What are outcome measures and their determinants?
An outcome is the consequence or end result of a process, which could be a treatment, programme or service or could simply be changes in the health condition over a phase or period of development. The outcomes measured are linked to either the effects of a particular intervention or service delivery programme or the impact of the disease or health condition. Outcome measures are tools that may be used to assess a change in particular attributes that are deemed meaningful to a person’s life over time. They may be differentiated from enduring attributes such as personality traits, expectations or demographic characteristics. This last factor may include important determinants of outcomes, and therefore can also be measured for either clinical or research purposes, along with outcome measures. A determinant may be any factor that causes or influences the outcomes of interest. The measurement of outcomes and their determinants is essential in informing decisions about treatment and targeting those who may benefit most. For example, the use of outcome measures as part of a randomized controlled trial may provide evidence that an injection of botulinum toxin for children with spastic cerebral palsy is effective in increasing range of motion, alleviating spasticity and improving movement quality and motor function. However, evaluation of the determinants of responsiveness, that is the factors associated with greater improvements following Botulinum toxin treatments, reveals that age, number of previous treatments and contextual factors such as the child’s mastery motivation and family stress levels can influence the degree of responsiveness to this expensive treatment modality. Some variables may be confounders that are correlated with both the independent variable (determinants, exposures) and the dependent variable (outcome). These need to be accounted for in the analyses of relationships between determinants and outcomes.

Knowledge of determinants is therefore helpful in targeting those children who may benefit most from particular treatments or services, with the aim of enhancing the outcomes of interest. Similarly, a range of outcome measures may be used to characterize the development and functioning of a specified group of children with a developmental disability; however, measurement of particular determinants is also very helpful in differentiating factors that may positively or negatively influence disability level. Cassidy presented a conceptual model that considered both medical factors (e.g. severity, comorbidities, aetiology) and demographic factors (e.g. age, sex, socioeconomic status) as important influences of outcomes. Similarly, the conceptual framework for the Canadian Health Measures Survey considers the dynamic interplay between non-modifiable and modifiable determinants together with health behaviours and characteristics as potential mediators (e.g. intermediary; intervenes between the exposure and the outcome) and moderators (e.g. alters the state or effect of a causal factor or variable) of health outcomes (Table 1.1). Whether in clinical practice or in the research setting, outcome assessment should involve consideration of factors that can potentially exert important influences on these outcomes. Therefore, appropriate tools need to be carefully selected to measure both outcomes and their determinants.

The measures used to assess outcomes and their determinants can be standardized, which implies that they are applied in a prescribed manner, as established by the developers of the instrument. Patient-centred outcomes are often measured using self-report questionnaires or interview formats to assess aspects of health and functioning from the individual’s own perspective. Often, clinical studies use surrogate end points (e.g. measurement of blood pressure and cholesterol) as a substitute for a more meaningful outcome (e.g. cardiovascular health/disease). Relationships are assumed and surrogates are used as they

are more objectively and easily quantified. Whether in clinical practice or research, the measures used to describe the outcomes of interest should be valid, objective measures of the constructs of direct interest. There are several factors that need to be considered when selecting outcome measures, and these are outlined later in this chapter.

It should be emphasized that quantification of outcomes and their determinants using objective, standardized measurement tools is only one way of depicting and characterizing this important information. Qualitative methods or mixed methods can also be used as processes of inquiry to gather a more in-depth understanding of the constructs of interest.

**Why are outcome measures important to use?**

Outcome measurement has important benefits to consumers of services (patients/clients and their family members), service providers, clinical managers, policy makers and researchers (Table 1.2). For patients or clients and their families, individualized application of outcome measures provides the family with objective, quantitative information regarding their child’s relative strengths and weaknesses in the area(s) assessed, demonstrates changes in performance over time and serves to foster achievement and motivate children to improve their abilities in the areas measured by the assessment tool.9,10 These measures would need to meet rigorous psychometric standards for use in individuals. Furthermore, outcomes data from research studies provide useful information regarding the natural history and developmental trajectories to be expected in particular childhood disability subgroups. Knowledge of these outcomes enables health service providers to more effectively counsel families regarding the outcomes to be expected for their child. It is important that parents and children have realistic expectations for the future and set personal goals that are attainable.

