1995). *Diffusion of Innovations*, 5th Edition. Simon and Schuster. ISBN 978-0-7432-5823-4.

- Sandhu, R., Baxter, J and Emsley, D., (2008). The balanced scorecard and its possibilities: The Initial Experiences of a Singaporean Firm. *Australian Accounting Review*. 18, 16-24.
- Scapens, R., (1990), "Researching Management Accounting Practice: The Role of Case Study Methods", *British Accounting Review*, 23: 3, 259-281.\
- Sharma, U and Lawrence, S., (2015). Power, politics and privatization: A tale of a telecommunications company. *Critical Perspectives on Accounting*. 28, 13-29.
- Suddaby, R., (2006). What grounded theory is not. *The Academy of Management Journal*. 49:4, 633-642.
- Tillema, S., (2010). Public sector benchmarking and performance improvement: what is the link and can it be improved? *Public Money & Management*. 30:1, 69-75.
- Townley, B., Cooper, D and Oakes, L., (2003). Performance measures and the rationalization of organizations. *Organization Studies*. 24:7, 1045-1071.
- Vagnoni, E and Maran, L., (2008). Public sector benchmarking: an application to Italian health district activity plans. *Benchmarking: An International Journal*, 15:3, 193-211.
- Van Helden., G.J and Tillema, S., (2005). In search of a benchmarking theory for the public sector. *Financial Accountability and Management*. 21:3. 337-359.
- Webster, C and Hoque, Z., (2005). Health sector performance: Cost information in the new public sector environment: Cost accounting change in a state-owned hospital. *Australian Accounting Review*. 15:3, 47-54.
- Wouters, M and Wilderom, C., (2008). Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society*, 33:4:5, 488-516.
- Wynn-Williams, K.L.H., (2005). Performance assessment and benchmarking in the public sector. *Benchmarking: An International Journal*. 12:5, 482-592.
- Yin, R., (2003), *Case Study Research: Design and Methods*, Sage Publications Inc., Thousand Oaks, CA.

Figure Legends

Figure 1: Change on Average Length of Stay-Overnight Episodes (1994/95=1.0)

Figure 2: Annual Bed Days Saved Using Percentage in Overnight Bed Usage between Consortium Average Organisations and National Average

Figure 3: Average Relative Length of Stay in 2001/2002 for Consortium participants

Figure 4: Average Relative Length of Stay in 2003/2004 for Consortium participants

Figure 5: Day of Surgery Admission Rates - Elective Multi-Day Surgery Only (Symbols represent individual hospitals)

Figure 6: Neurosurgery Day of Surgery Admission Rates - Elective Multi-Day Surgery Only Founding Chapter of Consortium (Symbols represent individual hospitals)

Figure 7: Total Bed Days Used by Chronic Obstructive Airways Disease Patients in Year Before, During, and Year After the July 1997 Admission

Figure 8: Average Emergency Length of Stay for Complex Medical Patients by Year (Founding Chapter Member Hospitals Only)

Table 1: Participants, their sphere of activity and position

Sphere of activity	No.
Third party Consortium Personnel	2
Clinical Personnel including Directors, Managers and Nurse Unit Managers	22
Hospital Administrative Personnel	16
Total	38

Table 2: Diffusion of Innovations Approach Adapted from Rogers (1995)

Compute: Focus on identifying differences which entails 3-4 months behind the scenes work designing survey instruments, data collection and adapting benchmark numbers to reflect local conditions.
Communicate: Participants from different hospitals attempt over two days to convince each other that their method or practice has benefits.
Assess: Each participant is free to choose which ideas or practices might be useful. No innovation is forced upon them.
Implement: Once a new idea or practice is verbalised, the Consortium shares implementation materials to assist implementations.
Follow up: The Consortium routinely follows up participants and collects evidence as to whether the new practice or innovation actually leads to measurable improvements.

Table 3: Hip Replacement Australian Refined Diagnostic Related Grouping 103C without Complications as Adapted by the Consortium

Elderly (over 80) , emergency admission, with commodities, discharge home:	10.1 days
Middle age (55-69) , elective admission, commodities, discharge home:	7.2 days