Literature Review:
MEASUREMENT OF CLIENT OUTCOMES IN HOMELESSNESS SERVICES

JANUARY 2011

Mark Planigale

HomeGround Services

Service Development and Research

MARK PLANIGALE

Research & Consultancy

results by design

PO Box 754

Macleod VIC 3085

Australia

Tel: 0429 136 596

Email: results@planigale.com

Web: www.planigale.com

HomeGround Services

we're housing people

HOMEGROUND SERVICES

1a/68 Oxford St

Collingwood VIC 3066

Australia

Tel: (+61) 3 9288 9600

Email: research@homeground.org.au

Web: www.homeground.org.au

Literature Review:
Measurement of Client Outcomes in Homelessness Services

Contents

HomeGround
Services

Literature Review:
Measurement of Client Outcomes in Homelessness Services

1. Introduction

1.1 ABOUT HOMEGROUND SERVICES

HomeGround Services is one of Melbourne's leading housing and homelessness organisations. HomeGround's vision is to end homelessness in Melbourne, and HomeGround's mission is to get people housed and keep people housed. HomeGround has a unique combination of experience delivering high volume housing and homelessness services to people in crisis, providing property and tenancy management services, and providing short and long-term support services to people who have experienced homelessness in the past.

HomeGround is a strong supporter of 'housing first' approaches to ending homelessness. HomeGround has played a leading role in introducing new housing and support models in Victoria, including the Private Rental Access Program and a range of Supportive Housing models. HomeGround has a strong reputation for quality service provision, effective advocacy, and for bringing private and community sector partners to the table to achieve change for people experiencing homelessness.

HomeGround has a long-standing commitment to sector development and a track record of involvement in significant research projects (see e.g. Chamberlain, Johnson et al. 2007; Johnson, Gronda et al. 2008).

1.2 ABOUT HOMEGROUND'S CLIENT OUTCOME MEASURES PROJECT

HomeGround is committed to ensuring that its approach to ending homelessness is underpinned by strong evidence. The organisation has identified that one of the most important sources of evidence is information on the outcomes achieved in partnership with clients.

HomeGround's Client Outcome Measures Project aims to establish the foundations of the systems needed to monitor outcomes for people engaged with HomeGround's services. The project activities include:
- Researching the types of indicators that can be used to measure housing and wellbeing outcomes
- Defining outcome measures for each of HomeGround's service streams
- Identifying or developing tools to collect client outcomes data
- Providing input to the development of electronic data systems that can store and report on client outcomes data
- Identifying strategies to embed outcome measurement in practice.

The project runs from January–July 2010 and is expected to be followed by a phase of piloting and implementation of outcome measures within the organisation, influenced by the findings of the project.

This Literature Review and the accompanying discussion paper 'A consistent set of casework domains for HomeGround' (Planigale 2010a) are the key outputs of Phase 2 (background research) of the project.

1.3 LITERATURE REVIEW: AIM AND KEY QUESTIONS

The aim of the literature review is to gain an overview of political, theoretical and logistical considerations related to the introduction of client outcome measurement systems within a homelessness organisation in Australia. The expected benefit of the literature review is that it will allow the process of developing and implementing measures to be well-informed, well planned, and more likely to be successful (avoiding known pitfalls, and enabling more useful measures to be developed). The

literature review document is intended to serve as a technical reference during the process of implementing an outcome measurement system.

The literature review was guided by the following key questions, grouped under the headings of context, system design, and implementation:

Context:
- What are the known benefits and disbenefits or risks of introducing client outcome measurement systems within a human service organisation?
- What current approaches to outcome measurement provide guidance or direction to HomeGround's efforts?

System design:
- What types of information can be produced by outcome measurement systems and what conclusions can this information support?
- What client outcomes are of relevance to homelessness services and how can they be conceptualised and categorised?
- What specific measures and measurement tools may be of relevance to homelessness services?
- What are the options for data collection processes (when should data be collected, from who, by who, and in what format)?
- How can results be presented to be of maximum benefit to clients and the organisation?

Implementation:
- What are the best processes for introducing outcome measurement systems within organisations?

The literature review document is organised into sections mirroring the above questions.

HomeGround Services

Literature Review:
Measurement of Client Outcomes in Homelessness Services

The document ends with a section summarising the main questions that an organisation may need to answer in developing an outcome measurement system.

Some elements of the literature review are specific to the context of the homelessness service system in Victoria. However, many of the issues discussed appear to be general themes within the outcomes measurement literature and are of wide applicability.

1.4 LITERATURE REVIEW: METHODOLOGY AND SCOPE

The literature review was conducted by Mark Planigale. The research process drew on two major sets of sources:
- Review of a range of source documents including books, journal articles, reports, government publications, web pages, and measurement tools; and
- Meetings with selected stakeholders including HomeGround staff and representatives of sector and government partner organisations.

The document review was 'seeded' using three publications already known to be of high relevance: Baulderstone and Talbot (2004), Spellman and Abbenante (2008) and MacKeith and Graham (2007). These publications, along with initial meetings with stakeholders, were used to derive a set of key themes and issues to be explored. These themes form the basis of the key questions noted above.

From the seed publications, other relevant sources were identified using a 'snowball' approach. As the research process identified areas where further depth was needed, other relevant sources were found either by internet search or through library catalogues.

For example, the area of evaluation theory was identified as an important input to the literature review, and the following sources were selected to provide a range of perspectives on this topic: Weiss (1972), Schalock (2001), Bloom et al. (2006), Wadsworth (1997), Patton (Patton 1987) and Rossi, Freeman et al. (1999).

There is a plethora of information available in the area of client outcomes measurement. Due to the limited time available for the literature review, many relevant areas were either visited only briefly, or not at all. Topics of interest that were not able to be adequately pursued within the literature review are labelled 'for further research' in this document.

Early in the literature review process, the topic of life domains was identified as a priority to both the Client Outcome Measures Project and to other work in progress at HomeGround. This topic was therefore given a stronger emphasis within the project. The discussion paper on casework domains (Planigale 2010a) should be read in conjunction with this literature review document.

1.5 NOTE ON LANGUAGE

In this paper the term 'client' is used to refer to a person who uses the services of a community or welfare organisation. This is in line with HomeGround Services' standard terminology. It is acknowledged that in other contexts, a range of other terms may be preferred including 'consumer', 'service user', 'patient' or 'participant'.

HomeGround Services

2. Benefits and challenges of client outcome measurement

The literature identifies a range of benefits associated with client outcome measurement, however it is also worth considering potential disbenefits, challenges and risks associated with outcome measurement systems. This allows a well-informed decision about whether, and how, to proceed.

2.1 BENEFITS

The principal benefit that tends to be cited in favour of client outcome measurement is that it focuses staff, organisations and service systems on the needs, goals and achievements of clients. Rapp and Poertner (1992: 16) articulate the central tenet that "the raison d'être of the social administrator is client wellbeing and that the principal task of the manager is to facilitate that wellbeing", and from this derive four principles of client-centred management, including 'Creating and maintaining the focus on the clients and client outcomes'. MacKeith (2007) notes that a strong focus on the desired outcomes of a service tends to positively change the way that staff and clients work together. Having a clear shared understanding of what the goals are (and a shared language for talking about them) can be an important basis for working together (Spellman and Abbenante 2008: 4).

A focus on client outcomes, together with the ability to measure them in a meaningful way, can have the following benefits (Burns and Cupitt 2003; MacKeith 2007; MacKeith and Graham 2007; Spellman and Abbenante 2008):

- Evaluative benefits (system level): outcomes measurement assists funders to assess the effectiveness of use of public funds and to consider how it may be targeted to maximise cost-effectiveness.
- Evaluative benefits (organisational level): outcomes measurement assists organisations to understand whether what they are doing is working, to what extent, and for which clients. It helps the organisation (or service or team) to answer the question of whether it is being successful in its mission, and can therefore play a role in guiding decision making. While outcomes measurement by itself cannot answer questions related to attribution, in combination with other data it can potentially provide some information on service effectiveness. Outcomes data can also at times provide a useful input to formal program evaluations.
- Evaluative benefits (individual level): information on individual client outcomes can provide useful data for reflection by individual staff members and clients on individual progress and the effectiveness of the services being provided to that individual.
- Assessment benefits: outcomes measurement can provide clients, service delivery staff, managers and Boards with an overview of how the clients' situations and needs are changing over time. This is important information that can help to drive future service delivery at the individual and program level.
- Quality improvement benefits: outcomes measurement can drive quality improvement, both by identifying what works, and by identifying interventions or approaches that are less successful and are in need of review.
- Motivational benefits: outcomes measurement can help both staff and clients to recognise progress and celebrate achievements. This can be a significant shift for organisations whose staff are constantly attending to the hard work of service delivery. "It can be de-motivating to always be travelling (i.e. focussed on delivery) and never arriving (i.e. recognising that a goal has been achieved)" (MacKeith 2007: 2). While the motivational benefits may vary across different groups of staff, organisations that involve staff in defining desired outcomes and measures often report an enthusiastic response (e.g. Hendrick 2010b).
- Advocacy benefits: outcomes measurement can assist in demonstrating the successful results of a program or intervention, as well as potentially demonstrating levels of client need. This information may help in the task of generating support (partnerships, public perception, funding).
- Knowledge building: outcomes measurement can contribute to research and evaluation, can generate hypotheses and questions for further research, and can contribute to drawing together learning from across many organisations to assist the development of evidence based service delivery.

One note of caution is that while these benefits are often asserted, there is little documented research that objectively demonstrates them (Booth and Smith 1997: 42). There is certainly anecdotal evidence that outcomes measurement can have positive effects on staff morale (Clements 2010; Hendrick 2010b), and there is some evidence that introduction of outcome measures by funders can positively affect service effectiveness (e.g. Wells and Johnson 2001: 194). Friedman et al. (Friedman, DeLapp et al. 2001a) cite a number of case studies in which outcomes measurement was important to demonstrating the population-level changes resulting from partnerships to improve wellbeing. It is unclear whether these change efforts were better implemented or more successful as a result of the use of outcome measurement, although it is clear that the use of baseline data was an important motivational tool in initiating change.

There is clearly scope for further research on the outcomes of client outcome measurement itself.

HomeGround
Services

2.2 DISBENEFITS, CHALLENGES AND RISKS

Potential adverse aspects of outcomes measurement are grouped here under headings that deal broadly with four categories of impacts:
- Resourcing impacts
- Staff impacts
- Impacts relating to the value of the information produced, and
- Impacts on service delivery.

See also Post, Isbell et al. (2005: 6-13).

Outcomes measurement can be expensive

Most authors agree that implementing outcomes measurement can be an expensive and time-consuming process. Resources are required for researching and developing outcome statements, measures and tools; training; developing or modifying data systems; collecting and entering data; supporting staff and trouble-shooting; analysing and reporting on data; and ongoingly reviewing the outcomes system itself. Rapp and Poertner (1992: 107) caution that managers typically underestimate the level of resources needed for data collection and data entry, while Berman and Hurt (1997: 87) note that outcomes data systems are more likely to contribute valuable information if they are adequately resourced, and operated by trained staff. Implementing outcomes measurement superficially in an attempt to minimise cost and avoid changes in organisational culture may actually be less cost-effective – MacKeith (2007: 4) suggests that organisations that approach outcomes measurement in this way will not achieve the benefits of increased focus, motivation and effectiveness.

Schalock (2001: 39) argues that despite the value of outcome measurement, it also needs to be balanced with other considerations, including its cost: "One needs to be cautious that the outcomes measurement system does not consume in resources more than its information is worth." Hudson (1997: 78) makes a related point, noting the distortion that can arise if the resources put into managing the performance of a program (including monitoring outcomes) exceed the resources actually provided for service delivery.

Outcomes measurement can be difficult to sustain over time

Anecdotal evidence (Clements 2010; Talbot 2010) suggests that one of the greatest challenges in outcomes measurement is sustaining the measurement systems over time. While appropriate resources and organisational focus may be provided initially, these may be impacted by the pressures of service delivery and the introduction of other organisational initiatives. Ongoing commitment and resourcing is required to ensure the systems function well.

Staff may feel threatened

It is possible that some service delivery staff may feel threatened by outcomes measurement systems (Rapp and Poertner 1992: 101). Staff can feel they are being scrutinised. Where services are delivered to clients with complex needs in resource-poor environments, there is a risk that managers or frontline staff will feel they are being held accountable for things that are outside of their control (Schalock 2001: 4, 39). It is essential that both those collecting and those interpreting the data understand the range of factors and constraints that affect outcomes, including the fundamental observation that outcomes are largely controlled by the client. It is possible that introduction of outcomes measurement may lead to the departure of a small number of staff who are not willing to make practice changes (Smith, Rost et al. 1997: 132).

The information produced may be of poor quality

There may be a variety of reasons why outcomes measurement systems do not produce useful information. Outcome measures (and measurement tools) are subject to the same range of threats to reliability and validity as any other psychometric measures. Some potential limitations and sources of bias or error that are particularly related to the organisational context may include low response rates (Hatry 1997: 18), administration of complex measures by staff with limited training or knowledge (Berman and Hurt 1997: 88), collector bias (especially where those responsible for ratings are the same as those delivering the service – cf. Rossi (1997: 31)), and use of ratings to achieve an instrumental purpose related to service delivery (e.g. to demonstrate client eligibility for certain resources or services) rather than as an accurate reflection of the client's status (Hudson 1997: 77). It is also possible that the selection of measures themselves can be subject to 'dumbing down' due to the political purposes to be served by the results. Segal (1997: 154-155) gives the example that stakeholders may shy away from the use of 'hard' outcomes measures as they are less likely to show change than 'soft' measures.

Another key challenge related to information quality is the adequacy of electronic systems for storing outcomes data. Most homelessness organisations operate with a suite of inconsistent and only partially functional client data systems that are not designed to capture or analyse outcomes data or integrate it with service delivery. Even with the best measures, if data systems are not available to process the data this can lead to a breakdown in the measurement process (Hendrick 2010b).

The information produced may be used poorly

Even where the information produced by outcomes measurement is of high quality and relevance, it is only beneficial to the extent that it is actually used to improve services for clients (Schalock 2001: 39). Berman and Hurt (1997: 87) comment that there is "no greater burden" than the collection of information that is not used. One danger is that the informational needs of some stakeholders (often service delivery staff) will be ignored in outcome measurement systems (Hudson 1997: 73, 76). A different problem is the use of outcomes data to motivate conclusions or decisions without understanding the limitations of the data. Booth and Smith (1997: 40) note that the end users of the results (often agency or governmental decision makers) are not necessarily familiar with technical aspects of research or evaluation design. This can lead to an inflated view of the reliability and validity of outcomes data in demonstrating causal links between programs and results. For this reason, among others, Wells and Johnson (2001: 193) advise caution in using outcome information to allocate resources.

The information produced may reflect adversely on services

Organisations typically hope that outcome measurement will demonstrate the success of their work. However, as Rossi (1997: 24) notes, program designers and operators often have exaggerated expectations in terms of the outcomes of their services. Outcome measurement may in fact suggest that a program is neither effective nor efficient - whether because the environment imposes severe constraints on what it is possible to achieve, or because the intervention itself is misguided. It is important that organisations entering into outcome measurement are prepared to face this possibility, and have a strategy for dealing with the political and

staff morale implications of data indicating low levels of success (Segal 1997: 155). Stakeholders may become dissatisfied with the length of time required to achieve change in outcomes; Wells and Johnson (2001: 195) suggest that achieving change tends to take much longer than expected.

One possible response is to use 'risk adjustment' or 'casemix' style approaches to balance outcomes against complexity and severity of presenting issues (e.g. Spellman and Abbenante 2008: 41-43). However, this in itself is complex and adjustment formulae tend to be controversial.

Adverse outcomes information can also have an effect in individual casework. Outcomes tools can show regression as well as progress (MacKeith, Graham et al. 2007: 13), and negative ratings by a case manager of a client's progress, if known to the client, may impact on the worker-client relationship (Love and Scott 2010). Careful thought should be given to the way that outcomes are discussed in these contexts.

Measurement may have adverse impacts on service delivery

The literature identifies two types of possible adverse consequences to service delivery of outcomes measurement. One is a distortion of the types of intervention provided by a measurement focus on certain outcomes. For example, in a child protection setting, a measurement focus on length of time to family reunification (with shorted durations understood as better) might lead to a pressure on staff to return children to their families while significant risks or instability still existed. Therefore, a counterbalancing measure of return to foster care within a specified period of time should also be used (Wells and Johnson 2001: 191). Wells and Johnson

argue that it is important to hold service providers accountable for both outcomes and process (quality assurance) – "Measuring one to the exclusion of the other can lead to inappropriate practices that focus on the measure rather than the experience of the child and family" (180). A related issue is the potential for outcomes measurement to lead to 'soft targeting' or to services focusing on countable outcomes to the exclusion of less easily measured goals (Wells and Johnson 2001: 195).

Another potential detriment to service delivery is that outcome measurement results may become just another way of categorising or labelling clients, and that this can become disempowering for clients. Pilon and Ragins (2007: 11) discuss an example of a mental health clinic that had been rating clients using the MORS recovery indicators and had begun using the milestones as a shorthand way to describe consumers: "She's a three." "He thinks he's a 7, but he's really a 5." The authors express concern at the dehumanisation implied by this use of language, and suggest that attention to agency culture is important to ensuring that outcomes tools are used in a strengths-based perspective.

2.3 BALANCING BENEFITS AND RISKS

Organisations considering introducing outcomes measurement systems may find it useful to weigh up the potential benefits and risks of such systems. Although the list of risks may appear daunting, the majority of these impacts can be mitigated through a well-planned and well-executed implementation process, including clear communication processes. This speaks to the need for adequate resourcing for implementation, and for timeframes that allow for meaningful consultation with a range of stakeholders.

A question of particular interest for community sector organisations is whether the benefits of outcomes measurement systems are worth the resources required. The answer to this question will depend on the context and informational needs of each individual organisation. However, it is useful to remember that:

- Organisations can choose the scope of the outcomes that they choose to measure, and find a scope that represents best informational value for money
- A staged approach to implementation can distribute the resource burden over time and also has other benefits, including enabling assessment of the usefulness of the initial stages prior to roll-out of subsequent parts of a package.

Baulderstone and Talbot (2004: vii) weighed up practical aspects of outcome measurement in Supported Accommodation Assistance Program (SAAP) services, concluding: "Of those that participated in the pilot instrument trials, some reported positive experiences and some negative. The project team found that where there was negative experience, the reasons for this were identifiable and could be dealt with, and that there was no fundamental barrier to outcome measurement implementation in most instances."

3. Homelessness sector:
approaches to defining and measuring client outcomes

3.1 AUSTRALIA – NATIONAL OUTCOMES AND INDICATORS

The overall framework for tackling homelessness in Australia is established by the National Affordable Housing Agreement (NAHA) (Council of Australian Governments 2009a), which aims to ensure that all Australians have access to affordable, safe and sustainable housing that contributes to social and economic participation. The NAHA is supported by three National Partnership Agreements: one on social housing, one on Indigenous Australians living in remote areas, and most relevant to the current paper, the National Partnership Agreement on Homelessness (NPAH) (Council of Australian Governments 2009b).

These government documents may be relevant to measurement of client outcomes in individual homelessness services in two ways:

- They provide an indication of government policy focus in terms of homelessness, which is one source of guidance in terms of the types of outcomes that services may pursue, and how outcomes could be framed for advocacy purposes.
- They provide an indication of government interest in particular data items, and therefore of potential future reporting requirements. While some of these are population outcomes (and therefore likely to be measured through census-type approaches rather than from agency data), others may need to be built into agency data collection.

The NAHA includes population-level outcomes, outputs and performance measures; selected items are summarised in Table 1.

Table 1:
National Affordable Housing Agreement:
excerpts from outcomes, outputs and performance indicators

Outcomes	• people who are homeless or at risk of homelessness achieve sustainable housing and social inclusion • people are able to rent housing that meets their needs
Outputs	• number of people who are homeless or at risk of homelessness who are assisted to secure and sustain their tenancies • number of people who are assisted to move from crisis accommodation or primary homelessness to sustainable accommodation
Performance indicators	• proportion of low income households in rental stress • proportion of Australians who are homeless • proportion of people experiencing repeat periods of homelessness

The NPAH is designed to contribute to the NAHA outcome 'People who are homeless or at risk of homelessness achieve sustainable housing and social inclusion'. Table 2 summarises the outcomes explicitly stated in the agreement; other outcomes implied in the agreement's outputs; and excerpts from the performance indicators specified in the agreement (the performance indicators included here are those focused on service effectiveness rather than service quality or effort). The performance indicators also have associated baselines and performance benchmarks: see Council of Australian Governments (2009b: 7-8).

Table 2: National Partnership Agreement on Homelessness: stated outcomes, implied outcomes and selected performance indicators

Stated outcomes	• Fewer people will become homeless and fewer of these will sleep rough; • Fewer people will become homeless more than once; • People at risk of or experiencing homelessness will maintain or improve connections with their families and communities, and maintain or improve their education, training or employment participation; and • People at risk of or experiencing homelessness will be supported by quality services, with improved access to sustainable housing
Implied outcomes	• Public and private tenants sustain their tenancies • Homeless people secure or maintain stable accommodation • Homeless people (including families) 'stabilise their situation'
Performance indicators	• Proportion of Australians who are homeless • Proportion of Australians who are experiencing primary homelessness (rough sleeping) • The number of families who maintain or secure safe and sustainable housing following family violence • Increase in the number of people exiting care and custodial settings into secure and affordable housing • Reduce the number of people exiting social housing and private rental into homelessness • The proportion of people experiencing repeat periods of homelessness • Number of young people (12 to 18 years) who are homeless or at risk of homelessness who are re-engaged with family, school and work

The Supported Accommodation Assistance Program (SAAP) has long maintained a data collection system. This system does include some data items which provide an indication of status maintenance and change, pre- and post- engagement, in areas such as accommodation, main source of income, labour force status, student status and living situation of clients. However, many of the changes achieved by clients are not reflected in SAAP data reporting (Baulderstone and Talbot 2004: 1, 5).

The SAAP V Multilateral Agreement (Commonwealth of Australia, State of New South Wales et al. 2005: 36-40) included a set of performance indicators, many of which are concerned with access, cost, and activities or outputs. Two of the current performance indicators and three of the indicators "being considered for future implementation" can fairly be seen as outcomes indicators. These are summarised in Table 3.

Table 3: SAAP V Multilateral Agreement: selected performance indicators related to client outcomes

Current indicators	Indicators being considered for future implementation
20. The extent to which clients' case management goals are achieved at case closure 37. Number and proportion of clients not returning to SAAP within 6 months	26. Number and proportion of clients at risk of homelessness who were assisted by SAAP to maintain their accommodation 27. Number and proportion of clients at risk of homelessness who were assisted by SAAP to maintain family links where their health and safety will not be endangered 28. Number and proportion of people at risk of homelessness who were assisted by SAAP to obtain appropriate accommodation upon their exit from an institution

3.2 VICTORIAN DEPARTMENT OF HUMAN SERVICES

The Victorian Department of Human Services (DHS) has an interest in the development of outcome measures for homelessness and other services. Prior to the 2009 re-organisation of the former DHS into the current Departments of Health and Human Services, the Department had engaged Allen Consulting to develop a set of draft Departmental, Program and aspirational outcomes (aspirational outcomes being those requiring the input of other stakeholders, such as other branches of government). Subsequent to the departmental re-organisation, these proposed outcomes are being revised to make them more suitable to the current focus of DHS. The process is being driven by the DHS Central policy team. However, these outcomes are not yet publicly available (Trinh 2010). There is also considerable work occurring within the Department in relation to development of outputs and outcomes for the new Victorian Homelessness Strategy (Homelessness 2020). Again the draft outcomes are not available for public comment at this stage, although they will be based partly on consultation that has already occurred.

Program evaluations of homelessness services contracted by the Department also provide an indication of the Department's focus of interest in terms of outcomes. Recent and current evaluation projects (including e.g. those of Elizabeth St Common Ground Supportive Housing, Kensington Redevelopment, J2SI, YHAP2) have focused on housing stability, social inclusion (including participation in Employment, Education and Training), and physical and mental health. Where children are involved (for example in evaluations of neighbourhood renewal programs), educational continuity, retention and attainment are also of interest (Trinh 2010).

Youth refuges: Support Assessment Framework

A pilot is currently underway of an outcomes measurement system for youth refuges in Victoria. The measures, recording tool and implementation process are being developed and coordinated by Sally Elizabeth for DHS, with considerable input from sector stakeholders. The tool is to be piloted from May 2010.

The Support Assessment Framework (SAF) is intended to meet a range of needs as an integrated assessment, planning, review, communication and outcomes measurement tool. The draft tool covers a range of domains including income, living skills, health, mental health, housing / homelessness and a series of others. For each domain, the tool has essentially a 5-point status maintenance and change scale. The young person is given a rating on each scale at intake, and again at exit. Each domain also has space for brief qualitative comments in relation to assessment of support required, support provided (action taken), and any action in progress at point of exit. The qualitative items are important because they help to validate the ratings and to 'unpack' the meaning of particular ratings at analysis stage. Some basic demographic data is also included. The tool allows for an index of the young person's overall situation (total of ratings at entry or exit), although the details of how this will be calculated are still being finalised. The tool is designed to reflect the short stays (average 6 weeks) and interim outcomes expected in a youth refuge service environment.

SAF data for all Victorian youth refuge clients will be collected by agencies in an Excel spreadsheet template. The data will be deidentified, aggregated and analysed on a quarterly basis by DHS. At this stage the estimate is that on average 1000 records of

data per quarter will be gathered. In addition to the sector-level statistics, it is intended that SAF will be useful at a variety of levels from individual client self-assessment, through worker use in case planning and review, to supervision and agency-level reflection on outputs and outcomes. In this way the data gathered may also be used by organisations to facilitate practice change.

3.3 SAAP PILOTS: BAULDERSTONE AND TALBOT (2004)

As noted above, SAAP performance measurement has primarily concentrated on equity and efficiency rather than on client outcomes. However, Baulderstone and Talbot (2004) conducted and reported on a pilot project that tested the applicability of a variety of outcome measurement tools across a range of SAAP service types. The intention was to develop a system of outcome measures that would be acceptable to SAAP agencies and that would allow aggregation of outcomes data to state and national levels. The project examined both housing and non-housing outcomes.

The project found that outcomes measurement is useful at a number of levels (client and worker use in case management; service management; program planning). However, the authors concluded that due to the great diversity in SAAP service types and service delivery across Australia, no one outcomes tool was applicable to all SAAP services (or even to all services within a service type) (2004: 37). Given this, the authors suggested that the needs of client service workers and managers need to be given priority in decisions on which forms of outcome measurement are undertaken. Another key finding was that staff skills and training were critical to the use of outcomes tools and the collection of reliable and valid data.

The main outcome measurement approaches trialled were Goal Attainment Scaling and Standard Goal Scaling. Several services trialled level of functioning scales and the BT Generic Outcomes Scale. Other approaches and tools were considered (a set of specific Status Maintenance and Change scales; pre-existing standardised scales), however these were not pursued further as consultation suggested they did not meet the needs of the stakeholders or provide a good fit with the service delivery context (Baulderstone and Talbot 2004: 9-10).

Mission Australia

During and following the Baulderstone and Talbot study, a range of Mission Australia services were involved in outcomes measurement pilots (Clements 2010; Talbot 2010). All Mission Australia NSW/ACT Community Services were expected to complete an outcomes measurement project in 2004/5; services in other parts of Australia also participated in pilots. A Tool Kit was developed to support these projects (Mission Australia 2005). Routine outcomes measurement was also seen as feeding into formal service evaluation processes and reflection on practice (Mission Australia 2006: 11).

The measurement approaches and tools promoted through the Tool Kit were Goal Attainment Scaling, Standard Goal Scaling and the BT Generic Outcomes Scale. The Tool Kit includes an extensive list of goals related to SAAP support areas but tailored to Mission Australia's services.