Measuring outcomes is important for health service providers as well. Outcome measures serve to identify
Selection and Use of Outcome Measures

areas of concern and to monitor changes over time. The
use of outcome measures can promote a more reflective
practice approach, thus providing clarity of purpose with
respect to goal attainment. Clinicians may carefully
consider whether or not goals of intervention are being
met, or whether new interventions or goals need to be
pursued. An outcomes-based framework for service
delivery is an important cultural shift for clinicians. The
focus is on achieving specified outcomes, with objective,
measurable treatment goals and commitment to long-
term changes. Objective data encourage clinicians to
be more accountable about the services they are pro-
viding, both to themselves and to others. Indeed, with
increasing fiscal constraints, rehabilitation specialists are
under greater scrutiny to provide high-quality care at the
lowest cost, and outcome measurement can be an effect-
ive strategy to validate the usefulness of particular pro-
grammes or interventions. For example, greater efforts
at restructuring service delivery by using dyads (two
children treated simultaneously by one health profes-
sional) or group treatment programmes, and by providing
consultative services in addition to direct interventions,
are cost-effective strategies that may need greater con-
sideration in practice.

Clinical co-ordinators or managers and other admin-
istrators of health services value outcome measurement
for a variety of reasons. First, interdisciplinary pro-
grammes are costly, and there is a need to demonstrate
that the outcomes of the programme align well with the
goals and objectives of the programme. For example,
if improving quality of life is the primary goal of a
social support group or assistive technology service for
children with developmental disabilities, it would be
important to use an outcome measure at baseline and
upon completion of the intervention to demonstrate such
an effect. The use of outcome measures enables admin-
istrators and managers to better appreciate which goals
are feasible and attainable for particular programmes or
services. Outcomes research has highlighted new popu-
lations who are at risk of developmental disability. For
example, outcomes data on infants with congenital heart
defects who require open heart surgery clearly demon-
strate that survival has improved dramatically; however,
it is not known whether these neonates are at high risk
for developmental and learning challenges as they grow
and develop. This emphasizes the need for new inter-
disciplinary health services for children with congenital
heart defects that are directed at the periodic surveil-
ance of developmental progress at key transition points as
the child matures, as are routinely provided for infants
who are born preterm. Therefore, new health service
programmes can be justified by outcomes research on
new target populations that are at risk for developmental
disabilities. As part of quality assurance, health admin-
istrators can use outcome measures to evaluate the quality
of services to include the structure of the services, the
process of care provided and the outcomes, and satisfac-
tion with the services. Performance measurement is
a process whereby indicators and assessment tools are
used to evaluate a programme’s mission, goals and target
outcomes. This may include the evaluation of structural
elements of a service such as access, frequency of treat-
ments, and qualifications and expertise of the service
provider. Furthermore, process elements may include
attributes of the treatment such as respectful and sup-
portive care and the appropriate selection of evidence-
based treatment modalities. Finally, evaluation inevitably
includes indicators and measures of the child’s outcome
in the domains in which improvements are expected.
Satisfaction with services by family members should
also be considered in programme evaluation and perfor-
ance measurement. This objective information not
only sets performance standards but is essential in order
for administrators and managers of health services to
make decisions regarding resource allocation that will
provide the best outcomes for children requiring these
services.

Public policy at the government level involves the
development and adoption of principles, programmes and
services in the health sector. Policy makers may take
advantage of outcomes data as evidence to develop new
policies or programmes for target populations in need
of services. Objective evidence provided by outcome
measures can identify where the needs are greatest for
services, resources and supports, which is required for
effective and judicious resource allocation.

Researchers use outcome measures to answer a
wide variety of questions related to childhood disability,
including elucidation of mechanisms or causes of dis-
ability, understanding patterns of recovery and reorga-
nization of the brain, validation of early identification
tools, verification of treatment effectiveness, recognition
determinants of disability, evaluation of health promo-
tion strategies, determination of quality of health services
and effective knowledge translation efforts. Invariably,
outcome measures are needed to quantify objectively par-
ticular attributes or characteristics of the individual with
or at risk for a disability, or aspects of their environment.

