Health Outcomes: An Overview from an Australian Perspective

Jan Sansoni
Director, Australian Health Outcomes Collaboration
Australian Health Services Research Institute, University of Wollongong

23 August 2016

Contents

List of figures

List of tables

Glossary of acronyms

ABS	Australian Bureau of Statistics
AHMAC	Australian Health Ministers' Advisory Council
AHOC	Australian Health Outcomes Collaboration
AHRQ	Agency for Healthcare Research and Quality
AIHW	Australian Institute of Health and Welfare
ACSQHC	Australian Commission on Safety and Quality in Health Care
AQoL	Assessment of Quality of Life
AMHOCN	Australian Mental Health Outcomes and Classification Network
AROC	Australasian Rehabilitation Outcomes Centre
CALD	Culturally and linguistically diverse
CATI	Computer-assisted telephone interviewing
CIHI	Canadian Institute for Health Information
CHSD	Centre for Health Service Development
DALYs	Disability Adjusted Life Years
EQ-5D	EuroQol five dimensions questionnaire
EQ-5D-5L	EQ-5D-5L five dimensions five-level questionnaire
FIM	Functional Independence Measure
HACC	Home and Community Care
HRQoL	Health-Related Quality of Life
HUI-3	Health Utilities Index Mark 3
HoNOS	Health of the Nation Outcome Scale
ICHOM	International Consortium for Health Outcomes Measurement
iPROM	Individual Patient-Reported Outcome Measure
K-10	Kessler Psychological Distress Scale
MHI	Mental Health Inventory
MH-NOCC	Mental Health National Outcomes and Casemix Collection

NATSIS	National Aboriginal and Torres Strait Islander Survey
NHS	National Health Service UK
NHIMG	National Health Information Management Group
NHMRC	National Health and Medical Research Council
NHISSC	National Health Information Standards and Statistics Committee
NHPC	National Health Performance Committee
QALYs	Quality Adjusted Life Years
QI	Quality improvement
QUM	Quality Use of Medicines
QWB	Quality of Well-Being
OARS-ADL	Older Americans' Resources and Services Activities of Daily Living
OARS-IADL	Older Americans' Resources and Services Instrumental Activities of Daily Living
PCOC	Palliative Care Outcomes Collaboration
PREMs	Patient-reported experience measures
PROMs	Patient-reported outcome measures
PROMIS	Patient-Reported Outcomes Measurement Information System
PROQOLID	Patient Reported Outcomes Quality of Life Database
RCT	Randomized Control Trial
PROs	Patient-reported outcomes
PROMs	Patient-reported outcome measures
PROM-PM	Patient-Reported Outcome - Performance Measure
SF-6D	Short-Form Six-Dimension utility index
SF-36 (V1 and V2)	Short Form Health Survey (Version 1 and Version 2)
WHOQOL-BREF	World Health Organization Quality of Life scale (short form)

1   Health Outcomes: A Framework

Health systems and hospitals in Australia and overseas have traditionally evaluated their activities on inputs and outputs and continue to do so. There has been a major focus on the efficiency of service throughputs rather than the effectiveness of interventions. It is only relatively recently that health system evaluation has also focused on evaluating processes and now, outcomes. While we are benchmarking hospital services with regard to cost, we may also wish to increasingly benchmark these services with regard to indicators of quality, and to undertake these activities in an integrated way. The assessment of quality and health outcome must become an integral feature of the model of care.

The health outcomes focus is about health system reform and a cultural change within the health system. It is concerned with the need to know which health interventions work, as against those that produce little or no health benefit, and to know which treatment alternatives for a condition are the most cost effective in producing health gains or increasing health value. This focus needs to be incorporated at all levels of program management; at health system, service and at clinical practice levels. It should also be noted that the 'outcomes' focus has spread to related sectors such as community services and allied health and thus one could increasingly speak of a health and community care outcomes focus. An integrated health system would require continuity of care to be provided across these sectors. Earlier government initiatives such as the Sharing Health Care and related projects (Bywood et al., 2011) and the rounds of the Australian coordinated care trials (Department of Health and Ageing, 2007) emphasised the need for continuity of care to be provided across sectors.

Assessing the relative effectiveness of alternative interventions for the same condition is extremely important when considering the allocation and distribution of health resources throughout Australia at the system or the service level. Similarly it is important to ascertain, in hospital systems, whether efficiency drives to reduce, for example, the length of stay are at the cost of quality or the longer term health outcomes of the patient.

1.1   The Outcomes Context

As elsewhere, the Australian focus on health outcomes has derived from a number of interrelated factors. These factors are as follows:

1. **The increasing proportion of expenditures going to health care and/or the need for cost containment.**

Cost containment has been more of a problem in the US than Australia, where health expenditure as a proportion of the economy has remained relatively stable in recent years. It was about 8.4% of GDP (AIHW, 1994; Ross et al., 1999) in the nineties and is now estimated to be about 9.7% of GDP (AIHW, 2014). Cost containment, however, may always be a necessity particularly given the ageing of the Australian population but it may be better to focus on 'value' created for patients rather than cost based on the volume of services provided (Porter, 2010, Porter et al., 2016).

The 'value' created for patients, is defined as outcomes achieved relative to the costs (Porter, 2010, Porter et al., 2016). The measurement of 'value' rather than 'volume' by health organisations implies the use of relevant health outcome performance indicators as well as patient-reported outcome measures (PROMs) which can assess the value to the patient of the outcomes of treatment. It may also imply consideration of bundled payment mechanisms across the continuum of care (e.g. an integrated practice unit model) rather than payment for discrete services as the drivers of quality improvement and health system reform (Porter et al., 2016).

There is a desire by health systems to allocate funds so that health gain for the community may be maximised. In order to do that, there is a need for health outcome and PROMs information concerning relative effectiveness and the costs of various health interventions. Changes in practice may reduce cost (e.g. reductions in length of hospital stay) but they could potentially reduce 'value' to the patient and the health system unless the outcomes of these changes are monitored and show that they maintain or improve outcomes for the patient. PROMs and other health outcome indicators can also contribute to understanding differences in health expenditure across providers or regions of a health system ensuring that 'efficiency' gains do not come at the cost of quality of care.

There are large unexplained differences in health expenditure by region without any observable differences in health outcomes. For example in Australia, the Hospital Utilisation and Costs Study (AHMAC, 1996) recorded more than two-fold variations in cost per bed-day for hospitals of similar size and apparently similar caseloads. Similar statistics, using a National Weighted Activity Unit as a comparative output indicator, have recently been reported for public hospitals by the National Health Performance Authority (2016). In 2013–14, the cost of providing an average service could be almost twice as high at one major metropolitan hospital ($6,100) compared to another ($3,100).

In those health systems where there has been a split in funding between purchasers and providers (e.g. UK) there has also been an increasing interest in outcomes of health services, in order to advise purchasing choices and to justify service provision (Allcock, 2015).

2. **The recognition of the serious limitations of available information about the effects of many services and treatments.**

There are claims that approximately 80% of commonly used medical interventions have never been demonstrated in control trials to provide benefits to patients (Kingman, 1994), and even fewer have been shown to be effective in routine practice.

The usefulness of randomised control trials as the 'gold standard' for medical practice has been questioned by Bunker (1988), Harvey (1995) and others. The criticism was two - fold: firstly, few treatments were evaluated because of the costs of the method and secondly, the populations on which they were conducted were almost always different to the groups for whom the treatments were intended, and there was generally little evaluation of how well treatments worked in routine practice. Gliklich et al. (2014) in a paper for the Agency for Healthcare Quality and Research also noted that many clinical trials are conducted under conditions that limit their generalisability or do not emphasise factors that are important to patients (e.g. those that may

be captured by PROMs), and/or clinicians, in the course of actual practice. They note that clinical registries can evaluate effects in a more real world population and can help fill the void of information on issues which include treatment options and responses, the natural history of the disease or condition, and quality of life. Given also the central role that registries can play in comparative effectiveness research they note it is important to collect PRO information in registries as they contribute to describing the natural history of the disease, measuring or monitoring safety or harm and measuring quality – which are the key functions of any registry.

There has also been increasing recognition that best practice 'consensus' statements may be a weak foundation on which to base practice (Reinhardt, 1990). Guidelines for the Development and Implementation of Clinical Practice Guidelines (1995) by the NHMRC Quality of Care and Health Outcomes Committee, differentiated between evidence-based and consensus-based guidelines as well as identifying the need for guidelines to indicate areas where no consensus may be reached. Both the revision of A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines (NHMRC, 1999) and the Health and Medical Research Strategic Review (NHMRC, 2000) suggested an increased emphasis on guideline implementation and dissemination and on applied health research to assist in transferring research knowledge into practice.

Clinical practice guidelines have the potential to translate findings from medical research into clinical practice, and when properly implemented have been shown to improve health outcomes (Grimshaw and Russell, 1993, Lugtenberg et al., 2009). A national NHMRC Clinical Practice Guidelines Portal was established in 2010 to enhance accessibility to guidelines produced for Australian practice and a recent report (NHMRC, 2014) reviewed the quality and coverage of 1,046 Australian practice guidelines. This report indicated serious and systemic problems in the ways guidelines are funded, developed and prioritised in Australia including the identification of a number of clinical areas where there was limited coverage (e.g. dementia, mental health conditions, musculoskeletal pain, and ischaemic heart disease). Only a minority of guidelines adequately reported their evidence review process, provided adequate information for guideline users or adequately acknowledged their funding sources. There are, however, current efforts being made to improve this situation and to establish a plan for investment in the development and implementation of prioritised clinical practice guidelines (ACSQHC, 2015a).

3. **The perception of large hospital variations in the use of medical procedures between geographical areas and between physicians.**

Examples would include the inability to explain the observed variations in surgery rates, hospital admissions and diagnostic interventions (Wennberg, 1987) and documented levels of inappropriate interventions for major types of surgery (Leape, 1989). In the USA it has been estimated that only a third of operations for endarterectomy, a half of coronary artery by-pass grafts and two fifths of pacemaker implantations are carried out on patients likely to benefit from the procedure (Leape, 1989).

In NSW it was noted there were variations in the frequency of common surgical procedures across hospitals, which also required some explanation (NSW Health Department, 1994). There have also been concerns expressed with regard to surgical rates for mastectomy (as contrasted

with the preferred procedure of lumpectomy relative to stage) as these rates are apparently much higher in rural than in urban regions (Craft et al., 1996). Similarly large differences in hysterectomy rates have been recorded across the States and Territories (Mishra, 1997) and between rural and urban areas (Reid et al., 1999).

A more recent Australian study examined variations in admission rates for several common hospital procedures across clusters of Medicare Locals (ACSQHC, 2014) and reported considerable variation (1.6-fold variation to 7.4-fold variation). The degree of variation was also found to differ between urban and rural areas and between public or private hospital sectors. Although this report does not explore to what degree such variations were unwarranted, and further research would need to examine this, this report raised issues concerning health care delivery and identified the need to address the lack of systematic monitoring of the outcomes of common health care interventions.

In 2015, the Australian Commission on Safety and Quality in Health Care also released the first 'Australian Atlas of Healthcare Variation', developed in collaboration with the Australian, state and territory governments, specialist medical colleges, clinicians and consumer representatives. The atlas is the first in a series, and explores the significant variation in health care provision across Australia identifying opportunities for improving health care delivery as well as the efficiency and effectiveness of the healthcare system.

Significant variations in surgical intervention rates across Australia were identified. For example, in some areas people 55 years and over had rates of knee arthroscopy that were more than seven times those than people living elsewhere. More than 33,000 operations were performed on this age group during 2012-2013. This was despite the evidence that suggests knee arthroscopy is of limited value for people with osteoarthritis and may cause harm (ACSQHC, 2015b). There were 17,000 lumbar spine admissions, including spinal fusion procedures, on average each year although there is limited evidence to support lumbar spine fusion surgery for those with painful degenerative back conditions and the outcomes of patients receiving these interventions is currently unknown.

Recommendations in Version 1 of the Australian Atlas of Healthcare Variation have noted that variation in the delivery of health care could be augmented by routine, nationally consistent use of PROMs for four particular conditions and procedures (radical prostatectomy, lumbar spine surgery, knee pain and cataract surgery).

4. Concerns as to whether new technologies are actually improving patients' well-being.

An example may be the earlier debate over whether medical rebates should apply to expensive technologies such as medical resonance imaging – whether this could be justified in terms of the technology's contribution to the health outcomes of patients when more widely available and less sophisticated technologies may achieve similar outcomes for less cost. Health technology assessment is an important arena for health outcomes research.

Another example concerned whether the increasing use of less invasive laser surgical techniques for some conditions may lower the threshold at which the surgeon decided to undertake surgery, which may lead to an increase in the number of surgical procedures. An

associated increase in surgical complication rates for some conditions (Hirsch et al., 1994) was earlier reported. A recent case study from the NHS PROMs Programme (NHS, 2016) linked PROMs and other data from the National Joint Registry and identified one implant brand had a significantly higher health gain for knee and hip replacement surgery outcomes.

A more recent presentation by Brooks et al. (2015) indicated the important role that Australian clinical registries can play in monitoring the outcomes of the various treatments, including newer treatments (e.g. robot assisted laparoscopic surgery) for prostate cancer, when combined with relevant data from clinical indicators, health outcome-related performance indicators and patient reported outcomes data.

5. Concerns about the quality of care.

As Harvey (1994) indicated, the quality of care provided has been called into question. Brennan (1991) found that 3.7% of patients hospitalised in New York State experienced adverse events due to medical care, over half of which were avoidable and over a quarter were due to negligence. In Australia, McLaws (1988a, b) found that in 1988, hospital acquired infections would have added $180 million to hospital costs and that 'clean wound' infection rates associated with surgery would have cost $60 million: in both cases significant reductions were possible. The Australian Hospital Care Study reported similar findings with respect to potentially avoidable death and disability (AHMAC, 1996).

Issues concerning safety and quality in Australian health care remain important in the Australian health context as a number of recent reports suggest (ACSQHC, 2010, 2014, 2015a, b, 2016; AIHW, 2009). The Australian Commission on Quality and Safety in Health Care has played an important role in numerous reports related to health outcomes assessment (e.g. those concerning variations in rates of surgical interventions, guidelines, clinical registries, and patient-reported experience measures etc.).

In 2010, Australian Health Ministers endorsed the Australian Safety and Quality Framework for Health Care (ACSQHC, 2010). The framework identified 'consumer-centred care' as the first of three dimensions (including 'driven by information' and 'organised for safety) required for a safe and high-quality health system in Australia. Including this dimension in the framework reflects a growing recognition of the importance of placing patients and consumers at the centre of the healthcare system. The current version of the National Safety and Quality Health Service (NSQHS) Standards, released in 2012 by the Australian Commission on Safety and Quality in Health Care, were developed by the Commission to improve the quality of health service provision in Australia and provide a nationally consistent statement of the level of care consumers can expect from health service organisations. They provide a uniform set of measures of safety and quality applicable across a wide variety of health care services.

Another way to ensure that health care is delivered in partnership with patients is to ask patients about their own perspective on the impact of treatments and care through routine use of PROMs. This should also assist in the evaluation of the outcomes of treatments and services and help to ensure that the outcomes evaluated include those that are most important to patients.

6. **The increasing empowerment of consumers/patients.**

Research into patient assessment of outcomes of care has found significant differences between patient and clinician assessment of outcomes (Wennberg, 1990) – contributing to a decline in the notion that 'the doctor knows best'.

Harvey (1995) also noted there has been a rise in 'consumerism' amongst patients in the USA and Australia. One of the manifestations has been the rapid growth in medical malpractice suits/settlements in both countries, and the assertion of patients' preferences for mode of treatment, initially in areas such as child birth. Many of these issues are concerned with the questioning of the scientific basis of medicine and in association with this, a questioning of the "medical model" of health care.

Entwistle (1995) from the NHS Centre for Reviews and Dissemination, University of York, also refers to the necessity of honesty concerning areas of certainty and uncertainty when communicating with the public on the effectiveness of health interventions. In our current models of care, we can no longer assume that the doctor knows best; it needs to be demonstrated that the health intervention does actually produce health benefit. This means that outcome measurement and monitoring should become a routine part of quality assurance activities and be integrated within the model of health care.

The indicators of the outcomes of care selected by services may not always include the assessment of outcomes that are most important to patients as providers tend to measure outcomes for the interventions and treatments they bill for rather than for the full care cycle (Porter, 2010). In Australia there is a now a renewed interest in the use of patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) combined with a focus on integrated models of care (Chen, 2015).

Lavallee et al. (2016) identify six examples of the use of PROs for patient centred care. Although PROs have traditionally been used to assess and the outcomes of treatment utilising the patient's perspective - other uses include the assessment of the severity of symptoms, informing treatment decisions, monitoring and tracking patient outcomes over time, prioritising patient-provider discussions and connecting providers to patient generated health data. They note that one way to encourage patient centred care is to incorporate PROMs into clinical settings.