For further research: insights of Mission Australia from their experience of the outcome measurement pilots and subsequent work in performance measurement

3.4 RESULTS-BASED ACCOUNTABILITY

Results-based accountability is a high-profile approach to managing for results that originated in the United States. In its broad sense the term implies at a minimum that "expected results are clearly articulated and that data are regularly collected and reported to determine whether results have been achieved" (Weiss 1997: 174). Results-based accountability has a particular emphasis on using outcomes definitions and measurements to focus service provision and to leverage collaboration among human services agencies and a broad range of partners who have the potential to impact on a problem.

Friedman's approach

One of the best known proponents of results-based accountability is Mark Friedman, who along with collaborators has developed a particular approach to applying results-based accountability (Friedman, DeLapp et al. 2001b; Friedman 2005). Results are understood as "population conditions of wellbeing for children, adults, families and communities"; results data provides information about whether the efforts being made to achieve results are succeeding (Friedman 2005: 11-13).

The core principles of results-based accountability are (Friedman, DeLapp et al. 2001b: 1.1):
1. Start with ends, work backward to means. What do we want? How will we recognise it? What will it take to get there?
2. Be clear and disciplined about language.
3. Use plain language, not exclusionary jargon.
4. Keep accountability for populations separate from accountability for programs and agencies.
   a. Results are end conditions of wellbeing for populations: children, adults, families and communities.

b. Customer or client results are end conditions of wellbeing for customers of a program or agency.
5. Use data (indicators and performance measures) to gauge success or failure against a baseline.
6. Use data to drive a disciplined business-like decision making process to get better.
7. Involve a broad set of partners.
8. Get from talk to action as quickly as possible

One of the key concepts of results-based accountability is that of baselines and of "turning the curve". A baseline is a representation of the current state of affairs - both a historical picture ("where we've been"), and a forecast ("where we're heading if we don't do something different") (Friedman 2005: 28). The expression 'turning the curve' is a way of describing the desired change in a particular condition with comparison to the baseline. Turning the curve involves doing better than the baseline forecast. One of the strengths of Friedman's approach is that it can be strongly motivating in bringing people together to identify which curves are the highest priority to turn, and then in working in partnership to "beat the baseline" on these curves (Cunningham-Smith 2010).

Friedman's approach involves clear processes for "getting from talk to action", and has a clear framework for identifying the most important types of performance measures that may be used to monitor progress (Friedman 2005: 72). However, it does not dictate any particular set of outcome measures and for this reason could be flexibly combined with other theoretical and practical approaches to outcomes measurement. For example, it would be possible for an organisation to monitor a range of outcomes, but to use Friedman's processes during planning, drawing on the monitoring data to drive a focus on particular results areas.

Melbourne Citymission: MORF

Melbourne Citymission (MCm) began introducing an outcomes measurement system in 2008. The organisation had already identified results-based accountability as the framework within which it wanted to develop its outcomes systems. Initially, Friedman's approach was piloted with seven services from MCm's Homelessness and Children & Disability Services portfolios. The pilot was highly successful in engaging staff and led to a decision to roll out the Measuring Outcomes and Results Framework (MORF) across the organisation from mid-2009 (Hendrick 2010a).

During the pilot MCm identified the need to augment Friedman's framework in several ways (Hendrick 2010b):
• Friedman's work tends to assume that the aims and objectives of services are already clear and can provide a starting point for defining client results – in fact, this is not always the case. The process of discussing outcomes with services needs to include a stage of clarifying program aims
• Processes for collecting, storing and analysing data need to be defined. Due to the complexity and diversity of the data systems in operation within MCm (as within most large human services organisations), this is a major challenge.

MCm continue to work on many areas of the implementation of MORF, including working with individual services on defining desired outcomes, standardising outcome statements within portfolios, and developing data systems. At this point, the expectation is that a set of standardised core outcomes will be defined for each portfolio; services within the portfolio will be expected to work towards one or more of these core outcomes, but may also have additional service-specific outcomes that are of importance in their context (Hendrick 2010b).

3.5 OUTCOMES STAR

The Outcomes Star approach was developed in the United Kingdom. The first version of the star tool was developed by Triangle Consulting in 2003 for St Mungo's, a London-based homelessness agency (MacKeith, Burns et al. 2008a: 7). Subsequently the star tool has evolved into a suite of related tools designed for use in different sectors: the Homelessness Star (MacKeith, Burns et al. 2008b), the Mental Health Recovery Star (MacKeith and Burns 2008), the Alcohol Star (Burns and MacKeith 2009a), and the Work Star (Burns and MacKeith 2009b).

All of the tools are based on a common approach which includes:
• An explicit 'journey of change' model with ten steps (ranging from "stuck" to "self-reliance")
• A set of life domains, which vary from tool to tool but with substantial similarities
• A single measure in each domain, reflecting a global measure of the person's "relationship" with that domain, including how well they are managing needs or problems in that domain
• A common visual presentation for the measurement tool and supporting information, using a star-shaped arrangement of axes representing different domains, and a ladder representing stages of change
• The intention that the tool be integrated with casework processes and that ratings be jointly agreed through discussion between client and worker.

The various versions of the outcomes star have been piloted and are currently being used in a range of services, mainly in the U.K. A web-based electronic data system has also been developed to capture and analyse data produced from the star tool.

Major strengths of the outcomes star include its thorough piloting process and appropriateness to the context of homelessness services. Potential limitations are its focus on 'soft' rather than 'hard' outcomes (for example, it would not provide a count of status change in respect to clients' housing situations), and the limited depth of measurement in any one domain. However, there may be potential to complement the 'soft' measures included in the tool with additional measures that help to fill out the picture of change for individuals.

Micah Projects, Brisbane

Micah Projects in Brisbane began piloting the Homelessness Star in 2009 (Stevens 2009). Initially the tool was trialled with one small team, with positive reactions from staff. Training in the use of the tool was provided for most of Micah's staff in the second half of 2009. Micah also intended to trial adapting the tool for use with homeless parents, adding an additional domain around parenting.

For further research: benefits and challenges of implementation of tool at Micah Projects

3.6 NATIONAL ALLIANCE TO END HOMELESSNESS

The United States-based National Alliance to End Homelessness (NAEH) strongly support outcomes measurement and have produced a Toolkit to support government and community organisations in measuring effectiveness outcomes (Spellman and Abbenante 2008). The approach is not prescriptive and does not provide a particular measurement tool.

Some features of the NAEH approach include (Spellman and Abbenante 2008; Barr 2009):
- Use of a program logic model to place desired and actual outcomes within the context of the problem or need that the program is designed to address, and the activities and outputs of the service
- A clear distinction between outputs and outcomes
- Analysis of outcomes within three distinct time frames (short-term, medium-term and long-term)
- A clear definition of the target population for a particular desired outcome (certain goals may only apply to a subpopulation such as persons with disabilities, or to those who have achieved earlier goals)
- A specific formula for calculating an actual outcome: the number of persons who achieved the desired goal, divided by the total number of persons in the target population (i.e. percentage of population achieving the goal)
- Setting meaningful outcome targets.

The NAEH approach is designed for use by funders and communities as well as by service delivery organisations; it includes suggested methods for comparing outcomes across services (including risk adjustment approaches), and for examining population-level changes.

A strength of the NAEH approach is that it provides clear mechanisms for aggregation and comparison of results; a weakness is that measuring outcomes solely in terms of percentage of target population achieving a goal can miss progress that clients make in other areas, and can obscure differences in the degree of change on particular measures. In addition, setting targets and comparing outcomes across services are often highly political processes, which may have unintended side-effects.

3.7 BENEFITS AND CHALLENGES OF A SECTOR-BASED APPROACH

One of the questions that hovers around outcomes measurement in the homeless sector is the extent to which it is possible to have an integrated approach to outcome measurement across the sector. There are potential benefits to an integrated approach, specifically:
- The ability to aggregate data to service system levels and to reflect on the effectiveness of the service system as a whole
- The ability to compare the performance of services and agencies, and to thereby gain an insight into factors influencing effectiveness
- Minimising data collection impositions on individual agencies and clients.

However, there are also complexities associated with implementing a sector-wide approach; these are briefly discussed here.

Developing common outcomes and measures

It is unclear to what extent desired outcomes are common to all stakeholders (services, funders, clients) within the homelessness sector, and to what extent they differ across service types and client groups. Baulderstone and Talbot (2004: 37) found that no one outcomes tool was applicable to all SAAP services, due to the diversity of service type and contexts across Australia. Even within a single organisation there can be tensions between standardisation versus customisation of tools and measures. These can reflect tensions between the needs of the organisation as a whole and the needs of its sub-units.

Putting aside the tools themselves, it is unclear to what extent different expressions of desired outcomes reflect differences in language, as opposed to differences in conception of service

goals. While there are certainly philosophical and political implications to the way outcomes are described, practice wisdom also suggests that at a practical level, service providers would be able to find areas of agreement about what constitute desirable as opposed to undesirable outcomes. One of the problems with such a process, however, may be that 'consensual' definitions of desired outcomes may be high-level and generic, making them difficult to operationalise. Another problem is that services may have different understandings of the process of change, leading to conflicting ways of describing the steps involved in progress (MacKeith 2007: 5).

A separate but related question is the feasibility of collecting particular core data items across all homelessness services, or service types (therefore providing a set of common core outcome measures). The more of a 'stretch' it is for service providers to collect data, the lower the reliability of the data is likely to be.

Exchange of outcomes information

One of the problems that can arise as more services begin to collect outcomes data is the duplication of data collection by multiple services. This particularly impacts on clients linked to a range of support services – for example, a client may have an outreach support worker, a clinical case manager, be linked to a community health centre, use a day program and receive occasional assistance from a crisis accommodation provider. If each of these services is gathering outcomes data, with no coordination between services, the client may be 'bombarded' by measurement requests using the same or different instruments (Love and Scott 2010). In addition to the annoyance caused to the client, this may well undermine the reliability of the data collected. The duplication of data collection also represents

wastage of resources within the service system.

Ideally, it would be possible for outcomes data to be collected by services who are best placed to obtain the information (i.e. services who have the expertise, the resources and the appropriate quality of relationship with the client), and shared appropriately with other services who are working in partnership with that client. This would fit with joint case planning and review processes (Hamilton 2010), and would also potentially allow responsibility for outcome measurement to be shared between services. For example, a clinical mental health service might collect one set of measures while a Psychiatric Disability Rehabilitation and Support provider would collect a different set.

Exchange of information in such a system would need to be governed by the informed consent of the client. Clear processes for seeking and recording this consent would need to be developed.

Burden of proof

MacKeith (2007: 5-6) differentiates the use of outcome measurement to evaluate the effectiveness of service provision models or interventions from its use by individual agencies to monitor their success. She argues that the "burden of proof" currently falls on each individual service to demonstrate that their intervention type is worth funding. Instead, she proposes following the health service model of clinical trials to assess the effectiveness of an intervention. Once the trials are complete, the value of the intervention is established and there is no requirement for individual services to prove the value of the intervention. Instead, agencies can focus on monitoring their own success compared with established benchmarks.

While this proposal has merit, one of the difficulties with it is that the robustness of generalisations about interventions in the homelessness field tend to be lower than those in the health field, owing to the types of research designs that can reasonably be pursued in homelessness settings.

4. Measurement of client outcomes as a research activity

Client outcomes measurement systems are usually intended as practical performance measurement and program management tools, rather than as vehicles for theoretical research. Nevertheless, they do involve gathering and analysing data to create new knowledge or insight, and therefore can be viewed as a form of research.

It is useful to consider outcomes measurement systems through the lens of research design for several reasons:

- To identify choices about the types of outcomes data that are desired, and how these may be gathered and analysed
- To identify ways in which measurement systems can be designed so as to provide the most reliable and valid data
- To identify the types of knowledge or insight that outcomes measurement systems can typically produce, and the level of certainty of the conclusions drawn. This may include understanding the limitations on the types of inferences that can reasonably be supported by the data.

Outcomes measurement also has a relationship with program evaluation. This section considers these relationships and explores the types of findings typically produced by outcomes measurement systems.

4.1 MONITORING AND EVALUATION

In general terms, evaluation refers to "the identification, clarification, and application of defensible criteria to determine an evaluation object's value (worth or merit) in relation to those criteria" (Fitzpatrick, Sanders et al. 2004: 5). Program evaluation typically involves the application of social research procedures to systematically develop a valid description of particular aspects of program performance, and the comparison of the performance to

selected criteria (Rossi, Freeman et al. 1999: 20-21). Some commentators distinguish evaluation from research on a range of considerations including the purpose of the study, the relationships of stakeholders to the study, and the types of standards used to judge the adequacy of the study (Fitzpatrick, Sanders et al. 2004: 6-7). Within the field of program evaluation, a wide spectrum of approaches can be used (e.g. Wadsworth 1997: 77-109).

Relationship between monitoring and evaluation

Measurement of client outcomes can be considered as an evaluative activity (Kimberley 2009) – that is, an activity that is designed to provide information that may assist in assessing the value of an intervention or program. However, there can be major differences in the way that this evaluative activity takes place in an organisational context. The key distinction that is drawn within the literature is between monitoring of client outcomes (as part of an organisational performance measurement system) and evaluation of client outcomes as part of a formal program evaluation.

Monitoring typically involves systematic, periodic collection and analysis of data to assess performance in relation to an agreed standard set of indicators. Monitoring systems are usually designed to be ongoing rather than time-limited. Monitoring provides succinct, regular feedback that can assist with accountability, quality improvement and responding to evolving trends in the environment.

Evaluation involves episodic, in-depth collection and analysis of information. Evaluation can draw on a broad range of data sources and methods, and can examine factors that are too costly or difficult to continuously

monitor. Evaluations are typically customised to the context of individual services and the time-specific needs of particular stakeholders. Evaluations can explore how and why certain effects occur and is suitable for exploring issues such as long-term impacts, causal attribution, cost-effectiveness and comparative effectiveness (McDavid and Hawthorn 2006: 293; Segone 2008: 101-103).

The borderline between evaluation and monitoring is not clearly defined and outlying forms of each approach may resemble the other. A detailed and sophisticated monitoring system that includes a wide range of performance measures and supplements this with a significant amount of qualitative data may in effect be an ongoing evaluation process. However, for most homelessness organisations such a system would be unaffordable. On the other hand, an effectiveness-focused evaluation with limited resources and a quantitative methodology may resemble a time-limited performance measurement system.

Some sources see both routine outcomes monitoring and in depth ad hoc evaluation as subtypes of program evaluation (e.g. Hatry 1997: 3-4). Some have argued that performance measurement can fulfil many of the same purposes that program evaluations have served, and that in-depth program evaluation is therefore an "expensive luxury". However, there are important differences between the typical forms that monitoring systems and program evaluations take, and the level of information that they are able to provide (McDavid and Hawthorn 2006: 293). The term evaluation is used in this paper for in-depth, episodic program evaluations in contrast to ongoing monitoring systems.

It is worth noting that the term monitoring evaluation is also used in the literature, however it typically refers to process-oriented program evaluations that focus on how the program is being delivered, rather than on outcomes (Fitzpatrick, Sanders et al. 2004: 21).

Advantages and disadvantages

Hatry (1997: 4) provides a useful summary of the relative advantages and disadvantages of evaluation and monitoring, which is adapted as Table 4 below. Monitoring is used to track performance against a limited set of measures, and for early identification of trends in the external or internal environment. Evaluation is used to guide significant organisational decisions or gain a deeper understanding of phenomena identified through monitoring.

Table 4: Advantages and disadvantages of monitoring compared to evaluation (adapted from Hatry (1997: 4) Table 1-1)

Type	Advantages	Disadvantages
In depth, ad hoc evaluation	• Identifies causes of outcomes (to some extent) • Can provide relatively strong evidence on outcomes and effect • Can provide guidance as to improvement action	• High cost • Covers few services • May take extended time to get results
Regular outcomes measurement	• Covers many or most agency programs • Provides information on a regular basis • Lower cost per program covered • Hints at improvement actions	• Provides little information on causes of outcomes • Provides little guidance on improvement actions • Subject to a variety of interpretations

For greatest benefit, organisations use a combination of ongoing monitoring and episodic evaluation. Data from monitoring can contribute to evaluations, by providing historical data streams (Hatry 1997: 4). Program evaluations provide a much richer understanding of context and of the factors that may be impacting on client outcomes.

Evaluation designs may also influence monitoring systems – for example, an evaluation might collect measures on particular domains, which are then partially carried on by the organisation in routine outcomes monitoring. Evaluations may also identify particular areas of strength or weakness which an organisation may wish to monitor in an ongoing way.

4.2 NATURALISTIC AND EXPERIMENTAL APPROACHES TO RESEARCH

There is a long-standing and sometimes fierce debate in the social sciences (and in science more generally) about which approaches to research are the best for studying and understanding the world. This "paradigms debate" is linked to philosophical positions about the nature of reality, truth, and of what constitutes credible evidence (Patton 1987: 165).

Experimental approaches to research are typically associated with the positivist tradition, which emerged from the physical sciences. These approaches tend to aim for objectivity, emphasise separation of the researcher from that being researched, and assume value-neutrality. They often rely on and valorise quantitative data (Patton 1987; Wadsworth 1997: 165). Within the evaluation literature there are significant critiques of this paradigm, which argue that subjective and objective meanings are socially constructed, are central to understanding human phenomena, and that all research is value-driven (Wadsworth 1997: 101).

In contrast, naturalistic approaches to research are typically associated with phenomenological and constructivist philosophical positions. These approaches do not attempt to artificially manipulate the environment or the phenomenon being studied, and typically proceed inductively from practice to theory. Naturalistic approaches aim to describe and understand naturally unfolding processes, including variations in experience, and tend to emphasise qualitative data (Patton 1987: 13-15; Wadsworth 1997: 95-96). Naturalistic approaches have been criticised at times for a perceived lack of objectivity (Patton 1987: 166).

At times these approaches and their associated methods have been characterised as fundamentally opposed and irreconcilable. However, more recently authors have argued that both approaches are useful and that the key challenge is to match appropriate approaches to particular research or evaluation questions and issues (e.g. Patton 1987: 169). Most evaluations use several designs or combinations of designs to address different questions. Mixed method designs that combine qualitative and quantitative approaches often yield richer insight and can increase validity (Fitzpatrick, Sanders et al. 2004: 263, 305).

The large majority of literature on client outcomes measurement assumes a quantitative methodology, and by association imports much of the focus of the positivist tradition on reliability and validity. A few sources do acknowledge the usefulness of qualitative data (e.g. Burns and Cupitt 2003), although generally as an adjunct to quantitative measures. This overwhelming focus on quantitative methods may reflect the informational needs that tend to drive monitoring systems (e.g. senior management desire for oversight of results for an entire cohort of clients), as well as the genesis of outcomes monitoring in governmental and business performance measurement systems (McDavid and Hawthorn 2006: 282-288).

Qualitative methods clearly have the potential to offer considerable insight into client outcomes, especially in programs in which outcomes are expected to be highly individualised rather than standardised. Qualitative methods have strengths in capturing diversity and subtlety of experience and of outcomes (Patton 1987: 24-30). 'Open inquiry' approaches may be of particular value in the early stages of developing outcomes measurement systems, when the range of

possible intended and unintended outcomes are being explored. The Most Significant Change approach (Davies and Dart 2005) is particularly relevant to this context. However, qualitative approaches will be less appropriate where the informational need is a broad understanding of overall patterns of outcomes across a large group of people.

Given the focus of the outcomes measurement literature, the bulk of this literature review focuses on quantitative methodologies for monitoring outcomes of large groups of people.

For further research: explore possible ways in which qualitative and naturalistic approaches could be used as the basis of an outcomes measurement system

4.3 MONITORING SYSTEMS AS SINGLE-SYSTEM RESEARCH DESIGNS

The literature distinguishes 'classical' experimental designs from quasi-experimental designs (including single system designs) (Cook and Campbell 1979: 4-6; Bloom, Fischer et al. 2006: 44-49):

- Experimental designs involve at least two groups of participants (a 'treatment' group and a 'control' group), with different treatment provided to the two groups. Random selection and random assignment are used to attempt to obtain equivalence (on average) between the two groups on all relevant characteristics. Data is aggregated within these groups and analysis of comparisons between the average scores of the groups are undertaken.
- Quasi-experimental designs depart from the experimental model in that they do not use random assignment to create the comparisons from which treatment-

caused change are inferred. Instead the comparisons depend on individual or groups for which equivalence either is not established, or is known to be absent.
- Single-system designs are those in which the comparisons are for one individual, group or collectivity, at different points in time (at a minimum, before and during application of the treatment).

It is often not feasible to use true experimental designs in human services evaluations, for a variety of reasons. In particular, there may be ethical issues and risks related to withholding treatment from a 'control' group, although in certain situations where demand outstrips supply a control group may be possible (Cook and Campbell 1979: 347-350; 374). In a monitoring system, with limited resources, it is very unlikely that an experimental design would be attempted.

Classic outcomes measurement systems are most closely aligned with single-system research designs. They share the defining characteristic of single system models: the planned comparison of a nonintervention ("baseline") period with observations of intervention period(s) or in some cases, a post-intervention period, for a single client or system (Bloom, Fischer et al. 2006: 322).

Within the area of single system designs, again there are a wide variety of different designs (Bloom, Fischer et al. 2006: 352). The basic single-system design is A-B (where A represents the baseline phase and B the intervention phase). More complex designs involve replication of original conditions and/ or treatment periods (e.g. A-B-A-B designs); others may involve successive or alternating interventions (e.g. A-B-A-C or A-B1-B2-B3 designs, where the subscripts represent varying intensity of the same intervention).

In general, it is suggested here that an outcomes monitoring system should be considered as tracking a large number of parallel instances of a basic A-B single system design. Even though services provided by homelessness agencies may often have multiple intervention components that might be applied at different times, these different components are not necessarily applied in isolation or in any particular planned sequence. In individual casework situations, it may be possible to discriminate different phases of service delivery, and therefore view the casework process with a single client as one of the more sophisticated designs with different intervention phases. However, from the aggregated perspective of a monitoring system, these different stages of interventions will tend to be invisible and the 'intervention phase' needs to be considered as a single B phase. The withdrawal of intervention (return to baseline) which is characteristic of some of the more powerful single system designs (such as experimental replication or successive intervention designs) would usually be seen as unethical or wasteful of resources in the context of homelessness service delivery.

It is possible that routine outcomes measurement might be considered as a multiple baseline design across clients, i.e. application of the same intervention at different times to different clients, comparing the results across clients. However, technically multiple baseline designs require that a stable baseline be obtained for all of the target problems, clients or settings, and that an intervention is introduced to only one of the targets at a time while the baselines of the other targets are continued (Bloom, Fischer et al. 2006: 421). This is clearly not the case in routine service delivery, where there may be no temporal overlap at all between baselines and service provision for different clients.

4.4 LIMITATIONS OF MONITORING SYSTEMS

A relevant concern of all forms of research is to understand the level of 'trustworthiness' or 'credibility' of the research findings. The traditional scientific paradigm tends to concentrate on reliability and validity as the two key measures of trustworthiness. Critiques of this paradigm suggest other indicators and techniques for demonstrating trustworthiness (Wadsworth 1997). Regardless of the paradigm, different research designs may lead to differences in the level of certainty or credibility of the findings, and therefore differences in the strength of the arguments that can be advanced based on the research (Schalock 2001: 68).

Given the primarily quantitative focus of outcome monitoring systems, it is useful to consider issues relating to the design validity of these systems – their ability to validly answer questions about change, causality and generalisability. (Validity of individual measures and tools is a related but separate issue and is discussed in Section 6). Table 5 summarises three fundamental questions that outcomes measurement might try to answer, and the types of validity that are relevant to these questions (Bloom, Fischer et al. 2006: 338-346). One way of conceptualising this is to ask whether the data provided by outcomes measurement is sufficient to reject the "null hypothesis" - i.e. the hypothesis that no change occurred, or that intervention and outcome are unrelated (de Vaus 2002). More powerful research designs are more able to exclude alternative explanations for their findings.

Table 5: Key questions that research may try to answer, and associated validity considerations

Key questions	Validity considerations	Threats to validity
Did change occur?	*Statistical conclusion validity* Are there sufficient grounds to believe that dependent and independent variables covary?	• Unreliable measures • Inconsistency in implementation of intervention • Random changes in the intervention setting • Small numbers of observation points
Was the change cause by the intervention?	*Internal validity* Are there sufficient grounds for ruling out the hypothesis that one or more extraneous variables influenced or produced the observed changes?	• Unrelated events occurring during the intervention period • Maturation – independent physical or psychological processes within the subject • Effect of previous testing • Instrumentation - changes in measurement tools or their uses • Attrition in sample • Regression of initial extreme scores to the mean • Diffusion or imitation of intervention

Key questions	Validity considerations	Threats to validity
	Construct validity Are there sufficient grounds to believe that the research is actually measuring what it intends to measure?	• Mono-operation and mono-method bias – use of only one measure and/or way of recording per target • Hypothesis guessing by participants • Valuation apprehension – people not representing themselves accurately • Participants able to perceive practitioner expectations • Interaction of multiple interventions
If the same intervention was applied under different circumstances, would the same change occur?	*External validity* Are there sufficient grounds to believe that the effect of the intervention can be generalised to other populations (population external validity) or other settings or environments (ecological external validity)?	• Interaction between the intervention and other variables related to client or environment • Individual styles of practice • Different conceptions of target variables in different studies • Interaction of history and intervention • Measurement differences between studies • Differences in characteristics of client group (representativeness of sample) • Interaction of testing and intervention • Reactive effects to evaluation – awareness by participants that they are involved in a study

Single system designs vary in their ability to deal with the 'threats to validity', and hence in the strength of the conclusions that they may support. A basic single system design (A-B), if well implemented has the potential to answer the first key question ("did change occur"), including identifying how likely it is that changes could have occurred by chance alone. However, basic single system designs are generally not powerful enough by themselves to provide strong evidence about causality, or to support generalisation of findings (Stedman, Yellowlees et al. 1997: 20; Bloom, Fischer et al. 2006: 351, 377).

One of the advantages of monitoring systems is that over time they potentially provide data on outcomes for a large cohort of clients. Clinical replication involves the repeating of an intervention package, applied by the same practitioner or group of practitioners to a series of clients in the same setting, with multiple problems that generally cluster together (Bloom, Fischer et al. 2006: 347). In some circumstances outcomes monitoring could be viewed as clinical replication; this would depend on the level of similarity of client characteristics, presenting problems, and the degree to which the interventions are applied uniformly and consistently across clients.

Successful replication strengthens the case for generalisability of results. It is unclear if it strengthens arguments for causal connections between interventions and effects.

Outcomes monitoring systems in service delivery settings have some particular limitations that further affect the strength of the conclusions that they support. These limitations are summarised below.

Attribution

Schalock (Schalock 2001: 141) points out that outcomes in the human services are integrative - that is, they reflect the contributions of all who provide care. Service delivery partnerships are a necessary part of doing business in the sector, and are widely seen as an important aspect of best practice (e.g. Department of Human Services (2006: 3.2.7, 3.2.9)). However, it can make it very difficult to disentangle the contributions of multiple practitioners and multiple services, especially in a monitoring system where (unlike in a program evaluation) there will generally not be sufficient data or time to enable in-depth consideration of other environmental factors.

Many authors have identified this problem. Rossi (1997: 25) distinguishes 'gross' from 'net' outcomes. Gross outcomes are the total change in the target problem, consisting of the effects (if any) of the program plus all of the other processes, including other interventions, the client's own independent processes, and environmental factors, that led to such outcomes. Net outcomes are those that would not have occurred without the program. Unfortunately, within a monitoring system it is usually difficult or impossible to establish the net effect (Baulderstone and Talbot 2004: 3-4).