Clearly, the potential benefits of application of out-
come measurement are broad and wide ranging. Therefore,
consideration should be given to applying multiple mea-
sures to capture the spectrum of outcomes of interests,
as well as possible indicators or determinants that may
influence the outcomes.
Factors to consider in the selection of outcome measures

Measures are widely used in clinical practice and research for a variety of purposes. Selection of the most appropriate outcome measures can raise a number of concerns and queries, for example:

- What are the best markers for success of your intervention?
- Which measures will enable you to specifically answer your research questions and test your hypotheses?
- Is the tool suitable for the population of interest (age, diagnosis, developmental level and abilities)?
- Will the findings on assessment provide useful clinical information?
- Will the evaluation tool accurately measure the ‘right thing’ (i.e. cohesive construct of interest) in your research or clinical hypotheses?
- Is the tool reliable, that is, is it consistent across multiple test administrations and raters?
- Is the measure affordable and easy to administer, and well tolerated by the participants being investigated or evaluated without the need for modification of the administration standards, which could jeopardize interpretability?
- Can you use only a portion of the measure without jeopardizing reliability and validity (do subscales stand alone)?

A number of criteria need to be considered in the selection of outcome measures, whether by clinicians or by researchers. Several authors9,19–21 (see also www.hta.ac.uk/fullmono/mon214.pdf) have provided checklists of these criteria, and there is consensus that the key factors to consider in the selection of outcome measures include the following and are described further below:

- measures are specifically measuring the outcome(s) of interest;
- measures are relevant to the children being evaluated and to their families;
- administration is feasible and practical;
- measures were developed with a particular purpose, for a specific target population; and
- the instruments chosen are psychometrically sound.

MEASURES SELECTED ARE MEASURING THE OUTCOMES OF INTEREST

Clinicians need to be sure that the measures they select are relevant to the goals of treatment, and therefore goals of intervention should first be established before the selection of outcome measures.9 What is the intended purpose of the intervention and what attributes of the child do you expect to change? If a treatment programme, for example, claims to improve developmental skills and enhance quality of life, then it is essential that the specific developmental domains targeted for improvement are measured using tools that are sensitive to change, and that quality of life is also measured. Therefore, clinicians need to carefully reflect on what outcomes are realistically likely to change in the time frame of the interventions being provided. Factors that might influence level of change (e.g. compliance with treatment) may need to be objectively quantified (measured) as well.1,10 Table 1.3 provides an example of some of the issues and clinical questions that may be raised and considered in the selection of outcome measures for

<table>
<thead>
<tr>
<th>Factors to consider</th>
<th>Clinical context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relate outcome measures to the programme objectives</td>
<td>What areas of development do you hope to change? Does the family have specific goals that need to be considered and measured?</td>
</tr>
<tr>
<td>Seek the best available instruments</td>
<td>Are the changes quantitative (better scores) and/or qualitative (a change in quality of a particular behaviour or attribute)?</td>
</tr>
<tr>
<td>Which tools will detect the degree or increment of change that you are hoping to accomplish?</td>
<td></td>
</tr>
<tr>
<td>Consider clinical significance and statistical significance of change</td>
<td>What level of change would be clinically meaningful for this child? To whom?</td>
</tr>
<tr>
<td>Will criterion-referenced or norm-referenced tools be more appropriate in measuring clinically meaningful changes?</td>
<td></td>
</tr>
<tr>
<td>Apply a conceptual framework that will help guide the selection of measures</td>
<td>Which measures will best explain the effects of the intervention on the child and family?</td>
</tr>
<tr>
<td>Which attributes or measures may help explain factors that are influencing level of improvement?</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 1.3

Example of factors to consider in the selection of measures for an early intervention programme11
an early intervention programme or for a research study evaluating the effectiveness of such a programme on a target population. Traditionally, many early intervention programmes have focused on measures of cognitive ability such as IQ; however, careful reflection a priori of the areas that should be measured will ensure that the effectiveness of such programmes is more broadly appreciated and understood.

Often, outcome measures are used for clinical accountability, so as to evaluate services and programmes. Outcomes are compared with baseline, and the proportion of patients or clients that meet or exceed treatment goals are determined. Increasingly, factors influencing ‘treatment success’ are considered, and may guide future service planning. Programme evaluation could include family functioning, and child’s self-concept and level of motivation, and programmes may therefore be interested in measuring these areas. In particular, these personal and environmental factors can potentially be modified by interdisciplinary teams, and therefore may be considered as part of the programme goals and interventions.