The increasing emphasis placed on consumer issues and consumer participation in health service management and evaluation can also be evidenced by efforts to increase consumer participation at the service management level and research to enhance consumer participation and decision making in health care consultations (Sansoni et al., 2015).

1.2 Definitions

Before the measurement of health outcomes commences, it is always useful to check on relevant definitions. In Australia the operational definition defines a health outcome as:

*A health outcome is a change in the health of an individual, or a group of people or population, which is **wholly or partially** attributable to an intervention or series of interventions. (AHMAC February 1993; Modified by NHIMG 1996)*

Given the key notion of attribution in the earlier definition (AHMAC, 1993) it might have implied to some that the only level of evidence that might be considered was that provided by a randomised control trial and as a result the National Health Information Management Group in 1996 added the words 'wholly or partially' to the original statement. Clearly there are different levels of research evidence that may be useful in examining health outcomes, particularly in routine health care settings. For example, these might include practice based evidence studies (Horn et al., 2007; 2012) and the observation data derived from clinical registries (Gliklich et al., 2014).

There are a number of other relevant definitions that apply to the measurement of health outcomes including the use of health outcome-related performance indicators and PROMs as Figure 1 below suggests.

Figure 1   A Health Outcomes Framework
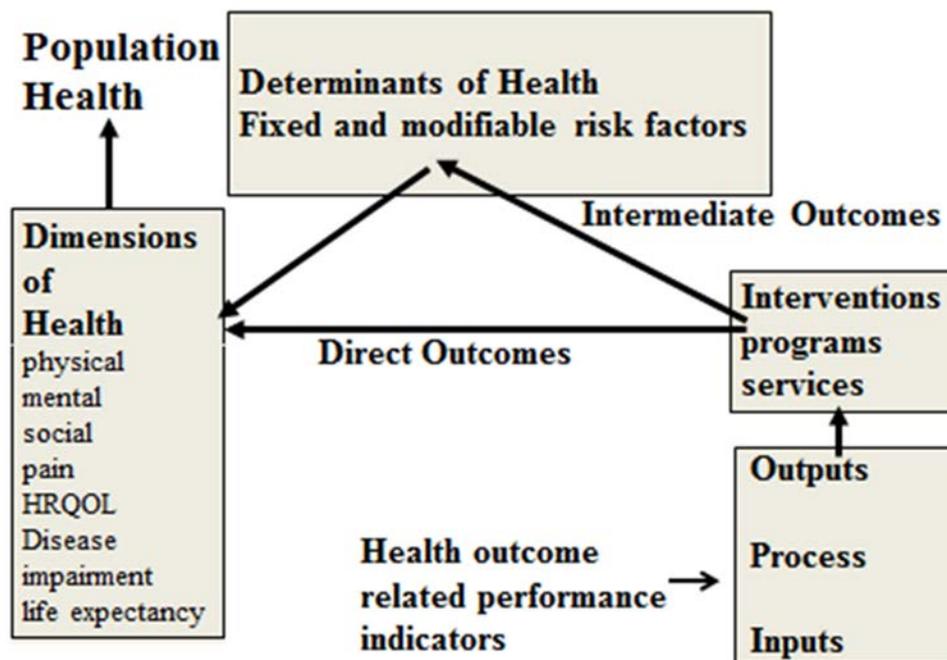
Direct measures of outcome can include clinical indicators, indicators related to survival and standardised measures of health status and function. It is also useful to differentiate between those interventions that have a direct effect on a health outcome (for example, the repair of a broken leg) as against those that may alter risk factors (for example, smoking rates). Where interventions have resulted in modifications to behavioural risk factors it is usually referred to as an intermediate outcome (e.g. reduction in smoking incidence and prevalence), as it will be some time before such population risk factor changes will result in reduced mortality or morbidity for particular diseases or conditions (see Figure 1).

Health outcome measures can include clinical/biomedical indicators, health outcome-related performance indicators, standardised clinical assessments, and PROMs. Increasingly outcome datasets, such as those developed by the International Consortium for Health Outcome Measurement (ICHOM) amongst others, use many of these measures of health outcome, including PROMs, in their datasets.

When the purpose is to gather information on health system or health service performance, health outcome-related performance indicators (sometimes also known as process outcomes or quality indicators) may be used. In these situations, when gaining information on the performance of health services (rather than the change in health status of individuals which can be aggregated), the following definition has been proposed by Bruce Armstrong:

> *An outcome-related performance indicator in the health and welfare field is a statistic or other unit of information which reflects, directly or indirectly, the performance of a health and welfare intervention, facility, service or system in maintaining or increasing the wellbeing of its target population. (Armstrong, 1994)*

Outcome-related performance indicators often measure aspects of process (e.g. practice variations). Health outcome-related performance indicators might include, for example, data on the rate of avoidable adverse events, hospital acquired infection rates, time to treatment rates, return to theatre rates, and unplanned readmission rates. National comparisons of hospital performance usually include such indicators. These indicators often focus on processes such as practice variations, some of which are important predictors of outcomes (e.g. 'time to receive treatment') but some process indicators (e.g. unplanned readmissions) may not always be clear indicators of outcome, since these may be potentially influenced or confounded by other factors (e.g. hospital policies on initial length of stay, bed availability, etc.).

Organisations may be collecting data that is related to health outcomes within their routine data collections and 'time to receive treatment' can be an important outcome-related indicator in cancer research (Tracey et al., 2012). Another example, concerning a tertiary rehabilitation service, found that there was considerable time lag (an average of eight years) between patients' discharge from hospital-based rehabilitation programs and their presentation to the tertiary rehabilitation service. In the intervening years these patients had often been unemployed and experienced less than optimum health and welfare outcomes, which may have been at least partly due to a lack of appropriate links between secondary and tertiary rehabilitation services. An examination of such data by organisations can often suggest effective ways to improve health service delivery which are related to outcome.

Standardised Measures of Health Status and Health-related Quality of Life

There are a range of standardised and validated measures of health status/health-related quality of life (HRQoL) that along with clinical and performance indicators can be used to assess the outcomes of treatment interventions. These measures can be clinician-rated (for example the Health of the Nation Outcome Scales, HoNOS; Wing et al., 1998) forming standardised clinical assessments or they can be patient-rated in which case they are known as PROMs.

Patient-reported outcomes (PROs) can be defined as follows:

A PRO is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. (Food and Drug Administration, 2009)

PROMs are the instruments used to measure PROs (Weldring and Smith, 2013). The reliance on PROs is driven by the view that patients are the best judge of their own welfare. PROMs usually focus on outcomes of care and may include for example, patients reporting on their symptoms, functional status and HRQoL prior to, during and after their treatment through the completion of standardised and validated questionnaires. By comparing patients' self-reported health before and after the intervention, the outcomes of the care they received can be assessed (Black, 2013).

These questionnaires can be generic and measure 'general' health status (generic PROMs) or be 'disease-specific' (e.g. for a type of cancer) or 'condition-specific' (apply to a service such as rehabilitation or mental health services or a population segment such as the elderly). In Section 2 a large number of these measures are discussed.

The National Quality Forum (NQF) in the US (2013) reports on the use of PROs in performance management and differentiates between a PRO (e.g. a patient reporting depression), a PROM which is a standardised measure of this (e.g. Patient Health Questionnaire-9; PHQ-9; Kroenke et al., 2001) and a PRO-Performance Measure (PRO-PM) (e.g. % of patients with diagnosis of major depression or dysthymia and an initial PHQ-9 score > 9 with a follow-up score of < 5 at six months). When PROs are aggregated across organisations or systems for the purposes of performance measurement, they are known as PRO-based performance measures (National Quality Forum, 2013). Creating appropriate PRO-PMs involves specifying relevant PROs, selecting appropriate instruments (PROMs) and collecting, aggregating and reporting them in standard ways to reflect an organisation's performance. An endorsed PRO-PM is one that meets stringent NQF standards and can be used as an endorsed measure of quality improvement and for accountability purposes.

PROMs data are also now being included in clinical registries to assist in identifying effective health care practice and to benchmark the performance of healthcare providers (Nelson et al., 2016). In Sweden the use of PROMs is linked with the national disease-specific clinical databases (quality registers) although PROMs were only included from 2000 onwards (Black, 2013). Sweden probably leads the world in the provision of comprehensive quality registries and there appear to be over 100 registries benchmarking costs and outcomes data for treatments for a broad range of diseases and conditions including such areas as dementia and mental health conditions. Over 90% of these registries now include PROMs data (Lundstrom and Karlskrona, 2015).

PROMs have been increasingly used in clinical trials to assess the health outcomes of interventions, but more recently are also being used as a quality improvement tool (Lavallee et al., 2016) to monitor outcomes of individual care.

PROMs are rarely used as stand-alone measures in clinical applications but are used in conjunction with other indicators. Increasingly health outcome datasets or measurement suites may include all of the types of measures and indicators discussed above including PROMs.

1.3   Health Status Monitoring and Health Outcomes

Health outcomes should be differentiated from the related area of population heath status monitoring – which is concerned with the provision of statistics about the health status of Australians over time but which is not directly concerned with the issue of attribution per se.

An example is the death rate for coronary heart disease from 1983-1993. There was a major reduction in the death rate for males (Abraham et al., 1995), but it was unclear which factors were responsible for this trend. It is quite possible that this may have been strongly associated with the decline in major risk factor determinants of health such as smoking rates. Some of this 'health gain', however, could also be due to lower fat diets, reduced serum cholesterol, and males exercising more. Without sophisticated causal modelling, or the evaluation of strategies used to modify such risk factors, the relative contribution of our health interventions to such health gains remains unknown. Changes in population health status over time are not examples of health outcomes research unless we can make some level of attribution to an intervention or interventions that may have been responsible for this change.

System wide approaches to health information such as the National Health Indicators Performance Framework (NHISSC, 2009) include health status measures, health outcome measures, measures of health determinants and health outcomes-related performance indicators within the framework, as clearly all such sources of information will be required to reflect on both population health and health system performance. However, when examining such 'system' frameworks it still remains important to distinguish between the different types of information each of these indicators can provide – some measures and indicators will be related to health outcome while others will not (see Figure 3). Many of these indicators at the national level will be population health status measures.

1.4   The Health Evaluation Cycle

Much of the confusion concerning health outcomes arises because this term is not placed in its evaluation context. Formative and summative evaluation and review mechanisms may be embedded within each phase of the evaluation cycle. It is important to continue to undertake evaluation activities at all phases of this cycle. The health outcomes focus is about including the evaluation of outcomes in evaluation activities, as for too long this has been ignored. Just as inputs, processes and outputs data can be examined at the system, service and practitioner levels so may outcomes data. The different indicators and measures currently utilised for health evaluation are placed within this framework so that outcomes are seen as one part of the evaluation cycle with outcomes, in turn, leading to inputs. Ultimately reliable outcomes information should be informing resource allocation (inputs) so that resources are allocated where they will produce the most health gain for the community.

The current challenge, however, is to develop valid, reliable and responsive health and welfare outcome measures and outcome-related indicators, and to endeavour to get some uniformity in the use of measures at system and service levels to allow for comparability of results.

Donabedian (1980, 1992) did not see much merit in debates of 'process versus outcomes' but saw them as necessary and related links in the chain as does Porter (2010). It is important to improve processes and to set standards and benchmarks by gaining evidence and/or consensus about best practice, but, this is not enough as outcome evaluation will also be required.

It can be seen that it is possible to have performance measures at various stages in the evaluation cycle (e.g. input, process and output indicators) but these should not be confused with measures of the ultimate outcome, which is a change in health status that can be attributed to an intervention. Where research is unclear or unavailable with respect to attribution, some outcome-related performance indicators may reflect upon, predict or be associated with outcome. It is desirable to measure outcomes/outcome-related performance indicators with respect to all phases of the treatment and service cycle from primary prevention through screening, diagnosis, treatment, rehabilitation, palliative care to community service use.

1.5 The Health Outcomes Jigsaw

The health outcomes approach is about integrating the elements for health improvement. There are many key elements that can contribute to health outcomes (e.g. a focus on quality improvement, addressing the continuum of care, guidelines development and application, benchmarking, consumer issues etc.) but there is a need to integrate these elements rather than to address such issues in isolation from each other.

1.5.1 Health Outcomes and Quality Improvement

As the purpose of health outcomes measurement is to ensure better health outcomes for our patients it is important to ensure health outcomes measurement activities are integrated with our quality improvement processes.

As an early example, in the United States, Batalden et al. (1994, 2007) noted that outcomes measurement, process improvement and continual quality improvement all have been used to improve quality in health care, but they also noted that these complementary approaches have been used primarily in isolation from each other. Batalden et al. (1994) suggested the use of a 'Serial V' approach as an integrative strategy to incorporate these approaches to create a comprehensive way to evaluate, institute and reflect on change aimed at improvement within the hospital sector. Such an approach integrates continuous measurement of outcomes, satisfaction and cost assessment throughout the clinical path of care. Wu et al. (2013) have recently provided case studies of a number of healthcare organisations in the US that are now integrating PROMs data collections with the electronic health record to both promote quality improvement in clinical practice (by outcomes benchmarking the results of providers) and for research concerning the effectiveness of interventions.

Outcomes Benchmarking Initiatives

In the UK, following a malpractice case, Bupa hospitals (now Spire Healthcare) started collecting PRO data in 1998 and they have now collected more than 100,000 patient episodes (Devlin and Appleby, 2010). Although initially data collection covered a wide range of procedures, the current focus is on one or two sentinel procedures for each specialty, using generic measures,

such as the SF-36, and condition-specific questionnaires (such as the Vision Function-14 for cataract surgery and the Oxford Hip and Knee scores for joint replacement procedures). Although partly driven by a desire to identify clinical 'bad apples', it was considered that the collection of PROMs also offered the potential for continuous quality improvement and to provide feedback to health professionals and patients (Devlin and Appleby, 2010; Vallance-Owen, 2008). One of the main benefits of the PROMs data collection included the identification and sharing of best practices. This prompted a number of quality improvement practice changes which included changes to clinical pathways, enhancing communication with patients concerning their procedure, and a greater focus on post-operative pain relief following hysterectomy. With the PROMs results (at hospital level) posted on their websites, Bupa Hospitals were also able to promote and market the HRQoL benefits of the interventions they provided (Devlin and Appleby, 2010).

Since 2009 it has been mandatory for providers in the UK to undertake PROMs data collection in four areas of elective surgery (hip and knee replacement, groin hernia repair, and varicose vein surgery) (Black, 2013). The preoperative questionnaire includes data on the patient's sociodemographic characteristics, the duration of their condition, their general health, any comorbidities, and whether they are undergoing a repeat/revision procedure. In addition, they are asked to complete a disease-specific PROM (Oxford Hip Score, Oxford Knee Score, or Aberdeen Varicose Vein Score) and a generic PROM (EQ-5D index and EQ-Visual Analogue Scale). A similar post-operative questionnaire is mailed to participants at a relevant time (three or six months) following their surgery. The PROMs data are linked to Hospital Episode Statistics and a regular analysis of each provider's preoperative patient characteristics (age, sex, severity) and the mean change in the PROM scores adjusted for casemix is provided. Providers are identified and compared by means of funnel plots that show whether or not any provider's outcome is significantly different from what would be expected.

As Black (2013) indicates this outcome benchmarking data can be used to identify ways to address practice variations in the outcomes from these procedures. Some examples of this are provided by the Health and Social Care Information Centre (2015). Given that the costs of such data collections are quite substantial (estimated at GBP 825K annually) the NHS England (2016) has recently conducted a consultation survey (early 2016) concerning the use and benefits of PROMs collection in the UK. In the interim a number of key orthopaedic groups and the National Joint Registry (2016) have placed their joint response on the web. This joint response strongly endorses the PROMs programme and it highlights the value that PROMs provides in both comparative data to support quality improvement and for research purposes regarding outcomes. Similarly, a major point raised by the Office of Health Economics (2016) response was that PROMs data are vital to understanding the effectiveness and cost effectiveness of NHS services. For this reason, they strongly urged NHS England to continue to field both a brief, generic PRO questionnaire in combination with a detailed, condition-specific measure, where available. It was stated that Generic PRO data provided the crucial, common denominator with which to measure outcomes across treatments and diseases. The final report and findings of the broader consultation process by the NHS will be of great interest.