HomeGround
Services

There may be ways to modify the design of the outcomes monitoring system so as to reduce the probability of alternative explanations for change:

- Collect and analyse additional quantitative data that indicates other potential causal factors operating in consumers' situations (Stedman, Yellowlees et al. 1997: 20). For example, there may be ways to analyse information on the number and types of support links that clients have to gain an understanding of the possible other service contributions to an outcome. If it is known that a client is not engaged with other support services, this rules out at least some alternative explanations for outcomes.
- Assess clients' and practitioners' opinions concerning potential change and the meaning of change scores (Stedman, Yellowlees et al. 1997: 20). For example, it may be possible to ask clients to what extent they think any change resulted from their involvement with the service, as opposed to external factors.
- Triangulate quantitative data from the monitoring system with other forms of data, including qualitative data that provides insight into program processes and staff and client perceptions of change. For example, the use of case studies in conjunction with monitoring data can greatly deepen the understanding of the processes that have led to observed outcomes.
- An immediate pretest-posttest design (Bloom, Fischer et al. 2006: 386), in which measures are taken immediately before and after a single session of intervention, can reduce the likelihood that other factors intervened in the causal relationship between intervention and effect. While this approach clearly has limited applicability (to interventions that

can be expected to have an immediate effect – perhaps providing material aid or emergency housing), it may be useful in some situations.

McDavid and Hawthorn (2006: 62-62, 361) discuss the concept of the 'core technologies' used by programs. They comment that human services programs tend to use 'low probability technologies' in which our knowledge of what works is often unclear, and in which results are heavily influenced by a wide range of variables which it is impossible to control. The corresponding argument is that attribution will always pose more challenges with low probability technologies. However, even within a single program it may be possible to discern a range of 'technologies' within use, with varying levels of probability.

For some categories of outcome (especially interim 'hard' outcomes concerning status change), with sufficient knowledge of the service it may be reasonable to conclude that the service was a necessary (if not the only) contributing factor to the outcome. For example, all clients of a particular service might be provided with Transitional Housing (as for example in HomeGround's Justice Housing Support Program). For clients that are homeless at assessment and housed in Transitional Housing during engagement with the program, the program can clearly claim a causal contribution to that interim outcome.

Other situations in which claims of causality may be more solid include:

- Situations where there is an existing body of evidence that demonstrates the effectiveness of the particular intervention or program being monitored
- Situations where the intervention is situated within a well-established theoretical framework and other aspects of the framework are supported by existing evidence.

In these cases, if outcomes data is consistent with the results expected according to pre-existing evidence or theory, this may bolster the argument that the program is operating as expected and that there is some causal connection between activities and results.

Another way to view questions of attribution may be to see the combined work of all services involved with a client as the 'intervention'. From this viewpoint, the issue of the differential contribution of various services become irrelevant. The attribution question then becomes, with what level of certainty is it possible to claim that change was caused by the actions of the service system, as opposed to other chance or environmental factors (economy, political changes, the client winning the lottery, ...). While in one sense this approach reduces the number of alternative explanations that must be considered, stakeholders may view it as unsatisfactory for two reasons:

- it still does not allow a well-founded claim that the service system caused the change, as opposed to other factors
- it does not meet the needs of individual organisations who wish to know how effective their programs are.

Depth of understanding of intervening variables

Intervening variables are those factors that affect the process of change, between inputs and outcomes. Weiss (1972: 48-49) categorises intervening variables into:

- program-operation variables – factors concerning the way particular interventions or services are provided, for example, frequency of contact, continuity of staff providing service, extent of coordination of services, etc.

- bridging variables – intermediate milestones or interim outcomes that (according to the explicit or implicit theory of the program) are necessary to achieve the desired end result; for example, change in attitude might be seen as a bridging variable for a program attempting to produce behavioural change.

Outcome monitoring systems may have limited information to draw on to assess the effects of intervening variables. This may be because only limited data is able to be collected or analysed on programs-operation factors; likewise, it may only be practical to collect data on a few (or no) interim outcomes. A related issue is that research designs (whether experimental or naturalistic) require the clearest possible definitions of the intervention, so that it is clear what was actually done and how it might be replicated. However, in a service delivery context, interventions may be quite amorphous, flexible, and/or may fluctuate in response to a wide range of factors over time and between staff members (Bloom, Fischer et al. 2006: 322, 329).

Careful selection and use of process measures in relation to quality of service delivery may allow some understanding of how program-operation variables affect outcomes.

Shortened research phases

In general, quantitative research designs support more robust conclusions about whether change occurred, and what caused the change, when the periods of data collection at baseline and during intervention are of sufficient length (Bloom, Fischer et al. 2006: 327). Sufficient length may be interpreted as involving a sufficient number of data points to provide a clear picture of a relatively stable pattern of events or trend. While there is no absolute number of data points required, in general three points are the minimum necessary to establish any sort of clear pattern, while having ten or more points enables better use of statistical techniques such as tests of significance (Bloom, Fischer et al. 2006: 364). For greatest validity it is ideal if baseline and intervention phases are roughly equal in length, as this allows external and internal factors to influence behaviours in baseline and intervention phases equally (Bloom, Fischer et al. 2006: 327).

In service delivery situations, especially for longer-term case management services, it is highly unlikely that baseline and intervention phases would be of equal length. Further, it is often the case that only a single baseline data point is available (at point of initial assessment or intake). It would generally be unethical to withhold service delivery until a stable baseline data picture can be established, especially where there was some imminent threat to the safety or health of the client or a third party. Clients are often in crisis at intake, and there is pressure to commence intervention immediately (Bloom, Fischer et al. 2006: 365). It should be noted, however, that for many longer-term services, service delivery does begin with a period of assessment and relationship establishment and this may provide an opportunity for collecting baseline data. In other cases, service delivery may only occur over a relatively short timeframe, and there may be only one data point available in the intervention phase also. Given only a single data point, it is impossible to tell from this information alone whether it represents the typical condition of the client at that time or not. For this reason, multiple assessments over time are preferred (Stedman, Yellowlees et al. 1997: 20).

It may be possible to obtain additional data points for the baseline phase using a reconstructed or retrospective baseline (Bloom, Fischer et al. 2006: 365). This may involve asking the client for their estimates or recollection of their condition over the months preceding intake. Alternatively, it might involve gathering historical data from other services involved with the client. Another possibility, for those services using waiting lists, would be to regularly collect assessment data from clients while they are on the waiting list. However, all of these options require additional data collection resources which may be unavailable in high-pressure service environments.

Follow up

Many authors emphasise the importance of gathering 'follow-up' data after the intervention (service delivery) has ended. This allows an understanding of whether any changes observed during engagement have been ongoingly maintained in the client's life, and hence whether the desired medium or long term outcomes have occurred (Bloom, Fischer et al. 2006: 337). This is particularly important given evidence that gains made by clients in human services programs may not be sustained (Rapp and Poertner 1992: 107). Booth and Smith (1997: 42) suggest that a follow-up rate of 80% of the original sample is required for generalisability of findings to the original population.

However, there are many logistical and resourcing challenges involved in follow-up of clients, especially those who are transient (Post, Isbell et al. 2005: 12). For short-term services (such as many SAAP services), the transitory nature of engagement further reduces the possibility of follow-up (Baulderstone and Talbot 2004: 3). The limited follow-up possible within an outcome monitoring framework limits conclusions about the stability of changes achieved during engagement.

Conclusion

Questions of causality are notoriously difficult in research in the social sciences. Even with the most robust research designs, conclusions about causality are still generally a case to be argued on the balance of probabilities, rather than a fact than can be definitively proven. Given the limitations of outcome monitoring systems, it is unrealistic to expect that these systems in themselves can support strong arguments for program effectiveness. However, when combined with other methods and sources of data they may form one plank of an argument for effectiveness.

It is worth remembering that in general it is not necessary or appropriate to try to demonstrate that the work of a particular program or agency was the sole cause of a client outcome. Apart from anything else, this ignores the central role played by the client in achieving change. It will generally be more appropriate to try to show that the program made a contribution to the outcome observed. However, even this lesser claim may be difficult to argue persuasively.

When well designed and implemented, outcome monitoring systems are potentially capable of providing valid and credible information regarding changes occurring for clients. This is valuable information for planning service delivery at the individual and program level.

It is also important to recognise that data does not have to achieve academic standards of rigour to be useful (Friedman, DeLapp et al. 2001b: 2.7). No matter what data collection techniques are used, there will always be limits on the reliability and validity of outcomes data. Within these constraints, outcomes measurement can still provide a useful gauge of what is happening for clients.

5. Conceptualising client outcomes

5.1 GENERAL DEFINITION

In its most general sense, the term outcome refers to a condition that arises from an action or set of actions. In the field of human services the term 'outcome' tends to be used in a more focused sense: a personal or organisational change or absence of change that results from an action or set of actions carried out by staff of the organisation (Schalock 2001: 3-4; Baulderstone and Talbot 2004). Similarly, Burns & Cupitt (2003: 6) define outcomes for homelessness services as "... the changes, benefits, learning or other effects that happen as a result of your activities."

This use of the term outcome is neutral with regard to whether the change or effect is regarded as positive or negative. An outcome of service delivery may be an improvement or decline in a client's wellbeing. In contrast, the term desired outcome can be used to describe those effects that a stakeholder positively values and aims to produce.

Effort and effect: organisation-focused vs. person-focused outcomes

The term outcome tends to be used in two different ways in outcomes measurement literature related to the human services. The distinction might be termed the 'broad' and 'narrow' uses of the term.

In the broad view, outcomes refer to a range of organisational and personal measures that reflect the quality of the organisation's service delivery across a range of dimensions including quantity and type of services provided, cost, satisfaction (of clients, staff and other stakeholders), equity, timeliness, relevance to client or community need, effectiveness, and other dimensions (Berman and Hurt 1997: 82-83).

In the narrow view, outcomes refer to effectiveness only - i.e. the change that occurs with respect to the client's problem that is the target of intervention. This may also be referred to as 'clinical outcomes' (Berman and Hurt 1997: 82-83).

The two different ways of using the term outcome are linked to the observation that effectiveness outcomes are only part of a larger picture of an organisation's performance, and that attention to inputs and to process is equally important to ensuring that an organisation is 'doing a good job' by providing quality services (Weiss 1972: 46; Planigale 2010b). Schalock (2001: 10-11, 18-19) provides a framework for outcomes (in the broad sense) which includes four quadrants, summarised as below. The distinction between performance measurement and value assessment in this model reflects a difference between 'hard' outcomes (role status and behaviour) vs. 'soft' outcomes (quality of life and personal experience). It should be noted that most authors use the term performance measurement more loosely to encompass any measure that provides an indication of the quality of the organisation's work.

Figure 1: Outcome categories as per Schalock (2001)

		Evaluation standard	
		Performance measurement	Value assessment
Evaluation focus	Organisational outcomes	*Organisational performance outcomes:* • Service coordination • Financial stability • Health and safety • Program data • Staff turnover	*Organisational value outcomes:* • Access to services • Customer satisfaction • Staff competencies • Family / consumer supports • Community support
	Individual outcomes	*Individual performance outcomes:* • Physical wellbeing • Material wellbeing	*Individual value outcomes:* • Quality of life • Emotional wellbeing • Personal development • Self-determination • Interpersonal relations • Social inclusion • Rights

Outcomes in the narrow sense (clinical effectiveness) are located in the bottom two quadrants.

HomeGround Services

Another way that clinical outcomes are distinguished from the broader picture of organisational outcomes is through the terms 'effort' and 'effect' (Friedman, DeLapp et al. 2001b: 1.1). Effort refers to the quantity and quality of service delivery activity, while effect refers to quantity and quality of change for the better that was produced by the activities. Similarly, Hudson (1997: 70) distinguishes outcomes with respect to the problem from outcomes relating to professional behaviour or activity of the organisation or its employees.

For the purposes of this paper, outcomes will be used in the narrow sense, i.e. to refer to changes in the presenting problem.

Connection between effectiveness outcomes and other organisational outcomes

Following the above discussion, performance measures such as client satisfaction or accessibility of services are not defined as outcome measures for the purposes of this paper. However, these aspects of service quality may still have some connection with effectiveness. In particular, there is some evidence that particular service delivery behaviours may tend to result in better outcomes than others (Rapp and Poertner 1992: 121-122). It may therefore be useful to use productivity measures as a proxy for outcomes in some circumstances (for example, in short-term services where it is very difficult to get follow-up data), although the assumed relationships between outputs and outcomes also need to be tested. Productivity (activity or output) measures can also help to measure fidelity in implementation of a service model (Schalock 2001: 22) and to identify opportunities for improvement in practice.

Client satisfaction measures, while useful to some extent as an indicator of service quality, are of little use in assessing client outcomes (and may in some cases actually be markers of poor outcomes such as dependence on a service). They are often positively skewed (Segal 1997: 155-156) and may be susceptible to validity problems depending on the scales used (Ware 1997).

Outcomes for groups and third parties

Some services may aim to achieve outcomes that are changes not in individuals but in larger groupings of people - for example, families, communities, other organisations (Burns and Cupitt 2003: 6-8). While the distinction between effort and effect is still relevant for work with these larger groups, the definitions of the types of outcomes that may occur becomes more complex. There are well-tested outcome definitions and measures for work with families, and to some extent community outcomes may be measureable through a census or population sampling approach.

For further research: outcomes and tools for groupings other than individuals, including tools for measuring changes in organisations

Population, program and individual outcomes

While homelessness agencies typically work to achieve positive outcomes for individual clients (or households), these individual changes also contribute to a larger picture of change at an agency, program, or community level (Weiss 1997: 175). This is not an issue about different types of outcomes, but about the way that outcomes are aggregated so that change (or lack of change) can be observed in larger populations.

Friedman, DeLapp et al. (2001b) distinguish results accountability and performance accountability:

- results accountability is accountability by the community (city, county or state) to the community (city, county or state) for the wellbeing (results) of a population (children, adults, families, all citizens...)

- performance accountability is accountability by managers to stakeholders for the performance of a program, agency or service system.

The distinction is crucial to the implementation of outcomes measurement in individual organisations. It prevents staff at an agency level from feeling that they are responsible for population outcomes (which individual staff cannot hope to impact on at a global level), while also making clear the contribution that an agency makes to population results (Cunningham-Smith 2010).

Some population results may reflect the direct aggregation of individual results. For example, an individual client achieving a better level of income may contribute to an agency percentage of clients with sufficient income, and to a population indicator around level of poverty in the community. Individual results may also contribute indirectly to other population-level results such as more appropriate use of welfare services, lowered crime rates, and so forth (Blunden and Johnston 2005: 1).

Change and maintenance

While outcomes are often thought of in terms of change, it is important to recognise that at times the desired outcome may seem to be a lack of change – i.e. maintenance of a person's current situation (Burns and Cupitt 2003: 6). An outcomes measurement system needs to allow for maintenance goals as well as change outcomes. However, maintenance goals may often be accompanied by a goal of reduction of risk, improvement of prognosis or avoidance of future negative consequences. Even a maintenance goal may therefore have some change indicators associated with it.

Intended and unintended outcomes

While human services programs are established with intended outcomes, they can also have a variety of unintended outcomes. These unintended outcomes may be positive or negative (and this may vary from the perspective of different stakeholders). The unintended outcomes may not necessarily be unexpected – i.e. they may be predictable by those with expertise in the field – but they may be 'side-effects' of the operation of the program. Rossi (1997: 23) gives the example of introduction of improved income support benefits for people in particular categories; an expected but unintended consequence was reduction in work effort. Blunden and Johnston (2005: 1) make a similar point in relation to provision of public housing. In some cases, the unintended consequences may be 'flow-on' effects to other parts of the system that the individual is a part of. For example, changes in the wellbeing of homeless people (a primary goal of homelessness services) may lead to different patterns of usage of a variety of other parts of the service system (Rossi 1997: 22).

By definition, unintended consequences will tend to be more difficult to detect, particularly in a monitoring (as opposed to evaluation) environment where the number of measures that can be tracked is limited. However, this does speak to the value of having a range of wellbeing measures in place rather than narrowly focusing measurement only on the specific goals of the program. In medical terms, just because a treatment is effective in controlling a particular condition, this does not mean it has improved a patient's overall quality of life (Ware 1997: 52-53).

Short, medium and long term outcomes

The question of stages of change is important to outcome measurement for a number of reasons. It is expected that different outcomes will take different lengths of time to occur, either because of the internal process that may need to occur within the individual who is changing, or because of the external process that needs to occur within the individual's environment.

One approach in the literature is to use a set of temporal durations to separate out sets of outcomes. For example, Spellman and Abbenante (2008: 12-13) categorise outcomes as:
- short-term (occurring within the first month of involvement in the program)
- intermediate (occurring within one year of commencement of involvement in the program)
- long-term (after three years or more from commencement of involvement with the program).

These timeframes appear appropriate to the types of outcomes that may occur in the environment of homelessness service provision – for example, some outcomes may well be observed within a month of engagement (provision of material aid), while others will take several years (or more) to 'mature' (e.g. obtaining public housing via an Early Housing Application). These timeframes may be dependent on the type of service provided – for example, in a Housing First approach, obtaining permanent housing would often be seen as a short-term rather than a long-term outcome.

Another approach would be to link sets of outcomes to stages within a theory of change model. Different types of outcomes and levels of outcome stability might be expected in early stages of change as opposed to subsequent stages.

Intermediate or interim outcomes are indicators of progress towards a final desired outcome. For example, if the desired outcome is housing stability for at least 6 months, an interim outcome would be placing a client in appropriate permanent housing (Spellman and Abbenante 2008: 16) – or assisting the client to have their public housing application approved. Given the difficulty of long-term follow-up with clients in some settings, interim outcomes may be feasible to measure and may provide some proxy indication of long-term outcomes. The difference between a short-term outcome and an interim outcome is not always clear – but in a sense a short-term outcome can be regarded as 'complete' whereas an interim outcome is always understood as a step on the way to somewhere else. Other authors use the terms 'proximal' and 'distal' outcomes, where proximal outcomes are similar to interim outcomes, i.e. outcomes that occur on the way to a final state and contribute incrementally to that state: "... it may be possible to measure outcomes that are consistent with [the program's] goal but that occur close enough in time to be practical to use in the measurement of that program's effectiveness and efficiency" (Rossi 1997: 29). Weiss (1972: 48-49) discusses 'bridging variables' – sub-goals on the way to achievement of a final goal – and notes that these bridging variables may affect the ultimate effectiveness of a program.

Interim outcome data can be useful for a number of reasons (Burns and Cupitt 2003: 8):
- it can demonstrate progress on the way to longer-term goals, providing encouragement to both clients and workers
- in situations where end goals may be difficult to reach, it can recognise the valuable contributions of the organisation to steps along the way
- it can help to acknowledge the contributions of different services to resolution of the same problem.

HomeGround
Services

A more subtle question is whether long-term outcomes are based on gradual change over time (linked to constant intervention - for example, in a long-term case management setting), or whether they may result from changes put in place by a relatively brief period of intervention, perhaps years earlier. Although behavioural theorists tend to be dismissive of 'sleeper' effects (Rossi 1997: 29), the example of a public housing application which results in an offer several years later is a practical example of this type of outcome in the homeless setting. It is also possible for medium-term successes to diminish over time, speaking to the value of follow-up measurements to establish long-term outcomes.

5.2 STAKEHOLDER PERSPECTIVES

Human services programs have multiple stakeholders – clients, service delivery workers, managers, funders, service delivery partner agencies, and communities or 'the public', amongst others. While it is easy to assume that everyone shares the same view of the aims of a program, often there are substantial differences in perspective (Weiss 1972: 27). This can be a hurdle in program evaluation, but also affects monitoring systems.

Stakeholder perspectives necessarily impact on decisions about what outcomes to measure, and about the interpretation of results. Whether a given outcome is positive or negative is a value judgement; the effects of programs may therefore be positive in the eyes of some actors and negative in the eyes of others (Rossi 1997: 22). Even where there is agreement on valuation of outcomes, there may be differences in views of priority. Government departments may have an interest in measuring participation in EET activities,

while service providers may see this as a relevant goal for only a minority of clients and may place more priority on physical health and safety.

A related distinction is that between client status and case status variables (Baulderstone and Talbot 2004: 5). Client status variables refer to information which indicates outcomes of direct relevance to clients, such as higher quality accommodation; case status variables refer to information about outcomes of direct relevance to agencies, often pertaining to how a client is categorised within the service system (e.g. removal of a client from an at-risk register).

At times there will be a disagreement between client and societal or program goals. A program may value the outcome of reduction in problematic drug use, while some clients might see an increase in drug use as positive. Involuntary services, such as Child Protection, often involve conflicts between the program and client goals. Rapp and Poertner (1992: 106) note that using ethical analysis to blend apparently conflicting goals is a standard social work skill, applied in practice daily. Blending and/or choosing between conflicting desired outcomes is a necessary activity that is the responsibility of those implementing outcomes systems; however, it can be supported by consultation and by clear articulation of the value basis for particular decisions.

5.3 OUTCOMES AS AN ELEMENT OF PROGRAM LOGIC

Program logic is a systematic way of documenting the connections between the various aspects of a program's operations, and in particular the connections between effort (inputs and activities) and effect (outcomes) (Spellman and Abbenante 2008:

10). Logic models are useful for purposes of communication, shaping of service delivery, and for evaluation, to guide the types of evaluation questions that may be asked and to interpret findings. There are three main types of program logic approaches: outcome approaches (based on an implicit theory of program functioning), activities approaches (documenting in detail the sequence of activities in program implementation), and theory approaches (documenting the theoretical and causal basis of the program's functioning) (DHS Evaluation Support Unit n.d.).

Logic models play an important role in contextualising outcomes measurement. Logic models commonly document the connections between inputs, activities, outputs, outcomes and impact. In this way they are useful in distinguishing client outcomes from other aspects of a service's operations, but also in pinpointing the ways that outcomes may be connected to or dependent on other elements of the model.

HomeGround has developed a program logic framework (Planigale, Kimberley et al. 2009); the use of terminology in that framework is consistent with its usage in this paper (including usage of the word outcomes) and discussion in this section should be referenced to that paper.

Outcomes and goals

A program logic model is usually linked to a program's goals, as expressed through its mission, aims and objectives. Goals may potentially be quite broad (e.g. "to improve family functioning"). To be practical for outcomes measurement, goals need to be clear, specific and measurable (Weiss 1972: 26). Rossi (1997: 28) suggests looking at the core activities and emphasis of casework within a particular service to determine which aspects

HomeGround
Services

of a general goal should be operationalised as outcomes. Program objectives may correspond to outcomes (e.g. to place clients in emergency accommodation) or to outputs (to provide three sessions of personal counselling to each client).

Outcomes and outputs

Many authors stress the difference between outcomes and outputs. The distinction is typically understood to be that outputs quantify the level and types of activities provided by a service, while outcomes indicate how the need/problem is affected by these activities. Outputs can be important as process measures, and help provide the context for explaining results, but do not measure whether the program is effective (Spellman and Abbenante 2008: 7, 16). Baulderstone and Talbot (2004: 3) define outputs as "the completed service transactions or immediate results created by the program often defined as units of service (e.g. support hours delivered, referrals made, beds provided)", which they note tend to be easier to count and measure than outcomes.

It is worth noting that not all sources embrace this distinction between outputs and outcomes. In particular, Duignan (2005) defines an outcomes hierarchy as including claims about a range of "causes" leading to higher-level outcomes – the lower-level causes, which might be termed outputs in some systems, are here regarded as lower-level outcomes.

Outcomes and activities

In general, the distinction between a program's activities and outcomes is relatively clear. The activity may be transporting a client to a dental appointment; the outcome may be an improvement in the client's dental health. However, there are some situations where an outcome may be defined in terms that are equivalent to the program's activities. An

example of such an 'endogenous' outcome might be a program that claims that one of its outcomes is avoiding the out-of-home placement of children while their families are participating in the program. If the program is structured so that families recruited into the program are offered participation as an alternative to out-of-home placement, then the outcome is essentially assumed as an element of the program's activities. "The fact that a child … is not taken into custody while in the program restates the existence of a moratorium and is not itself an outcome of the program" (Rossi 1997: 30).

Outcomes and targets

Targets set particular (generally numerical) benchmarks for the level of particular types of outcomes that the program wishes to achieve (Spellman and Abbenante 2008: 27). Targets are commonly set in relation to the percentage of a given client population who will achieve a particular type of change, however they might also be set in relation to the degree or quality of the change (for example, to remain stably housed for at least 12 months, to improve at least three points on a particular scale).

5.4 DOMAINS

Many approaches to outcome measurement organise outcomes and measures into domains. Within evaluation and outcome measurement literature, the concept of domains is typically assumed to be useful and valid. While individual domains are discussed in detail, the concept of domain itself is rarely defined or critiqued. From a practical point of view, this paper also assumes the concept of domains to be useful. However, it is worth defining the concept clearly and pointing out some of its limitations and complexities. Among other factors, complexities can relate to 'boundary questions' (where is the dividing

line between one domain and another) and 'subsumption questions' (is a domain independent, or part of a higher-level domain).

For the purposes of this literature review, the term domain is used to refer to an aspect of human existence that is understood to be relevant to most or all people within a particular community or population – for example, physical health, material wellbeing, and so forth.

Domains (in the sense used in this paper) are also referred to as:
- 'life domains' – especially within the Quality of Life literature (e.g. Felce and Perry 1995)
- 'life areas' – especially within social work / practice focused literature (e.g. Kaufman 2007)
- 'outcome areas' – especially within literature focusing on outcome measurement (e.g. MacKeith and Graham 2007).

Domains as cultural constructs

Often, domains are thought of as universals - aspects of human life that are common or relevant to all people. For example, the domain of physical health is often asserted to be universally applicable to human beings. One implication of this statement is that it should be possible to measure the physical health status of any person. However, it is also possible to have a domain which is only of relevance to a particular community, population or society. For example, employment is a domain often used to assess outcomes of social service programs within industrialised societies. Labour markets and employment relationships are cultural phenomena and are not necessarily found within all human societies.

In the context of health care, Booth and Smith (1997) and Ware (1997) discuss the

difference between generic and specific measures, and generic and specific health concepts. Specific measures are those that are targeted to specific conditions or diseases; while the presence or absence of these conditions may be generally applicable as a population measure, tools focused on specific condition domains will only be applicable to the sub-population of people who experience the condition. Generic measures relate to generic components of health or wellbeing that are understood to be of relevance to an entire population.

An analogous distinction can be applied outside of the realm of health care. For example, parenting and child wellbeing are examples of domains that are specific to families; measures in these domains are therefore specific rather than generically applicable to the entire population (or to an entire client group).

Some suggested domains may be contested either in terms of their universality, or their validity as a model of some aspect of human life. For example, some Quality of Life and outcome measurement tools include domains or items relating to spirituality. The Ridgway Recovery-Enhancing Environment Measure (Campbell-Orde, Chamberlin et al. 2005) includes the item 'I have a positive spiritual life/ connection to a higher power'. Some authors might dispute the universality of spirituality as a domain of human experience, or model it as an aspect of emotional wellbeing.

Even where a convincing argument can be made for the general applicability of a particular domain, there will be cultural differences in the way these domains are conceptualised in different populations. The way that emotional or mental health, for

example, is understood in different societies (or even by different sub-groups within a single society) can be widely different. Any set of data items designed to measure a person's status or functioning within a particular domain therefore expresses a particular model or conception of the domain. These models are grounded in particular cultures and professions, and should therefore be applied with caution (if at all) to other cultural settings.

A related point is that the relative importance and significance attached to particular domains varies from society to society, and from individual to individual. This observation is particularly relevant to tools such as Quality of Life scales that aggregate scores from a number of domains to reach an overall rating. When assessing the relative contribution of each domain to the aggregate score, it is useful to weight the domains to reflect individual and societal appraisals of the importance of this life area (Felce and Perry 1995; University of Toronto Centre for Health Promotion 2010).