Similarly, researchers need to carefully select the outcome measures that will specifically address their research questions, and can be used to test the proposed hypotheses. It is often challenging to find an appropriate criterion standard measure of particular areas of behaviour and development, thus limiting the research questions that can be investigated. Indeed, often one of the first phases of research programmes involves the development of new measures that will then be utilized to pursue the research questions of investigators.

Measures are Relevant to Children and to their Families
The evaluation tools chosen for use must be acceptable to the children being assessed and to their families in terms of administration format and test requirements. Furthermore, the areas selected for measurement should be of importance to the child and family, in line with health and developmental outcomes of relevance to them. Family-centred care involves a partnership between health service providers and families, whereby the parents and children are actively involved in the selection and prioritization of treatment goals, and the care provided must be respectful and supportive. Family-centred approaches to care have been found to enhance developmental outcomes for the child, improve family adjustment and coping, and increase satisfaction with services. Active involvement of families in the selection of pertinent outcomes will undoubtedly reinforce the positive attributes and benefits of family-centred service.

Practicality and Feasibility of the Measure
Measures should be ‘user friendly’, that is, easy to administer by the evaluator. The measure should have a clear (standardized) administration protocol. Indeed, some tools may require specific training or special qualifications to ensure standardized administration, scoring and interpretation. It is critical to follow the standardized administration and scoring procedures as outlined in the test manual rather than attempt to simplify the test administration, as deviation from the designated standardized approach can increase error and bias and jeopardize interpretation of the results. Often, the time to complete the measure is a factor in its successful uptake as a clinical or research tool, with short administration times being an asset for children with disabilities. Cost is another factor to consider, and investment in an expensive test kit is only likely to be necessary if there is broad applicability and utility. In summary, practical considerations in the selection of outcome measures include cost, time, ease of use, portability and availability of the instrument. Often, clinicians will use what is already available in their department; however, consideration should be given to periodic investment in new outcome measures that may provide accurate new information that is not collected using existing measures.

The Purpose of the Measure is Considered
Tools are developed with a particular purpose. They may be discriminative tests, which provide information about children with disabilities compared with typically developing children (i.e. norm-referenced tests). Predictive measures and screening tools are used to identify children with or at risk of a particular attribute (now or in the future). Screening tests provide preliminary information, which is then followed up with a more detailed assessment in those who fail or obtain borderline results on the initial screen. Evaluative measures can detect changes over time. These tools are often criterion referenced (judged against a criterion standard, not compared with the ‘norm’) and thus are more likely to be sensitive to small changes in performance. When selecting a particular measure, it is essential that evaluators verify the sample of children in whom the measure was standardized in order to be sure that it is representative and therefore appropriate for the children for whom it is intended. Thus, when selecting measures either for clinical practice or for research, it is important that the tools chosen are being used as intended. For example, if the intent is to demonstrate the effectiveness of a new aid or adaptation (e.g. built-up handles to facilitate grasp) in promoting independence in self-feeding, personal care and hygiene, then a measure such as the Pediatric Evaluation of Disability Inventory, which has been shown to reliably perceive small improvements (or
deteriorations) in self-care activities, should be selected.8 These tests must demonstrate that they are responsive tools that can be used to detect clinically important changes over time. Furthermore, the tool should be sensitive to change at various levels of performance. If the test is too difficult and scores tend to cluster at the lowest values (floor effect) or, conversely, if the test is too easy and scores cluster at the highest values (ceiling effect), then detecting small changes may not be possible.20

The Measure Is Psychometrically Sound
Psychometric theory is the science of assessing the measurement properties of a tool. Collectively, this can include (1) data quality such as the impact of missing scores or non-response to items; (2) scaling assumptions and weighing of items to create scales; (3) acceptability of score distributions in representing the construct of interest; (4) reliability or consistency and reproducibility of the scores; (5) validity or the extent to which the tool is measuring the construct it claims to measure; and (6) responsiveness or the extent to which the measure can detect clinically important changes over time.8