As another international example, Porter (2016) also cites the case of OrthoChoice in Sweden. The Stockholm County Council since 2009 has reimbursed total hip and total knee replacement surgeries using bundled payments (for both public and private providers) for the care cycle as a

means to include responsibility for avoidable complications. The PROMs included functional outcomes for patients along with other outcome-related indicators. In the first two years complications fell by 25%, length of stay reduced by 16% and cost was reduced by 17% but the functional outcomes for the patients remained constant. This example shows how 'value' in health care is increased by such quality initiatives. Currently there are eight bundled payment initiatives being tested in Sweden: osteoarthritis, spine surgery, obstetrics, obesity - bariatric surgery, stroke, diabetes, osteoporosis and breast cancer (Wohlin, 2014). Some other countries, such as the US are currently undertaking experiments with bundled payments (Centers for Medicare and Medicaid Services, 2015) to facilitate more integrated care across sectors and which also provide incentives for the use of PROMs to assess and monitor quality.

There are a number of disease-specific clinical registries in Australia that are beginning to benchmark performance of relevant services on clinical and health outcomes-related performance indicators. Relatively few have incorporated the use of PROs data to date although this is an increasing trend as can be seen by the inclusion of PROMs data in prostate cancer registries (Brooks et al., 2015).

Related service sector registries such as the Australasian Rehabilitation Outcome Centre (AROC) and the Palliative Care Outcomes Collaboration (PCOC) were established to undertake outcome benchmarking activities for these service sectors (Marosszeky and Eagar, 2001; Eagar, 2014) and the Australian Mental Health Outcomes and Classification Network (AHMOCN) is fulfilling a similar role for the mental health sector (see Section 2.6.1). Some more recent benchmarking and patient decision making systems (e.g. DiscoverQuick, http://www.discoverquick.com/) now make use of web-enabled intelligent knowledge management systems which have the capacity to provide real-time feedback to participating clinicians or groups (Sansoni et al., 2013b).

Individual Patient Monitoring

In research settings there has been a recent interest in evaluating the use of PROMs in quality improvement activities within routine care. Lavallee et al. (2016) noted that PROs have traditionally been used to assess the outcomes of treatment from the patient's perspective (e.g. in comparative and clinical effectiveness research). Information from such research can be used to inform patients about the likely outcome of their medical treatments and can assist with informed decision making. However, Lavallee et al. (2016) identified some more recent uses for quality improvement which included the assessment of the severity of symptoms, informing treatment decisions, monitoring and tracking patient outcomes over time, prioritising patient-provider discussions and connecting providers to patient generated health data.

Boyce and Browne (2015) noted that reviews of these QI interventions demonstrated that the use of PROMs can improve patient-clinician communication and the processes of care for individual patients, but also noted they have also consistently shown minimal influence on patient health status outcomes. However, as QI interventions are really patient management interventions, rather than treatment interventions, might it may be unrealistic to expect such changes in health status? Sansoni (2015) indicated even well-validated generic PROMs have been shown to be somewhat insensitive when used to assess 'management style' interventions such as the coordination of care (Department of Health and Ageing, 2007). Thus, the outcomes indicators selected to assess these QI interventions should be clearly related to the

hypothesised effects of the intervention which might well include such factors as improved communication and better detection of symptoms etc. However, the research on the use of PROMs with regard to QI style interventions is still at a relatively early stage of development and Boyce and Browne (2013) noted that many of the RCT studies included in their review were rated as being of relatively poor quality. Further consideration may be required regarding the appropriate outcomes measures for such applications of PROMs. Gonçalves Bradley et al. (2015) outline a protocol for a Cochrane Collaboration review which includes consideration of the appropriate endpoints for such studies.

By contrast to the studies reviewed above, a study by Shadbolt et al. (2015) indicated the usefulness of PROMs for surveillance and patient monitoring applications. This study examined the factors that affected the risk of subsequent joint arthroplasty in patients who have had primary hip or knee arthroplasty. They found that the surgical trajectory for osteoarthritis patients after their first lower limb arthroplasty was significantly dependant on the joint (hip or knee), patient's age, one year post-operative Oxford Hip Score (OHS) and the pre-operative Oxford Knee Score (OKS). This study identified the use of these PROMs (OHS and OKS), with associated defined score cut-points, as important markers for joint arthroplasty failure. Such an application of PROMs not only can inform surgeons and patients about the likelihood of future joint failure but it also clearly identifies the need for enhanced surveillance of 'at risk' patients.

However, the research on the use of PROMs with regard to QI style interventions is still at a relatively early stage of development. There is also a need for further consideration of the best ways to integrate health outcomes assessment within the quality of care improvement cycle and to also consider this across the continuum of care.

1.5.2 Continuum of Care

Similar approaches have also been expanded to cover treatment across the continuum of care, as has been exemplified by the natural history of disease approach. A promising approach that was developed in NSW examined a framework for the natural history of the disease or condition (e.g. the diabetes prototype) and identified interventions and their processes and outcomes at each stage of care from primary prevention through to palliative care (NSW Health, 1994).

Porter (2010) outlined a similar approach and noted that outcome measurement has tended to focus on the immediate results of particular procedures or interventions rather than the success of the full care cycle for medical conditions. For this reason Porter advocates the use of bundled payment mechanisms across the continuum of care (e.g. an integrated practice unit model) rather than payment for discrete services as the drivers of quality improvement and health system reform (Porter et al., 2016).

There is a need to identify outcomes indicators across the continuum of care and not to focus only on outcomes of acute care in isolation. It is for this reason that models of care that focus on coordination / integration of care / shared care became increasingly popular and of interest to funders, as was indicated by the national coordinated care trials and the establishment of a range of chronic disease self-management projects. A recent review by Chen (2015) examined

the potential use of PROMs and PREMs with reference to the NSW Health integrated care strategy.

Asthma is another example where effective links between acute and ambulatory care can assist in reducing the number of avoidable admissions to hospital (which is both in the interests of the individual and the community). This is an example where a health outcomes focus can lead to cost savings and greater efficiency as well as effectiveness.

A major study in the ACT by Shadbolt et al. (1996b, 1997) tracked a large sample of patients from admission through discharge to six months post discharge in the community. This study examined both service utilisation and the HRQoL outcomes of approximately 6,000 patients. Assessing the health outcomes of health interventions will often necessitate tracking patients post discharge from hospital, which will increasingly require coordination between community and acute services, common data definitions and compatible information systems.

1.5.3   Guidelines

Similarly, there has been a growing emphasis on the implementation of evidence based best practice guidelines to improve practice (NHMRC, 1999). As Harvey (1995) indicated, the Guidelines for the Development of Guidelines (NHMRC, 1995) document also recommends that known adverse outcomes from treatments should be documented and should be incorporated into specific quality assurance processes by hospitals, colleges and others who certify and re-certify health professionals. It recognised that unless guidelines are used, and the results of their application are collected and reviewed, practice will not improve.

The monitoring of practice variations in relation to guidelines can be viewed as outcome-related clinical benchmarking. For example, Craft et al. (1996) reported there was a higher rate of mastectomy for earlier stages of breast cancer in the rural as opposed to the urban areas of Australia, when lumpectomy should be the preferred procedure. The colleges may wish to address such issues through the monitoring of practice and then provide educational programs to ensure greater compliance with the guidelines.

The uptake of guidelines will often be influenced by both their credibility (commercial or government sponsored) and whether key stakeholders have been involved in their development. In the USA the implementation of guidelines was found to be much better for asthma specialists than for general practitioners, as the latter had been excluded from the development process and thus found some aspects of the guidelines impractical for local implementation. Similarly as Ken Harvey (2003) suggested some of the increase in pharmaceutical expenditure in Australia may partly be due to inappropriate prescribing patterns and again there is a need for GPs to become involved in guideline development and implementation through initiatives such as the Quality Use of Medicines (QUM) strategy. It is also important that up to date guidelines for prescribing are integrated within prescribing software systems.

Implementation of the evidence-based medicine and the health outcomes approaches are also influenced by information accessibility. In 1997, the US Agency for Healthcare Research and Quality (AHRQ) established the National Guideline Clearinghouse (www.guideline.gov). This

clearinghouse serves as an international electronic repository of clinical guidelines, although most guidelines currently included have been developed in the United States. The mission of the National Guideline Clearinghouse has been to provide physicians and other healthcare professionals, healthcare providers, health plans, integrated delivery systems, purchasers and others an accessible mechanism for obtaining objective, detailed information on clinical practice guidelines and to further their dissemination, implementation and use (Slutsky, 1998). The Clearinghouse provides structured abstracts, tabular comparisons of abstracts, syntheses of clinical guidelines on similar topics, full text of guidelines or ordering information, electronic discussion groups and annotated bibliographies. To satisfy inclusion criteria all guidelines must meet the Institute of Medicine definition of a clinical practice guideline (Field and Lohr, 1992), show proof of a substantial literature search and a review of the current scientific evidence and have been developed or revised in the last five years. As long as guidelines meet the evaluation criteria they are placed on the site; no other formal endorsement procedure exists for those guidelines.

A national NHMRC Clinical Practice Guidelines Portal was established in 2010 to enhance accessibility to guidelines produced for Australian practice. The NHMRC website currently provides access to a range of guideline documents developed in association with the NHMRC in Australia, and State and Territory Health Department websites also contain guideline information.

1.5.4   Consumer Issues

Although the best practice guidelines under development are now insisting upon the adequate provision of information to patients to inform their choice of treatment alternatives, practitioners and health professionals often state that many patients neither want this information nor desire to make the choice about treatment alternatives. It has been claimed that patients rarely have the required medical knowledge to evaluate the information they may be given. One of the choices that a patient can make, however, is to be guided by their practitioner's view if they wish. However, this choice should be overt and clear, and for those patients wanting further information it is necessary that this is provided in a way that can be understood.

Dr Vikki Entwistle (NHS Centre for Reviews and Dissemination, University of York) has been involved in trials concerning the evaluation and development of information provided to consumers in the UK. Entwistle (1995), as indicated earlier, stressed the need to involve consumers in making choices and to be honest about the areas of certainty and uncertainty concerning interventions, whereas frequently health professionals want control over what information is given to the consumer. For example, if information concerning the pros and cons of various treatments for glue ear is given to patients/carers then it may be that surgical rates will decline which may then make surgeons somewhat unwilling to distribute this information.

The selection of outcome indicators to evaluate treatments for health conditions should focus on those outcomes that are most relevant or important to patients, be comprehensive across the full care cycle for their condition, and not just be those indicators that are easiest to measure (Porter, 2010). The increasing use of PROMs internationally reflects the view that patients may be the best judges of their own welfare.

Traditional survival, disease, and physiological outcomes may demonstrate the physiological benefits of treatment; however, the patient perspective provides a more holistic interpretation and a comprehensive assessment of the benefits of the treatment under investigation (Black, 2013). There is substantial body of evidence to support the use of PROMs in comparative effectiveness research to evaluate the outcomes of treatment interventions (Weldring and Smith, 2013) and clearly this information can be provided to patients to assist with decision making concerning medical treatments.

Some recent research and consumer initiatives have also examined ways to support patient-centred care, and to enhance consumer participation and shared decision making in healthcare consultations (Sansoni et al., 2015).

1.5.5   Initiatives for Integration

To incorporate a health outcomes focus in quality improvement activities the following will be required in Australian hospitals and health services:

- a client-centred consumer focus;

- staff training in health service evaluation;

- good information technology systems and common data definitions;

- a commitment to patient-based assessment of health outcome, e.g. HRQoL and patient experience rather than just survival;

- a commitment to undertaking some patient follow-up post hospital discharge;

- greater continuity and coordination of care across the continuum of care (e.g. across the primary, acute care, rehabilitation and palliative care settings);

- a commitment to evidence based practice and to change practice when there is evidence that an alternative treatment/intervention produces greater health benefits; and

- a preparedness to assess the relative cost effectiveness of new technologies compared with other methods rather than to adopt such technologies without appropriate evaluation.

1.6   Some Other System Initiatives

The National Health Performance Committee published the National Health Performance Framework Report (NHPC, 2001), which was derived from the Canadian Health Information Roadmap Initiative Indicators Framework (CIHI, 2000). A revision of this framework was published in 2009 (NHIS&S). The stated goals of this report were to promote benchmarking based on national health performance indicators; to improve the quality of care of health services; and to extend the national performance indicators framework to include such areas as community health, general practice and public health. The framework is presented below. The framework proposes a range of indicators in such areas as health status and health outcomes, health determinants and system performance indicators – but only some of these indicators will be outcome related.

Figure 2   The National Health Performance Framework 2009

Health Status
How healthy are Australians? Is it the same for everyone?
Where are the best opportunities for improvement?

Health conditions
Prevalence of disease, disorder, injury or trauma or other health related states

Human function
Alterations to body, structure or function (impairment), activity limitations and restrictions in participation

Wellbeing
Measures of physical, mental and social wellbeing of individuals

Deaths
Mortality rates and life expectancy measures

Determinants of Health
Are the factors determining good health changing for the better?
Where and for whom are these factors changing? Is it the same for everyone?

Environmental factors
Physical, chemical and biological factors such as air, water, food and soil quality

Community and socioeconomic
Community factors such as social capital, support services, and socio-economic factors such as housing, education, employment and income

Health behaviours
Attitudes, beliefs, knowledge and behaviours such as patterns of eating, physical activity, smoking and alcohol consumption

Bio-medical factors
Genetic-related susceptibility to disease and other factors such as blood pressure, cholesterol levels and body weight

Health System Performance
How does the health system perform? What is the level of quality of care across the range of patient care needs? Is it the same for everyone?
Does the system deliver value for money and is it sustainable?

Effectiveness
Care/intervention/action provided is relevant to the client's needs and based on established standards. Care, intervention or action achieves desired outcome

Safety
The avoidance or reduction to acceptable limits of actual or potential harm from health care management or the environment in which health care is delivered

Responsiveness
Service is client orientated, clients are treated with dignity, confidentiality, and encouraged to participate in choices related to their care

Continuity of Care
Ability to provide uninterrupted, coordinated care or service across programs, practitioners, organisations and levels over time

Accessibility
People can obtain health care at the right place and right time irrespective of income, physical location and cultural background

Efficiency & Sustainability
Achieving desired results with most cost effective use of resources. Capacity of system to sustain workforce and infrastructre, to innovate and respond to emerging needs

Community and Health System Characteristics

Many of the higher level indicators (e.g. prevalence, incidence, mortality, life expectancy) will be population health status measures for both descriptive purposes and for making comparisons between the states and territories which might identify variations that may warrant further investigation. Indicators relating to treatment effectiveness such as variations in medical interventions or practice (e.g. rates of surgical interventions) can also be examined across states and territories or clusters of Medicare Local areas (ACSQHC; 2014, 2015b). Variation can be driven by a range of complex and interacting factors which can be both warranted (e.g. reflecting differences in population need, patient preferences or innovative

practice) or unwarranted (e.g. some patients may be receiving unnecessary and potentially harmful care or not receiving required care). It is important to undertake outcomes research to determine the factors that are associated or may cause unwarranted variation before improvements to quality of care can be undertaken.

Many health system performance indicators relate to structure, process, and output / throughput variables, and only some aspects within each cell may be related to outcome. Clearly 'effectiveness' of interventions is primarily outcome focused, whereas many of the 'efficiency' indicators will be output focused. While there is general acceptance that health determinants influence health status and health outcomes, the causal pathways are not always clear. Further work is also required to develop appropriate indicators in such areas as community capacity and health inequalities.

2   How Do We Measure Outcomes?

2.1   Typology of Outcomes Measures and Instruments

Table 1   Typology of Health and Associated Outcome Measures

| Quantity of Life

Mortality | Process Based Outcome Measures | Quality of Life
Health-related quality of life
Dimensions of health | | | Satisfaction with Health Care - Patient satisfaction surveys |
|---|---|---|---|---|---|
| | | Physical | Social | Mental | |
| Avoidable premature mortality | Re-admission rates
Relapses
Complications | Measures of impairment
Disease specific measures
Pain scales
Measures of functional status
Measure of handicap | Measures of social support
Measures of disability | Measures of depression
Measures of social adjustment | |
| | | Multi-dimensional health status profiles and indexes cover many of these dimensions in a single instrument | | | |

*(This diagram appeared in the article 'Issues in Outcome Measurement' in Outcomes Briefing, UK Clearing House for Information on the Assessment of Health Outcomes, Nuffield Institute for Health, Introductory Issue, Spring 1993, p. 11)

The UK Clearing House suggested the typology of health and associated outcomes measures depicted above. Porter (2010) has also developed an Outcomes Measurement Hierarchy which includes similar types of outcomes measures but these are examined across the continuum of care for particular diseases (e.g. breast cancer) and this article and associated appendices is highly recommended to readers.

Mortality and period of survival are commonly used population health outcome indicators but they are not particularly responsive to the change in delivery of health care, as for many conditions it may take some years for the reduction in mortality, or increase in years of survival, to become apparent.