Sub-domains

Many authors (particularly in the Quality of Life literature) group domains in a hierarchical fashion, with a small number of top level domains, each of which includes a number of sub-domains. For example:
- The World Health Organisation's WHOQOL-BREF is based on four domains (physical health, psychological, social relationships and environment) - each of these has a number of sub-components (for example, physical health is subdivided into activities of daily living, dependence on medicinal substances and medical aids, energy and fatigue, mobility, pain and discomfort, sleep and rest, and work capacity) (World Health Organisation 1996)
- The University of Toronto's Quality of Life

model has three top-level domains (being, belonging and becoming), each of which have three sub-domains (for example, the being domain is divided into physical being, psychological being and spiritual being) (University of Toronto Centre for Health Promotion 2010).

Other approaches (particularly those with a practice rather than theoretical orientation) enumerate a range of domains with minimal attempt to group these together into larger units. For example:
- Kaufman (2007) lists 16 life domains ranging from 'housing and basic needs' to 'spirituality and values' - these are not grouped into larger units;
- The BT Generic Outcome Scale (Baulderstone and Talbot 2004: 65-74) includes 17 domains related to 'knowledge, feeling and behavioural change' (including two which are identified as aspects of health), and three related to 'client-environment change'. The division into personal vs. environment-focused change could be regarded as a top-level domain division. However, the intent of the tool is that relevant domains can be excerpted for use with particular clients without affecting the validity of the tool. From this point of view it is an advantage that the domains are arranged in a flat structure rather than being embedded in larger units.

Where domains are arranged into larger groupings, in some cases these groupings are based on empirical evidence – for example, through use of statistical techniques such as factor analysis (DeCoster 1998); see also World Health Organisation (1996: 6). In other cases, the grouping is based either on the author's own preferred grouping, or on a meta-analysis of common themes and approaches across other models (e.g. Felce and Perry 1995: 60).

A further issue is that some approaches to assessing status or measuring outcomes include a range of indicators but do not attempt to group these into domains at all. For example, the Illness Management and Recovery Scales include 15 items that each relate to a different aspect of illness management and recovery, with no explicit attempt to group these in any way (Campbell-Orde, Chamberlin et al. 2005: 32).

The discussion paper 'A consistent set of casework domains for HomeGround' (Planigale 2010a) looks at examples of a variety of domains 'sets' or 'trees' that are in use and relevant to outcomes measurement in homelessness services.

5.5 LOCUS OF CHANGE

Another way of 'dividing up' outcomes is according to what might be called the 'locus of change' – what is it about the client or the situation that actually changes? For example, does the client feel better about their situation, are they behaving differently, or do they have some physical resource now which they did not have previously?

In understanding the types of changes that social service programs aim to produce, the relationship between life domains and locus of change can be confusing. This section discusses different taxonomies of 'locus of change' and then looks at how these can be integrated with the concept of life domains.

Locus of change taxonomies

One well-known categorisation of locus of change is that of Rapp and Poertner (1992: 108), who identify five major categories in which social work interventions are designed to work with clients to produce outcomes:

- affect – the way people feel about something; their emotional response to a situation; their attitudes

- knowledge – acquisition of information, understanding, self-awareness, insight
- behaviour – skills (a person's ability to do some behaviour); performance (a person's use of particular knowledge and skills in their life)
- status – categorisation of someone's situation into one of a set of mutually exclusive categories (for example, 'homeless' and 'housed' might be examples of high-level status categories)
- environment – the relationship between a person and their social and physical context (including e.g. a person's access to resources, relationships with caregivers, degree of control or choice).

This taxonomy can be refined in a number of ways. Rapp and Poertner acknowledge that the change category of 'environment' is the least well-developed and least clearly conceptualised (1992: 199-120). In Rapp and Poertner's formulation, environmental change includes a number of areas that could be viewed as specific life areas rather than general types of change (e.g. adequacy of a person's residence, food and finances). It is suggested here that these are better viewed as domains (particularly in the case of residence / housing). However, the other four change categories (affect, knowledge, behaviour and status) are primarily person-focused. To discard a focus on the environment altogether would be problematic in ignoring a range of factors that often have a major impact on the work that occurs between staff and clients, and also have a major impact on outcomes.

The suggested resolution is to replace the change category of environment with two more precisely defined loci of change:
- resources – the person's access to resources (broadly defined – including assets, access to services and other supports)

- environmental structure – in particular, policies and constraints imposed by the environment that impact upon the person.

For further research — explore the concept of environmental structure / constraints and look at how it could be defined more clearly

Rapp and Poertner's category of 'affect' includes a number of areas that other authors categorise separately. Weiss (1972: 39) separates attitudes, values and personality variables, while Ware (1997: 59) expands on the concept of values to include how people as evaluators rate themselves – a concept related to Quality of Life approaches such as Multiple Discrepancies Theory that consider how people perceive themselves in relation to a variety of reference points (Hubley, Russell et al. 2009). Schalock (2001: 23) distinguishes adaptive behaviour (self care, language, mobility, independent living) from role status (a set of valued activities considered normative for specific age group). Ware (1997: 59) also touches on the concept of role performance and role disability. Finally, in the area of status it is possible to look at both a person's current status, and their potential future status (prognosis or risk).

Burns and Cupitt (2003: 6) also present what is essentially a taxonomy of locus of change, however it combines elements of both life domains and locus of change. The discussion below suggests that these are better kept separate.

Table 6 presents a taxonomy of locus of change based on Rapp and Poertner's model but integrating the refinements noted in this section.

Table 6: Integrated taxonomy of locus of change

Broad category	Specific locus of change
Affect	Emotion / mood
	Attitude
	Value (including self-evaluation)
Knowledge	Information
	Insight
Behaviour	Skill
	Performance - adaptive behaviour
	Role performance
Status	Status
	Risk
Resources	Physical resources
	Support
Environment	Structure (especially constraints)

'Hard' and 'soft' outcomes

Some authors use the terms 'hard' and 'soft' to broadly categorise outcomes and measures. Hard outcomes are those that are concrete, countable and externally observable; for example, an increase in the number of days that someone was in paid employment, a move from one housing tenure to another, or the occurrence of a particular behavioural pattern. Soft outcomes are related to a person's internal state and skills, and are typically measured by self-assessment on a scale with relative response categories (e.g. 'very satisfied' through 'very dissatisfied') (Segal 1997: 153; Butcher and Marsden 2004; Anderson 2008).

Soft and hard outcomes (or measures) could be seen as one way to divide up the locus of change set, where the first two locus of change categories (affect and knowledge) can be seen as soft outcomes and the remaining four as more associated with hard outcomes.

Segal (1997: 151) suggests that hard and soft outcomes are not mutually exclusive - i.e. that it does not necessarily make sense to only measure hard outcomes or soft outcomes. The two types can be closely intimately connected; if the objective is improving a person's sense of optimism, the "hard-core resources" for this process may be an apartment of the person's choice, a job that accommodates a disability, or the opportunity for social relationships.

Relationship between
locus of change and domains

It is useful to have a model of how domain and locus of change fit together. The suggestion made here is that locus of change cross-cuts the domain set. In other words, in any domain it may be possible to observe changes in one or more locus of change. For example, in the life domain of employment, staff and clients may work together to produce changes in affect (how the client feels about their employment situation), knowledge (for example, understanding of expectations within an employment environment), behaviour (job search skills), status (employed full time, part time, unemployed), resources (an up-to-date CV) and environment (cultural or policy change within a workplace to better accommodate the needs of the client).

One way to think about the distinction between domains and locus of change might be in terms of content and process. Domains relate more to the "subject area" of change, while locus relates more to process – 'how' the client is changing in relation to these domains. Table 7 shows this diagrammatically with a sample set of domains.

Table 7: Matrix illustrating cross-cutting relationship between domain and locus of change

Domains	Locus of change					
	Affect	Knowledge	Behaviour	Status	Resources	Environment
Housing						
Physical health						
Mental health						
Legal						
etc.						

This cross-cutting arrangement appears useful, as it allows both workers and clients to be clearer about the types of change they might pursue within particular life areas, and to identify and celebrate a range of achievements within life areas, rather than limiting the focus to (for example) status change.

One area that needs clarification in this model is where a domain appears to 'cover the same territory' as one of the loci of change, or a domain is conceptualised in terms of a locus of change. For example, the domain of mental health could be seen as occupying similar territory to the affect locus of change. There are two possible responses to this:

- Define the affect locus of change as pertaining to a person's emotional response or attitude towards a particular life area, rather than their affect in general. In the domain of mental health, therefore, affect change would relate to how a person feels about or regards their mental health. A person's global mood condition would then be seen as fitting within the status locus of change within the mental health domain.
- For the mental health domain, the measurement of a person's affectual state (whether global or specific to a particular area) is seen as fitting within the affect locus of change.

I would suggest that the first approach is clearer as it fits with the approach that locus of change is about a person's capabilities in a particular area. Thus, affect locus of change is always seen as a person's affect in relation to a particular life area.

A similar issue exists with regard to the domain of education and the knowledge locus of change. Following the above suggestion, the knowledge locus of change would be seen as relating to a person's knowledge in relation to a particular life area (housing, physical health, or in this case the process of education). Measurement of change in terms of a person's understanding of the content of education would fit either with measurements within the status locus of change for the education domain (such as whether a person had completed and passed a particular course), or in the knowledge domain for the life area corresponding to the content of the course.

5.6 POINT OF INTERVENTION (PREVENTION VS. AMELIORATION)

In recent years, government policy has raised the profile of efforts to prevent homelessness for at-risk groups, placing this on a par with the importance of assistance for people already experiencing homelessness (e.g. Commonwealth of Australia 2008).

The distinction between prevention and amelioration relates to the timing of service provision in relation to a potential or actual episode of homelessness. This is a separate dimension that can be added to the domains and locus of change model without requiring changes to the model. For example, a preventative intervention may be in relation to a threatened tenancy breakdown. In addition to the domain of housing, an assessment may explore some of the reasons why the tenancy has become at risk, for example in domains of a person's mental health, family situation or employment. Support and resources might be provided in each of these areas, and outcome measurement would also focus on these areas.

Specific goals and outcomes measured may differ, however it is suggested that the framework of domains and locus of change is equally applicable to work at any stage along the timeline of an episode of homelessness. Table 8 below includes some examples of outcomes that might be observed in the domain of housing, for preventive- and amelioration-focused service delivery.

Table 8: Possible outcomes of prevention and amelioration activities, categorised by domain and locus of change

Domain	Locus of change	Prevention	Amelioration
Housing	Affect	Person less anxious about their housing situation	Person feels more satisfied with their housing situation
	Knowledge	Person has better understanding of tenancy rights and responsibilities	Person has increased knowledge of housing options
	Behaviour	Person pays rent more regularly	Person able to complete private rental application form
	Status	Continuation of tenancy Reduced risk of tenancy breakdown	Positive change in housing status (e.g. primary homeless > transitional, transitional > housed)
	Resources	Person linked to tenants' legal / advocacy service	Person linked to private rental support worker
	Environmental structure	Landlord agrees to arrears repayment agreement that is reasonable for the tenant to maintain	Improved standards in rooming houses used by the person, due to enforcement of standards

See also (Culhane, Parker et al. 2007; Gray 2008) re: outcomes for prevention and early intervention programs.

5.7 HOUSING AND NON-HOUSING OUTCOMES

Conceptualising and categorising outcomes in terms of domains and loci of change can give the impression that multiple outcome types for an individual are independent of each other. To some extent this is true; as Segal (1997: 150) notes, "given the complex hostile environment of each case, it would be unusual to find all outcomes moving in the same direction at the same time. Some good things and some bad things will result simply from the nature of the environment..."

However, it is also true that outcomes in particular domains may influence outcomes in other domains. One example is the relationship between housing and non-housing outcomes. Hulse and Saugeres (2008: 39-42) found that housing insecurity is closely related to other types of insecurity (for example, insecurities in family life and employment insecurities), and impacts on physical and mental health. Blunden and Johnston (2005: 6, 34) argue that provision of public housing can dramatically improve other outcomes for people who are homeless and/or who suffer from mental illness. However, it is also the case that housing stability outcomes may be dependent on support and intervention in other life domains including mental health and substance misuse. For this reason, housing first programs build in case management of varying levels of intensity and duration following a person's access to permanent housing (National Alliance to End Homelessness 2006).

The example of public housing also provides a caution about causal inferences. Living in public housing is correlated with poorer socioeconomic outcomes across a range of domains, when compared with other tenure groups. However, this is predominantly not a result of public housing tenure itself, but an 'allocation effect' – i.e. a result of eligibility criteria which target public housing to people with low socioeconomic status and complex needs (Blunden and Johnston 2005: 33-35).

See also (Phibbs 2002).

HomeGround Services

6. Measures and measurement tools

Measures are the dependent variables of study (Weiss 1972: 34). Outcome measures "describe observable, measurable characteristics or changes that represent achievement of a desired outcome. Outcome measures specify exactly what is going to be measured (indicators) and units of measurement used to determine the extent to which desired outcomes are attained – e.g., HbA1c level (<7.0%) as an indicator of diabetes control" (Post, Isbell et al. 2005: 4).

Measurement tools are instruments that enable collection of data pertaining to a set of measures. Such tools allow data to be collected in a standardised, consistent way (MacKeith and Graham 2007: 3), and are typically accompanied by instructions for analysis and interpretation of the data.

This section examines criteria for selecting outcome measure and tools, and then looks at a range of measures and tools that may be applicable in the context of homelessness services.

6.1 CRITERIA FOR SELECTING MEASURES AND TOOLS

Many authors discuss criteria that can be used to select outcome measures for particular contexts. In general, the criteria apply to both individual measures and to measurement tools, although in some cases the interpretation may differ (for example, 'feasibility' with respect to an individual measure may involve consideration of whether the item is too invasive to be appropriate in the service delivery environment; with respect to a tool, it may involve consideration of the number of items included). Some criteria relate to innate characteristics of the measures themselves (e.g. validity, sensitivity) whereas others relate to the tool's fit with the context of use (e.g. acceptability to stakeholders, relevance to service provided) (MacKeith and Graham 2007: 8).

Table 9 lists a range of criteria for selection of measures and tools, based on NMHWG Information Strategy Committee Performance Indicator Drafting Group (2005: 98-101), Spence, Donald et al. (2002: 24-29), Stedman, Yellowlees et al. (1997: 13-15), and MacKeith, Graham et al. (2007). Evans, Greenhalgh et al. (2000: 383) provide a general framework for organising the criteria into four categories: psychometric validation, stakeholder perspectives and user-centeredness, feasibility and utility. These have been adapted as the general categories for Table 9.

Table 9: Criteria for selection of outcome measures and tools

Category	Criterion
Utility	• Acceptability to stakeholders • Informational value to stakeholders • Comparability of findings • Value added to service delivery
Relevance and user-centeredness	• Relevance to service provided • Appropriateness to client diversity • User friendliness • Sensitivity
Psychometric validation	• Reliability • Validity
Feasibility	• Cost • Complexity and user competence • Length • Invasiveness

The text below briefly discusses each of the criteria listed.

Barr (Barr 2008; Barr n.d.) also provides a range of tips for selecting performance measures (including outcome measures) and a rigorous framework for defining and documenting measures through her PuMP process.

Acceptability to stakeholders

Schalock (2001: 134) notes that it is important that stakeholders "buy into" the measures selected so that the information produced will be seen as trustworthy and will be well-used. Stakeholder acceptance in part reflects the suitability of the measures on a range of the other criteria articulated here - i.e. measures are more likely to be acceptable to stakeholders if they are perceived as valid, reliable, cost-effective, relevant to the service provided, and so forth. However, there may also be other factors that influence stakeholder perceptions of measures. Stakeholders' previous experience with particular measures or tools, and association of measures with particular sectors, organisations, or theoretical orientations, may influence acceptance.

Informational value to stakeholders

Fundamentally, an outcome measure or tool should provide additional information to stakeholders that is not already available (Evans, Greenhalgh et al. 2000: 383). However, different stakeholders are likely to have different needs. For example, service delivery staff are likely to need detailed information on a client's current level of need and functioning in a range of areas. Management and funders are more likely to be interested in headline indicators that can provide a picture of overall progress within a cohort of clients. A combination of measures that meet the informational needs of all stakeholders is required.

It is worth noting that informational value also interacts with other criteria, in particular psychometric validity and reliability. Scales with relatively low reliability and validity may not provide useful information about individual cases, and may only be useful at a system level for aggregating outcomes over thousands of cases (Hudson 1997: 73). In a service delivery environment, it is preferable to use measures and tools that are sound for work with individuals.

Another issue related to informational value is what level of information a particular result on a measure provides. For example, if a client scores '3' on a measure of self-care, what does this indicate in terms of the behaviour, attitudes and skills the client shows? Scales that have clearly described anchor points, or that are based on concrete questions, are likely to provide clearer answers to these questions (MacKeith, Graham et al. 2007: 9).

Comparability of findings

It is essential that measures selected allow for aggregation in a meaningful way (Baulderstone and Talbot 2004: 5). This may have a number of dimensions:

- Aggregation within a single service unit allows for examination of patterns of change within the clients who access that service
- Aggregation on an organisational level is required for assessment of the performance of the organisational as a whole
- Aggregation across a service system (at program or total cohort level) allows for an understanding of the overall condition of clients served by that system, and may allow for comparison of the performance of service delivery organisations.

The basis for aggregation beyond the level of the individual service is the use of consistent measures across different services. At an intra-agency level, this requires agreement on 'standard' outcomes and measures that are collected by all service units (Hendrick 2010b). At an inter-agency level, it requires negotiation of a general sector framework or approach, possibly including an agreed theory of change (MacKeith 2007: 5).

An additional consideration is the availability of normative data for a population or social group, i.e. availability of information on mean values and standard deviations for a representative sample (Spence, Donald et al. 2002: 27). This potentially allows comparison between the data that a given agency or service collects, and a wider population. Normative data tends to be available for widely used standardised scales. Use of measures that have been used in program evaluations may also enable some comparability (Spence, Donald et al. 2002: 28), although attention needs to be paid to service, environmental or process factors that can reduce comparability.

Value added to service delivery

MacKeith and Graham (2007: 5) distinguish between tools that are primarily intended to provide evidence of outcomes for purposes of accountability, advocacy and knowledge-building, and tools that both provide evidence, but also support casework. Hudson (1997: 72-73) warns of the problems that arise when introduced measures are of little direct use to practitioners or clients. In an outcomes monitoring system, measures that are of 'clinical' relevance and add value to daily service delivery work are likely to have higher acceptance and greater benefits to stakeholders. Baulderstone and Talbot (2004: 37-38) also argue that the needs of clients and workers be prioritised in outcome measurement systems.

Tools and measures may add value to service delivery where they:

- Provide indicators of risk
- Facilitate regular assessment of areas of need that might otherwise not be discussed
- Provide immediate summary information to clients and service delivery staff, for example in an immediately available report or visual summary (MacKeith and Graham 2007: 12)
- Can be used to identify and celebrate achievements
- Are based on an explicit model of change that can help clients to understand the possible stages of their 'journey' of change (MacKeith and Graham 2007: 12)
- Have a direct link with action planning.

Relevance to service provided

Measures should reflect the nature of the service provided, the types of outcomes sought, and dimensions of importance to clients and of use to staff (Baulderstone and Talbot 2004: 5). Relevance may have a number of levels: measures selected should be relevant

HomeGround
Services

to the organisation's mission and strategic goals, should relate to the desired outcomes identified for particular services, and should also be related logically to what the specific service does and has control over (Schalock 2001: 32, 134). Stedman, Yellowlees et al. (1997: 15) suggest that measures should cover "the most important and frequently observed symptoms, problems, goals, or other domains of change for the [client] group(s)."

Measures should also reflect aspects of the service model including the duration of engagement, nature of relationship established between client and worker, and expectations as to potential interim outcomes that may occur during the period of engagement (Elizabeth 2010). Measures for short-term or crisis services may be different to those for long-term support services.

Appropriateness to client diversity

Measures and tools may be generic (applicable to a wide range of client groups), or may be intended to be used with particular groups (people with alcohol or drug addictions. people over the age of 65, ...). It is important that outcomes and measures are relevant to the range of diversity expected within the client population in terms of socio-demographic factors such as age, gender, household groupings, functional levels, and other factors (Schalock 2001: 32). In some cases, it may be necessary to have several versions of a measurement tool that are tailored for use with particular client groups (Wells and Johnson 2001: 191-192; MacKeith and Graham 2007: 9).

In addition to their basic applicability to different groups, the design of measures and tools also needs to be assessed for its cultural appropriateness and how it handles diversity (Schalock 2001: 135). For aggregation purposes, measures should be used that are

equally applicable across cultures, and that do not show different psychometric properties across cultures (Spence, Donald et al. 2002: 27). Measures and approaches that do not reflect 'cultural competence' are likely to produce information that is of little value (Mertens 2008), and may have other negative impacts. Issues to be aware of include:

- Differences may occur in understanding of particular concepts across cultures - for example, 'family' may have different meanings in different cultures. This potentially affects validity of measures.
- The medium may need to be adapted to suit particular groups (including those with sensory disabilities or limited literacy) (Mertens 2008: 53)
- The language may need to be modified to suit particular groups (Baulderstone and Talbot 2004: 7)
- Some groups may be suspicious of information-gathering processes due to historical events and recent traumatic experiences (Baulderstone and Talbot 2004: 7).

User friendliness

In a service delivery (as opposed to research) setting, measurement tools are most likely to be used regularly and successfully where they are presented in a 'user friendly' format. MacKeith and Graham (2007: 12) note the importance of client friendly language rather than technical jargon, attractive clear presentation of the tool itself, and clear format for any report or summary.

Sensitivity

Sensitivity refers to the measure's ability to indicate whether a clinically significant change has occurred for a client over consecutive administrations of the measure (Stedman, Yellowlees et al. 1997: 14). Eardley, Thompson et al. (2008: 7) note that because

of the multiple and complex difficulties facing homeless people, "what may appear to be very small changes can, for some, still be highly significant." For this reason, it is vital that measures are sensitive enough to detect small changes in a person's situation. Post, Isbell et al. (2005: 24) note that some standardised instruments designed for use with mainstream populations may not be sufficiently sensitive for use with homeless populations – the large majority of homeless clients may fall within the 'low functionality' segment of the scale without sufficient detail as to the reasons for this result.

MacKeith and Graham (2007: 13) note that measures with three or five point scales will tend to have low sensitivity, and recommend scales of ten points.

Reliability

Reliability is a general term for the consistency of measurements; unreliability refers to inconsistency due to random measurement errors (Bloom, Fischer et al. 2006: 68-69). More reliable measures are those that demonstrate consistency in results when applied to the same person or thing repeatedly and independently, under the same circumstances.

Reliability of measures and tools has a number of aspects (Spence, Donald et al. 2002: 24-25; Bloom, Fischer et al. 2006: 70-73):

- Test-retest reliability – the level of consistency in results when a measure is applied on different occasions under the same circumstances. If the measure is not 'stable' and fluctuates in response to random changes in the individual or the environment, it will have lowered reliability.

- Alternate forms reliability – the extent to which responses to two forms of a measurement tool, which differ only in the content of the individual items, are correlated – this indicates the extent to which the different measures do measure the same concept.
- Internal consistency reliability – the extent to which parts of a measurement tool (e.g. different items within a tool) are homogeneous or measure the same thing. Tools with high internal consistency reliability will contain a set of items that are highly correlated with each other, indicating that the tool measures a single concept.
- Interobserver reliability – the extent to which two or more observers from the same group are consistent with each other when they independently observe or rate a particular thing. This is particularly relevant to a routine monitoring setting where multiple staff members (or clients) independently complete particular tools. Interobserver reliability may also be affected by the extent of information and training provided to users of the tool.
- Cross-informant agreement – the extent to which observers from different groups (e.g. a client, a housing worker and the client's carer) agree in their ratings on a particular measure. Low cross-informant agreement is common and it can be difficult to tell whether the differences are due to inconsistent application of the measure, or differences in behaviour in different environments.

Reliability of measures or tools can be tested and rated as a correlation between 0.0 and 1.0, where 1.0 represents perfect reliability. In general, correlations of at least 0.70 are recommended in order for the measure to be useful, however reliability of 0.80 or above is preferable (Bloom, Fischer et al. 2006: 74).

Reliability of measures depends on a range of factors, including the types of items included in the tool and the way that they are worded and presented. One key issue is the extent to which users of the measure have different understandings of the language used. In some cases, measures may be inadequately defined so that users end up providing a rating based on their 'best guess' rather than with clear guidance (cf. the example of Wells and Johnson (2001: 181-182) in relation to measures of re-referral in child protection settings). Training and access to a data dictionary can help to increase inter-observer reliability.

Validity

Validity refers to the extent to which a measure or tool actually measures what it is intended to measure (Bloom, Fischer et al. 2006: 75). Validity therefore affects whether, and to what extent, one can trust the results of particular measures, as interpreted for specific purposes and with reference to specific concepts or constructs (Cook and Beckman 2006). Measures are not valid in the abstract, but for particular populations and settings.

Validity also has a number of aspects (Spence, Donald et al. 2002: 25-26; Bloom, Fischer et al. 2006: 76-83), although more recent views suggest that these could all be grouped together under the overarching framework of construct validity (e.g. Cook and Beckman 2006: 166.e110).

- Face validity – the opinion of some observer that 'at face value' the measure appears to measure what it is designed to measure. While face validity can aid user acceptance of the measure, it is not quantifiable and is subject to differing judgements of differing observers, hence some sources argue that it should not be considered a form of validity at all (Cook and Beckman 2006: 166.e112).

- Content validity – the extent to which the particular questions or items selected for measurement are representative and adequately cover the domain being assessed, or are a biased or limited sample of what the tool is intended to measure.
- Criterion validity – the extent to which an individual's rating on a measure is predictive of the individual's performance or status on certain related outcomes (the criterion against which the measure is assessed). The validity of the criterion must already be established. The higher the correlation between measure and criterion, the better the criterion validity.
- Construct validity – the extent to which empirically observed relationships among measures of concepts agree with the inter-relationships among concepts predicted by theory. This may include convergent validity (measures co-vary in the predicted manner with other variables that they should correlate with in theory), and discriminant validity (measures do not correlate with variables that theory predicts they should not correlate with).

Some issues that need to be considered in relation to validity are the reactivity of measures (the extent to which changes may be caused by the client's awareness of the measurement process itself rather than by the intervention), and observers' biases – e.g. the desire to present oneself in a socially desirable manner, or expectations about particular people based on their socio-demographic background (Bloom, Fischer et al. 2006: 81).

Both reliability and validity are partly a reflection of the development process of the tool. An iterative process of development with repeated cycles of testing and revision allows for improvement over time of the tool's reliability and validity as well as other aspects of its design (MacKeith and Graham 2007: 10).

Cost

Operation of an outcomes monitoring system has a range of costs. While there will always be a baseline investment needed to operate such a system, the actual cost may vary widely depending on the particular measures or tools selected. It is important that the measures selected be affordable for the organisation (Schalock 2001: 32, 134) and represent 'value for money'. Analysis of outcomes in some domains may require a depth of information collection that is unaffordable (Rapp and Poertner 1992: 104).

Specific costs to consider include:
- Licensing costs associated with standardised tools - some tools are available free of charge, whereas others are licensed on a per-service or per-client basis (MacKeith and Graham 2007: 11)
- The 'administrative load' (Pleace 2008: 64) on service delivery staff of collecting and entering outcomes data, or alternatively additional staff resources to collect and enter data
- The costs of staff time to clean, aggregate, analyse and report on data (Wells and Johnson 2001: 192-193)
- Computer equipment and software required to store and process data
- Extent of training required for staff who will be using the measures and tools (Wells and Johnson 2001: 192-193).

Cost will be affected by a range of factors including the number and complexity of the measures collected.