Whether in clinical practice or in research, careful attention should be given to the psychometric properties of a measure to ensure that the measure selected was appropriately scaled and can accurately measure the construct of interest. Scoring the results must be done using the prescribed manner in which the tool was developed, including the interpretation of missing values and the ability to use only particular subscales. Reliability estimates the extent to which the scores produced are free from random error and are stable and accurate. Reliability can be estimated across time (test–retest), across raters (interrater) and within the measure (internal consistency).21 Test–retest reliability verifies whether the score is consistent on repeated evaluation over a time frame in which you would not expect to observe any alteration in the results. For example, if too short a time interval is used, learning or practice effects may influence item execution. If the test is repeated right after completion of the first test, fatigue may influence performance. If too wide a time interval is used, there may be actual maturational or other (e.g. disease progression) changes in abilities. Interrater reliability ensures that any trained rater will achieve the same results when evaluating a group of individuals. Reliability decreases as a function of greater measurement error (i.e. the extent to which observed scores will vary from the true scores). Both the experience or training of the tester and the child’s level of attention or compliance can contribute to measurement error. Indeed, the evaluator can introduce personal biases in the administration of the test and interpretation of the results, and therefore the use of ‘blind’ evaluators in research, who are unaware of the study hypotheses or information about the research participants, is necessary to minimize evaluator bias. Reliability is typically represented statistically as a coefficient between 0 and 1.0, with 1.0 meaning no measurement error; the closer to 1.0, the more confident you can be about the stability of the scores produced. Correlations provide incomplete data, as they reflect only the degree of association and not the level of agreement. For example, one evaluator may consistently score children 10 points higher than another evaluator, and the resultant correlations between the two evaluators would be highly associated, but clearly with low agreement. The intraclass correlation coefficient (ICC) is a more appropriate analytical approach for continuous data to measure both association and level of agreement between multiple raters or scores obtained over repeated measurements. Kappa should be used for categorical data and weighted kappa should be used for ordinal data in order to determine the reliability between raters or tests. This provides the level of agreement as a ratio of observed agreement compared with the agreement expected by chance. Generally, ICC or kappa should be > 0.70 or 0.75 (for group data; it should be higher for use with individuals), with scores of 0.5 to 0.75 considered moderate correlation. Confidence intervals (range of values of the actual reliability, as low as ‘x’ or as high as ‘y’) of these estimates of reliability are helpful to examine as well, as the narrower the confidence interval, the more precise the estimate of the true value.20,27,28 When examining reliability estimates for a measure, it is critical to verify which group of participants (e.g. typically developing individuals or children with a particular diagnosis or health condition) were used in the reliability studies as the results cannot be generalized to other populations.

Validity refers to the extent to which a tool is measuring what it is designed to measure. First, in consultation with experts, the content of the tool is carefully evaluated to ensure that it includes all attributes of the phenomenon of interest and that it is logical and meaningful to the tester and the individuals being tested (face and content validity). The test can then be compared with a criterion standard measure which assesses the same characteristics or construct (concurrent validity). Another validation approach is to ascertain whether the measure correlates with scores obtained on a clinically relevant criterion standard in the future (predictive validity), thus verifying an association with behaviours in the future that you would expect it to predict. For these statistical comparisons, Spearman’s or Pearson’s correlation coefficient is generally used, and the closer the score is to 1.0, the stronger the association. Criterion standards often do not exist, and therefore the level of correlation is not perfect as the two
measures being compared may be measuring similar but not identical attributes. Validity can also be estimated by demonstrating that the tool behaves as predicted in differentiating groups of individuals (discriminant validity) with high and low performance levels in known groups, for example, by using t-tests. Evidence that the tool relates to measures of a similar construct (convergent validity) and does not relate to measures of a different construct (divergent validity) can also be verified statistically, thus providing additional evidence for construct validity. In addition, the domains or subscales within the measure should aggregate well together to represent particular attributes of interest (internal consistency), and this can be verified using Cronbach’s alpha. Factor analysis and principal components analysis are other statistical approaches that are used to verify that, conceptually, several items or subscales collectively explain the variance in scores.20,22,27,28

As discussed above (i.e. the purpose of the measure), responsiveness of an instrument refers to its ability to detect clinically meaningful changes over time. This is particularly important to verify in the case of clinicians who want to use the tool to quantify the effectiveness of a treatment, or for researchers who are carrying out an intervention trial. A variety of statistical approaches may be used to evaluate the responsiveness of a tool, although often the effect size is reported, with an effect size >0.2 indicating a small effect, >0.5 a moderate effect and >0.8 a large effect.20,27,28

For a more detailed description of psychometric theory, Streiner and Norman27 and Portnoy and Watkins28 are excellent reference texts.

REFERENCES