Various Australian surveys and studies have included items on morbidity and changes in behaviour related to risk factors (e.g. incidence, prevalence and levels of smoking and drinking behaviours, dietary factors). Given the substantial evidence associating smoking as a causal agent in a variety of health conditions the reduction in the incidence and prevalence of smoking, for example, could be seen as a proxy or an intermediate outcome reflecting on the ultimate outcome of improved population health.

Process-based outcome measures (health outcome-related performance indicators; HORPI) that may also be of relevance might include hospital acquired infections, complication rates, admissions and readmissions (recurrences) to hospital or health services and time to receive appropriate treatment. Stage of disease at diagnosis could be a measure of earlier intervention associated with improved quality of care, as could the reduction in later stage complication rates for some diseases (e.g. limb amputation for diabetes). Similarly, a reduction in infectious disease notifications and an increase in immunisation rates might be a product of better coordination of care processes.

More recently there has been greater interest in the quality of life following a health intervention. One may be cured of a disease, but have residual disabilities and handicaps, and these may have incurred costs to the individual and society. There are a range of instruments and measures (PROMs) that are used to assess such factors which are discussed in the following section.

Patient satisfaction with the process of care and service experience can contribute, or be related, to outcomes (e.g. a health outcomes related performance indicator) but is not a true/direct measure of health outcome. Patient satisfaction surveys may include items which cover hospitality, friendliness of staff, amenities etc. which may reflect on service experience (Porter, 2010) but these elements may only be marginally related to health outcomes. For example, one may experience poor health outcomes from a given treatment but one may still be satisfied with the care provided. Patient compliance with treatment protocols may have a more major influence on outcomes but it is rarely measured (Porter, 2010).

There are also well-being outcomes other than health that one may wish to consider, for example well-being indicators such as housing, employment, service utilisation and access, satisfaction with coordination of services and degree of participation by the community.

2.2 Health, Health-related Quality of Life, Quality of Life and Well-being

Historically, health has usually been referred to negatively as the absence of death, disease and illness. The World Health Organisation (WHO) has recommended the development of measures of positive health and defined health as:

> *A state of complete physical, mental and social well-being, and not merely the absence of disease or injury (WHO, 1981).*

Although this definition is somewhat broad and idealistic, and has been criticised because of the difficulty of defining and measuring 'total' well-being, it has also focused attention on a more positive concept of health. In order to conceptualise and measure positive health, researchers have developed multi-dimensional models of health, involving more than one health concept or 'dimension' (Kaplan et al., 1976; Rosser, 1988; Ware and Sherbourne, 1992). Health concepts most frequently included in such models are:

- morbidity (disease or impairment);
- limitations to functional abilities (disability);
- role limitations because of health problems (handicap);

- bodily pain;

- mental health (psychological distress and psychological well-being);

- vitality (energy/fatigue); and

- general perception of health (e.g. excellent/good/fair/poor).

Although some of these concepts relate to personal, economic and social well-being, broader concepts of well-being have not usually been included in such multi-dimensional models of health or the instruments developed to measure them. Such concepts are dimensions of another imprecisely defined term 'quality of life' (Harvey, 1991; Kolstad, 1994). Quality of life is generally understood as a broader concept than health, but the concept means different things to different people, reflecting differences in experience, perceptions and values. It is equivalent to the term total well-being and may include:

- health;

- social well-being;

- economic well-being;

- environmental well-being (sustainability);

- life satisfaction;

- spiritual or existential well-being; and

- other characteristics valued by humans.

These dimensions overlap with each other, therefore in measuring life satisfaction one may also measure aspects of health and social well-being, and in measuring social well-being there are large overlaps with health and economic well-being.

It has become common to equate non-clinical dimensions of health (such as disability, handicap, psychological well-being, general perception of health) with 'quality of life'. This is likely to be a source of conceptual confusion (Bowling, 2005; Harvey, 1991), and it is recommended that, if the term must be used in relation to aspects of well-being that relate to health, then the terms 'health-related quality of life' or 'health status' be preferred. There is an increasing trend to also refer to instruments that measure such aspects as 'patient-reported outcomes measures' (PROMs) or 'patient-reported health outcomes measures'.

2.3   Well-being or Health-related Quality of Life?

For practical reasons, instruments of perceived health status or HRQoL usually measure the domains of health mentioned above. Instruments that focus on quality of life more broadly, such as those developed by Becker et al. (1993) and Cummins (1993), include a greater coverage of the social, economic and employment domains of well-being, but they usually do so at the cost of having a less extensive coverage of the health domain. Focusing on the health aspect of total well-being, however, should not mean the other aspects of total well- being are ignored. Clearly an impact of a disability may be to reduce one's material well-being, which may in turn have a profound effect on quality of life. Having reduced material circumstances (e.g.

loss of employment) can clearly affect one's health either 'directly' or 'indirectly' through lesser access to treatment (equity issues). Many of these welfare outcomes may be particularly important with reference to conditions resulting in long-term residual disabilities.

One of the first decisions that need to be made when endeavouring to measure quality of life outcomes for particular conditions is whether to focus on a broad measure of quality of life, or a measure of HRQoL. A broader measure may tap some aspects of health and welfare outcomes for health conditions and these may be of particular significance for some conditions, or may be particularly related to the questions of the study.

Figure 3   Some different dimensions relating to the focus and type of HRQoL measures

Physical Mental Social	Impairment Disability Handicap	Disease, Symptom, Condition Specific, Generic Measure	Single Dimension, Multi-dimensional, Profiles/Indexes

Generally researchers choose to measure HRQoL, as invariably the choice of instruments involves a trade-off between depth, breadth and utility. As mentioned earlier, instruments addressing quality of life more broadly usually have very few items addressing health, which is obviously limiting when evaluating medical/health interventions.

The next question that needs to be considered is whether to use a 'generic' measure of perceived health status/HRQoL, a disease- or condition-specific measure or a combination of both.

2.4   Types of Measures for Health Outcomes Evaluation

Measures can be disease-specific measures which focus on particular aspects of health (e.g. symptoms) and HRQoL that are relevant to a particular disease. Examples of the latter instruments include the Arthritis Impact Measurement Scale (Meenan et al., 1980), Living with Asthma Questionnaire (Hyland, 1991), the Functional Living Index for Cancer (Schipper et al., 1984), and may include symptom checklists such as the Rotterdam Symptom Checklist (de Haes et al., 1990). Condition-specific measures may include instruments that apply to a broad class of conditions (e.g. mental health conditions) or a population segment (e.g. the elderly, people with a number of chronic conditions). HRQoL and health status measures (when used in an outcome evaluation application) are usually classified as 'generic' instruments, namely a common measure that can be used for comparisons across conditions, such as the multi-dimensional profiles (e.g. SF-36, Nottingham Health Profile, Sickness Impact Profile) and indices (e.g. Quality of Well-Being Scale, Australian Quality of Life Scale, EQ-5D-5L, SF-6D).

The limitation of disease-specific measures is that they can only be completed by those with the disease concerned, whereas the advantage of generic PROMs is that they allow comparison across conditions concerning the impact or burden of the disease or health condition. Using both types together can be complementary, as they provide different insights, and this is a common practice.

McDowell (2006) provides useful reviews of many of these instruments. Summaries of many measures can also be found in the Patient Reported Outcomes Quality of Life Instrument Database (PROQOLID[TM]) operated by the Mapi Research Trust (Emery et al., 2005).

These types of measures are outlined in the following sections below.

2.5   Disease-Specific Measures

Measures can be disease-specific – for example those developed for diabetes or asthma or a type of cancer. These types of measures are useful for more detailed measurement of symptoms and the impact of symptoms for particular health conditions, but have the obvious limitation that comparisons cannot be made across health conditions. Disease-specific measures can include clinical indicators and physiological measures such as blood pressure or serum cholesterol, but they can also include patient and clinician reported symptom checklists (e.g. the Functional Living Index for Cancer; Schipper et al., 1984) and 'Type' indicators (Radosevich and Husnik, 1995). Patient-reported disease specific measures are increasingly being used in clinical registries internationally and in Australia.

However, some of these instruments exemplify the recent trend for disease-specific measures to try to cover as many dimensions of health as general health profiles, as well as including condition specific items or symptom checklists. Clearly there are ranges of disease specific clinical outcome indicators and symptoms that need to be assessed along with aspects of perceived health status, but to incorporate these elements in the one instrument may lead to other psychometric inadequacies (e.g. inadequate item representation on generic domains) and prohibit comparisons across conditions.

Increasingly such 'blended' instruments may be displaced by modular packages combining a general health profile with a complimentary disease- /condition-specific element in an assessment battery which may also contain sets of condition- or disease-specific clinical indicators, and a core set of other relevant indicators (demographics, co-morbidities etc.).

2.6   Condition-Specific Measures

Condition-specific measures might include a range of functional status or disability measures used to assess the health of a particular population group such as the elderly or those with mental health problems. Some brief mental health measures such as the self-reported Kessler-10 Psychological Distress Scale have also been used in population mental health surveys (Kessler, 1997; Kessler et al., 2002; ABS, 1997) and in clinical monitoring in this sector. Condition-specific measures do not focus on a particular disease but on a broader health condition or state or they could apply to a population segment (e.g. frail elderly). For example, an Australian project (Thomas et al., 2006) was undertaken to develop an outcomes measurement suite for incontinence conditions and some components were field tested by Sansoni et al. (2013a).

In Australia a number of service sector registries include condition-specific measures in their outcomes benchmarking activities. The Australian Mental Health Outcomes and Classification Network (AMHOCN) monitors and benchmarks the outcomes of patients in public mental health services in Australia. The Australasian Rehabilitation Outcomes Centre benchmarks data

across providers using condition-specific measures as does the Palliative Care Outcomes Collaboration. Some of these examples are discussed briefly below.

2.6.1 Mental Health

As part of the AMHOCN initiatives (National Mental Health Information Development Expert Advisory Panel, 2013) public mental health services across Australia collect a range of outcome-related indicators (Mental Health National Outcomes and Casemix Collection, MH NOCC) which include health outcome-related performance indicators (e.g. readmission rates, length of stay), a condition-specific clinical rating (Health of the Nation Outcome Scales (HoNOS); Wing et al., 1998) and a consumer-reported outcome measure which is selected by each state or territory.

Consumer or PRO measures used included the Behaviour and Symptom Identification Scale (BASIS-32; McClean Hospital) and the Kessler - 10 Psychological Distress Inventory (K-10; Kessler and Mroczek, 1994) and the Mental Health Inventory (MHI; Viet and Ware, 1983). These measures could be classed as condition-specific measures for mental health as they may be used for a variety of mental health conditions; however, they are not specific to any particular psychiatric disorders such as schizophrenia, depression or anxiety disorder. As clinical ratings (e.g. HoNOS), patient self-report measures (e.g. K-10) and other clinical indicators are utilised it allows for a triangulation of perspectives on patient outcomes.

This data collection (MH NOCC) with the associated casemix adjustment is used for the benchmarking of casemix adjusted treatment outcomes across public mental health services in Australia. This enables services not only to routinely monitor the outcomes of their patients but also allows them to compare the patient outcomes from their service with similar services and thus can provide very useful feedback to assist quality improvement. However, to compare services it is necessary to standardise outcome results to take into account the mix of patients/consumers – for example some services may have patients with more severe problems than other services. Thus services are compared in relation to 'Casemix Adjusted Relative Mean Improvement Scores'.

A report on the benchmarking demonstration project (National Mental Health Strategy, 2009) indicated for this pilot only health outcome-related indicators, costs data and an 'outcome readiness' indicator were utilised and they flagged that in future benchmarking activities PRO data needed to be included in a comprehensive performance management framework.

2.6.2 Functional Limitations and Activities of Daily Living

A number of instruments are used to assess activities of daily living (e.g. self-care tasks such as bathing, dressing, toileting, grooming, feeding etc.) and instrumental activities of daily living (e.g. shopping, transport and housekeeping). Many of these instruments are used to assess the maintenance/improvement/deterioration of such skills in the elderly or for those with chronic conditions or those undergoing rehabilitation and could be considered to be condition-specific instruments. These instruments include both clinical rating scales (such as the Functional Independence Measure; Hamilton et al., 1987) and PROMs (e.g. the disability dimension of the Health Assessment Questionnaire; Fries, 1980).

Some of the commonly used instruments are the Barthel Index (Mahoney and Barthel, 1995) and the more recent modifications of this; the Functional Independence Measure (FIM) (Hamilton et al., 1987; Stineman et al., 1997, UDSMR see www.udsmr.org), the OARS-IADL scale and the OARS Physical ADL scale (Fillenbaum and Smyer, 1981; Fillenbaum, 1988) and the Index of Independence in Activities of Daily Living (ADL) (Katz et al., 1959), although the latter instrument is somewhat dated. The disability dimension of the Health Assessment Questionnaire (Fries et al., 1980; Ramey et al., 1995) has been used in chronic disease management studies (Lorig et al., 1996; Stanford Sample Questionnaire, 2000). The Australian Community Care Needs Assessment battery, which contains the CHSD Functional Profile (Eagar et al., 2001; Green et al., 2006; Sansoni et al., 2013b) has been used in aged care assessment for Home and Community Care Settings in Australia and some of these datasets would have a potential application for outcomes monitoring and benchmarking (Sansoni et al., 2013b).

The Australasian Rehabilitation Outcomes Centre (2016) makes use of FIM ratings in their outcome benchmarking of rehabilitation services across Australia. Australian normative data is available for this instrument and its psychometric properties are good. However, this instrument may be better suited to inpatients or those with fairly severe conditions, as there are likely to be ceiling effects (scoring at the top of the scale) in outpatient populations such as tertiary rehabilitation. Many of these instruments are also used with the elderly or people with disabilities to assess need for services and for some instruments there are options for administration which include self-report/ interview, or as a rating scale.

It should be noted that the FIM is a clinician rated instrument and not a PROM. For groups that may experience difficulties using self-report measures (PROMs) such as the frail elderly, those with severe disabilities or mental health conditions, and those with cognitive impairment the outcome dataset may commonly include clinical rating scales. This raises the issue that clinical perspectives, as measured by standardised clinical rating scales (e.g. the FIM and HoNOS are commonly used in Australia), may also be useful to include in outcomes monitoring and benchmarking activities.

2.7 Generic Measures

Instruments designed to measure general health status are usually profiles which cover several dimensions of health which previously might have been measured using separate instruments (e.g. pain or mental health). They usually include items on physical functioning, role functioning, mental health perceptions and pain, but some instruments include additional domains such as sleep, social and sexual functioning. As they are not disease specific they can be used to compare the burden of illness across diseases and conditions.

Some generic profiles such as the multi-dimensional indices or multi-attribute utility measures are largely used for economic evaluation. Both types of generic measures are discussed below.

2.7.1 Multi-dimensional Health Profiles

Currently many researchers are using multi-dimensional health status measures/profiles, for example the SF-36 (Ware and Sherbourne, 1992), the Nottingham Health Profile (Walker and Rosser, 1993) and the Sickness Impact Profile (Bergner et al., 1981). These are used in conjunction with other condition- or disease-specific clinical measures, as generic instruments

may miss critical factors for individual conditions, interventions or patient groups. Some earlier comprehensive reviews of these instruments have been undertaken by Marosszeky in Thomas et al. (2006) and these are available at http://ahsri.uow.edu.au/ahoc. Given the widespread use of the SF-36 in Australia, Appendix 1 provides an update on the use of this measure in Australia.

Recently a new generic measure, the PROMIS[TM] (Patient-Reported Outcomes Measurement Information System) Global Health scale has been developed using both classical test theory and modern psychometric methods (e.g. item response theory, IRT). Approaches such as IRT can be used to refine measures and to make them suitable for computerised adaptive testing (Ware, 2003, Cella et al., 2012). In the latter situation patients/clients are only given the minimum number of items that are necessary, through statistical inference, to determine their final score and thus in these situations respondent burden is far less likely. The PROMIS[TM] measures which include both dynamic (versions for use with computerised adaptive testing) and static versions (paper and pencil questionnaires) are a useful development. However, as the Global Health-10 instrument has only been recently released there will be less data available pertaining to the validation of this measure than for the more established generic measures.

There are also a number of generic PROMs known as individual PROMs (iPROMs) or patient generated outcome measures. These include, for example, the Schedule for the Evaluation for Individual Quality of Life (SEIQoL; O'Boyle et al., 1993; O'Boyle 2005) and the Patient Generated Index (Ruta et al., 1994). Unlike the standardised measures mentioned above where all patients rate the same domain of health, these instruments allow each respondent to individually define or select the most relevant domains and assign their weighting. As the domains will vary across individuals that are not considered to be as suited to comparative effectiveness research but their use for individual patient monitoring within the clinical consultation may be a more relevant application, although further research and evidence is required to support the use of iPROMs (Browne, accessed 2016).