Complexity and user competence

Tools that are complex and difficult to use or score present serious challenges, especially when data collection is to be undertaken by service delivery staff, who may have a wide variety of backgrounds, skills and levels of formal education and training (Baulderstone and Talbot 2004). Some tests require a high level of training in administration, scoring and/ or interpretation (Spence, Donald et al. 2002: 26-27). In addition to increasing training costs, complex tools are likely to be less reliable when administered by multiple staff. Whichever tool is used, the availability of training, support and documentation will be important (MacKeith and Graham 2007: 10).

Some tools are restricted to purchase and use by specific professions, such as registered psychologists (Spence, Donald et al. 2002: 26-27) and would therefore be inappropriate for most homelessness services.

Length

Measurement tools for routine monitoring need to be feasible in terms of their length (number of items) and the time required to complete them (MacKeith and Graham 2007: 10). However, what is feasible will depend on the nature of the service and the relationship between staff member and client: services with longer-term engagement are likely to be able to use a lengthier tool. Tools that are overly cumbersome are likely to have a low rate of completion.

Invasiveness

The nature of the particular items included in a tool also needs to be commensurate with the setting and relationship between staff and clients. Invasive or highly personal or sensitive questions may not be appropriate in settings where only brief or superficial contact occurs. Some questions (e.g. in relation to sexuality) may also be perceived as inappropriate or raising a safety risk in some situations (e.g. between a female worker and male ex-prisoner client). In general, questions in relation to sensitive areas should only be asked if the service has the capacity to follow up with support in that area if needed.

6.2 HOW MANY MEASURES?

In the absence of detailed information on the service and organisational context, it is not possible to specify an absolute number of measures as the 'right number' to use. However, the literature does provide guidance on factors to keep in mind in relation to the number of measures implemented. Two distinct themes emerge from the sources reviewed.

One strand of thinking emphasises the importance of limiting the number of measures: "select a minimum number of outcomes that are relevant and obtainable" (Schalock 2001: 32). McDaniel (1996) advises "don't measure it unless you plan to change it"; Berman & Hurt (1997: 87) emphasise that organisations should only measure what they will use. A number of authors suggest keeping the number of measures small initially and letting the system evolve with experience (e.g. NMHWG Information Strategy Committee Performance Indicator Drafting Group 2005: 99). The main benefits cited for the "minimum measures" approach are that the system is more likely to be manageable, will be less confusing for staff, will impose less administrative burden on the organisation, and will therefore be more likely to be sustainable. Rapp and Poertner (1992: 103) suggest that having too many measures tends to dilute organisational focus and lead to paralysis, because it gives the message that everything is equally important.

Another strand of thinking points out the benefits of having a greater number of measures, and/or more detailed measures. More detail provides greater information on the areas where services are succeeding or failing (Pleace 2008: 64). To minimise error and/or bias, Stedman, Yellowlees et al. (1997: 20) recommend using multiple assessment methods, multidimensional assessment

methods, and if possible using more than one rater for each measure or domain (see also Bloom, Fischer et al. (2006: 307-308)). To the extent possible, measures should examine the perspective of all those involved in the service (client as well as service providers). Similarly, Schalock (2001: 135) comments that human behaviour is not singular, and that measures should reflect the complexity of the human condition and of service's desired outcomes. Multiple measures also may make it more difficult to 'fudge' the data - i.e. for those collecting data to hide failures or fake successes (Pleace 2008: 61). At times, multiple measures may be needed to 'balance' each other – to avoid the distortion of service delivery which can follow a narrow focus on a single measure of success (Rapp and Poertner 1992: 102).

Integration of the two strands suggests a need to balance the "specificity and rigour" of the measurement process with the expense of data collection, to ensure that the usefulness of information gained is not sacrificed to false economy, while at the same time avoiding measures that are too resource intensive to be implemented effectively (Wells and Johnson 2001: 192; Pleace 2008: 64). Ultimately the decision depends on context (Burns and Cupitt 2003: 27).

Breadth vs. depth

Given limits on the number of measures that can feasibly be collected, there will usually be trade-offs between breadth and depth of measurement. These tradeoffs are a feature of any evaluative activity (Patton 1987: 46-48) but can be particularly heightened in the resource-limited environment of outcomes monitoring systems.

There are advantages to the 'whole client' approach, which seeks to cover a wide range of potential concerns or aspects of a client's condition, can give a more rounded picture of a client's situation, and may indicate not just whether a particular condition has changed but whether the client's overall quality of life has improved. The Outcomes Star suite is one example of a measurement tool that prioritises breadth over depth, with only a single measure in each domain. The disadvantage of this type of approach is that the level of insight into any one domain (and potentially the level of reliability and validity) is limited. Approaches that focus on only a few outcomes areas, but examine them in more depth, can provide much more useful information in these domains but may miss changes in other areas of a client's life (Hatry 1997: 17-18).

It is possible to make an organisational decision to opt for either breadth or depth, but there are also a few other possible responses to this issue.

- Use a 'modular' system of in-depth measures for particular domains, in conjunction with a generic suite of 'core' indicators covering the full breadth of issues at a shallow level. For example, multi-item modules might be developed in the areas of housing, mental health, physical health or legal issues if these were seen as relevant to the delivery of particular services. However, these modules might only be implemented in certain teams, or in respect to certain clients, where the measures were relevant and feasible. Other services might use only the generic core indicators. The BT Generic Outcomes Scale (Baulderstone and Talbot 2004: 65-74) provides one example of a modular tool where blocks of questions can be extracted for use with particular clients.
- Use of 'tracer' conditions. In health care, a tracer condition is a single condition that is measured throughout an organisation and is assumed to be generalisable to the broader system of care. For example, in primary care, asthma, diabetes or hypertension might be chosen as a tracer condition. Findings concerning processes and outcomes of care for the tracer condition have implications for the entire organisation and are assumed to reflect the functioning of the entire system of care (Booth and Smith 1997: 38-39). By analogy in a homelessness setting it might be possible to argue that a particular domain or domains were 'tracer' domains that reflect the quality and effectiveness of the organisation's services overall. Research would be required to establish the validity of the assumption that it is possible to generalise from the tracer domain to the entire system of care.

For further research: does the suggestion of a 'tracer' domain model hold up, has anyone applied this outside of primary health, and how could the validity of generalisations be established?

When opting for a 'breadth' approach, it will usually be necessary to select only one or a few measures for each domain. One possibility is to select items from various standardised instruments that load most strongly on the dimension being measured. Factor analysis may be required to establish the strength of loadings. However, Spence, Donald et al. (2002: 28-29) note that the psychometric properties of the original instrument/s no longer apply, and the revised measures should be piloted and shown to have acceptable reliability and validity. Psychometric expert advice is recommended in this process.

6.3 TYPES OF DATA AND TYPES OF MEASURES

There are a variety of different types of measures which can be used to assess outcomes, and a variety of sources of data which can be used as the basis for completing the measure. Table 10 summarises a range of measure types which could be used in outcomes monitoring; the information is integrated from discussion in Bloom, Fischer et al. (2006: 169-313), Burns and Cupitt (2003: 22), Rapp and Poertner (1992: 106-107) and Baulderstone and Talbot (2004).

Underlying all measure types, Bloom, Fischer et al. (2006: 63-64) identify four levels of measurement, which provide increasing levels of information about the thing measured:
- Nominal level – assignment to one of a set of mutually exclusive categories (e.g. male / female)
- Ordinal level – as per nominal-level measures, plus information about the relative amount of one category compared with another (categories are rank-ordered on a scale – e.g. never, sometimes, often, always)
- Interval level – as per ordinal-level measures, plus that adjacent intervals are equal so it is possible to determine how much more one category is compared with another (e.g. Celsius scale for temperatures). It is often reasonable to use standardised scales with the assumption that they are interval-level measures.
- Ratio level – as per interval-level measures, plus the categories have a defined zero point with an intrinsic meaning (e.g. age).

Table 10: Selected measure types and associated data sources

Measure type	Locus of change	Data source/s	Measurement / recording tool
1. *Individualised rating scale* – typically rating frequency, extent or severity of some condition, satisfaction with some aspect of a situation, ability or agreement with some statement	Any	Self-administered questionnaire; client response to interview question; rating by a practitioner, relevant other party, or independent evaluator	Typically paper-based or electronic questionnaire
2. *Standardised scale* – combines a set of individual items / measures which may be of various types including rating scales, concrete questions etc. Usually have known reliability and validity level and may be normed. May provide an overall score or 'index'	Any – depends on the specific scale	Client self-report, including self-administered questionnaire (e.g. Rapid Assessment Instrument); rating by a practitioner, relevant other party, or independent evaluator	Typically paper-based or electronic questionnaire

HomeGround Services

Measure type	Locus of change	Data source/s	Measurement / recording tool
3. *Status maintenance and change scale* – assigning a client's status as one of set of designated categories, normally arranged in a progression from undesired to desired status	Status, Resources	May be client self-report but usually provided by practitioner or other party on basis of knowledge of client's situation. Data may already be present in routine agency data collection (e.g. housing status)	Various – may be included in questionnaire or recorded separately
4. *Level of functioning scale* – rating client's functioning or performance in particular domains on a scale from low level functioning to competent, independent functioning	Behaviour	Client self-report (self-administered questionnaire, response to interview question); practitioner or third party rating	Typically paper-based or electronic questionnaire; Outcomes Star
5. *Goal attainment scale* – individualised rating scale anchored by a set of 'expected', 'more than expected' and 'less than expected' outcomes which are jointly agreed by client and practitioner at baseline	Any	Rating negotiated between client and worker, or may be independently completed by both and then compared	Paper-based or electronic goal scaling tool
6. *Standardised goal scale* – rating scale from 'no progress' to 'fully achieved', with goals referenced to standard goal list	Any	Client self-rating or practitioner rating, or independent ratings by both	Paper-based or electronic goal achievement form
7. *Behavioural count* – counts of frequency and/or duration of specific behaviours	Behaviour	Observation – either self-monitoring (e.g. occurrence of self-deprecating thoughts) or direct observation of overt behaviours	Various – paper-based checklist, small objects, electronic counters, etc

HomeGround Services

Measure type	Locus of change	Data source/s	Measurement / recording tool
8. *Simple observation measure* – qualitative information and/or quantitative ratings	Any – but inferences become more problematic with 'internal' loci such as affect	Simple observation e.g. of physical attributes, expressive movements, physical location or language behaviour. Physical traces (erosion / accretion) – practitioner or third-party observation	Paper-based or electronic
9. *Qualitative information* about the nature, extent and context of client needs, problems and progress	Any	Client's written self-report; client's verbal self-report, practitioner observation, third party report	Client log; paper-based or electronic case notes and client files
10. *Case status count* – e.g. number of re-referrals, number of clients on at-risk register, etc.	Tends to be linked to behaviour and status	Organisational records	Various

Measurement tools and packages may include a variety of different types of measures, depending on the purpose for which they are designed and the types of change they aim to measure. Using a combination of measure types and data sources can strengthen findings and may also have beneficial effects on practice (Segal 1997: 157; Burns and Cupitt 2003: 28).

Qualitative and quantitative measures

Quantitative measures are typically used as the basis of outcomes monitoring systems, in part because of the ease of aggregation they provide. However, a number of sources also suggest supplementing these with qualitative information (Spence, Donald et al. 2002: 29; Burns and Cupitt 2003: 16; Elizabeth 2010). Qualitative data can provide rapid feedback, identify areas of change that "fall through the gaps" of the quantitative measures, establish the context for change and interpretation of results, and provide a more authentic view of the client's lived experience. To make analysis feasible qualitative data needs to be brief, focused and/or readily searchable for keywords.

Lead vs. lag indicators

Lead indicators are measures that can be used to predict change in another performance measure (the lag indicator) (Barr n.d.: 13). They typically measure results of earlier stages of a process that eventually result in the outcomes tracked by the lag indicator. A possible application of this in the homelessness service context would be to establish lead indicators around interim outcomes (e.g. successful public housing applications; successful negotiation of arrears repayment agreements) that predict medium-to long-term outcomes. Barr suggests that lead indicators need to be tracked with greater frequency than lag indicators.

Vantage point

Vantage point refers to the party or parties who are the source of a specific rating – typically the client, the practitioner, or a relevant third party such as a parent, teacher or spouse. It should be noted that vantage point is not necessarily only that of a single person (ratings could be agreed jointly between client and worker, for example). Vantage point may also be different from recorder role (for example, a worker may complete a measurement tool on the basis of ratings provided by the client).

Vantage point is a key consideration in outcomes measurement because it is common to find a lack of agreement in ratings by clients, staff and independent observers on effectiveness measures (Rapp and Poertner 1992: 105). In the homelessness sector it is generally accepted that client involvement in outcomes ratings is important, and that the assessment of client outcomes without their knowledge may be questionable ethically (Baulderstone and Talbot 2004: 7). Any measure that attempts to assess client quality of life, satisfaction, affect or perception must be based on the client's viewpoint (Hudson 1997: 76; Ware 1997: 51-52).

However, multiple vantage points can provide a more reliable estimate of change (Rapp and Poertner 1992: 105). In particular, practitioner

observation can provide a useful addition to self-report for clients who are under the influence of alcohol or drugs, have mental health issues, have low self-awareness or little insight into their own behaviour. In such circumstances, clients may either over-estimate or under-estimate their difficulties (Burns and Cupitt 2003: 28).

Type of anchor points; theories of change

Measures using scales require anchor points to allow users of the measure to select the appropriate rating. In general, the more clearly defined the anchor points are, and the more direction that raters have as to the factors to be considered and their relative importance, the more reliable the measure is likely to be (Weiss 1972: 41).

Subjective scales that relate to how the client feels about an area of their life (e.g. very satisfied, somewhat satisfied, neither satisfied or dissatisfied, somewhat dissatisfied, very dissatisfied) do not have clearly defined anchor points. On the other hand, defined scales have clear descriptors of behaviours or circumstances associated with each step of the scale. These scales can be useful in improving reliability, providing more objective criteria for ratings and therefore allowing more meaningful comparison between the ratings for different clients (Segal 1997: 153; MacKeith and Graham 2007: 5-8).

Tools using defined scales may also be based on an explicit theory of change, which includes descriptions of the different points on the journey towards the desired end outcome. MacKeith and Graham (2007: 12) argue that tools which use an explicit journey of change are more powerful as casework tools because they "help clients to understand the change process they are engaged in and the particular challenges and rewards of different stages in the journey."

Generic vs. condition specific or program specific

Generic tools are intended to be used across a wide spectrum of client groups, whereas condition-specific or program-specific tools focus on specific indictors relevant to the needs and condition of a particular client group. Generic measures are valuable in allowing comparisons between different client groups (for example, across services or organisations), however they may be more error prone (Pleace 2008: 60). Specific measures may have higher user acceptance and utility in the service delivery environment but make comparisons difficult (Stedman, Yellowlees et al. 1997: 16).

Standardised vs. self-developed tools

There is mixed opinion in the literature on the relative merits of using existing standardised scales, as opposed to an organisation developing their own measures and tools. Key advantages of using standardised tools include (Rapp and Poertner 1992: 105; MacKeith and Graham 2007: 9-10):

- Developing and piloting new tools is a resource-intensive process
- Existing tools have generally already been tested and revised to iron out problems
- The validity and reliability of most standardised scales is known. There is evidence that agency evolutions of tools can sometimes lead to instruments that have poor psychometric properties (Culhane, Parker et al. 2007: 12.21)
- Existing tools may come with documentation, training and/or software to support data collection
- Standardised tools are more likely to support comparison across client groups and with normative data
- Standardised tools may carry more credibility with funders.

However, there are also a number of potential disadvantages to using standardised scales in the homelessness context (Weiss 1972: 36; Baulderstone and Talbot 2004: 10; Post, Isbell et al. 2005: 23-24):

- Some standardised scales are costly to use due to licensing fees
- Standardised scales may not cover all the domains of interest to a service
- Use of multiple standardised scales can be time consuming (although in some cases there are shortened versions of the scales available that can be used as Rapid Assessment Intstruments)
- Standardised scales are often associated with particular conceptual or service contexts; use of tools outside of the contexts for which they were intended can rely on a series of unproven assumptions which cast doubt on the validity of the measures in the new setting.
- Standardised scales may be designed for use with the general population and may not be sensitive enough when used with clients with complex or severe issues.

MacKeith and Graham (2007: 9-10) point out that developing a new tool (or adapting an existing tool) can help to engage stakeholders and build an outcomes orientation within an agency.

Goal scaling

Goal scaling (Rapp and Poertner 1992: 114; Baulderstone and Talbot 2004) is different to other approaches to outcome measurement in that instead of attempting to measure the client's 'condition', it measures the client and/ or practitioner's assessment of the degree of progress or goal attainment that has occurred. Goal attainment scaling is content-rich with specific goals and outcome expectations at the level of client and worker conversations. However, reported ratings are 'content free'

in the sense that they summarise the extent of progress but do not indicate the domain, focus of change, or how one client's progress or condition compares to another's. Standard goal scaling extends basic goal attainment scaling by linking progress ratings to specific goals from a standard goal list, thus providing information about domain and potentially locus of change. However, ratings are still relative to the client's starting point, thus making meaningful comparison between clients or services difficult.

Goal scaling approaches tend to integrate well into casework and are popular with service delivery staff as they have a natural fit with action planning. Their weaknesses in terms of comparison and lack of defined external anchors could potentially be reduced by combining them with other measures such as status maintenance and change scales (Talbot 2010).

6.4 POSSIBLE TOOLS AND MEASURES

This section lists a variety of tools and measures that are of potential utility to homelessness service agencies in measuring client outcomes. It is understood that while the lack of permanent safe housing is the common thread that links all people experiencing homelessness, homelessness services work with clients on needs, problems and goals in a wide range of life domains: physical health, mental health, relationship difficulties, legal issues, problematic substance use, and so forth. In the SAAP context, self-reliance includes a wide range of aspects ranging from accommodation, income and self-care through to connectedness and relationships with other people, and access to appropriate forms of support (Eardley, Thompson et al. 2008: 14). The discussion here therefore covers a range of domains.

The first two sub-headings (quality of life and 'wide-spectrum' homelessness tools) include tools which aim to establish a rounded picture of a person's situation across multiple domains. The remaining areas focus on measures and tools targeted to more specific domains and/ or client groups. However, many of the more specific tools are also influenced by 'whole person' models of health or wellbeing, and bridge a number of domains.

While the focus is mainly on complete measurement tools, in a few instances individual measures are included where they appear to fill a gap where I have not found a suitable existing tool. In general these measures are drawn from program evaluations or other academic studies.

Several sources provide a comparison or 'audit' of measurement tools across a range of dimensions (e.g. Stedman, Yellowlees et al. 1997; Spence, Donald et al. 2002; Campbell-Orde, Chamberlin et al. 2005; MacKeith, Graham et al. 2007; Anderson 2008). These provide a useful reference when considering the benefits of particular standardised tools. Another resource is Schalock's (2001: 156) list of generic measures which are applicable to a wide variety of contexts.

Quality of life

There are a variety of approaches to measuring Quality of Life (QoL), and many different instruments have been developed. Table 11 lists several QoL tools of relevance to homelessness services.

Table 11: Selected Quality of Life tools of relevance to homelessness services

Name of tool	Domains / areas covered	Measures	Notes
WHOQoL-Bréf	Physical health, psychological, social relationships and environment	26 items, 5-point rating scales, self-report	Produced by the World Health Organisation. A condensed version of the 100-item WHOQoL tool. Has been validated in many different cultural settings. Australian versions are available. An 8-item version is also available. Registration is required prior to use. (World Health Organisation 1996; University of Melbourne 2010)
Lancashire Quality of Life Profile (LQoLP)	General wellbeing, work and education, leisure and community participation, religion, finances, living situation, legal and safety, family relations, social relations, health, self-conflict	Mix of quantitative concrete questions (e.g. current accommodation) and 7-point satisfaction rating scales	Free to use. Information available from Institute of Psychiatry, Kings College, London.

Name of tool	Domains / areas covered	Measures	Notes
Quality of Life for Homeless and Hard to House Individuals (QoLHHI)	Health, health care system, place where you live or stay, living conditions, financial situation, employment situation, social and support services, recreational and leisure activities, spiritual life, romantic situation, family, and friends	Includes an impact survey (mix of category selection and 7-point rating scales) and MDT tool (4-point and 7-point rating scales) for each major domain	Can be used in modular form. (Hubley, Russell et al. 2009)
Manchester Short Assessment of Quality of Life (MANSA)	Life in general, health, work and education, finance, leisure, social, safety, living situation, family	43 items, combination of 7-point rating scales and yes/no questions	Developed by the Institute of Psychiatry, Kings College, London. Questions are framed to be applicable to a range of living situations including homelessness. (Priebe, Huxley et al. 1999)

Homelessness (wide spectrum)

A number of tools provide a broad overview of a person's situation across multiple domains relevant to people experiencing homelessness. Often the focus is on functioning – i.e. how well a person is 'managing' particular areas of their life. The CANSAS tool, while developed for use in mental health services, provides such a broad overview of relevant life areas that it is also included here.

Table 12: Wide-spectrum outcomes tools of relevance to homelessness services

Name of tool	Domains / areas covered	Measures	Notes
Homelessness Star	Self care and living skills, managing tenancy and accommodation, managing money, social networks and relationships, drug and alcohol misuse, physical health, emotional and mental health, motivation and taking responsibility, meaningful use of time, offending	10-point defined scale based on explicit journey of change	(MacKeith, Burns et al. 2008a)
BT Generic Outcomes Scale / Environmental Outcomes Scale	Housing, Govt allowance, Financial counselling / support, Gambling, Education / training, Employment, Incest / sexual assault, DV, Family / relationship, Pregnancy, Family planning, Living skills / personal development, Legal issues, Recreation, Health (General), Health (Mental Health), Emotional support / other counselling, Drug and alcohol issues, Material Goods, Community Resources, Safety	3-point importance rating scales, 5-point change rating scales	(Baulderstone and Talbot 2004)
Camberwell Short Assessment of Need (CANSAS)	Accommodation, Food, Looking after the home, Self-Care, Daytime activities, Physical Health, Psychotic symptoms, Information on condition and treatment, Psychological distress, Safety to self, Safety to others, Alcohol, Drugs, Company, Intimate relationships, Sexual Expression, Child Care, Basic Education, Ability to use telephone, Transport, Money, Benefits	22 items, 3-point category choice for each item (no need, met need, unmet need)	Developed by the Institute of Psychiatry, Kings College, London. A variety of versions are available oriented to research, clinical and forensic settings. CANSAS-P also available - self-administered version (Trauer, Tobias et al. 2008)

Name of tool	Domains / areas covered	Measures	Notes
Draft service outcomes return (substance misuse and homelessness)	Experience of homelessness, substance misuse, social support, activity, health and wellbeing, housing situation	Various - primarily yes / no questions and frequency rating scales	(Pleace 2008: 110-111)
Colorado Coalition for the Homeless Consumer Outcome Scales	Activities of daily living, legal status, quality of life, housing quality, employment / school, access to benefits, physical health, mental health, substance use, social resources, basic needs, family / household communication, youth risk, childcare	A few concrete questions, plus 7-point scales oriented to functioning and status, to be completed by practitioner during interview with client	(Post, Isbell et al. 2005: Appendix 1)
Outcomes field test - Montana, 2005	Life skills, mental health, substance abuse, family relations, mobility, community involvement	5-point rating scales oriented to status, resources and functioning	(Post, Isbell et al. 2005: Appendix 4)
ERoSH Tenant Self-assessment questionnaire	Quality of life, health, social networks and involvement, skills and hobbies, environment, dealing with finances and administration, cultural and religious needs	26 questions, 5-point rating scales Oriented to tenants living in sheltered housing.	(ERoSH 2010)

Housing

The literature review has not identified any existing standardised tools that focus on the domain of a person's experience of and functioning in relation to their housing. However, there is a substantial body of research and evaluation literature that has measured different aspects of housing outcomes. The key dimensions that tend to be measured are housing type, tenure, affordability, stability, and quality (adequacy and appropriateness) (Kolar 2004; Blunden and Johnston 2005: 1).

Housing status (e.g. primary, secondary or tertiary homeless, interim housed, permanently housed, housed at risk) can be used as an index drawing together a number of these dimensions. Housing status could be measured at a point in time, or as a proportion of time spent in various housing status categories during a particular period (Clark and Rich 2003: 79-80).

Table 13 provides some sample measures and approaches to measuring these dimensions of a person's housing situation.

Phibbs (2002: 4-5) provides a useful model of the different scopes to consider when assessing housing: tenure, dwelling, neighbours, area, and community.

Table 13: Example measures related to housing

Dimension	Possible measures
Housing type and tenure	Typically nominal status categories are used, in some cases an ordinal scale from 'worst' to 'best' housing options although this can be problematic as perceptions of what is better can vary and are context-dependent
Affordability	Typically considered as a proportion of a person's income – 30% is conventionally considered the limit of affordability (Kolar 2004) although for people with complex needs on low incomes, the figure may need to be lower
Stability	Stability has retrospective, prospective and perceptual components • Retrospective: housing mobility (e.g. number of moves in previous given time period) • Prospective: risk of tenancy breakdown – one approach is to use a list of 'warning signs' (e.g. arrears, subject to Notice to Vacate, has received breach notice etc) (Pleace 2008: 70-71) • Perception: rating scale – self-report of how stable a person believes their housing is (Kolar 2004)
Quality	Can measure the existence of specific negative aspects of housing (e.g. mould, infestation, etc - Pleace (2008: 70)); can measure overall satisfaction with housing (rating scale); or could develop a tool rating satisfaction with a range of aspects of housing (condition of the building, location, health impact, neighbour relations, etc)

HomeGround
Services

Psychological wellbeing

There is a vast array of standardised scales available in the area of mental health. Spence, Donald et al. (2002) provide a detailed list of measurement tools and scales covering both generic and diagnosis-specific aspects of mental health. NMHWG Information Strategy Committee Performance Indicator Drafting Group (2005) provides a general framework for performance measurement in the mental health field in Australia (including measurement of client outcomes).

Mental health tools can be divided into two main groups: clinically-oriented tools and recovery-oriented tools. Table 14 lists a number of clinically-oriented tools; the first five are the most widely-used in mental health services in Australia. Stedman, Yellowlees et al. (1997) field tested and compared these tools in a variety of settings. See also (Eagar, Buckingham et al. 2001; Siggins Miller Consultants 2003).

Within Victoria, certain tools are mandated by the Department of Health for collection of outcomes data by clinical and PDRSS mental health services. Clinical services are required to use the HoNOS and LSP tools, while BASIS-32, WHOQoL and CANSAS are recommended for use by PDRSS services (Department of Health 2009).

The short form (SF-36) is listed under the general health heading below rather than in this table.

Table 14: Selected clinically oriented mental health outcomes tools

Name of tool	Domains / areas covered	Measures	Notes
Behaviour and Symptom Identification Scale (BASIS)	5 sub-scales: Relation to self and others, daily living and role functioning, depression and anxiety, impulsive and addictive behaviour, psychosis	32 items, 5-point scales, client-rated	Not in the public domain - copyright McLean Hospital although an arrangement exists through DHS for use of BASIS-32 in PDRSS services in Victoria. (McLean Hospital 2010). BASIS-24 also available
Health of the Nation Outcome Scales (HoNOS)	Behavioural problems, impairment, symptomatic problems, social problems	12 items, 5-point scale, clinician-rated	(Buckingham, Burgess et al. 1998b: 112-118) HoNOSCA / HoNOS 65+ also available
Life Skills Profile (LSP)	Self-care, non-turbulence, social contact, communication, responsibility	39 items, 4-point scales completed by clinician or family member	LSP-16 (abbreviated version) also available (Buckingham, Burgess et al. 1998b: 119-122)

HomeGround
Services

Name of tool	Domains / areas covered	Measures	Notes
Role Functioning Scale (RFS)	Working, independent living and self care, immediate social network relationships, extended social network relationships		(Stedman, Yellowlees et al. 1997: 29)
Kessler 10 Scale (K-10)	Focus on psychological distress or symptoms, particularly of anxiety and depression / dysphoria	10 items, 5-point rating scales completed by client	Also 6 item and 15 item versions available. (Department of Health and Ageing 2003: 57-60)
Depression Anxiety and Stress Scale (DASS)	Depression, anxiety and stress	42 items, 4-point rating scales completed by client	(University of New South Wales 2010) Public domain. 21 item version also available.
General Health Questionnaire (GHQ)	Somatic symptoms, anxiety and insomnia, social dysfunction, severe depression	Available in 12-, 28-, 30- and 60-item versions; four-point scales (self-report)	Not public domain. Focus on non-psychotic psychiatric disorders. 60-item version covers physical illness as well.
Clinical Outcomes in Routine Evaluation (CORE) Outcome Measure	Subjective wellbeing, problems / symptoms, life functioning, risk / harm	34 items, 5-point scales, self-rated	(Core Systems Group 1998)

In recent years a range of recovery-oriented measurement tools have emerged. In contrast with clinically-oriented tools these tend to focus more on the client's experience and may cover a broader range of life areas. Campbell-Orde, Chamberlin et al. (2005) is a useful resource on recovery tools. Table 15 lists a few examples.

(Schalock 2001: 143-144) also considers a range of mental health oriented measures, and notes that mental health issues may have significant impacts on families and carers as well as the primary client. It would be useful to consider how these broader outcomes could be measured.

Table 15: Selected recovery-oriented outcomes tools

Name of tool	Domains / areas covered	Measures	Notes
Recovery-Enhancing Environment Measure (REE) Consumer Self-Report Tool	Various	28 items, select all that apply	REE is a suite of tools designed to assess organisational performance, includes a consumer self-report tool. Copyright Priscilla Ridgway. (Campbell-Orde, Chamberlin et al. 2005: 75)
Recovery Assessment Scale	Coping, empowerment and wellbeing, personal relationships	41 items, 5-point scales	(Siggins Miller Consultants 2003: 87)
Carers' and Users' Expectations of Services (CUES) Service User Questionnaire	Where you live, finances, use of time, family, social life, information and advice, access to services, choice of services, relationships with workers, consultation and control, advocacy, stigma and discrimination, medication	16 items, combination of three-category questions and open-ended qualitative questions	Plain language, clear presentation. (Rethink n.d.)
Milestones of Recovery Scale (MORS)	Stages of recovery	Single item, 8 categories (select one), clinician rated	Underlying concepts of level of risk, level of engagement, and level of skills and supports. (Pilon and Ragins 2007)
Recovery Advisory Group – Recovery Model Structure	Internal (cognitive, emotional, spiritual, physical), external (activity, self-care, social relations, social supports)	8 dimensions, choice of 6 stages of recovery for each dimension	Recovery Measurement Tool also developed based on this model. (Ralph n.d.)
Stages of Recovery Instrument (STORI)	Hope, identity, meaning, responsibility	50 items, 6-point scales	5-stage model of recovery. (Andresen, Caputi et al. 2006)

HomeGround Services

General health

General health measurement tools tend to cover a range of dimension such as mental health and social relationships in addition to physical health. Table 16 provides information on the SF scale, one of the most commonly used general health scales.

Table 16: General health outcomes measurement tool

Name of tool	Domains / areas covered	Measures	Notes
Short Form (SF)	Physical health (physical functioning, role-physical, bodily pain, general health), mental health (vitality, social functioning, role-emotional, mental health)	36 items, mostly 5-point scales	Multiple translations available. Not public domain. Shortened versions also available - SF-12, SF-8. (QualityMetric 2010)

In addition to generic scales, there are a wide variety of condition-specific scales oriented to particular acute or chronic health problems such as diabetes, asthma, cardiovascular disease, depression, and cancer (Post, Isbell et al. 2005: 22). However, it is suggested that while these scales might be of relevance to clinical services working closely with people experiencing homelessness, they are likely to be too specific and require too much expert knowledge to be used within generalist homelessness services. Health Care for the Homeless (1998) summarises the results of a range of pilot projects of health-related outcome measures with people experiencing homelessness.

The World Health Organisation framework of pathology, impairment, functional limitation and disability may provide a useful conceptual model for thinking about outcomes around specific disabilities. In this model, a functional limitation (such as low intelligence) becomes a disability only when it impacts on or interferes with a person's social role or functional level (Schalock 2001: 146). Such impacts are mediated by a range of factors including the person's physical and social environment. One way to measure progress may be in terms of the intensity of support needs, which should gradually decrease over time if services are effective (Schalock 2001: 147).

Addiction / substance misuse

Table 17 lists a number of assessment and outcome measurement tools that have a focus on substance use, and/or are specifically designed for working with people with problematic substance use issues.

Table 17: Substance misuse outcome measurement tools

Name of tool	Domains / areas covered	Measures	Notes
Maudsley Addiction Profile (MAP)	Substance use, health risk behaviour, physical and psychological health, and personal/social functioning	60 items, variety of scales (largely frequency-related)	(Marsden, Gossop et al. 1998)
Christo Inventory for Substance-misuse Services (CISS)	Social functioning, general health, sexual / injecting risk behaviour, psychological, occupation, criminal involvement, drug/alcohol use, ongoing support, compliance, working relationship	10 items, 3 point scales	Free to use. (Christo 2008)
Treatment Outcomes Profile (TOP)	Substance use over 4 weeks, legal issues, general and mental health	23 items, various scales	(National Treatment Agency for Substance Misuse 2007; Pleace 2008: 67, 71-74)
AUDIT alcohol dependence scale	Alcohol consumption, drinking behaviour and dependence, consequences or problems related to drinking	10 items, primarily 5-point scales	(Babor, Higgins-Biddle et al. 2001)
Severity of Dependence Scale (SDS)	Severity of dependence on opioids	5 items, 4-point scales	(World Health Organisation 2010)

Participation in society

Employment and educational outcomes are typically assessed using status maintenance and change scales (e.g. full time employed, part time employed, casual employed, looking for work, not in the job market). Children's educational outcomes measures might include school attendance / absenteeism and school retention.

Table 18: Measurement tools for services assisting people to obtain employment

Name of tool	Domains / areas covered	Measures	Notes
Work Star	Job-specific skills, aspiration and motivation, job-search skills, stability, basic skills, social skills for work, challenges	7 items, 10-point scales	(Burns and MacKeith 2009b)
The Employability Map	Motivation, lifestyle and social skills, basic skills, work related skills	5 items, 9-point scales	Currently only available within OSW network. (OSW 2004)

There is a larger picture a person's reintegration into the community, of which employment, education and leisure activities might be considered part (Busch-Geertsema 2005). I have not found standardised tools that measure degree of community participation, although Burke and Hulse (2002) used a range of measures of community engagement and participation.

Social wellbeing

Social wellbeing has a number of dimensions, covering for example family functioning and various types of social support (esteem support, informational support, social companionship, instrumental support) (Pleace 2008: 75).

Available tools include generic Quality of Life tools (which tend to include domains related to number and quality of relationships), or standardised sections of surveys such as the British Household Panel Survey (Pleace 2008: 76). Calsyn and Winter (2002) looked at the relationship between social support, mental illness and stable housing. The measures of social support that they used included: natural support (frequency of contact), perceived support (number of people available to help), satisfaction with support (feelings about social support), and support from professionals.

Legal issues

I did not locate any standardised tools in the area of assessing outcomes with respect to legal issues (whether offending behaviour or other engagement with the legal system). One approach is to ask people to indicate whether they have been subject to particular legal orders or events within the previous given time frame (e.g. Pleace 2008: 77).

Material wellbeing

Material wellbeing outcomes tend to be measured using concrete and status measures (change in gross income, change in benefit status). However, there are a range of other dimensions that are relevant including the extent to which people can afford basic necessities, and people's perception of their financial situation, in general or in relation to that of others (Pleace 2008: 79). The QoLHHI module on income provides a standardised set of items that provide good coverage of this territory.

Activities of daily living

Activities of daily living can include both self-care (personal hygiene, dressing and undressing, eating, continence, elimination) and instrumental activities of daily living (light housework, preparing meals and clean-up, taking medications, shopping, communication, literacy and numeracy, use of technology – e.g. telephone, ATM).

The literature review did not have time to investigate scales for measuring level of functioning in these areas. Table 19 provides information on one tool that has been reviewed in the literature on mental health outcomes measurement.

Table 19: Activities of daily living assessment tool

Name of tool	Domains / areas covered	Measures	Notes
Resource Utilisation Groups – Activities of Daily Living Scale (RUG-ADL)	Bed mobility, toileting, transfer, eating	4 items, primarily 4-point scales	(Department of Health and Ageing 2003: 40-41)

For further research: other assessment or outcome tools relating to activities of daily living

Safety

Safety is a significant concern of many homeless people, however the literature review did not identify any outcomes tools that focus solely on assessing a person's level of safety. However, vulnerability indices often include one or more items related to safety (Common Ground n.d.; Downtown Emergency Service Centre n.d.). Measurement of safety could focus on people's perceptions of their safety, and/or on a history of violent or abusive incidents.

For further research: personal safety assessment tools; consider literature relating to family violence, child abuse and neglect

HomeGround
Services

7. Measurement processes

7.1 SAMPLING

One decision point for organisations implementing client outcome measurement systems is what approach to take to sampling. In general, there are two approaches: to try to measure outcomes for all clients, or to select a subsample. Sampling all clients has advantages:

- It is likely to provide greater precision and validity in the results (Hatry 1997: 18)
- It can be logistically simpler in terms of workflow and management practice to survey every client routinely rather than (for example) one in five (Booth and Smith 1997: 38)
- If outcomes measurement is integrated with casework (e.g. in assessment, case review and action planning) then it is seen as a standard good practice element of service delivery with every client.

Larger samples do tend to support more sophisticated statistical analysis, especially where a population is divided relatively equally on points of interest or where differences are more subtle.

However, surveying every client also requires greater resources and effort. Particularly for services with high throughput and/or demand, it may be impossible within the given resources to attempt outcomes measurement with every client. Even where resources do potentially allow for data collection with the entire client population, this may be seen as wasteful (Booth and Smith 1997: 38).

If a sample is used, it is preferable that a probability sample be used if possible (Booth and Smith 1997: 38). Probability samples provide greater ability to generalise to an overall population (Trochim 2006b) – for example, to generalise from a sample to the entire group of current clients of a service or

organisation. For human service organisations, the sampling frame (list of clients of the organisation) may closely approximate the population of interest (all clients of the organisation). The question of whether any generalisation is possible to broader populations outside of the organisation (for example, to homeless men in the greater Melbourne area) is complex, and relates to the level of representativeness of the organisational client population with respect to the broader population. The fact that the organisation's client population is not a random sample of the broader population immediately reduces the likelihood of representativeness.

One issue that may arise for organisations with a diverse service portfolio is whether to sample on a whole-of-organisation basis, or a service-by-service basis. If services have widely varying numbers of clients (for example, a case management service with a group of 50 long-term clients as opposed to a drop-in service seeing thousands of clients per year), simple random sampling from a whole-of-organisation client list may lead to few clients being sampled from the smaller services. One possible response is to use stratified random sampling, with disproportionate samples drawn from small services, and to adjust for this over-representation in whole-of-organisation analysis (Trochim 2006b). If there are other sub-populations of particular interest to the organisation (for example, survivors of domestic violence or recent migrants) it would also be worth considering these groups within the stratified random sampling mechanism.

Sampling techniques are also complicated by the fact that organisational client populations are dynamic rather than static, and that in general it will be desired to monitor clients until they exit the service. This means that the sample will progressively change over time,

with some clients remaining in the sample from one reporting period to the next, and others being removed or added.

Due to the need to collect baseline data, it is necessary to select clients for inclusion in the sample at the point of assessment or intake to the service. This could be achieved in several ways:

- Using a 'quota' system where each service is always monitoring outcomes for a certain minimum number or proportion of their client population – as the number drops below the quota, more clients are selected into the sample
- Using a variation on a systematic random sampling approach (Trochim 2006b) where (for example) every 5th client accepted into a service is selected
- Using a time-limited snapshot (Burns and Cupitt 2003: 19) where all clients accepted into the service within a particular time period are selected

Each of these suggestions affects the randomness of the sample to some extent, and may therefore lead to some bias of the sample with respect to outcome variables.

There may be times when a nonprobability sampling approach might be appropriate (Burns and Cupitt 2003: 19). For example, purposive sampling with a heterogeneity focus (Trochim 2006a) might be used where the organisation is interested in information about the full range of client experiences and outcomes (including 'outlier' outcomes), as opposed to an understanding of average or typical outcomes. Regardless of whether random sampling is used, it is important to sample with clear defensible criteria. Convenience sampling should be avoided where possible.

HomeGround
Services

If a sub-sample is used, a decision is also necessary on the desired proportion of the total population to sample. Berman and Hurt (Berman and Hurt 1997: 87) suggest that, if faced with a choice between a high response rate in a single setting (e.g. a single service) and a low response rate across an entire organisation, it is preferable to concentrate on the single setting. A 20% sample across an entire organisation is likely to "provide distorted or incomplete information" that is not reliable as an aid to decision making.

Regardless of the sampling method chosen, some level of non-response is also to be expected (de Vaus 2002: 152). Baulderstone and Talbot (2004: 42) note that in the homelessness service context it will never be possible to obtain a 100% completion rate for outcomes measurement. Clients may decline to participate, may exit the service at short notice, or their circumstances may mean that it is "difficult or even inappropriate for workers to engage them in formal measurement processes."

7.2 CONSENT

Consideration is needed to the nature of consent that would be required to collect and use outcomes data. If outcomes measurement is seen as an integral part of service delivery, then a general consent to collection of personal information would cover outcomes data in addition to all other client-related information held by the organisation. It is unlikely that additional, specific consent is required for an organisation to collect outcomes data. In the context of Australian privacy legislation, use of client outcomes data for internal service management and quality improvement purposes would be seen as a directly related secondary purpose of the data collection and would also not require additional consent.

However, it would be expected that clients would be informed about the use, storage of and access to outcome measurement data (Baulderstone and Talbot 2004: 41).

Some clients may wish to receive a service, but to opt out of outcomes measurement processes. There is no clear-cut answer on how services should respond to this. Participation in client self-report measures is clearly voluntary, however some outcomes data may be generated by practitioners. As noted above, a non-response rate is to be expected for a variety of reasons; services may see 'opt out' as one reason. Where outcomes measurement is integrated within casework processes, but clients are offered the choice to opt out, staff will need clear guidance on what alternative processes are to be used for assessment and review.

Organisations may wish to communicate outcomes information externally, for example:
- Aggregated de-identified reports to funders
- Use of aggregated outcomes results in advocacy materials, tenders, annual reports etc.
- Exchange of individual client outcome information with service delivery partners involved with that client.

These forms of use may require additional specific consent from clients. Where clients provide general consent for exchange of information between two services or agencies, this may include exchange of outcomes information, however again best practice suggests that this should be made explicit as part of the informed consent process.

It should be noted that while information can be de-identified in aggregate reports, the stored outcomes data itself must remain identifiable in order to be able to track

progress for individual clients over time. Assignment of a unique identifier such as a statistical linkage key may assist in providing some level of de-identification during analysis processes.

7.3 TIMING AND FREQUENCY

At a minimum, outcomes data should be collected twice for each client – once at intake (or at the onset of a particular service), and once at exit or completion of the particular service (Rapp and Poertner 1992: 107), to give an estimate of total 'distance travelled'. Outcomes data can also be collected at multiple points during the period of service provision. Whether this is appropriate depends on the nature of the service and time frame of engagement. Crisis services are unlikely to have time to collect more than the pre- and post-service measures. Long-term services are oriented to goals that are cumulative and are likely to want regular updates on progress (Berman and Hurt 1997: 85-86).

Burns and Cupitt (2003: 26-29) suggest that regular monitoring is useful as it helps to combat attrition of the sample. If a client exits the service in an unplanned or rapid manner and the service is unable to collect exit data, the most recent set of regular monitoring data can be used instead. Burns and Cupitt suggest a frequency of between one and three months for regular data collection, depending on the service type.

Follow-up
Many sources recommend collecting follow-up data if possible, at some point after the client has exited the service (e.g. Rapp and Poertner 1992: 107; Hatry 1997: 18). There are two key reasons for this:
- There is evidence that positive gains made by clients while engaged with humans services programs may decay over time (Wells and Johnson 2001: 180)

- Some long-term outcomes may only be visible or measurable at some point after the completion of service delivery.

However, there are considerable logistical difficulties involved in tracking clients. The level of transience of many people experiencing homelessness means that it can be very difficult to locate them one they have disengaged from the service (Post, Isbell et al. 2005: 12). Researchers recommend that considerable effort be put into recording as much contact information about the client as possible, and seeking consent to contact others who may be able to assist in locating the client (Pleace 2008: 62).

Collecting follow-up information can be a time consuming and costly exercise (Hatry 1997: 18). It is questionable whether service delivery staff would have time to undertake this function; if not, specialist staffing resources would be required for data collection. Another option is 'remote monitoring' – receiving information from other parties involved with the former client about their progress or about the occurrence of particular positive or negative outcomes. For example, a service provider could be notified by a public housing provider if a former client was allocated a property, or abandoned their tenancy. However, this approach creates substantial ethical and legal issues in relation to privacy and data protection (Pleace 2008: 79).

There are also some technical issues in relation to the collecting and analysing follow-up data:

- The possibility of selective attrition – this could be in a number of directions. Clients with worse outcomes are more likely to be transient and harder to find. However, clients with good outcomes are less likely to re-contact the service of their own accord than clients who are still in need (Post, Isbell et al. 2005: 12; Pleace 2008: 62)

- The optimal follow-up interval is not necessarily obvious. Longer intervals may demonstrate more significant changes (whether positive or negative), however longer intervals may also increase the chance of attrition and make the connection between intervention and outcome more tenuous. Hatry (1997: 18) suggests 3, 6, 12 or 24 months as typical follow-up intervals.

7.4 GATHERING AND RECORDING DATA

The two key questions in relation to collection of outcomes data are:
- *Who* will collect the data?
- *How* will they collect the data?

Who will collect and record the data?

Decisions about who should collect data are closely related to what measurement tools are used – client self-administered questionnaires will obviously be completed by the client, while some level of functioning scales are designed to be completed only by the practitioner. However, many tools are designed to be completed in an interview-type situation, with discussion between the client and another party.

In deciding who will be involved in such conversations, human services organisations have two basic options. The first is to use service delivery staff such as case managers or housing workers; the second is to use specialist data collection personnel.

There are several advantages to having service delivery staff collect outcomes data. These staff have an existing relationships with clients. Within the context of such a relationship, clients may be more willing to participate in outcomes measurement, and may also be more willing to discuss more personal or sensitive aspects of their lives. Where outcomes

measurement involves a conversation between client and worker, this conversation may lead to a more balanced picture of an individual's situation or progress – challenging overly negative or unrealistically positive views of outcomes. These conversations can provide workers with a valuable assessment of a client's current situation, can feed directly into case planning and review, and can also be a transformative element of casework in themselves (MacKeith and Graham 2007: 12). Finally, having service delivery staff collect outcomes data is an inexpensive option as it does not require employment of additional staff (Pleace 2008: 79).

Despite these advantages, there is a downside to having service delivery staff collect outcomes data. One problem is the impact of data collection on the service delivery role; every additional data collection requirement impacts on time available for other aspects of service delivery (Rapp and Poertner 1992: 107). Another disadvantage is the potential for data quality to be lower than that collected by specialist personnel. Where measurement is undertaken by a large number of staff with varying levels of skill, this can reduce the reliability of the data; regular thorough training is likely to be required to ensure a minimum level of consistency in use of the measurement tools. There is also the potential for validity to be affected by collector bias. Where the measurements recorded can result in consequences for the individual or service collecting the data (for example, more or less favourable treatment of particular staff members or clients; continuation or withdrawal of funding for a service), there is a risk that this may lead to falsified or fraudulent reporting (Rossi 1997: 31-32; Pleace 2008: 60). Even where no overt consequences exist, staff members may be influenced by the context and goals of their relationship with particular

clients, and may see improvement where none exists, or place high value on relatively minor shifts (Weiss 1972: 41).

Some authors suggest the use of audits as a means of monitoring collector bias and discouraging fraudulent reporting (Rossi 1997: 32; Pleace 2008: 60). However, auditing and inspection require additional resourcing and may only detect bias in a haphazard fashion.

Another approach is to use specialist personnel for collection of outcomes data. For example, an organisation might employ research staff whose role includes collection of outcomes data. This may potentially increase reliability and validity, but it does require substantial extra resourcing, especially for larger organisations who may wish to collect outcomes information for thousands of clients. Data collection tools that are more complex or that require more technical knowledge to administer may also require specialist personnel for effective use.

Another approach might be to make use of trained volunteers or client representatives, either for data collection or data entry (Rapp and Poertner 1992: 107). Involvement of client representatives would require consideration of privacy issues and awareness of pre-existing relationships among the service's client group.

At a broader level, there are service system issues around collection of outcomes data where clients are involved with multiple services or agencies. To avoid over-surveying, it may be appropriate to negotiate (or have protocols in relation to) which service(s) will monitor outcomes in such situations.

How will data be collected and recorded?
Anecdotal evidence suggests that it is likely to be easier to regularly and systematically

collect practitioner-rated outcomes data than client self-report data (Love and Scott 2010). Completion of practitioner-rated tools is typically quicker, more flexible in timing and more under the control of service-delivery staff than completion of client self-report tools. However, in the homelessness service environment, client perceptions of progress tend to be highly valued.

Client self-administered questionnaires tend to suffer from low response rates. Many client self-report measures can best be completed through an interview or conversation between client and worker. This can address literacy barriers, clarify the meaning of particular items and enhance the casework process. However, this requires allocation of sufficient time for the conversation (which may be a challenge for staff with high caseloads), and requires an additional level of competence from the practitioner to be able to facilitate conversations about outcomes.

These challenges point to the need for services to provide adequate training for staff collecting data, and to think carefully about how data collection will fit into other service delivery processes. Baulderstone and Talbot (2004: vii) concluded that outcome measurement "is only feasible and practical where it is properly integrated with the case management process." Outcomes conversations have a natural fit with intake, exit and case review processes and should ideally occur at these points rather than as a separate 'add-on'. Similarly, Burns and Cupitt (2003: 19) suggest that outcomes monitoring should be integrated into the organisation's other monitoring systems. Outcomes questions are more likely to be completed regularly and well where they are part of the overall service monitoring system rather than an additional parallel system.

Outcomes conversations may occur in person or over the phone. Cultural considerations are relevant to how such conversations are conducted. For example, with indigenous clients it may not be appropriate to sit down and work through a paper-based or electronic survey; data collection may involve an informal conversation which is later recorded on paper by the staff member (Hamilton 2010).

In general efficiency of recording is a key consideration, and data entry directly into a database (rather than onto paper) can save duplication of effort (Barr n.d.: 19). Some software packages allow for automated data entry; these typically rely on specially designed paper-based forms which are completed by hand and then optically scanned by a scanner or fax (Berman and Hurt 1997: 88-89). If the forms are customisable, these systems are potentially quite flexible and efficient, however the cost and/or development effort may be beyond the reach of many organisations.

7.5 IT SYSTEMS

IT systems are vital to the efficient management, analysis and reporting of outcomes information. It is not uncommon for organisations to begin implementing outcome measurement processes, and then to find that their client data systems do not have the capacity to store the information being generated (Parkinson 2005: 4; Hendrick 2010b). Parkinson (2005) provides an overview of options for organisations in managing outcomes data, including modifying existing data systems, developing new systems or using off-the-shelf products. Each option has advantages and disadvantages.

The nature of the outcomes and other data to be collected will have a major impact on the design of data system. Ideally, clarity should be

achieved on reporting requirements, measures to be used and data collection processes, prior to commencement of development (Parkinson 2005: 15-16).

Gaining an understanding of stakeholder requirements is a crucial early step. Hudson (1997: 74-77) argues that many data systems are designed with insufficient attention to the day-to-day realities of practice and the needs of clients and service delivery staff. In the worst case this may result in a "DINO" (data in, nothing out) system that provides little useful information for staff. Staff buy-in is crucial to data quality. Availability of real-time reporting on outcomes or other data for individual clients or groups of clients greatly enhances the value of the system to staff and clients (Berman and Hurt 1997: 87). Early consultation with stakeholders is likely to increase buy-in and lead to a set of specifications that more fully meets the organisations' needs (Parkinson 2005: 18-19).

Berman and Hurt (1997: 87) articulate several other key criteria for client data systems, including data entry mechanisms that minimise impact on staff time and maximise accuracy and throughput; management of data capture with checks and reminders; and storage in a way that allows aggregate analysis and reporting.

Parkinson (2005: 30-31) provides basic spreadsheet designs for capturing hard and soft outcomes data. There are a number of existing software packages designed for the human services which are oriented to capture of client outcomes data. For example:

- "Efforts to Outcomes" from U.S. company Social Solutions integrates case management and outcomes recording functionality (Social Solutions 2010)
- The Outcomes Star System is the online version of the Outcomes Star and includes a free web-based database for capturing and reporting on Outcomes Star data (Homeless Link 2010).

8. Use of outcomes data

8.1 REPORTING

Client outcomes may be aggregated and monitored at a variety of levels including the individual client or household, service delivery team, service, program (which may be delivered across a number of organisations), organisation, and geographic region (NMHWG Information Strategy Committee Performance Indicator Drafting Group 2005: 19). Separate aggregation chains can be identified based on geographic areas and on the functional organisation of the service system; these are diagrammed in Figure 2. Population outcomes are generally referenced to geographic aggregation at the levels of Region or above. Service provider organisations will generally be unable to collect population data (although they may be able to monitor census data aggregated by other organisations), and will be unable to aggregate beyond the level of service or organisation on the function-based aggregation chain.

Figure 2: Function-based and geography-based aggregation chains for client outcomes data. Arrows point in direction of increasing aggregation
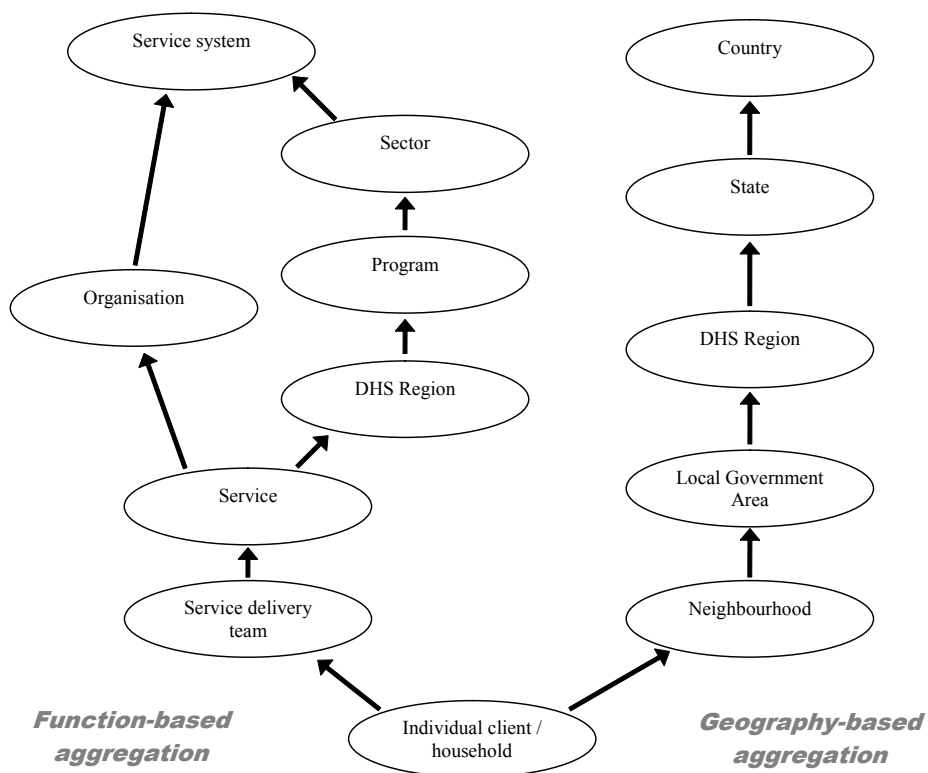
A variety of stakeholders will also be interested in the findings from outcomes measurement, including clients, service delivery staff, organisational management, Boards, service delivery partners, funders, and the broader public (Office of Housing n.d.: 6). Systems of feedback and/or reporting may need to be tailored to the different needs of these stakeholder groups; different report formats and levels of detail will be appropriate for different groups (Wells and Johnson 2001: 193). Many sources suggest that accessibility of outcomes results to staff is essential to maintain commitment to data collection and to improve practice (e.g. Hudson 1997; Clements 2010; Elizabeth 2010).