2.7.2   Multi-dimensional Indices

The health outcomes approach also has a major economic focus, although we are some way as yet from resource allocation based on health outcomes evaluation. While it is important to know which interventions lead to an improvement in health status, it is also useful to know the relative costs and benefits in comparison to alternative treatments for the same condition. Such information is extremely useful in setting priorities and directions for health expenditure. This has led to the development of multi-dimensional indices (also known as multi-attribute utility measures). Multi-dimensional profiles need to be distinguished from multi-dimensional indices, as scores on the latter are aggregated to form a single index or number, whereas this is not the case for multi-dimensional profiles. Some commonly used measures are the Australian Quality of Life Scale (Hawthorne et al., 1999, Hawthorne et al., 2000a, Hawthorne et al., 2000b) the EQ-5D (formerly the Euroqol) (EuroQol Group, 1990; Kind, 1996), Health Utility Index 3 (HUI3) (Feeny et al., 1996a; Feeny et al., 1996b, Torrance et al., 1995), 15D (Sintonen and Pekurinen, 1993; Sintonen, 1995, 1994, 2001), Quality of Well-Being (QWB) (Kaplan, 1993; Kaplan et al., 1996), Rosser Index (Rosser, 1993), and the SF6D (Brazier et al., 2002; Brazier et al., 1998).

If data are to be used to estimate relative costs and benefits of alternative treatments (as in comparative effectiveness research), a number of these multi-attribute measures may be used. There has been much research on multi-dimensional indices which may be used to generate QALYs (Quality-Adjusted Life Years) and DALYs (Disability-Adjusted Life Years) which, together with information on costs, enable cost-utility analyses and thus cost and benefit comparisons across conditions (see box below). Given that health resources may need to be rationed, such approaches compare the health gains to be made and their community valuation in relation to costs across health interventions, and claim to generate more rational models for health service resource allocation than is true of current resource allocation. QALY units 'integrate side effects and benefits of treatment by combining, into a single number, mortality, morbidity, and duration of each health state' (Kaplan, 1993). For example Harvey (1991), commenting on the costs associated with breast cancer screening, noted that if women were screened annually rather than biennially that an additional 14% of life years could be saved for a 70% increase in expenditure.

Health State Valuations and QALYs

- A generic HRQoL measure that describes a range of health states between life and death is needed. For example, the EQ-5D has five dimensions or questions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) and three levels of response possible (1 = no problems, 2 = some problems, 3 = extreme problems. If a person scored 1 (no problems) on all five dimensions the health state would be classified as 1,1,1,1,1 whereas if a person experienced major problems on all dimensions the health state would be classified as 3,3,3,3,3. Thus the EQ-5D has 243 health states that can be so described.

- Each of these health states can be valued on a scale between 0 (death) and 1 (life) – usually using preference methods such as standard gamble or time trade off. For time trade off a person is asked how many years of life they are prepared to give up for a treatment that will return them to full health from this health state. These preference techniques provide the utility value for each health state.

- Improvement gained by a treatment can be classified on the same metric, for example a treatment moves incontinence patients from a health state valued at 0.6 to 0.7 (this is their improvement in HRQoL).

- The quality adjusted life year (QALY) is a measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life. One QALY is equal to one year of life in perfect health.

- The improvement gained needs to be adjusted for the period of survival/ life expectancy – let's assume 10 years for the example of incontinence.

- Thus the treatment has gained the patients one QALY (10 years survival*.10 improvement gained = 1 QALY) for their incontinence treatment – this is a measure of health gain.

- Costs data are added to this – let's say $10K per QALY for the specified incontinence and this can be compared with the cost per QALY for a different treatment for incontinence or a treatment for asthma or diabetes.

- QALYs may assist in choosing between various treatments for a condition or to compare the costs and benefits of treatments across conditions and so may advise resource allocation decisions.

The base scales for health status assessment used in such multi-dimensional indices usually contain about five to twelve questions with each question representing one domain (e.g. physical mobility, role functioning, pain, mental health), with the aim of such indices being to compress this information into a single number which can be related to costs data. While economists may see this as desirable, psychologists and others may wish to question how representative each of these questions can be of the domain they supposedly represent. Porter (2010) notes these measures (both QALYs and DALYs) collapse quality of life into a single number despite the fact that it is inherently multi-dimensional and relevant dimensions vary by medical condition. Thus the issue of the extent to which items are representative of the health domains chosen is even more pertinent to the multi-dimensional indices than for the multi-dimensional profiles. Cadet (1994) also noted there are many other concerns with regard to QALYs (especially surrounding methods for valuing health states and the use of QALYs for resource allocation).

Reports by Hawthorne in Thomas et al. (2006) and Sansoni et al. (2007) compared a number of these health utility indexes with reference to a range of evaluation criteria for the purposes of selecting a utility index for studies of a) incontinence and b) dementia. These instruments were further assessed in relation to incontinence status in a special edition of the SA Health Omnibus Survey (Harrison Health Research, 2004) and the recommendation was to use either HUI 3 or the AQoL for cost utility studies in the field of incontinence (Hawthorne and Sansoni, 2004). With regard to dementia the AQoL and the EQ-5D were preferred (Sansoni et al., 2007) although there were caveats raised concerning their use. It was noted that the EQ-5D had issues concerning competing scoring algorithms, ceiling effects, inconsistent utility scores and poor score distribution. These earlier studies also noted that the apparent sensitivity of these instruments varied significantly with the measurement method and by disease area and this has also been recently reported by Richardson et al. (2016). Notwithstanding the above Richardson et al. (2016) found that the 15D, AQoL-8D and the SF-6D generally achieved better tests results on their rating and review criteria than the QWB and the EQ-5D-5L. It should be noted the EQ-5D has been widely used by the NHS in the UK as a measure of health gain and by some Swedish registries (NHS, 2016; Lundstrom and Karlskrona, 2015). In more recent times in Sweden (from 2013) it appears the EQ-5D-5L is now preferred as it is less prone to ceiling effects and there are now Time Trade Off values for the health states derived from the EQ-5D-5L (Garellick et al., 2014). In the UK values for the EQ-5D-5L have recently been developed by the Office of Health Economics (Devlin et al., 2016; Feng et al., 2016)

There has been interest in developing and refining models that make use of QALY algorithms. One such modelling exercise was undertaken by Professor Hindle and associates in the Illawarra, New South Wales (Cromwell et al., 1995). Such work raises interesting issues as to the prioritisation of caseloads for hospital systems. Is it more cost effective to spend your resources on six tattoo removals or on one coronary artery bypass graft, and what are the relative health benefits to be gained by the region?

There has been much interest in the notion of DALYs which could also form a base for more rational priority setting and funding at the national or system level than has been the case for historical approaches (Murray and Lopez, 1996). Some research has been undertaken in Australia to examine this burden of disease approach in the Australian context (Mathers, 1999; AIHW, 2014).

2.8 Outcome Measurement Suites

A more recent approach has been to develop outcome measurement suites for a range of conditions (e.g. chronic disease management, dementia, incontinence conditions, mental health, assessment and monitoring of the elderly and asthma) or for particular situations (assessment and monitoring in primary and community care). Outcome measurement suites are a collection of measures and other information items that are seen as relevant for the outcomes monitoring of these conditions. They will usually contain patient information items, medical history, medicinal use, service use, clinical indicators and generic and disease/condition specific measures.

The Stanford Sample Questionnaire (2000) is an example of an outcomes measurement suite for chronic disease self-management studies and the AMHOCN data collection provides a suite of measures for mental health. An initial needs assessment battery for primary care has been collated (CHSD, 2001; Sansoni et al., 2013b) which has the potential to be used as a health outcomes monitoring suite for this area. Outcome measurement suites for incontinence conditions (Thomas et al., 2006) and dementia and associated conditions have been developed in Australia (Sansoni et al., 2007). A similar approach was undertaken by the Health Outcomes Institute in the US (Radosevich and Husnik, 1995) in the development of their TYPE modules. These modules contain sets of condition or disease specific clinical indicators, a generic HRQoL instrument and a core set of other relevant indicators (demographics, co-morbidities etc.). More recently the International Consortium for Health Outcomes Measurement has developed over twenty standardised sets for health outcomes measurement across a range of health conditions and population segments with a view to promoting international benchmarking on the outcomes of care and treatment. The Agency for Clinical Innovation in Australia has recently become a strategic partner with ICHOM in these developments.

It is important to consider the issue of redundancy in the development of outcome measurement suites, as many of the PROM scales may have overlapping items. It is desirable to avoid respondent burden for the clients in such studies, as otherwise one is likely to find both sample attrition and a preponderance of missing data. Approaches such as Item Response Theory can be used to refine measures and to make them suitable for computerised adaptive testing (Ware, 2003) which can reduce respondent burden. These approaches hold great promise for the refining of many commonly used standardised measures in the future. Recently, the Patient-Reported Outcome Measurement Information System (PROMIS[TM]) has developed a considerable number of PROs in physical, mental, and social health for adults and paediatric samples with chronic conditions (Cella et al., 2012). The PROMIS measures which include both dynamic (versions for use with computerised adaptive testing) and static versions (paper and pencil questionnaires) appear to be a useful development. The Agency for Clinical Innovation (NSW) is using the PROMIS 10 in some demonstration projects concerning the use of PROMs currently.

Often a tiered approach, a computerised adaptive testing approach or a decision tree model may be useful in the development of outcomes measurement suites.

2.9 Criteria for Selecting Instruments and Measures for Outcomes Evaluation

It is important to use valid, reliable and appropriate instruments when using PROMs. The dimensions below are those one must consider in selecting instruments and measures for outcome evaluation.

RELIABILITY: consistency of measurement, e.g. internal consistency and test/retest reliability.

VALIDITY: does the instrument measure what it claims to measure? There are different types of validity – content, construct, criterion, concurrent, convergent, discriminant etc.

    DISCRIMINATORY POWER /discriminant validity: is the instrument able to discriminate well between groups, for example, healthy public versus terminally ill?

RESPONSIVENESS/SENSITIVITY to CHANGE: can the instrument detect change in health status over time?

AVAILABILITY OF COMPARATIVE DATA: are there norms and clinical reference datasets available for comparison purposes?

TYPE OF INSTRUMENT: generic health status measure, condition or disease specific measure, profile or index.

STYLE OF INSTRUMENT: for example, is it better to use a self-report instrument or a rating scale or a combination of both? Is a self-report inventory the best instrument to use with severely disturbed patients?

PRACTICAL UTILITY: is the instrument too long/short, is it easy to administer and use, is it easy to score, will there be respondent burden, etc.?

FREEDOM FROM CONFOUNDING FACTORS: for example, social desirability of responses, inappropriate questions associated with missing data, literacy level of the survey etc.

RELEVANCE and SUITABILITY OF APPLICATION: for example, whether the generic and/or disease specific measures adequately capture the relevant domains for the condition or disease concerned.

MODE OF ADMINISTRATION: self-reported or structured interview, telephone administration etc.

CULTURE, GENDER and AGE APPROPRIATENESS: are there translations/adaptations for other cultural groups, are all the items suitable for both genders, and are there versions suitable for use with children/adolescent? Some instruments need linguistic validation for use in the Australian context.

An examination of the above criteria indicates that some expertise in psychological measurement may be required in selecting measures, in administering measures and in interpreting the data that is derived from such measures. Those that are involved in routine data collection need adequate training and briefing concerning the purpose and the proposed methods of the PROMs data collection.

Lavallee et al. (2016) noted that there are some challenges for PROMs data collection which include logistical concerns such as workflow barriers including the increased burden on staff and patients to collect PROMs data. There are ways to address these challenges such as the use of tablets and online data collection methods (Lavallee et al., 2016; Schuler and Miller, 2014; www.discoverquick.com). Electronic data capture systems can also provide 'real time' feedback concerning PROMs to both the provider and the patient.

The use of item response theory and computer adaptive testing (Ware, 2003, Cella et al., 2012) are used to ensure that patient needs to only answer the minimum number of questions to assess their health status (computer assisted 'dynamic measures'). Such methods show promise

but are not without cost, and again require training of the staff involved. Feedback to patients must also be provided in a way they can understand. It must also be remembered that using online and electronic technologies with all patients may not be possible as lack of access to technology may also be a barrier for some patients and for some patients paper and pencil forms will still be required (Lavallee et al., 2016).

In Australia insuring that PROMs are appropriate for different cultural groups such as Aboriginal and Torres Strait Islander peoples and those from culturally and linguistically diverse (CALD) populations is an issue that needs to be addressed (see Section 2.1.12).

2.10 Some Selected Applications of the Health Outcomes Approach

The health outcomes approach has already gained some acceptance in the acute care sector. However, there are a number of issues that need to be addressed when translating this approach to other sectors, such as primary and community care and allied health.

2.10.1 Applications to Primary and Community Care

Professor Allen Hutchinson (1998), with reference to outcomes monitoring in general practice, strongly recommended the KISS approach, or 'keep it simple stupid!' His view is that it is unrealistic to expect that busy general practitioners would have the time to routinely monitor the outcomes of their patients using instruments that assess HRQoL. Ware (1996, 2003), however, has indicated that a number of GPs in the USA do in fact use the SF-36 to monitor individual patients. Earlier, Sansoni (1995) and Dixon et al. (1994) suggested that the psychometric properties of the instrument would not recommend its use for individual patient monitoring although increasingly it does seem a fairly robust and responsive instrument for both group comparisons and for monitoring patient groups pre and post intervention.

A more realistic approach in general practice may be to use such instruments with particular groups of patients with similar needs or where a specific intervention is being assessed, rather than as a matter of routine. Sometimes some of the preferred instruments may be too long for use in routine care but now approaches such as Item Response Theory and computerised adaptive testing (Ware, 2003; Cella et al., 2012) can be used to shorten measures (see Section 2.7.1).

The use of health outcomes performance indicators as proxies for the outcomes of patients and of patient management may be a more practical approach to adopt initially in primary care settings. Indicators might include such items as the proportion of patients that have been appropriately immunised, the proportion of female patients that have been appropriately screened for cancer of the cervix, the number of hospital admissions for patients with asthma, the presence and routine monitoring of asthma management plans, and the appropriateness of prescriptions with respect to best practice guidelines. PROMs can be used in primary and community care settings but to date in Australia their use appears to be largely limited to demonstration projects.

A similar approach could also be undertaken in community care and allied health settings. Health outcome-related performance indicators pertaining to the particular service can be more informative concerning patient outcomes than the measures of throughput and output

(such as the number of occasions of service performed) that are now used as performance indicators by health departments.

Output measures may reflect efficiency and the volume of service, but do not give any indication of the quality or appropriateness of the services provided. One way to address this issue is to examine the desired outcomes of the particular strategy or health intervention and then pose the question as to what readily available information might reflect whether the intervention is obtaining the desired outcome. For example, a tertiary rehabilitation service could examine indicators concerning the proportion of clients returning to paid and unpaid work or study. Nancarrow (1999) and Rubenach (1999) provide some examples of the application of this approach in a range of community care and allied health settings and an application to allied health is outlined in Appendix 2

2.11 Some Selected Issues and the Health Outcomes Approach

2.11.1 Cultural Applicability of Instruments

In 1996 AHOC was asked to convene an expert working group to recommend measurement instruments for both the Aboriginal and non-Aboriginal components of the Co-ordinated Care Trials.

The SF-36 was recommended for the non-Aboriginal component of the Co-ordinated Care Trials. This was because it was assessed as being the most valid, reliable and responsive of the generic measures of health outcome of those available (Shadbolt, 1996a). As there are a number of standardised versions for other languages it was also considered to be appropriate for use with people from a non-English speaking background.

However, the expert group workshop considered that the SF-36, as it is, would not be suitable for Aboriginal and Torres Strait Islander populations, especially those in remote areas. An example of the cultural inappropriateness of the questions is illustrated by question 3a, which asks if your health limits you in activities such as climbing a flight of stairs. Obviously this question is inappropriate in communities where buildings do not have stairs. Other inappropriate questions mention activities such as vacuum cleaning, playing golf and bowls which are also unlikely to be routine activities in remote rural communities. Thus it is most important to find instruments that are appropriate to the group being assessed.

An individual's perception of his/her own health status may also be influenced by the general health status of the community in which they live. If the community as a whole has a poor health status, then a person may see themselves as being healthier than their companions and rate themselves accordingly, when their actual health status, as seen by an outsider is low. This factor was observed in the National Aboriginal and Torres Strait Islander Survey of 1994 (Australian Bureau of Statistics, 1994). The NATSIS Survey indicated that a similar proportion of Aboriginal peoples reported their health as being good to excellent as for a sample from the general Australian population, yet if objective criteria are used (incidence and death rates associated with various diseases) it is clear that the health of the indigenous population is not as good as that for the Australian population overall. This would tend to suggest that cultural relativities may be taken into account when people complete global ratings on 'perceived' health status items. Whilst perceived health status measures are often good proxies for direct

measures of 'health' and are an important component of the construct of health, 'perceived' and 'actual' health status may be not quite the same thing.