Analysis and reporting of outcomes data typically relies on specialist staff (Schalock 2001: 35), and can be a resource-intensive activity in itself. Trade-offs may be required between the number of different ways in which data is de-aggregated, the number of different reports produced, and the level of detail of each report.

Many sources support the importance of regular outcomes reporting, with timing of reports linked to the organisation's planning and performance management timetable so that up-to-date outcomes information is available to support decision-making (Schalock 2001: 35; Wells and Johnson 2001: 193; Burns and Cupitt 2003: 34). MacKeith and Graham (2007: 16) suggest that after a number of periodic reports have been produced, it may be possible to establish benchmarks for performance. They also recommend that data be analysed more comprehensively on a less regular basis (e.g. every 1 – 3 years) to provide deeper insight into how outcomes related to interventions, and how outcomes compare across client groups and service types.

Report formats

Outcomes reports should be designed for maximum clarity of communication. Guidelines for report design include (Rapp and Poertner 1992: 103; Burns and Cupitt 2003: 33; Barr n.d.: 27-31):

- Keep it simple – highlight the key information (particularly key trends) and avoid irrelevant detail
- Use well-designed graphs to enable comparison of results over time and across groups
- Place outcomes in context of the client group and service environment (especially for reports intended for external audiences)
- Explain the meaning of numerical scores or particular categories when reporting results on particular scales
- Comment on the reliability and validity of the data and note any factors which might affect this.

Use of standardised formats for reports will make both the production and use of reports more efficient. Two typical standardised formats are report cards (e.g. for inclusion in an annual report), and longitudinal comparisons (Schalock 2001: 35).

Hatry (1997: 14-15) provides an example of a tabular report format comparing actual vs. target outcomes on a range of measures for a single service or organisation; this is adapted as Table 20.

Table 20: Sample tabular format for comparison of actual and target outcomes (adapted from Hatry (1997: 15)). + indicates actual performance better than target, - indicates actual performance worse than target

Outcome indicator	Last period			This period		
	Target	Actual	Difference	Target	Actual	Difference
Percentage of children returned to home within 12 months	35	25	-10	35	30	-5
Percentage of children with more than two placements within the past 12 months	20	20	0	15	12	+3
Percentage of clients reporting satisfaction with their living arrangements	80	85	+5	80	75	-5

The Outcomes Star chart provides an example of a graphical report format for individual client outcomes data, enabling comparison of results for one client at different time periods (MacKeith, Burns et al. 2008a: 13).

Client access to outcomes information

Outcomes information can be shared with a client in two ways: discussion of their individual outcomes, and discussion or provision of de-identified aggregated outcomes data (for a service or the organisation). Both approaches may have benefits in motivating clients to continue providing outcomes information (Burns and Cupitt 2003: 27).

Under Victorian privacy legislation (Information Privacy Act 2000 - Information Privacy Principle 6), clients have a right to access and correct personal information that the organisation holds about them. This would be expected to include their outcomes data, and it is unlikely that any of the exceptions to the access right under IPP 6 would apply to outcomes data. However, in some circumstances service delivery staff may be reluctant to share outcomes data with clients, particularly in the case of practitioner-rated measures. Staff may perceive that sharing the

HomeGround
Services

ratings would constitute negative feedback to the client that may impact on the client's self-esteem or motivation, and may jeopardise the therapeutic relationship. Sharing outcomes ratings, particularly where they differ from a client's perceptions of themselves, may also involve a lengthy conversation requiring considerable skill on the part of the practitioner (including for example, motivational interviewing skills). Some staff may feel a pressure to rate more positively than they otherwise would, if they know that ratings will be shared with clients (Love and Scott 2010).

Given the right of clients to access their outcomes information, it appears important that organisations provide clear guidance and training for staff about how outcomes ratings are to be generated and shared with clients.

8.2 ANALYSIS OF AGGREGATE DATA

Methods for analysis of outcomes data can range from basic to sophisticated, depending on the questions that the organisation wishes to answer, and the resources available.

An initial consideration for many organisations is reconciliation of data from multiple data collection systems, including multiple client databases. Some form of unique identifier such as a statistical linkage key is essential to linking together data on a single client held in multiple repositories. Often a client may have contact with more than one service within an organisation, and conventions are needed for the process by which data is combined and discrepancies are resolved.

Once individual client data has been cleaned, aggregate analysis can take place. For many purposes, simple percentage-based analyses are sufficient. Percentages are typically calculated by dividing the number of clients

to achieve a particular outcome in a given period, by the total number of clients of the service who were in the "target population" for that outcome during that period (Friedman, DeLapp et al. 2001b: 3.7-3.8; Spellman and Abbenante 2008: 24-26). The target population may be the entire client group of the service, or a sub-group. For example, if the desired outcome was that an individual maintain stable housing for at least 6 months, clients who had been engaged with the service for less than six months would not be included in the target population for calculation of the percentage of clients who achieved the outcome.

Even with simple percentage-based analysis, a number of decisions may need to be made:

- Are point-in-time or improvement-over-time findings more useful? Point-in-time measures reflect the proportion of the client population who meet or exceed some target in the period – for example, the proportion of clients in permanent housing in a particular 3-month period. Improvement-in-time measures reflect the proportion of clients whose outcomes data for a given period showed improvement compared to some previous point in time (for example, compared to their outcomes data for the previous period, or their baseline data) (Friedman, DeLapp et al. 2001b: 3.7)

- Is the target population drawn from all clients of the service during the period, or only those exiting? Basing it on the total client population (including clients who continue with the service as well as those who have exited) provides a much broader picture of the types of changes that are occurring for clients of the service. However, analysing outcomes for clients who have exited the service helps to focus on the "net effect" of the service, comparing the client's situation before and

after the period of engagement, rather than the details of the many 'ups and downs' that may occur during the period of engagement. It may be useful to undertake both analyses.

There is some debate in the literature about the reliability of simple change scores (assessment rating at time 2 minus assessment rating at time 1) as measures of change. Some sources recommend the use of more complex methods including analysis of variance and covariance and regression techniques, however Stedman, Yellowlees et al. note a growing body of literature suggesting that simple change scores are valid indicators of change (1997: 19-20).

Where a series of data points are available over time (for example, a series of outcome ratings for an individual client, or a series of percentage values for a client group), more sophisticated analyses become possible including calculation of mean, median, range and standard deviation, and identification of trends over time (Bloom, Fischer et al. 2006: 540, 555-559). Barr (n.d.: 26) notes that calculation of mean values becomes more reliable with a greater number of data points, and recommends a minimum of 20 data points if possible. However, even with a smaller number of data points, where statistics are triangulated with other forms of evidence (including the knowledge of practitioners) they may provide useful information (Friedman, DeLapp et al. 2001b: 2.11).

Identification of 'desired', 'typical' and 'undesired' ranges for particular measures may be useful, and there are a variety of techniques for generating these, including statistical process control methods such as the X-moving-Range Chart (Bloom, Fischer et al. 2006: 582-583).

Tests of statistical significance may be of relevance to outcomes analysis in several ways:

- In comparing a series of outcome results over time, tests of statistical significance can help to establish whether a particular data value is likely to represent a genuine change in the mean of the results, or may be due to chance (Bloom, Fischer et al. 2006: 582)
- When aggregate outcomes results are based on a probability sample from the total client group, tests of statistical significance can help to establish whether the aggregate results can be generalised to the broader client population, or may be affected by sampling error (de Vaus 2002: 169-171).

It is important to balance consideration of statistical significance with clinical or practical significance. Clinical significance refers to the extent to which the parties involved in the service delivery situation (including client, service delivery staff, carers) believe that meaningful change has occurred (Bloom, Fischer et al. 2006: 516). Clinical significance is usually referenced to the expectations and goals of clients, and/or to comparison of the client's functioning against some standard (Stedman, Yellowlees et al. 1997: 19-20). Outcomes results may show statistically significant change but not clinically significant change, or vice versa.

Where samples are used, it will usually be valuable to test for and adjust for sample bias (de Vaus 2002: 152-165). In particular, in the homelessness setting it will be useful to look at non-response groups, including those who declined to participate in outcomes measurement and those where it was not feasible to collect the data for other reasons. Further investigation might show particular patterns of outcomes for these groups.

One issue that may arise is the extent to which diverse measures can be combined into a composite index giving an overall rating of change or current situation. Indexes might be used in a variety of situations:

- Some measurement tools calculate an overall score (e.g. an overall risk, severity or complexity rating) based on combining values from a variety of measures
- Organisations might wish to combine a number of measures within a single domain into an overall index for that domain – for example, to combine measures relating to housing tenure, affordability, stability and quality into an overall 'proportion of clients whose housing situation improved'
- It might also be possible to combine measures across different domains to give a sense of overall improvement or deterioration in the situation of a person or group, or overall level of success of a service in achieving positive change.

The advantage of an index is its communicative power – its ability to summarise complex information in a form that can be quickly communicated and understood. However, indexes can also be problematic. Their validity depends on a range of design issues including the choice of measures to include, avoiding repetition of the same factor or dimension, decisions on the relative importance of different measures, and statistical procedures to represent the relative values of measures that use different scales. Indexes may also over-simplify complex situations and mask the upward and downward movement of separate indicators (Weiss 1972: 37; Ware 1997: 63-64).

It is useful to consider what other information may be combined with quantitative analysis of client outcome measures. Process measures may illuminate some of the service delivery factors that impact on client outcomes. Where

qualitative information is available from staff or clients, this may provide a much richer understanding of the process of change and of barriers and enablers to positive outcomes. Qualitative information will need to be themed which can be a complex process in itself (Burns and Cupitt 2003: 32).

Recognising client and service differences

In addition to the type and quality of the service intervention itself, a wide variety of other factors influence client outcomes. These 'outcome predictors' include factors related to the client (e.g. socio-demographic variables, co-morbidities and personal biography) and factors relating to the environment (e.g. level of resourcing of the service itself, accessibility of other resources and supports in the local community) (Berman and Hurt 1997: 86).

Inclusion of these variables in outcomes analysis is important for several reasons:

- When comparing outcomes results from services of the same type, it is important to be able to adjust for differences in the client groups served by the services. In particular, it is useful to adjust for the severity and complexity of clients' presenting issues. This enables fair comparisons across services and also avoids creating a disincentive for services who work with clients who are less likely to achieve desired outcomes (Hatry 1997: 14-16; Spellman and Abbenante 2008: 41). Adjustment of outcomes results with respect to client need, complexity and/or severity of issues is usually referred to as 'casemix adjustment' or 'risk adjustment' (Booth and Smith 1997: 41-42; Spellman and Abbenante 2008: 41).
- Comparing outcomes across different sub-groups of the client population of a service or organisation can help to identify differences in the effectiveness of services

for particular groups (Johnson 2010). Some client groups may show unusually positive or negative results on a particular outcome indicator; this can provide a prompt for further investigation and for development of new service approaches (Berman and Hurt 1997: 86; Hatry and Lampkin 2003: 25).

While casemix adjustment is important, casemix methodologies are still in development (Booth and Smith 1997: 42). Casemix adjustment depends on a complex set of decisions regarding which characteristics are chosen as the basis for adjustment, and how they are weighted (Spellman and Abbenante 2008: 43). As these decisions affect how the performance of individual services is viewed relative to others, they can become a point of conflict (Rossi 1997: 26).

Buckingham, Burgess et al. (1998a) undertook a major study that developed the first version of a casemix classification system for specialist mental health services in Australia. The study focused on the relationship between client need and cost of service provision. While the study demonstrated that it was possible to identify a set of empirically validated casemix categories, both the process and the resulting set of categories were complex, and it was not always easy to separate client from practitioner variation (p.264). No single measure was found suitable to determine need, and the classification used a combination of instruments and measures that assessed demographic, clinical and level of functioning characteristics (p.271). The study concluded that accurate measurement of client need would require the introduction of new measurement tools. However, one of the advantages of casemix adjustment for outcomes is that (depending on the outcome measures chosen) the baseline outcomes data itself may also provide the required data for casemix categorisation.

Hatry and Lampkin (2003: 18) suggest that it will typically be useful for human service organisations to analyse outcomes data with respect to client factors including:
- gender
- age group
- race/ethnicity
- income group
- type of disability
- educational level
- housing status.

Personal biography, specifically a person's pathway into homelessness, has also been shown to have a profound impact on experiences while homeless and the type of outcomes that may be expected (Johnson, Gronda et al. 2008).

It is also relevant to consider service factors such as:
- the specific team, office or facility
- the specific types of service provided (did the client receive support plus transitional housing, or support only?)
- the amount of service provided (e.g. number of hours or visits).

Hatry and Lampkin (2003: 18) also suggest that reports for use by Coordinators or Managers within individual teams could be de-aggregated by the specific caseworker, however this may have potential disbenefits including increasing the temptation for staff to falsify data.

HomeGround
Services

Hatry and Lampkin (2003: 25) provide a sample tabular format for de-aggregating data by client characteristics; this is adapted as Table 21. Hatry (1997: 16) provides a sample tabular format for comparing outcomes by service unit and difficulty of problems at intake; see Table 22.

Table 21: Sample format for summarising results on a specific outcome indicator by demographic or service characteristics (adapted from Hatry and Lampkin 2003: 25)

Respondent characteristics	Clients Reporting Change in Their Problem Condition from the Time They Entered Service to Time of Measurement				
	Total responding (n = 625)	Extent of Improvement (%)			
		None (n = 50)	A Little (n = 83)	Somewhat (n = 429)	Considerable (n = 63)
Sex and race					
White male	265	7	11	71	10
White female	284	8	13	68	11
Non-white male	36	11	30	53	6
Non-white female	40	10	17	65	8
Age					
18-34	272	13	16	58	13
35-49	125	6	11	75	8
50-64	105	3	11	80	5
65 and over	123	6	11	74	9
Total	625	8	13	69	10

Table 22: Sample format for comparing outcomes by organisational unit and difficulty of pre-service problems (adapted from Hatry 1997: 16)

Difficulty of problems at intake	Percentage of clients who maintained stable housing for at least 6 months after housing allocation			
	Service Delivery Unit 1	Service Delivery Unit 2	Service Delivery Unit 3	Total
Minor	52	35	56	47
Moderate	35	30	54	39
Major	58	69	61	63
Total	48	44	57	50

Triangulating with other data sources

It will often be useful to combine quantitative outcomes data with other information including data about processes, resources and the broader context. Triangulation helps to enhance the credibility of findings, and to build a much deeper understanding of underlying processes and causal factors. Sources of additional data may range from anecdote and opinion of staff and clients, to outputs from formal program evaluations.

8.3 USING THE FINDINGS

One of the least explored topics in the literature is how outcomes information can best be used by organisations.

Service development and planning
An important use of outcomes information is in service development. Both positive and negative results can provide valuable prompts for the evolution of particular interventions (Hatry and Lampkin 2003: 27). Some organisations routinely use an outcomes-focused approach to planning, including reflecting during strategic or business planning sessions on client outcomes data (Burns and Cupitt 2003: 33; Talbot 2010). However, the loop of feedback from outcomes information to organisational and service change can also be part of a broader, ongoing process of continuous quality improvement (Burns and Cupitt 2003: 34) as management and staff identify opportunities for improvement.

Hatry and Lampkin (2003: 28) suggest holding "How Are We Doing?" sessions with managers and staff using outcome reports as a basis for discussion. The group can explore why certain functions have been going well and how successful strategies can be extended to other areas. The group can also attempt to identify reasons for poorer outcomes and suggest ways

to improve the situation. Team meetings could be one forum for such discussions.

Schalock (2001: 35-37) notes that many factors impact on the extent to which organisations actually implement improvements in response to findings of evaluation or outcomes monitoring. These include the organisation's "personality", operating environment and other priorities. Success factors for internal utilisation of outcomes findings include level of stakeholder engagement, credibility of the results, authority and leadership, and level of resourcing for implementing changes.

Outcomes data may also be used for advocacy and tendering purposes. Burns and Cupitt (2003: 35) suggest that apart from making funding applications more effective, sharing outcomes results with funders can improve their understanding of the nature of the client group and the process of change, and help to build realistic expectations about what can be achieved with the available resources.

Using measures in individual casework

Relatively little has been written on the way in which outcomes data can be used to support individual casework. Literature in the area of mental health suggests that it may be useful to compare current with previous episodes of service delivery to identify patterns of change: improvement (greater positive change from admission to discharge than in previous episodes), stability (same level of change from admission to discharge), deterioration (smaller level of improvement per admission), and to understand patterns of relapse (Department of Human Services 2009: 60-61). This model could be extended to understand patterns of change for clients with high levels of transience who frequently present within the homeless service system.

Dealing with unexpected or negative findings

Aggregated data sets are likely to show unexpectedly high or low outcomes for some indicators, either across an organisation or for a particular service or client group. Some of the most useful learning can stem from these outlying values. Hatry and Lampkin (2003: 23) advise that organisations should establish a routine process for examining the findings from outcomes reports and identifying unusual values. Further investigation will often be required to understand whether these values represent inconsistencies in the measurement process, or in fact represent unusual outcome patterns experienced by clients. If it does appear that the data represents unusually positive or negative outcomes, organisations will generally be keen to understand the factors that may have contributed to these outcomes.

Hatry and Lampkin (2003: 23) suggest the following strategies for investigating unusual outcomes results:

- Discussing the results with staff and asking for their interpretation of why outcomes outperformed or under performed expectations
- Forming a staff working group to examine the reasons for the observed outcome levels
- Consulting with clients about the outcomes results, for example via a focus group
- Conducting a more formal research or evaluation process in conjunction with an appropriate academic partner organisation or consultant.

It is possible that outcomes results may be perceived by some stakeholders to reflect poorly on the effectiveness of the service or organisation. This may potentially impact negatively on staff motivation. The literature offers little advice on processes for handling

such situations. However, Burns and Cupitt (2003: 32) note that negative outcomes do not necessarily reflect the level of effort and skill of staff, but may be connected to a variety of organisational and external factors. The following questions may help to identify some of these factors:

- Did the level of resources, for example, the level of money or staffing devoted to the project, affect the outcome?
- Was the outcome affected by the way you delivered your service (for example, by service policy, procedures or particular interventions selected)?
- Did external factors, such as lack of housing stock, prevent you achieving your outcome?
- Were the aims as originally stated realistic or appropriate, given the nature of the client group or the services delivered?

It may also be useful to consider the distinction between program failure and theory failure (Weiss 1972: 38). Program failure occurs when the program activities did not achieve the proximate goals or conditions that are understood to lead to the desired outcomes (for example, program activities did not lead to clients being linked with health services, and therefore did not improve clients' health status in the longer term). Theory failure occurs when the program did achieve the expected proximate goals or conditions, but these did not lead to the desired final outcomes (e.g. the health service linkages established did not in fact lead to any improvement in client health). Distinguishing these two explanations for negative outcomes requires that interventions are in fact based on a clear theory of change, and that data is available to establish whether interim conditions were met or not.

9. Process of introducing outcome measurement

Introduction of outcomes measurement within an organisation can mean significant changes for many areas of the organisation. MacKeith and Graham (2007: 14) comment that organisations typically underestimate the magnitude of impact of outcome measurement on service delivery (both in changes to practice, and in benefits). It can be a major cultural shift for an organisation. Outcomes measurement also needs to integrate with service planning and review activities, and with IT systems.

There is no definitive roadmap for introduction of outcomes measurement, and it is likely that every organisation's journey will be unique, with much learning along the way (Hendrick 2010b). However, there are also common themes in the literature around key processes that support implementation, and some of these are summarised here.

9.1 STAKEHOLDER INVOLVEMENT

Most authors emphasise the importance of involving a range of stakeholders throughout the process of implementation of outcomes measurement, including the early stages of development of the system. This is seen as having a number of benefits:

- Involvement of stakeholder increases "buy-in", i.e. level of stakeholder commitment and understanding, and therefore increases the chances of successful implementation (Wells and Johnson 2001: 177-178). In particular, Burns and Cupitt (2003: 36) advocate that staff should be "brought on board early on, rather than be presented with a monitoring tool and required to gather information."
- Stakeholder perspectives are vital to defining the types of outcomes that should be measured. Stakeholder involvement helps to ensure that stakeholders'

informational needs can be met. (Wells and Johnson 2001: 182)
- Outcomes-based accountability can involve shifts in agency culture at all levels (Wells and Johnson 2001: 177); stakeholder involvement helps facilitate this cultural shift and allows space for concerns and issues to be raised and worked through.

There can be a tension, however, between inclusive process and speed of implementation: "Many states and localities are using an inclusive process in the development and implementation of outcome initiatives. For example, they begin by organising work groups or engaging a broad advisory committee. The more people involved, the longer the process will take. Conversely, the fewer the people, the more quickly the process may proceed. However, without strong commitment and understanding, the process may be derailed by lack of agreement" (McDaniel 1996).

Bottom-up vs. top-down implementation
It is evident that both bottom-up and top-down approaches can be used to the development of desired outcome statements and measures. Bottom-up approaches start with individual teams or service units discussing their goals and valued outcomes, and how they might measure achievement of these (Hendrick 2010a). A key advantage of bottom-up approaches is that they tend to generate a lot of enthusiasm and commitment from service delivery staff, who have a strong sense of ownership of the outcomes being measured.

Top-down approaches tend to be driven centrally by management, defining overarching organisational outcomes and indicators (perhaps with reference to desired population outcomes) and then looking at how these can be devolved to individual service streams, and then to individual teams, often becoming more

specific in the process. A key advantage of top-down approaches is that they can provide a consistent language and conception of high-level outcomes, with the assurance that outcomes information gathered from individual service units will then be able to be easily aggregated and fitted into this framework. Burns and Cupitt (2003: 15) note that large organisations often find it useful to identify a set of high-level outcome areas common to all services, to provide this type of overall framework.

The two approaches are not mutually exclusive, and tend to be combined in varying degrees in implementations of outcomes systems. The approach used may also fluctuate over time in response to the stage of implementation. For example, Melbourne Citymission found it useful to begin in 2008 with a bottom-up approach, to introduce the concepts of outcome measurement to frontline staff and generate statements of key outcomes for particular services. The balance has now shifted to incorporate more top-down work, looking to standardise a core set of outcomes and measures to enable aggregation of data across the organisation (Hendrick 2010b).

9.2 KEY STAGES IN IMPLEMENTATION

A number of authors outline a series of key steps or stages in the selection and implementation of outcome measures. In some cases, these appear to be intended as a linear sequence to follow (e.g. Wells and Johnson (2001)). In other cases, they appear to be intended more as key areas that need to be covered off during the development and implementation process, although not necessarily in a linear progression.

HomeGround
Services

While different sources vary on the details, the approaches tend to have many common elements. These are summarised in Table 23, which draws on five sources: Wells and Johnson (2001: 182-193), McDaniel (1996), MacKeith and Graham (2007), Hatry and Lampkin (2003) and Burns and Cupitt (2003).

Table 23: Key stages in development and implementation of an outcomes measurement system

Stage	Elements
Define purpose and audience	• Identify the primary reasons why outcome measurement is being introduced (decision-making, accountability, improving effectiveness, as a casework tool, ...) • Identify the intended audiences for outcomes information (including both internal and external audiences) • Identify which programs are to be included
Identify desired outcomes	• Review and clarify program aims and areas of responsibility, drawing on program documentation, observation of practice, and stakeholders' views • Involve a wide range of stakeholders (clients, staff, management, funders, sector partners, the community) in discussing the key results that the program aims to produce. Explore concerns and questions • Identify areas of agreement and seek a resolution to areas of disagreement regarding desired results (it is unlikely to be possible to satisfy everyone) • Define a set of outcome domains and within these, statements of key desired outcomes, linked to the program aims • Prioritise among multiple outcomes: determine those that are most important
Select measures and measurement tools	• Determine the types of measures that will be used to assess each outcome. Add indicators systematically based on priority of the outcome and feasibility and cost of collection • For each outcome, examine how setting performance standards may cause unintended consequences. Create checks and balances as necessary • Identify key client and service characteristics to be linked to outcome information • Identify whether there are pre-existing measurement tools which cover the types of outcomes and measures under consideration, and whether these are viable in the service context • Where pre-existing tools are not available or cannot be used, identify required sources of data for each measure • Examine currently available data to determine whether any of it is directly relevant to desired outcomes. Begin to use this data for performance monitoring • Where required, develop measurement instruments that will enable collection of data for the desired outcomes • Pilot and revise measurement tools to ensure they are user-friendly and are gathering in the required information • To the extent possible, examine the validity, reliability, sensitivity and cultural appropriateness of proposed measures and tools

Stage	Elements
Implement the measurement process	• Develop an implementation plan that is flexible to adapt to ongoing learning • Allocate resources and responsibility to undertake development, training, data collection, data analysis and reporting, and review of the measurement processes • Pilot the measurement processes on a limited scale before rolling out on an agency-wide basis • Ensure that measurement processes are integrated into the other processes of the organisation (especially service delivery) • Provide training to those collecting and using the data • Revise pre-existing forms and data collection instruments where required to increase consistency and minimise duplication • Revise electronic data systems to support data capture
Implement feedback systems	• Develop reporting templates and produce reports • Seek explanations for unusual or unexpected findings • Ensure that data is fed back to teams (e.g. through team meetings or in supervision) • Ensure that data is used as an input to organisational planning and quality improvement activities
Create an outcomes-oriented culture within the organisation	• Demonstrate clear and ongoing executive commitment to outcomes measurement • Encourage all staff to see outcomes (and their measurement) as central to their work • Celebrate successful outcomes at an individual, team and organisational level • Reflect on opportunities for improvement and areas where the organisation wishes to "turn the curve" – to make a substantial difference to its effectiveness

9.3 TIMEFRAME

As every organisational context is different, it is clearly impossible to give a definitive estimate of timeframe for introduction of an outcome measurement system. However, a number of authors agree that it tends to be a long-term process, due to the many philosophical, technical and political challenges associated with the work. It could be hypothesised that timeframe of implementation will be dependent partly on level of resourcing for the developmental work, and partly on the size of the organisation and the diversity of its operations.

McDaniel (1996) provides a ballpark estimate for the initial outcomes development stage, stating that "many states and local agencies are taking up to two years in the beginning stages of outcome development. The challenge is to balance the necessary process with the need to maintain momentum and interest." Hatry and Lampkin (2003: 4) also map out a two-year process for initial introduction of measures, including a pilot phase of around 12 months. Burns and Cupitt (2003: 38) suggest between 6 and 18 months for implementation depending on the size and complexity of the organisation.

HomeGround
Services

9.4 DEFINING AND ORGANISING MEASURES

Organisations introducing a package of client outcome measures (or other performance measures) need to be able to clearly define the measures, organise the definitions and update them as required. Barr (n.d.: 15) suggests adopting a single, standard organisation-wide template for defining performance measures. She also suggests that organisations maintain a 'dictionary' summarising their performance measures, which will provide users of the measures (including those coordinating measurement systems and those analysing the data) with up-to-date information on the measures. A database could be used for this purpose.

9.5 SUSTAINING OUTCOME MEASUREMENT SYSTEMS

Anecdotally, there is evidence that sustaining a measurement system is one of the key challenges in the outcome measurement field (Talbot 2010). Often the initial implementation of a measurement system receives appropriate resourcing and is accompanied by a burst of enthusiasm as staff focus on the goals of their work with clients and begin recording outcomes. However, maintaining staff motivation over time amidst the pressure and stress of service delivery can be difficult. As other organisational priorities and quality improvement initiatives are introduced, outcomes systems may be 'put on the backburner' and lose impetus. This can become a self-reinforcing process as data quality drops and the information produced becomes less meaningful.

Sustainable measurement systems are those that are embedded so that they become a normal and routine part of operations. However, few authors have concrete suggestions about how this can be achieved.

The following suggestions indicate some elements that help to sustain outcome measurement systems over time:

- Understand the embedding of outcomes measurement as an ongoing process that will require ongoing attention (Burns and Cupitt 2003: 37)
- Strong commitment from senior management to outcomes measurement over the long term, and clear communication of that commitment and vision to the rest of the organisation (Wells and Johnson 2001: 177-178)
- A preparedness to respond appropriately to potential "bad news" (including apparent short-term program failure) (Wells and Johnson 2001: 178)
- A commitment to recurrent allocation of resources to support the measurement system
- Staff performance expectations tied to completion of outcomes measurement for an appropriate proportion of clients (although not expectations tied to the outcomes themselves).

Outcomes measurement systems, like all performance measurement systems, need to recognise and respond to the changing informational needs of stakeholders over time (McDavid and Hawthorn 2006: 328-). A regular process of review of the measures will help to keep the outputs relevant and aligned with desired client outcomes.

10. Key questions to resolve in developing an outcome measurement framework

A wide range of questions need to be answered in developing a comprehensive framework for measuring outcomes within a human service organisation. The list of questions below are informed by the considerations outlined by Berman and Hurt (1997: 84-85), but also by the issues identified throughout this literature review.

The list appears daunting, and does in fact represent a large amount of work which will be required in developing a well-thought-out measurement system. However, there is some logical flow through the questions so that answering earlier ones will help to answer later ones. For example, identifying which domains and loci of change are of interest will help to answer questions about the appropriate measures and tools, the required sources of data, and who will be in the best position to gather this data.

What are the key outcomes that the organisation wishes to monitor?

- Who are the key stakeholders for outcomes measurement and what are their needs?
- Which domains are of interest?
- Which loci of change are of interest?
- Which time frames (short, medium, and/or long term)?
- Which interim outcomes are important?
- Which outcomes are consistent across the organisation and which are specific to particular services or client groups?
- What unintended outcomes (if any) should the organisation seek to monitor?

How do outcomes and measures fit with the organisation's theory of casework?

- What is the connection between outcomes measures, assessment information, action plan goals and action plan progress reviews?

- How do desired outcomes, and agreed outcome measures, fit with program logic models of the organisation's services?
- What theory of change underlies the organisation's approach to casework and to outcome measurement?

Which measurement approaches, tools and measures will be used to monitor these outcomes?

- Goal attainment scaling? Standardised scales? Status maintenance and change scales?
- Whose vantage point should be prioritised?
- Are there pre-existing measurement tools available that are suited to the outcomes that the organisation wishes to monitor, or does the organisation need to develop its own?
- Which measures should be applied across all services and which measures should apply only to selected services, functions or client groups?
- Should the measures and tools be designed in modular formats to allow use of sections of measures for particular contexts?
- What process, activity or output measures will be used in conjunction with the outcome measures?
- What breadth and depth of measures are cost-effective for the organisation to apply?
- What breadth and depth of measures are feasible to collect in particular services?
- How can additional measures be added for particular purposes (e.g. for a snapshot during a particular time period)?
- How can the organisation maximise the potential for consistency of approach with sector partners?

Whose outcomes will be measured?

- Aim for entire client population, or a sample? What sampling procedure will be used?
- Individuals or households / families?
- What processes will be used to seek consent to collect, use and exchange outcomes data?
- How will service delivery be adapted (if at all) for clients who opt out?

At what times will outcomes be measured?

- Ongoingly or episodically (snapshot)?
- Linked to key service events for individual clients (e.g. entry, exit, housing transition) or regular cyclical measurement (e.g. monthly, quarterly)?
- How will outcome measurement be timed for services with brief or unpredictable client contact (e.g. Initial Assessment and Planning)?
- Will follow-up data be collected, and if so, at what duration(s) after exit?

How will outcomes data be gathered and stored?

- What are the key sources of data for the measures selected (self-administered questionnaire, observation, records, ...)?
- Who should undertake data collection (service delivery staff, service development staff, volunteers, ...)?
- How (if at all) will outcomes measurement processes be integrated into casework processes such as assessment, action planning and review?
- Where will data be collected (in the office, on outreach to clients' accommodation, ...)?
- In what format will data be collected (paper, electronic, Rickter scale, verbal then later recorded, ...)?
- In what format will data be stored for analysis (database, spreadsheet, ...)?
- What modifications or developments need to occur to IT systems to allow this?
- Who will enter the data?

HomeGround
Services

How will outcomes data be used?

- What aggregate reports are required, for which stakeholders, in what formats, and at what times?
- What analysis techniques and tools will be used to generate these reports?
- How (if at all) will outcomes be adjusted for 'casemix' (severity and complexity of presenting issues)?
- How can limitations on internal and external validity be made clear to users of these reports?
- How will aggregate reports be used in team, business and strategic planning?
- How will outcomes data for individual clients be made available to those clients and the staff working with them, in an accessible and user-friendly format?

What resources will be allocated to sustain outcome measurement within the organisation?

- What resources will be allocated for development and review of the system (IT development, development of measures and tools, review of pilots...)?
- What resources will be allocated for operation of the system (data collection, data entry, data analysis and reporting, ...)?
- What documentation and training will be provided to staff in order for them to operate the system to a high standard?
- What incentives (if any) will be provided ?

References

Anderson, C. (2008). An Overview of Soft Outcome / Informal Learning Measurement Tools. Norwich, The Research Centre, City College Norwich.

Andresen, R., P. Caputi, et al. (2006). "The Stages of Recovery Instrument: Development of a measure of recovery from serious mental illness." Australian and New Zealand Journal of Psychiatry 40: 972-980.

Babor, T. F., J. C. Higgins-Biddle, et al. (2001). AUDIT: The Alcohol Use Disorders Identification Test. Guidelines for Use in Primary Care. Geneva, World Health Organization, Department of Mental Health and Substance Dependence.

Barr, S. (2008). PuMP® Implementer E-Workbook: active ready-to-use templates to design and bring to life your performance measures.

Barr, S. (2009). Case Study Notes: Measuring Outcomes to End Homelessness: Meaningful Measures for Program Outcomes from the National Alliance for Ending Homelessness. Stacey Barr's Measures & More Mastermind.

Barr, S. (n.d.). 202 Tips for Performance Measurement: Practical ideas to choose, create and use meaningful performance measures with lots of buy-in!

Baulderstone, J. and C. Talbot (2004). Outcome Measurement in SAAP Funded Services: Final Report, Flinders Institute of Public Policy and Management.

Berman, W. H. and S. W. Hurt (1997). Developing Clinical Outcomes Systems: Conceptual and Practical Issues. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 81-97.

Bloom, M., J. Fischer, et al. (2006). Evaluating practice: Guidelines for the accountable professional. Boston, Pearson / Allyn & Bacon.

Blunden, H. and C. Johnston (2005). Public housing and nonhousing outcomes – a background paper. Shelter Brief. Sydney, Shelter NSW.

Booth, B. M. and G. R. Smith (1997). Outcomes Measurement: Where We Are. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 35-43.

Buckingham, B., P. Burgess, et al. (1998a). Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report. Canberra, Commonwealth Department of Health and Family Services.

Buckingham, B., P. Burgess, et al. (1998b). Developing a Casemix Classification for Mental Health Services. Volume 2: Resource Materials. Canberra, Commonwealth Department of Health and Family Services.

Burke, T. and K. Hulse (2002). Sole parents, social wellbeing and housing assistance. Final Report. Melbourne, Australian Housing and Urban Research Institute.

Burns, S. and S. Cupitt (2003). Managing outcomes: a guide for homelessness organisations. London, Charities Evaluation Service / London Housing Foundation.

Burns, S. and J. MacKeith (2009a). Alcohol Star: the Outcomes Star for alcohol misuse. User Guide., Alcohol Concern and Triangle Consulting.

Burns, S. and J. MacKeith (2009b). Work Star: The Outcomes Star for work and learning. Organisation Guide. London, Triangle Consulting.

Busch-Geertsema, V. (2005). "Does Re-Housing lead to re-integration? Follow-up studies of re-housed homeless people." Innovation: The European Journal of Social Science Research 18(2): 205-226.

Butcher, B. and L. Marsden (2004). Measuring Soft Outcomes: A Review of the Literature. Norwich, The Research Centre, City College Norwich.

Calsyn, R. and J. Winter (2002). "Social Support, Psychiatric Symptoms, and Housing: A Causal Analysis." Journal of Community Psychology 30(3): 247-259.

Campbell-Orde, T., J. Chamberlin, et al. (2005). Measuring the Promise: A Compendium of Recovery Measures Volume II. Cambridge MA, Human Services Research Institute.

Chamberlain, C., G. Johnson, et al. (2007). Homelessness in Melbourne: Confronting the Challenge. Melbourne, RMIT University Press.

Christo, G. (2008). "Christo Research Systems." Retrieved 17 June, 2010, from http://users.breathe.com/drgeorgechristo/.

Clark, C. and A. R. Rich (2003). "Outcomes of Homeless Adults With Mental Illness in a Housing Program and in Case Management Only." Psychiatric Services 54(1): 78-83.

Clements, D. (2010). Personal communication: Meeting at HomeGround Services, 26/03/2010.

Common Ground. (n.d.). "Vulnerability Index: Prioritizing the Street Homeless Population by Mortality Risk." Retrieved 16 April, 2010, from http://www.commonground.org/wp-content/uploads/2008/04/Vulnerability%20Index%20101.pdf.

Commonwealth of Australia (2008). The Road Home: A National Approach to Reducing Homelessness. Canberra, Department of Families, Housing, Community Services and Indigenous Affairs.

Commonwealth of Australia, State of New South Wales, et al. (2005). MULTILATERAL AGREEMENT IN RELATION TO THE SUPPORTED ACCOMMODATION ASSISTANCE PROGRAM.

Cook, D. A. and T. J. Beckman (2006). "Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application." The American Journal of Medicine 119: 166.e167-166.e116.

Cook, T. D. and D. T. Campbell (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston, Houghton Mifflin.

Core Systems Group (1998). CORE System User Manual. Leeds, Core Systems Group.

Council of Australian Governments (2009a). National Affordable Housing Agreement.

Council of Australian Governments (2009b). National Partnership Agreement on Homelessness.

Culhane, D. P., W. D. Parker, et al. (2007). Accountability, Cost-Effectiveness, and Program Performance: Progress Since 1998. 2007 National Symposium on Homelessness Research: 12-11 to 12-42.

Cunningham-Smith, V. (2010). A journey keeping our focus on making sure people are better off as a results of our services. The search for consumer outcomes in health and welfare services. Powerpoint presentation. Melbourne, Inner East Community Health Service.

HomeGround Services

Davies, R. and J. Dart (2005). The 'Most Significant Change' (MSC) Technique: A Guide to Its Use.

de Vaus, D. (2002). Analyzing Social Science Data: 50 Key Problems in Data Analysis. London, SAGE Publications.

DeCoster, J. (1998). "Overview of Factor Analysis." Retrieved 6 April, 2010, from http://www.stat-help.com/factor.pdf.

Department of Health. (2009). "About Outcome Measurement: Victorian Collection of OM in PDRSS." Retrieved 19 May, 2010, from http://www.health.vic.gov.au/mentalhealth/outcomes/about2.htm.

Department of Health and Ageing (2003). Mental Health National Outcomes and Casemix Collection: Overview of clinician-rated and consumer self-report measures, Version 1.50. Canberra, Department of Health and Ageing, Mental Health & Suicide Prevention Branch.

Department of Human Services (2006). Homelessness Assistance Service Standards. Melbourne, Victorian Department of Human Services.

Department of Human Services (2009). Opportunities in practice: Outcome measurement in mental health. Revised edition, March 2009. Melbourne, Victorian Department of Human Services Mental Health and Drugs Division.

DHS Evaluation Support Unit (n.d.). Evaluation frameworks: A development guide. Melbourne, Department of Human Services.

Downtown Emergency Service Centre. (n.d.). "Vulnerability Assessment Tool." Retrieved 16 April, 2010, from http://elearning.ahc.org/files/DESC_Vulnerability%20Assessment_Tool%202003.pdf.

Duignan, P. (2005). "Outcomes hierarchy definitions." Retrieved 21 April, 2010, from http://www.parkerduignan.com/ot/untitled_text_page.html.

Eagar, K., W. Buckingham, et al. (2001). The Victorian Mental Health Outcomes Measurement Strategy: Final Report on the Implementation of Outcome Measurement In Adult Area Mental Health Services. Melbourne, Victorian Department of Human Services.

Eardley, T., D. Thompson, et al. (2008). Measuring the Impact of SAAP-funded Homelessness Services on Client Self-reliance: Final Report. Sydney, Social Policy Research Centre, University of New South Wales.

Elizabeth, S. (2010). Personal communication: Meeting between Sally Elizabeth and Mark Planigale to discuss Youth Refuges outcome measurement pilot. Meeting at HomeGround Collingwood, 14/04/2010.

ERoSH. (2010). "Publications." Retrieved 17 June, 2010, from http://www.shelteredhousing.org/information/publications.

Evans, S., J. Greenhalgh, et al. (2000). "Selecting a mental health needs assessment scale: guidance on the critical appraisal of standardized measures." Journal of Evaluation in Clinical Practice 6(4): 379-393.

Felce, D. and J. Perry (1995). "Quality of life: Its definition and measurement." Research in Developmental Disabilities 16(1): 51-74.

Fitzpatrick, J. L., J. R. Sanders, et al. (2004). Program Evaluation: Alternative Approaches and Practical Guidelines. 3rd edition. Boston, Pearson.

Friedman, M. (2005). Trying Hard Is Not Good Enough: How to Produce Measurable Improvements for Customers and Communities, FPSI Publishing / Booksurge.

HomeGround Services

Friedman, M., L. DeLapp, et al. (2001a). The Results and Performance Accountability Implementation Guide: Case Studies.

Friedman, M., L. DeLapp, et al. (2001b). The Results and Performance Accountability Implementation Guide: Questions and Answers about How to Do the Work.

Gray, J. (2008). Outcome measures for preventing homelessness through mediation, Framework.

Hamilton, L. (2010). Personal communication: Meeting between Lorrinda Hamilton and Mark Planigale, 68 Oxford St Collingwood, to discuss outcomes measurement in the context of the local Mental Health Alliance. Meeting date 31/03/2010.

Hatry, H. and L. Lampkin (2003). Key Steps in Outcome Management. Series on Outcome Management for Nonprofit Organizations. H. Hatry and L. Lampkin, The Urban Institute.

Hatry, H. P. (1997). Outcomes Measurement and Social Services: Public and Private Sector Perspectives. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washongton, DC, NASW Press: 3-19.

Health Care for the Homeless (1998). Health Care for the Homeless Outcome Measures... A Chronicle of Twenty Pilot Studies. October 1998, Health Care for the Homeless.

Hendrick, E. (2010a). MORFing @ MCM: The MORF: Measuring Outcomes & Results Framework. Powerpoint presentation, 31/03/2010. Melbourne, Melbourne Citymission.

Hendrick, E. (2010b). Personal communication: Meeting between Elaine Hendrick and Mark Planigale to discuss MORF. Meeting at 79 Little Oxford St Collingwood, 16 April 2010.

Homeless Link. (2010). "The Outcomes Star System." Retrieved 12 May, 2010, from http://www.outcomesstarsystem.org.uk/.

Hubley, A. M., L. B. Russell, et al. (2009). Quality of Life for Homeless and Hard-to-House Individuals (QoLHHI) Inventory: Administration and Scoring Manual. Vancouver, University of British Columbia.

Hudson, W. W. (1997). Assessment Tool as Outcomes Measures in Social Work. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 68-80.

Hulse, K. and L. Saugeres (2008). Housing insecurity and precarious living: an Australian exploration. AHURI Final Report No. 124. Melbourne, Australian Housing and Urban Research Institute.

Johnson, G. (2010). Personal communication: Comments by Guy Johnson at Outcomes Project Reference Group Meeting, 3 May 2010.

Johnson, G., H. Gronda, et al. (2008). On the outside: pathways in and out of homelessness. North Melbourne, Vic., Australian Scholarly Publishing.

Kaufman, M. C. (2007). Durham System Of Care: Life Domain Guide for Child and Family Teams, The Durham Centre.

Kimberley, S. (2009). Monitoring and Evaluation: Definitions. Melbourne, HomeGround Services.

HomeGround
Services

Kolar, V. (2004). Home First: A longitudinal study of outcomes for families who have experienced homelessness - Final Report. Melbourne, Hanover Welfare Services.

Love, A. and M. Scott (2010). Personal communication: Meeting between Mark Planigale, Anna Love and Matthew Scott at Clarendon Clinic, 19/03/2010.

MacKeith, J. (2007). 'User-focussed' – what does it really mean? London, London Housing Foundation.

MacKeith, J. and S. Burns (2008). Mental Health Recovery Star. User Guide. London, Mental Health Providers Forum and Triangle Consulting.

MacKeith, J., S. Burns, et al. (2008a). The outcomes star: Supporting change in homelessness and related services, 2nd edition. Organisation Guide. London, London Housing Foundation and Triangle Consulting.

MacKeith, J., S. Burns, et al. (2008b). The outcomes star: Supporting change in homelessness and related services, 2nd edition. User Guide. London, London Housing Foundation and Triangle Consulting.

MacKeith, J. and K. Graham (2007). A Practical Guide to Outcomes Tools. London, London Housing Foundation.

MacKeith, J., K. Graham, et al. (2007). Review of Outcomes Tools for Homelessness Sector, London Housing Foundation.

Marsden, J., M. Gossop, et al. (1998). The Maudsley Addiction Profile: A brief instrument for treatment outcome research. Development and User manual. London, National Addiction Centre / Institute of Psychiatry.

McDaniel, N. C. (1996). "Key Stages and Common Themes in Outcome Measurement." The Evaluation Exchange 2(3) Retrieved 29 March, 2010, from http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/results-based-accountability-2/key-stages-and-common-themes-in-outcome-measurement.

McDavid, J. C. and L. R. L. Hawthorn (2006). Program Evaluation & Performance Measurement: An Introduction to Practice. Thousand Oaks, California, Sage Publications.

McLean Hospital. (2010). "Measuring Patient Self-Reported Outcomes." from http://www.basissurvey.org/basis32/.

Mertens, D. M. (2008). Stakeholder Representation in Culturally Complex Communities: Insights from the Transformative Paradigm. Fundamental Issues in Evaluation. N. L. Smith and P. R. Brandon. New York, The Guilford Press: 41-60.

Mission Australia (2005). NSW/ACT Community Services Outcomes Measurement Tool Kit. Version 1. November 2005. Sydney, Mission Australia, NSW/ACT Community Services.

Mission Australia (2006). Internal Service Evaluation. Template Version 4. June 2006. Sydney, Mission Australia, NSW/ACT Community Services.

National Alliance to End Homelessness. (2006). "Frequently Asked Questions about Housing First for Individuals and Families." Retrieved 22 April, 2010, from http://www.endhomelessness.org/content/article/detail/1424.

National Treatment Agency for Substance Misuse (2007). Treatment Outcomes Profile v1.0, National Treatment Agency for Substance Misuse.

NMHWG Information Strategy Committee Performance Indicator Drafting Group (2005). Key Performance Indicators for Australian Public Mental Health Services: Report of Performance Indicator Drafting Group. ISC Discussion Paper No.6. Canberra, Australian Government Department of Health and Ageing.

Office of Housing (n.d.). Measuring outcomes: Rationale, ideas and resources. A resource kit for Victorian agencies implementing the Homelessness Assistance Service Standards and the National Community Housing Standards. Melbourne, Victorian Department of Human Services, Office of Housing.

OSW (2004). The Employability Map: A tool for measuring progress to employability, Off the Streets and Into Work.

Parkinson, D. (2005). Putting outcomes into practice: A guide to developing IT-based outcome data management. London Housing Foundation IMPACT Programme briefing paper, London Housing Foundation & Charities Evaluation Service.

Patton, M. Q. (1987). How to Use Qualitative Methods in Evaluation. Newbury Park, California, Sage Publications.

Phibbs, P. (2002). Housing assistance and non-shelter outcomes. Sydney, AHURI.

Pilon, D. and M. Ragins (2007). Milestones of Recovery Scale (MORS).

Planigale, M. (2010a). Discussion paper: A consistent set of casework domains for HomeGround. Melbourne, HomeGround Services.

Planigale, M. (2010b). Measuring Service Performance – the Big Picture. Melbourne, HomeGround Services.

Planigale, M., S. Kimberley, et al. (2009). HomeGround Services – Program Logic Model – version: 18 Feb 2009. Melbourne, HomeGround Services.

Pleace, N. (2008). Effective Services for Substance Misuse and Homelessness in Scotland: Evidence from an International Review. York, The Centre for Housing Policy, University of York / Scottish Government Social Research.

Post, P. A., F. E. Isbell, et al. (2005). Developing Outcome Measures to Evaluate Health Care for the Homeless Services. Nashville, National Health Care for the Homeless Council.

Priebe, S., P. Huxley, et al. (1999). "Application and results of the Manchester Short Assessment of Quality of Life (MANSA)." Int J Soc Psychiatry 45(1): 7-12.

QualityMetric. (2010). "Generic Health Surveys." Retrieved 17 June, 2010, from http://www.qualitymetric.com/WhatWeDo/GenericHealthSurveys/tabid/184/Default.aspx.

Ralph, R. O. (n.d.). "At the Individual Level: A Personal Measure of Recovery." Retrieved 17 June, 2010, from http://www.nasmhpd.org/spec_e-report_fall04measures.cfm.

Rapp, C. A. and J. Poertner (1992). Social administration: A client-centered approach. New York, Longman.

Rethink (n.d.). CUES: Carers' and Users' Expectations of Services. London, Rethink.

Rossi, P. H. (1997). Program Outcomes: Conceptual and Measurement Issues. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 20-34.

Rossi, P. H., H. E. Freeman, et al. (1999). Evaluation: A Systematic Approach. 6th edition. Thousand Oaks, California, Sage Publications.

Schalock, R. L. (2001). Outcome-Based Evaluation. New York, Kluwer Academic / Plenum Publishers.

Segal, S. P. (1997). Outcomes Measurement Systems in Mental Health: A Program Perspective. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 149-159.

Segone, M. (2008). Bridging the gap: the role of monitoring & evaluation in evidence-based policy making. Geneva, Unicef.

Siggins Miller Consultants (2003). A report to the Mental Health Branch, Department of Human Services: Consumer self-rated outcome measures in mental health.

Smith, G. R., K. M. Rost, et al. (1997). Assessing Effectiveness of Mental Health Care in Routine Clinical Practice. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 124-136.

Social Solutions. (2010). "Human Services Software." Retrieved 12 May, 2010, from http://www.socialsolutions.com/Human-Services-Software/Performance-Management.aspx#.

Spellman, B. and M. Abbenante (2008). What Gets Measured, Gets Done: A Toolkit on Performance Measurement for Ending Homelessness, National Alliance to End Homelessness - Homelessness Research Institute.

Spence, S. H., M. Donald, et al. (2002). Outcomes and indicators, measurement tools and databases for the National Action Plan for Promotion, Prevention and Early Intervention for Mental Health 2000. Brisbane, School of Psychology and Centre for Primary Health Care, University of Queensland.

Stedman, T., P. Yellowlees, et al. (1997). Measuring Consumer Outcomes In Mental Health: Field Testing of Selected Measures of Consumer Outcome in Mental Health. St Lucia, Qld, University of Queensland.

Stevens, B. (2009). Personal communication: Email from Briannan Stevens re: Pilot of Outcomes Star at Micah Projects, Brisbane. Email dated 4 April 2009.

Talbot, C. (2010). Personal communication: Meeting between Chris Talbot and Mark Planigale, Fitzroy, Friday 9 April 2010.

Trauer, T., G. Tobias, et al. (2008). "Development and evaluation of a patient-rated version of the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS-P)." Community Mental Health Journal 44(2): 113-124.

Trinh, Q.-T. (2010). Personal communication: Meeting between Mark Planigale, Quynh-Tram Trinh and Trish Watson at Department of Human Services, Melbourne, 17/03/2010.

Trochim, W. M. K. (2006a). "Nonprobability Sampling." Retrieved 10 May, 2010, from http://www.socialresearchmethods.net/kb/sampnon.php.

Trochim, W. M. K. (2006b). "Probability Sampling." Retrieved 10 May, 2010, from http://www.socialresearchmethods.net/kb/sampprob.php.

University of Melbourne. (2010). "WHOQoL Instruments downloads." Retrieved 17 June, 2010, from http://www.psychiatry.unimelb.edu.au/qol/whoqol/whoqol-insruments.html.

University of New South Wales. (2010). "Depression Anxiety Stress Scale." Retrieved 17 June, 2010, from http://www2.psy.unsw.edu.au/groups/dass/.

University of Toronto Centre for Health Promotion. (2010). "QOL Concepts: The Quality of Life Model." Retrieved 6 April, 2010, from http://www.utoronto.ca/qol/concepts.htm.

Wadsworth, Y. (1997). Everyday Evaluation on the Run. 2nd edition. Crows Nest, NSW, Allen & Unwin.

Ware, J. E. (1997). Health Care Outcomes from the Patient's Point of View. Outcomes Measurement in the Human Services: Cross-Cutting Issues and Methods. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 44-67.

Weiss, C. (1972). Evaluation research: Methods for assessing program effectiveness. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.

Weiss, H. B. (1997). Results-Based Accountability for Child and Family Services. Outcomes Measurement and Social Services: Public and Private Sector Perspectives. E. J. Mullen and J. L. Magnabosco. Washington, DC, NASW Press: 173-180.

Wells, S. J. and M. A. Johnson (2001). "Selecting Outcome Measures for Child Welfare Settings: Lessons for Use in Performance Management." Children and Youth Services Review 23(2): 169-199.

World Health Organisation (1996). WHOQOL-BREF: Introduction, administration, scoring and generic version of the assessment. Field trial version. Geneva, World Health Organisation Programme on Mental Health.

World Health Organisation. (2010). "Severity of Dependence Scale (SDS)." Retrieved 17 June, 2010, from http://www.who.int/substance_abuse/research_tools/severitydependencescale/en/index.html.

CONTACT US

COLLINGWOOD
68 Oxford Street
Collingwood 3066

Housing/Support	9288 9611
Freecall	1800 048 325
Outreach	9288 9650
Fax	9288 9602
Corporate enquiries	9288 9600
Fax	9288 9601

ST KILDA
122 Chapel Street
St Kilda 3182

Housing/Outreach	9537 7711
Fax	9537 7722
Support	9537 7888
Fax	9537 7991

CITY
660 Elizabeth Street
Melbourne 3000

Supportive Housing	8359 0200
Fax	8359 0201

PRAHRAN
PO Box 354
Prahran 3181

Tenancy & Property	9537 7999
Housing Options	9932 8368
Fax	9537 7797

PRESTON
600 Plenty Road
Preston 3072

Support	9471 4522
Freecall	1300 844 125
Fax	9471 4533