One could possibly modify the items of the SF-36 to make them more comprehensible (Henderson and Gray, 1994) or substitute items such as going hunting and fishing as examples of moderate physical activities, change the time frame, etcetera as has been suggested. Making such modifications, however, would mean that the reliability and validity of this modified instrument would need to be re-assessed. There is also the problem that the construct of health implicit in this scale is based on the perceived dimensions of health that are relevant to middle class Americans. Studies by Senior (1999) and Scott et al. (2000) raise serious questions about such assumptions. Tinkering at the edges will not solve the issue of cultural inappropriateness and insensitivity.

Senior (1999) indicated that the construct of 'health' may have a different meaning and different dimensions than is the case for middle class non-Aboriginal people on whose health construct such standardised scales have been developed. For example, Aboriginal and Torres Strait Islander communities might consider such factors as environmental sustainability, spirituality and kinship, and a sense of self-esteem and self-control as major components of their health construct (Senior, 1999). Senior's interviews about domains of quality of life in Lulaluk and Minmarama Park strongly indicated there was a need to develop an entirely new measure of quality of life for the Aboriginal people in this community. Senior (1999) found that even the WHOQOL-BREF (World Health Organization, 1996), an instrument designed to be applicable across different cultural groups, omitted domains relevant to quality of life for these groups, and various items were also subject to misinterpretation by the community participants.

In New Zealand there have been some initiatives (Kingi and Durie; 1998; 2002) to develop some culturally-specific PROMs for Māori and similar initiatives are reported on the Te Pou website (Te Pou, 2016). There have been initiatives in Australia to develop culturally specific cognitive assessment tools (e.g. standardised clinical assessments) such as the Kimberley Indigenous Cognitive Assessment (LoGuidice et al., 2006; Jackson Pulver et al., 2012) but the development of culturally-specific PROMs for Indigenous peoples particularly in remote locations needs further exploration.

The Mapi Research Trust has a particular interest in the cross cultural translation and validation of measures and their PROQOLID™ database provides information about the translations available for many PROMs.

Although this brief discussion has focused on issues relevant to Indigenous peoples there are many other pertinent examples. Consider the assessment of those with dementia conditions from culturally and linguistically diverse (CALD) backgrounds. It is suspected that later diagnosis and misdiagnosis is more likely to occur due to communication difficulties, cultural misunderstandings, culturally inappropriate assessment tools and the lack of available interpreters (Black et al., 2001). It is also noted that second language skills also tend to be lost early in Dementia. Thus cultural appropriateness will be an important consideration for a broad range of groups.

2.11.2 Other Issues Concerning the Applicability of Measures

When selecting measures for health outcomes evaluation some other applicability issues are relevant and one needs to consider whether the instrument is fit for the purpose and whether it will be appropriate to the sample identified. For example, this may relate to the age appropriateness of the instrument (e.g. is the instrument appropriate for use with such groups as children or the elderly and has the instrument been designed for use with such groups). Patients with disabilities such as limited mobility or visual impairment may also find it difficult to complete PROMs and care should be taken in designing systems and measures to ensure they can participate fully (Lavallee et al., 2016).

Similar issues relate to whether the language used in the instrument can be comprehended by those with a low level of literacy or those who have received limited education. This factor can also be relevant when proposing to use electronic PROMs completed on electronic devices – there may be cultural differences that affect access to and comfort with the electronic tablet technology (Schamber et al., 2013). For example, it was found that white Caucasian participants were more likely to complete an electronic tablet survey than other cultural groups (Altreja and Rizk, 2012). It must be remembered that using online and electronic technologies with all patients may not be possible as lack of access to technology may be a barrier for some patients and for some patients paper-and-pencil forms will still be required (Lavallee et al., 2016).

The gender appropriateness of items is also an area that has sometimes been ignored by those developing standardised instruments used to assess HRQoL (Sansoni, 1996). It is also important in developing health outcome data collections to ensure that the data can be disaggregated by factors such as age, gender and educational level.

3   Conclusions

In conclusion, it would seem there are some promising approaches to developing more integrated and coordinated approaches to health care and to the routine assessment of patient based health outcomes. It is to be hoped that the re-emerging health outcomes focus in Australia may act as a catalyst for us to integrate the various efforts that are already being made to improve the quality of our health systems and hospitals. It is only by evaluating our services that we can become both more effective and more efficient in the delivery of health care. It is also essential that health outcomes evaluation should become integrated within our quality improvement systems – there is not much point in evaluating patient outcomes unless we use this information for service improvement and to improve patient outcomes.

3.1   The Australian Health Outcomes Collaboration

The Australian Health Outcomes Clearing House was established in 1994, and served as a site for the dissemination of information about health outcomes research, practice and policy in Australia and overseas. In 1997 the Clearing House became the Australian Health Outcomes Collaboration (AHOC) at the Australian Health Services Research Institute.

The primary aims of the AHOC are to:

- Disseminate information about health outcomes research

- Maintain an active network of collaborators in health outcomes research

- Provide advice on the selection of measures for health outcomes assessment

- Provide health outcomes education and training (e.g. workshops and a Master of Public Health unit)

- Organise national and international conferences and seminars

- Facilitate health outcomes research throughout Australia.

The AHOC staff and associates also participate in a range of health outcomes related research projects which include the development, cultural and linguistic adaptation and revision of PROMs and the review and development of health outcomes related performance indicators across a wide range of areas (for example, Sansoni, 2016, Sansoni et al., 2007; 2010, 2011, 2013a, 2013b, 2015, 2016). A forthcoming report on the use of PROMs in Australia will be available shortly (Williams et al., 2016) Visit the website at www.ahsri.uow.edu.au/ahoc for further information about the AHOC and health outcomes workshop activities. The AHOC can also provide advice and assistance concerning the implementation of the health outcomes approach and the measurement tools that might be used.

4 References

Abraham B, d'Espaignet E and Stevenson C (1995) Australian Health Trends 1995. Australian Institute of Health and Welfare.

Altreja A and Rizk M (2012) Capturing patient reported outcomes and quality of life in routine clinical practice: ready for prime time? Minerva Gastroenterologica e Dietologica. Marc, Vol. 58, No.1, pp.19-24.

Allcock C (2015) Outcomes-based commissioning – much promise, but is it something that CCGs can actually deliver on? The Health Foundation, 24 September 2015. Accessed at http://www.health.org.uk/blog/outcomes-based-commissioning-%E2%80%93-much-promise-it-something-ccgs-can-actually-deliver.

Armstrong B (1994) Getting health outcomes into state and national health policy, a national perspective. NSW Health Outcomes Conference, Sydney, 12-13 August 1994.

Australasian Rehabilitation Outcomes Centre (2016) AROC Reports. Available from: http://ahsri.uow.edu.au/aroc/reports/index.html accessed 23 August 2016.

Australian Bureau of Statistics (1994) National Aboriginal and Torres Strait Islander Survey: Details Findings. ABS Catalogue No. 419.0. ABS, Canberra.

Australian Bureau of Statistics (1997) 1995 National Health Survey: SF-36 population norms, Australia. Cat No 4399.0. ABS, Canberra.

Australian Commission on Safety and Quality in Health Care and Australian Institute of Health and Welfare (2010) Australian Safety and Quality Framework for Health Care. ACSQHC, Sydney.

Australian Commission on Safety and Quality in Health Care (2012) National Safety and Quality Health Service Standards (September 2012). ACSQHC, Sydney.

Australian Commission on Safety and Quality in Health Care and Australian Institute of Health and Welfare (2014) Exploring Healthcare Variation in Australia: Analyses Resulting from an OECD Study. Sydney: ACSQHC.

Australian Commission on Safety and Quality (2015a) Establishing National Priorities for Clinical Practice Guidelines 2015. Discussion paper prepared for the Australian Commission on Safety and Quality in Health Care.

Australian Commission on Safety and Quality in Health Care and National Health Performance Authority (2015b) Australian Atlas of Healthcare Variation. ACSQHC, Sydney.

Australian Health Ministers' Advisory Council (1993) As quoted in: Health Outcomes Bulletin. No. 1, February 1994, 5.

Australian Health Ministers' Advisory Council (1996) The Final Report of the Taskforce on Quality in Australian Health Care. AGPS, Canberra.

Australian Institute of Health and Welfare (1994) Health expenditure highlights. Health Expenditure Bulletin, No. 10, p.1.

Australian Institute of Health and Welfare (2009) Towards national indicators of safety and quality in health care. Cat. no. HSE 75. Canberra: AIHW.

Australian Institute of Health and Welfare (2014) Health expenditure Australia 2012-13. Health and Welfare Expenditure Series no. 52. Cat. no. HWE 61. Canberra: AIHW.

Batalden PB and Davidoff F (2007) What is "quality improvement" and how can it transform healthcare? Quality and Safety in Health Care. Vol. 16, No. 1, pp.2–3.

Batalden P, Nelson E and Roberts J (1994) Linking outcomes measurement to continual improvement: The serial "v" way of thinking about improving clinical care. Journal of Quality Improvement. Vol. 20, No. 4, pp.167-180.

Becker M, Diamond R and Sianfort F (1993) A new patient focused index for measuring quality of life in persons with severe and persistent mental illness. Quality of Life Research. Vol. 2, pp.239-251.

Benson DS (1992) Measuring Outcomes in Ambulatory Care. American Hospital Publishing, Inc.

Bergner M, Bobbitt RA, Carter WB and Gilson BS (1981) The Sickness Impact Profile: Development and final revision of a health status measure. Medical Care. Vol. 19, pp.787-805.

Black K, Lo Giudice D, Ames D, Barber B and Smith R (2001) Diagnosing Dementia. Alzheimer's Association of Australia.

Black N (2013) Patient reported outcome measures could help transform healthcare. British Medical Journal. Vol. 346, pp. f167.

Bowling A (2005) Measuring Health: A Review of Quality of Life Measurement Scales (3rd Ed.). Milton Keynes, Open University Press.

Boyce MB and Browne JP (2013) Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. Quality of Life Research. Vol. 22, No. 9, pp.2265-78.

Boyce MB and Browne JP (2015) Clinician Responses to Using Patient-Reported Outcome Measures for Quality Improvement. National Quality Measures Clearinghouse, Agency for Healthcare Research and Quality, 19 January 2015.

Boyce MB, Browne JP and Greenhalgh J (2014) Surgeon's experiences of receiving peer benchmarked feedback using patient-reported outcome measures: a qualitative study. Implementation Science. Vol. 9, pp.84

Brazier J, Roberts J and Deverill M (2002) The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. Vol. 21, pp.271-292.

Brazier J, Usherwood T, Harper R and Thomas K (1998) Deriving a preference-based single index from the UK SF-36 Health Survey. Journal of Clinical Epidemiology. Vol. 51, No.11, pp.1115-1128.

Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, Newhouse JP, Weiler PC and Hiatt HH (1991) Incidence of adverse events and negligence in hospitalised patients: results of the Harvard Medical Practice Study. New England Journal of Medicine. Vol. 324, No. 6, pp.370-376.

Brooks A, McParlane A and Tuess S (2015) Introducing the NSW arm of the Prostate Cancer Outcomes Registry Australia and New Zealand (PCOR-ANZ). Urological Society of Australia and New Zealand NSW Section ASM.

Browne J (2016) Individualised Patient Outcomes. Available from: http://www.kingsfund.org.uk/sites/files/kf/media/Professor%20John%20Browne%20-%20Individualised%20Patient%20Outcomes.pdf accessed 23 August 2016.

Bunker JP (1988) Is efficacy the gold standard for quality assessment? Inquiry. Vol. 25, No. 1, pp.51-8.

Bywood PT, Jackson-Bowers E, Muecke S (2011) Initiatives to integrate primary and acute health care, including ambulatory care services. PHCRIS Policy Issue Review. Adelaide: Primary Health Care Research and Information Service.

Cadet B (1994) History of the construction of a health indicator integrating social preference: The Quality of Well-Being Scale. Seventh Meeting of the International Network on Health Expectancy (REVES). Canberra, 23-25 February 1994.

Canadian Institute for Health Information (CIHI) (2000) Canadian Health Information Roadmap Initiative Indicators Framework 2000. Canadian Institute for Health Information and Statistics, Canada.

Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ and Rothrock N (2012) Methodological Issues In The Selection, Administration And Use Of Patient-Reported Outcomes In Performance Measurement In Health Care Settings. Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago.

Centres for Medicare and Medicaid Services (2015) CMS Bundled Payments for Care Improvement Initiative Fact Sheet. Available from: https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-08-13-2.html accessed 23 August 2016.

Centre for Health Service Development (CHSD) (2001) Pilot Testing of Tools: Guidelines for Undertaking Initial Needs Identification and developing a Service Coordination Plan. Centre for Health Service Development, University of Wollongong and Australian Institute for Primary Care, La Trobe University. Available from: http://aipc.latrobe.edu.au/pcpssc/.

Chen JJ (2015) Integrated Care: Patient reported outcome measures and patient reported experience measures - a rapid scoping review. NSW Agency for Clinical Innovation: Sydney.

Craft P, Primrose J, Lindner J and McManus P (1996) Patterns of surgical treatment of primary breast cancer in Australian women in 1993: Analysis using medicare statistics. 23rd Annual Scientific meeting of the Clinical Oncological Society of Australia inc. Brisbane Convention and Exhibition Centre, 27-29 November 1996.

Cromwell D, Haball J, Viney R and Hindle D (1995) Illawarregon, Development of a Model to Assist Priority Setting by an Area Health Service, Centre for Health Service Development, Wollongong.

Cummins RA (1993) Comprehensive Quality of Life Scale for Adults (4th Ed.). Deakin University, Deakin.

De Haes JC, Van Knippenenberg FC, Neijt JP (1990) Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. British Journal of Cancer. Vol. 62, pp.1034-8.

Department of Health and Ageing (2007) The National Evaluation of the Second Round of Coordinated Care Trials Final Report -Australian Coordinated Care Trials 2007: Part 1 – Executive Summary - Coordination of Care and Efficiency of Healthcare: Lessons from the Second Round of Australian Coordinated Care Trials ISBN: 1-74186-288-4 Online ISBN: 1-74186-289-2 Publications Number: P3 - 2052.

Devlin N, Shah K, Feng Y, Mulhern B and van Hout B (2016) Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England. Research Paper 16/01, Office of Health Economics, London.

Devlin NJ and Appleby J (2010) Getting the most out of PROMs: Putting health outcomes at the heart of NHS decision-making. The King's Fund, London.

Dixon P, Heaton J, Long A and Warburton A (1994) Reviewing and applying the SF-36. Outcomes Briefing. Vol. 4, pp.3-24.

Donabedian A (1980) Basic approaches to assessment: Structure, process, and outcome. In Donabedian A (1980) The Definition of Quality and Approaches to its Assessment: Explorations in Quality Assessment and Monitoring. Vol. 1. Health Administration Press.

Donabedian A (1992) The role of outcomes in quality assessment and assurance. Quality Review Bulletin. November, pp.356-60.

Eagar K, Owen A, Green J, Cromwell D, Poulos R, Gordon R, Quinsey K, Adamson L and Fildes D (2001) A National Measure of Functional Dependency for Home and Community Care Services in Australia: Stage 2 report of the HACC dependency data items project. Centre for Health Service Development, University of Wollongong.

Eagar K (2014) Measuring outcomes in pain management. Pain and Palliative Care Conference, Southbank Institute of Technology, Brisbane, 3-5 December 2014.

Emery M-P, Perrier L-L, Acquadro C (2005) Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID): Frequently asked questions. Health and Quality of Life Outcomes. Vol. 3, pp.12.

Entwistle V (1995) Paper delivered to Cochrane Collaborative Review Group on Communicating Effectively with Consumers, sponsored by the Public Health Division, Melbourne, 2-3 May 1995.

EuroQol Group (1990) EuroQol: a new facility for measurement of health-related quality of life. Health Policy. Vol. 16, pp.199-208.

Feng Y, Devlin N, Shah K, Mulhern B and van Hout B (2016) New Methods for Modelling EQ-5D-5L Value Sets: An Application to English Data. Research Paper 16/02, Office of Health Economics, London.

Feeny D, Furlong W and Torrance G (1996a) Health Utilities Index Mark 2 and Mark 3 (HUI2/3) 15-item questionnaire for self-administered, self-assessed usual health status, Centre for Health Economics and Policy Analysis. McMaster University, Hamilton.

Feeny D, Torrance G and Furlong W (1996b) Health Utilities Index In Spilker B (ed.). Quality of Life and Pharmacoeconomics in Clinical Trials. Lippincott-Raven Publishers, Philadelphia.

Field M and Lohr K (1992) Guidelines for Clinical Practice: From Development to Use. Institute of Medicine, National Academy Press, Washington DC.

Fillenbaum GG (1988) Multidimensional Functional Assessment of Older Adults: The Duke Older Americans Resources and Services Procedures. Lawrence Erlbaum Associates, Hillsdale, NJ.

Fillenbaum GG and Smyer MA (1981) The development, validity and reliability of the OARS Multidimensional Functional Assessment Questionnaire. Journal of Gerontology. Vol. 36, pp. 428-34.

Food and Drug Administration (2009) Guidance for Industry – Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. U.S. Department of Health and Human Services, Food and Drug Administration.

Fries JF, Spitz P, Kraines RG and Holman HW (1980) Measurement of patient outcome in arthritis. Arthritis and Rheumatology. Vol. 23, No. 2, pp.137-45.

Garellick G, Kärrholm J, Lindahl H, Malchau H, Rogmark C, Rolfson O (2015) The Swedish Hip Arthroplasty Register: Annual Report 2014.

Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEcIDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014.

Gonçalves Bradley D, Gibbons C, Ricci-Cabello I, et al (2015) Routine provision of information on patient-reported outcome measures to healthcare providers and patients in clinical practice (Protocol). Cochrane Database of Systematic Reviews. 2015; Issue 4.

Green J, Eagar K, Owen A, Gordon R and Quinsey K (2006) Towards a measure of function for Home and Community Care Services in Australia: Part 2 – Evaluation of the screening tool and assessment instruments. Australian Journal of Primary Health. Vol. 12, No.1, pp. 82-8.

Grimshaw JM and Russell IT (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. Lancet. Vol. 342, pp.1317–22.

Hamilton BB, Granger CV, Sherwin FS, Zielezny M and Tashman JS (1987) A uniform national data system for medical rehabilitation. In M. Fuhrer (Ed.), Rehabilitation outcomes: Analysis and measurement (pp. 137–147). Baltimore: Brookes.

Harrison Health Research (2004) South Australian Health Omnibus Survey (HOS) Report. South Australia Department of Health, Adelaide.

Harvey K (2003) "Saving" the Australian Pharmaceutical Benefits Scheme. Monash University, Health Economics Unit.

Harvey R (1991) Making it Better: Strategies for Improving the Effectiveness and Quality of Health services in Australia , Background Paper No. 8, National Health Strategy, Melbourne.

Harvey R (1994) A Pilot Study to Integrate Health Outcomes and Health Interventions, AIHW.

Harvey R (1995) Health Outcomes, quality assurance and information: 'Doing the do-able' or 'You've got to start from here!' part 1. Health Outcomes Bulletin. Spring, No. 6, p.22.

Hawthorne G (2006a) Multi-attribute utility measures of quality of life. In Thomas et al. (2006) Continence Outcomes Measurement Suite, DHA.

Hawthorne G (2006b) Measuring incontinence in Australia. Australian Government Department of Health and Ageing, Canberra.

Hawthorne G, Osborne RH, Taylor A and Sansoni J (2007) The SF36 Version 2: critical analyses of population weights, scoring algorithms and population norms. Quality of Life Research. Vol. 16, No.4, pp.661-73.

Hawthorne G, Richardson J and Day N (2000a) A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. XII Medical Symposium 'Quality of Life Measurement in Clinical Studies', Helsinki, Finland, pp.358-3760.

Hawthorne G, Richardson J and Osborne R (1999) The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life. Quality of Life Research. Vol. 8, pp.209-24.

Hawthorne, Richardson J, Osborne R, Day N and McNeil H (2000b) Construction and Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument. Centre for Health Program Evaluation, Melbourne.

Hawthorne G and Sansoni J (2004) Incontinence and quality of life. Health Outcomes 2004: Perspectives on Population Health Conference. Canberra, 15-16 September 2004.

Health and Social Care Information Centre (2015) Summary – Benefits case study for Patient Reported Outcomes Measures (PROMS) outputs. NHS,UK.

Henderson G and Gray A (1994) The Tharawal Aboriginal Family Environmental Health Survey. Australian Institute of Aboriginal and Torres Strait Islander Studies, ANU, Canberra.

Hirsch N et al. (1994) Minimal Access Surgery: an Update. A Discussion Paper, AGPS.

Horn SD, DeJong G and Deutscher D (2012) Practice-Based Evidence Research in Rehabilitation: An Alternative to Randomized Controlled Trials and Traditional Observational Studies. Archives of Physical Medicine and Rehabilitation. Vol. 93 (Suppl. 1), pp.S127-S37.

Horn S and Gassaway J (2007) Practice-Based Evidence Study Design for Comparative Effectiveness Research. Medical Care. Vol. 45 (Suppl. 2), pp.S50-S57.

Hutchinson A (1998) Primary care outcome assessment- 'the keep it simple' approach. National Health Outcomes Conference: Implementing the Health Outcomes Approach, Canberra, 7-8 August 1998.

Hyland ME (1991) The Living With Asthma Questionnaire. Journal of Respiratory Medicine. Vol. 85 (Suppl. B), pp.13-16.

Jenkinson C, Stewart-Brown S, Petersen S and Paice C (1999) Assessment of the SF-36 Version 2 in the United Kingdom. Journal of Epidemiology and Community Health. Vol. 53, No. 1, pp.46-50.

Kaplan RM (1993) Quality of life assessment for health resource allocation. Harkness Health Conference. Canberra, 8-9 December 1993.

Kaplan RM, Alcaraz JE, Anderson JP and Weisman M (1996) Quality-adjusted life years lost to arthritis: effects of gender, race, and social class. Arthritis Care Res. Vol. 9, No.6, pp.473-82.

Kaplan RM, Bush JW and Berry CC (1976) Health status: Types of validity for Index of Well-being. Health Service Research. Vol. 11, No. 4, pp.478-507.

Katz S et al. (1959) Index of independence in activities of daily living (ADL). In Wilkin D, Hallam L, and Doggett MA (eds.) (1992) Measures of Need and Outcome for Primary Health Care. Oxford University Press, UK.

Kessler R (1997) Kessler's Psychological Distress Scale. Department of Health Care Policy, Harvard Medical School, Boston.

Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, Walters EE and Zaslavsky AM (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychological Medicine. Vol. 32, No. 6, pp.959-76.

Kessler R and Mroczek D (1994) Final Version of our Non-specific Psychological Distress Scale [memo dated 10/3/94]. Ann Arbor (MI): Survey Research Center of the Institute for Social Research, University of Michigan.

Kind P (1996) The EuroQol Instrument: An index of health- related quality of life. In Spilker, B (ed.) Quality of Life and Pharmacoeconomics in Clinical Trials. Lippincott-Raven Publishers, Philadelphia, pp.191-201.

Kingi T and Durie MH (1998) A framework for measuring Māori mental health outcomes (TPH 97/5). Palmerston North: Te Pūmanawa Hauora, School of Māori Studies, Massey University.

Kingi T and Durie MH (2000) Hua Oranga: A Māori Measure of Mental Health Outcomes (TPH 00/01). Palmerston North: Te Pūmanawa Hauora, School of Māori Studies, Massey University.

Kingman S (1994) Quality control for medicine. New Scientist. Vol. 143, pp.22-6.

Kolstad A (1994) Quality of Life: The concept and how to measure it. Australian Association for Social Research Conference. University of Tasmania, 25-27 January, 1994.

Kroenke K, Spitzer RL and Williams JB (2001) The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine. Vol. 16, No.9, pp.606-13.

Lavallee DC, Chenok KE, Love RM, Petersen C, Holve E, Segal CD and Franklin PD (2016) Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. Health Affairs. Vol. 35, No. 4, pp.575-582.

Leape LL (1989) Unnecessary surgery. Health Services Research. Vol. 24, No. 3, pp.351-407.

Lee C (2002) Women's Health Australia, seven years on: Contributions to the evidence base for health outcomes policy and practice. Health Outcomes 2002: Current Challenges and Future Frontiers. Canberra, November 2002.

LoGiudice D, Smith K, Thomas J, Lautenschlager NT, Almeida OP, Atkinson D and Flicker L (2006) Kimberley Indigenous Cognitive Assessment tool (KICA): development of a cognitive assessment tool for older indigenous Australians. International Psychogeriatrics. Vol. 18, No.2, pp.269-80.

Lorig K, Stewart A, Ritter P, Gonzalez V, Laurent D and Lynch J (1996) Outcome Measure for Health Education and Other Health Care Interventions, Sage Publications, 1000 Oakes CA.

Lugtenberg M, Burgers JS and Westert GP (2009) Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. Quality and Safety in Health Care. Vol. 18, No. 5, pp.385-92.

Lundström M and Karlskrona R (2015) Three National PROM-projects in Sweden. PROM Seminar, Stockholm 2015.

Mahoney FI and Barthel DW (1965) Functional Evaluation: The Barthel Index. Maryland State Medical Journal. Vol. 14, pp.61-65.

Marosszeky JE and Eagar K (2001) The Australian Rehabilitation Outcome Centre (AROC). Health Care in Perspective 2001 – Incorporating the 13th National Casemix Conference. Hobart, 16-19 September 2001.

Marosszeky N (2006) Reviews of SF-12, SF-36, WHOQOL-100, WHOQOL-Bref. In Thomas et al. (2006) Continence Outcomes Measurement Suite Project: Final Report, DHA.

Mathers C (1999) Australian Burden of Disease – implications for priority setting. Conference Proceedings - Health Outcomes: Integrating the Elements. Australian Health Outcomes Collaboration, Canberra (CD-ROM).

Meenan RF, Gertman PM and Mason JH (1980) Measuring health status in arthritis: The Arthritis Impact Measurement Scale. Arthritis Rheumatism. Vol. 23, No. 2, pp.146-152.

McDowell I (2006) Measuring Health: A Guide to Rating Scales and Questionnaires. Oxford University Press, edition 3.

McLaws ML, Gold J, King K, Irwig LM and Berry G (1988a) The prevalence of nosocomial and community acquired infections in Australian hospitals. Medical Journal of Australia. Vol. 149, No. 11-12 pp.582-590.

McLaws ML, Irwig LM, Mock P, Berry G and Gold J (1988b) Predictors of surgical wound infection in Australia: a national study. Medical Journal of Australia. Vol. 149, No. 11-12, pp.591-595.

Mishra G (1997) Hysterectomy and incontinence in middle age women. Conference Proceedings: Managing and Measuring Health Outcomes: From Policy to Practice. Canberra, 31 Oct – 1 Nov, 1997.

Murray CJL and Lopez AD (1996) The Global Burden of Disease. The Harvard School of Public Health on behalf of the World Health Organisation and The World Bank, Harvard.

Nancarrow S (1999) Health Outcomes: Issues for community care. Workshop: Health Outcomes and Community Care. Canberra, 26 February 1999.

Nelson EC, Dixon-Woods M, Batalden PB, et al (2016) Patient focused registries can improve health, care, and science. British Medical Journal. Vol. 354, pp.i3319.

National Health Information Standards and Statistics Committee (NHISSC) (2009) The National Health Performance Framework (2nd Edition).

National Health and Medical Research Council (1995) Guidelines for the Development and Implementation of Clinical Practice Guidelines. 1st Ed., Commonwealth of Australia, AGPS.

National Health and Medical Research Council (1999) A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. Commonwealth of Australia, AGPS.

National Health and Medical Research Council (2000) Health and Medical Research Strategic Review. Implementation of the Government's response: Final Report. Commonwealth of Australia, AGPS.

National Health and Medical Research Council (2014) Annual Report on Australian Clinical Practice Guidelines. Canberra: National Health and Medical Research Council.

National Health Performance Authority (2016) Hospital Performance: Costs of acute admitted patients in public hospitals from 2011-12 to 2013-14 (In Focus).

National Health Performance Committee (NHPC) (2001) National Health Performance Framework Report, Queensland Health, Brisbane.

National Health Service (NHS) England (2016) National Patient Reported Outcome Measures (PROMs) Programme Consultation (Publications Gateway Reference: 04478). Available from: https://www.engage.england.nhs.uk/consultation/proms-programme accessed 9 August 2016 accessed 9 August 2016.

National Joint Registry (2016) NHS England's national PROMs programme consultation: Response from the British Orthopaedic Association (BOA), British Hip Society (BHS), British Association for Surgery of the Knee (BASK), British Elbow and Shoulder Society (BESS), British Orthopaedic Foot and Ankle Society (BOFAS) and the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR). Available from: http://www.njrcentre.org.uk/njrcentre/NewsandEvents/NJRrespondstoNHSEngland%E2%80%99sPROMsconsultation/tabid/407/Default.aspx accessed 9 August 2016.

National Quality Forum (2013) Patient Reported Outcomes (PROs) in Performance Measurement. Available from: http://www.qualityforum.org/Publications/2012/12/Patient-Reported_Outcomes_in_Performance_Measurement.aspx accessed 9 August 2016.

National Mental Health Information Development Expert Advisory Panel (2013) Mental Health National Outcomes and Casemix Collection: NOCC Strategic Directions 2014 – 2024. Commonwealth of Australia, Canberra.

National Mental Health Strategy (2009) National mental health benchmarking project evaluation report July 2009. Canberra: Dept. of Health and Ageing

NSW Health (1994) Report on the 1993 NSW Diabetes Outcomes Workshop (NDOW), NSW Health, Sydney.

Marin T, Taylor A and Gill T (2009) A Population Profile of Quality of Life in South Australia - Population Norms for 2008 using the SF36 Version 2. Population Research and Outcome Studies Unit. South Australian Department of Health, Adelaide.

O'Boyle CA, Browne J, Hickey A, McGee HM and Joyce CRB (1993) The Schedule for the Evaluation of Individual Quality of Life (SEIQoL): a Direct Weighting procedure for Quality of Life Domains (SEIQoL-DW): Administration Manual. Department of Psychology, Royal College of Surgeons in Ireland, Dublin.

O'Boyle C and Ring L (2005) Individual Quality of Life (IQoL): Theory and Measurement. 11th Annual National Health Outcomes Conference, 17-18 August 2005, Canberra.

Porter ME (2010) What is value in health care? New England Journal of Medicine. Vol. 363, pp. 2477-81.

Porter ME (2016) The Strategy to Transform HealthCare and The Role of Outcomes. Fourth ICHOM Conference, London, UK, 17 May 2016.

Porter ME, Larsson S, Lee TH (2016) Standardizing Patient Outcomes Measurement. New England Journal of Medicine. Vol. 374, No. 6, pp.504-6.

Jackson Pulver L, Broe GA, Grayson D, Chalkley S, Flicker L, Daylight G, Mack H, Radford K (2012) Dementia screening for urban Aboriginal Australians: the modified Kimberley Indigenous Cognitive Assessment (mKICA) Sydney: Dementia Collaborative Research Centres.

Radosevich D and Husnik M (1995) An abbreviated health status questionnaire: HSQ-12. Update. Vol. 2, No. 1.

Ramey DR, Raynauld J-P and Fries JF (1992) The Health Assessment Questionnaire 1992: Status and review. Arthritis Care and Research. Vol. 5, pp.119-29.

Reid B et al. (1999) Relative Utilisation Rates of Hysterectomy and Links to Diagnoses. DHAC, Canberra.

Reinhardt UE (1990) The social perspective. In Heithoff KA.and Lohr KN (eds.) (1990) Effectiveness and outcomes in health care. National Academy Press, Washington DC.

Richardson J, Iezzi A, Khan MA, Chen G, Maxwell A (2016) Measuring the Sensitivity and Construct Validity of 6 Utility Instruments in 7 Disease Areas. Medical Decision Making. Vol. 36, No. 2, pp.147-59.

Ross B et al. (1999) Health Expenditure: Its Management and Sources (Occasional Papers: Health Financing Series Vol 3). DHAC, Canberra.

Rosser RM (1988) A health index and outcome measure. In Walker SR and Rosser RM (eds.) (1988) Quality of Life: Assessment and Application. MTP, Lancaster.

Rosser R (1993) A health index and output measure. In Walker S and Rosser R (eds.) Quality of Life Assessment: Key Issues in the 1990s. Kluwer Academic Publishers, Dordrecht.

Rubenach S (1999) Health outcomes for the elderly: community and domiciliary care. Workshop: Health Outcomes and Community Care. Canberra, 26 February 1999.

Ruta DA, Garratt AM, Leng M, Russell IT, MacDonald LM (1994) A new approach to the measurement of quality of life. The Patient-Generated Index. Medical Care. Vol. 32, No.11, pp.1109-26.

Sansoni J (2016) The Australian Dementia Outcomes Measurement Suite. Poster presentation at ICHOM Conference, 16-17 May 2016, London.

Sansoni J, Williams K, Morris D and Thompson C (2016) Development of prioritised patient experience indicators for the Patient Centred Quality Cancer System Program: Technical Report. Centre for Health Service Development, Australian Health Services Research Institute, University of Wollongong.

Sansoni J (2015) Health Outcomes: An Overview from an Australian Perspective. Australian Health Outcomes Collaboration, Australian Health Services Research Institute, University of Wollongong.

Sansoni J, Grootemaat P and Duncan C (2015) Question Prompt Lists in health consultations: A review. Patient Education and Counselling. Vol. 98, No.12, pp.1454–64.

Sansoni J, Hawthorne G, Fleming G and Marosszeky N (2013a) The Revised Faecal Incontinence Scale: A clinical validation of a new short measure for assessment and outcomes evaluation. Diseases of Colon and Rectum. Vol. 56, No.5, pp.652-9.

Sansoni J, Samsa P, Duncan C, Westera A, Grootemaat P, Shadbolt B and Eagar K (2013b) Final project report on the validation and field trials of the assessment framework and tool for aged care. Centre for Health Service Development, Australian Health Service Research Institute, Wollongong, Australia.

Sansoni J, Hawthorne G, Marosszeky N, Moore K, Fleming G and Owen E (2011) Validation and Clinical Translation of the Revised Continence and Patient Satisfaction Tools: Final Report. Centre for Health Service Development, University of Wollongong.

Sansoni J, Marosszeky N, Fleming G and Sansoni E (2010) Selecting Tools for ACAT Assessment: A Report for the Aged Care Assessment Program (ACAP) Expert Clinical Reference Group. Centre for Health Service Development, University of Wollongong. Report for the Aged Care Assessment Program, Australian Government Department of Health and Ageing.

Sansoni J, Marosszeky N, Jeon Y-H, Chenoweth L, Hawthorne G, King M, Budge M, Zapart S, Sansoni E, Senior K, Kenny P, Low L (2007) Final Report: Dementia Outcomes Measurement Suite Project. Centre for Health Service Development, University of Wollongong.

Sansoni J (2004) The use and value of HRQOL measurement for the Australian health system. Harmonizing International Health-Related Quality of Life (HRQOL) Research, 11th Annual Conference of the ISOQOL, October 16-19, Hong Kong.

Sansoni J and Costi J (2001) SF-36 Version 1 or Version 2: The need for Australian normative data. Proceedings of Health Outcomes 2001 - The Odyssey Advances: 7th Annual National Health Outcomes Conference. Canberra, 27-28 June 2001.

Sansoni J (1997) The use of the SF-36 in population health surveys: The Australian experience. In Proceedings: State of the Art health Outcomes Conference. MOT, San Francisco, May 1997.

Sansoni J (1996) The Centrality of Health Outcomes Measurement. In Integrating Health Outcomes Measurement in Routine Health Care Conference Proceedings. Australian Institute of Health and Welfare, Canberra, 13-14 August 1996.

Sansoni J (1995) Quality of life: Measure for Measure. In Health Outcomes and Quality of Life Measurement: Conference Proceedings. Australian Institute of Health and Welfare, Canberra, August 14-15 1995.

Schamber EM, Takemoto SK, Chenok KE and Bozic KJ (2013) Barriers to completion of Patient Reported Outcome Measures. J Arthroplasty. Vol. 28, No.9, pp.1449-53.

Schipper H, Clinch J and McMurray A (1984) Measuring the quality of life in cancer: the Functional Living Index-Cancer, development and validation. Journal of Clinical Oncology. Vol. 2, pp.472-483.

Schuler T and Miller AA (2014) PROsaiq: A Smart Device-Based and EMR-Integrated System for Patient-Reported Outcome Measurement in Routine Cancer Care. Journal of Radiation Oncology Informatics. Vol. 6, No.1, pp.111-31.

Scott KM, Sarfati D, Tobias MI and Haslett SJ (2000) A challenge to the cross-cultural validity of the SF-36: A comparative factor analysis in Maori, Pacific and New Zealand European ethnic groups. Social Science and Medicine. Vol. 51, pp. 1655-64.

Senior KA (1999) Cross cultural issues in measuring health and community care outcomes. Workshop: Health Outcomes and Community Care. Canberra, 26 February 1999.

Shadbolt et al. (1996a) Selecting Instruments and Measures for the Co-ordinated Care Trials Workshop. Canberra, July 1996.

Shadbolt et al. (1996b) First Report from the ACT Continuum of Care Project. ACTH&CC.

Shabolt B, McCallum J and Singh M (1997) Health Outcomes by Self-report: Validity of the SF-36 among Australian Hospital Patients. ACT Department of Health and Community Care, Canberra.

Shadbolt B, Merefield S, Wang V, Smith P (2015) Life after arthroplasty. OrthoACT, The Trauma and Orthopaedic Research Unit, Centre for Advances in Epidemiology & IT, The Canberra Hospital, 4 December 2015.

Sintonen H (1994) The 15D measure of health-related quality of life: reliability, validity and sensitivity of its health state descriptive system. National Centre for Health Program Evaluation, Melbourne.

Sintonen H (1995) The 15D‑measure of health‑related quality of life. II. Feasibility, reliability and validity of its valuation system. National Centre for Health Program Evaluation, Working Paper 42, Melbourne.

Sintonen H (2001) The 15D instrument of health-related quality of life: properties and applications. Annals of Medicine. Vol. 33, pp.328-336.

Sintonen H and Pekurinen M (1993) A fifteen-dimensional measure of health-related quality of life (15D) and its applications. In Walker S and Rosser R (eds.) Quality of Life Assessment: Key Issues in the 1990s. Kluwer Academic Publishers, Dordrecht.

Slutsky J (1998) Guidelines and outcomes: AHCPR's National Guideline Clearinghouse. In Sansoni J and Tilley L (Eds.) Implementing the Health Outcomes Approach (Conference Proceedings). AHOC, November 1998.

Stanford Patient Education Research Center (2000) Sample Questionnaire for the Chronic Disease Self-Management Program. Palo Alto, CA.

Stineman MG, Jette A, Fiedler R, Granger C (1997) Impairment-specific dimensions within the Functional Independence Measure. Archives of Physical Medicine and Rehabilitation. Vol. 78, No. 6, pp.636–43.

Te Pou (2016) Mental health outcome measures. Available from: http://www.tepou.co.nz/outcomes-and-information/mental-health-outcome-measures/28 accessed 9 August 2016.

Thomas S, Nay R, Moore K, Fonda D, Hawthorne G, Marosszeky N and Sansoni J (2006) Continence Outcomes Measurement Suite Project (Final Report). Department of Health and Ageing, Canberra.

Torrance GW, Furlong W, Feeny D and Boyle M (1995) Multi-attribute preference functions. Health Utilities Index. Pharmacoeconomics. Vol. 7, No.6, pp.503-20.

Tracey EA, Roder DM, Currow DC (2012) What factors affect the odds of NSW cancer patients presenting with localised as opposed to more advanced cancer? Cancer Causes and Control. Vol. 23, No. 2, pp.255-62.

Vallance-Owen AJ (2008) PROMs promote health gain and patient involvement. British Medical Journal. Vol. 336, No.7640, pp.344.

Viet CT and Ware JE Jr (1983) The structure of psychological distress and well-being in the general population. Journal of Consulting and Clinical Psychology . Vol. 51, No. 5, pp.730-42.

Walker SR and Rosser RM (eds.) (1992) Quality of Life Assessment, Key Issues in the 1990s. Kluwer Academic Publishers, Hingham, USA.

Ware J (2003) Standardizing health metrics: The SF-36 Health Survey and beyond. In Sansoni, J and Tilley L (Eds.) Conference Proceedings: Health Outcomes 2003: The Quest for Practice Improvement. Canberra, 20-21 August, 2003.

Ware JE et al. (1996) The SF-36 Health Survey Annotated Bibliography: First Edition (1988-1995). The Health Institute, NEMC, Boston.

Ware JE Jnr, Kosinski M, Keller SD (1996) A 12-item short form health survey: Construction of scales and preliminary tests of reliability and validity. Medical Care. Vol. 34, No.3, pp.220-33.

Ware JE and Kosinski M (1996) SF-36 Health Survey (Version 2.0) Technical Note. The Health Institute, Boston, MA.

Ware JE, Kosinski M and Dewey JE (2000) How to Score Version 2 of the SF-36 Health Survey. QualityMetric Inc., Lincoln, RI.

Ware JE and Sherbourne CD (1992) The MOS 36-item short form health status survey (SF-36): 1, Conceptual framework and item selection. Medical Care. Vol. 30, pp.473-83.

Weldring T and Smith SMS (2013) Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Service Insights. Vol. 6, pp.61-68.

Wennberg JE (1987) The paradox of appropriate care. Journal of the American Medical Association. Vol. 258, No. 18, pp.2568-9.

Wennberg JE (1990) On the need for outcomes research and the prospects for evaluative clinical sciences. In Andersen TF and Mooney G (1990) The Challenges of Medical Practice Variations. McMillan Press, London.

Williams K, Sansoni J, Morris D, Grootemaat P and Thompson C (2016) Patient-reported outcome measures: Literature review. Centre for Health Service Development, Australian Health Services Research Institute, University of Wollongong.

Wing JK, Beevor AS, Curtis RH, Park SB, Hadden S, Burns A (1998) Health of the Nation Outcome Scales (HoNOS). Research and development. British Journal of Psychiatry. Vol. 172, pp.11-8.

Wohlin J (2014) SVEUS – National collaboration for value-based reimbursement and monitoring of healthcare in Sweden. OECD Expert group meeting on payment systems, 7 April 2014. IVBAR, Institute, Stockholm.

World Health Organisation (1981) Program Evaluation: Guiding Principles. Health For All. Series No 6, WHO, Geneva.

World Health Organization (1996) WHOQOL-BREF, Programme on Mental Health, WHO Geneva.

Wu AW, Jensen RE, Salzberg C, Snyder C (2013) Advances in the Use of Patient Reported Outcome Measures in Electronic Health Records. Center for Health Services and Outcomes Research, Johns Hopkins Bloomberg School of Public Health, Baltimore.

Appendix 1 The SF-36

There has been much outcomes work using multi-dimensional profiles, particularly the SF-36 Version 1 (Ware and Sherbourne, 1992). This instrument was utilised with a 50% sample in the 1995 National Health Survey (ABS, 1997), which enabled the linking of data on health differentials to this measure of perceived health status. It has also been used in a variety of population health surveys across the United States and has also been trialled in a variety of clinical settings (Sansoni, 1995, 1997, 2004). Critical reviews of this instrument have been prepared by Sansoni (1995), Dixon et al. (1994) and Marosszeky (2006). Generally the SF-36 is considered to be one of the better multi-dimensional health profile instruments as there is substantial evidence of its reliability, validity and responsiveness to change in a range of studies both in Australia and overseas (Shadbolt et al., 1996a, 1996b, 1997). There are a range of standardised versions available for languages other than English.

It was recommended as the generic measure of health status for the non-Aboriginal National Co-ordinated Care Trials. It was utilised in the ACT Continuum of Care Study which tracked the health status of a sample of 6,000 patients from the commencement of treatment (Shadbolt et al., 1996b) through to six months post-discharge. There are extensive, long term studies that have used the SF-36, such as the Australian Longitudinal Study on Women's Health, which originally involved a longitudinal study of 1,500 women in each of three age groups over a period of five years (Lee, 2002) – although this study is still continuing today. There is Australian normative data for Version 1 of this instrument (ABS, 1997) and a substantial collection of data pertaining to condition profiles and self-reported morbidity. As Australian normative data, until 2006, was only available for Version 1 much Australian research has made use of this version.

Concerning the use of the SF-36 Version 1 in Australia a number of points emerge:

- the SF-36 has been used in a variety of population groups including non-English speaking groups, elderly patients and patients with psychological disorders;

- the SF-36 is frequently used in conjunction with disease or condition specific measures.

The SF-36 has been administered in a number of ways, for example a self-administered postal questionnaire, by personal interview and by computer assisted telephone interview (CATI). Care needs to be taken, however, to check on the mode of administration before making any comparisons between datasets as some subtle differences by mode of administration have been found.

Although the SF-36 has been found to be responsive to changes in health status in a range of acute care clinical settings including both surgical and medical treatments (Shadbolt et al., 1996b; 1997) it may not be sufficiently sensitive to detect change arising from more subtle health care interventions such as a change in health care management practices (e.g. coordination of care). It is more likely that disease, symptom-specific, patient knowledge and satisfaction measures may be more sensitive to change in these contexts.

A revised version of this instrument is now available (Version 2, Ware and Kosinski, 1996; Ware et al., 2000). There have been minor changes made to the wording to develop a more 'international' version, the layout has been refined and made more user friendly and there are

more levels available in the response categories for the role functioning scales (5 point versus 2 point response categories). Thus the role functioning scales are now less prone to floor and ceiling effects. This is relevant to its use with some population groups (elderly, those with chronic disease and multiple morbidities) as elderly patients with multiple morbidities may be more likely to score near the floor of some items (reflecting poorer function). Norm based scoring has also been introduced to ease clinical interpretation. Sansoni and Costi (2001) outline the changes made and discuss the pros and cons of using the different versions of this instrument in the Australian context.

The revisions encompassed in Version 2 appear to have led to a greater precision of measurement particularly for the role functioning scales (Jenkinson et al., 1999; Ware et al., 2000; Sansoni and Costi, 2001). The Version 2 made available for Australian use is the same as the 'international version' with only changes to the metric (kilometres and metres vs. miles and yards) for some items (Sansoni and Costi, 2001). In 2004 the SF-36V2 was included in a special SA Health Omnibus Survey (Harrison Health Research, 2004; Hawthorne, 2006) to collect normative data for norm comparison purposes (Hawthorne and Sansoni, 2004; Hawthorne, 2006; Hawthorne et al., 2007) and additional normative data for 2008 is now available (Marin et al., 2009). However, due to more recent concerns about registration costs associated with the use of the newer members of the SF family of instruments (e.g. the later versions marketed by Optum.com) some researchers are now using the RAND 36-Item Health Survey 1 Questionnaire Items - from which the SF-36 was derived (it has an identical descriptive system). The RAND instrument is in the public domain and is available at www.rand.org/health/surveys_tools/mos/mos_core_36item_survey.html. However, some adaptation for Australia would be required as occurred with the SF-36 Version 1 for Australia and New Zealand. Although it is unclear from the Optum website as to whether registration fees are required for this earlier 'developmental' version of the SF-36 Version 1 for Australia and New Zealand, a response from a recent query suggests that this remains the case.

Appendix 2 An Example of Health Outcomes Assessment in Allied Health

Nancarrow (1999) applied Benson's (1992) Ambulatory Classification Model in a podiatry application known as The Footpath Project. This project was a primary health initiative designed to prevent unnecessary lower limb ulceration or amputation, primarily in people with diabetes, but also in people with peripheral vascular disease or other causes of foot problems.

Table 2   **Health Outcome Measures for the Footpath Project**

Outcomes Type	Foot Health Indicators	Effectiveness Goal	Quality Action Point	Data Source	Frequency of Review
Disease specific outcome	Prevention of pressure (neuropathic) foot ulceration	90% of all patients with peripheral neuropathy remain free of pressure ulcers	80%	Medical Records	Annually
General health outcome	Foot health status questionnaire	90% of patients report that their foot status is excellent or very good	70%	Questionnaire to 33% of patients who receive foot health care	6 monthly
Patient performance outcome	Patient knowledge of risk factors for lower limb ulceration	90% of patients able to demonstrate understanding of lower limb factors	80%	Chart audit – 1% medical records (or patient survey)	6 monthly
Patient satisfaction outcome	Patient satisfaction with level of care they receive for foot complications	90% of patients express satisfaction with the level of care they receive for foot complications	80%	Survey administered to 100% of patients	6 monthly

- The *indicator* is based on what the intervention is trying to achieve within the scope of the service. A number of indicators may be identified, and these should be prioritized into those aspects of care which are most important for the service.

- The *effectiveness goal* is the level to which the organization is going to aim to achieve the chosen indicator. The *effectiveness goals* and *quality action points* are arbitrary and ideally, should be based on the evidence of the effectiveness of interventions as shown by research.

- The *quality action point* is a predetermined threshold that is used to flag the need to introduce quality improvement activity to improve performance on the indicators.

- The *data source* requires careful consideration to provide the level of information required in the most effective way. Data collection can form a major cost in a quality improvement process. Therefore consideration must be given to the availability of the data, the method of data collection, how much data is required to provide meaningful results and, the value of the data in terms of providing useful information.

It can be seen the advantage of this model is that indicators can be built into many areas of routine clinical practice.