

# AACPDM Methodology for Developing Evidence Tables and Reviewing Treatment Outcomes Research

---

## Inside

- **Step-by-Step Process for Reviewing Treatment Outcomes Evidence**
- **How to Code Outcomes by Dimensions of Disablement and How It Changed in 2003**
- **How to Code Studies for Strength of Evidence**
- **How the AACPDM Levels of Evidence Relate to Sackett's Levels of Evidence**
- **How to Structure Evidence Tables**
- **How to Write and Submit a Review Article for Publication under the Auspices of the AACPDM**

*This procedures manual was authored by Charlene Butler in 1998-1999 and approved by the 1998-1999 AACPDM Treatment Outcomes Committee to standardize the development and presentation of evidence tables and review articles about treatment outcomes. Special thanks goes to Peter Rosenbaum, MD, Suzann Campbell, PhD, and Owen White, PhD for their extensive help. Our additional goal has been to produce a universally accessible and up-to-date manual.*

*Charlene Butler, Ed.D., Committee Chair 1995-1998*

*Hank Chambers, M.D., Committee Chair 1998-1999*

*Murray Goldstein, D.O., M.P.H.*

*Richard Adams, M.D.*

*Lisa Samson-Fang, M.D.*

*Johanna Darragh, Ph.D.*

*Maureen O'Donnell, M.D.*

*Judy Leach, P.T.*

*Terrance Edgar, M.D.*

*Mark Abel, M.D.*

*Michael Msall, M.D.*

**Committee-approved Modifications: August 1999; July 2001; January 2003**

## Table of Contents

<b>Table of Contents</b>	<b>2</b>
<b>Limitations of this Methodology</b>	<b>4</b>
<b>Developing the Evidence Tables and Review of Literature: Step-by-Step</b>	<b>5</b>
<b>Appendix 1: Study Summary Form</b>	<b>10</b>
<b>Appendix 2a: Coding Dimensions of Disablement; Replaced January 2003 with Appendix 2b</b>	<b>12</b>
<b>Appendix 2b: Coding Dimensions of Disablement (Beginning 2003) Using WHO 2001 Classification (Replaces Appendix 2a)</b>	<b>34</b>
<i>Classification</i>	35
<i>Sticky Wickets in Coding</i>	35
<i>Differentiating Manifestation of Dimensions, Treatment Outcomes, and Outcome Measures</i>	37
<i>Impairment of Body Function</i>	38
Manifestation	38
Treatment Outcomes	39
Outcome Measures	40
Examples of Impairment of Body Function	40
<i>Impairment of Body Structures</i>	41
Manifestation	41
Treatment Outcomes	42
Outcome Measures	42
Examples of Impairment of Body Structure	43
<i>Activity Limitations and Participation Restriction</i>	43
Manifestation	43
Treatment Outcomes	45
Outcome Measures	46
Examples of Activity and Participation	46
<i>Environmental Factors</i>	48
Manifestation	48
Treatment Outcomes	50
Outcome Measures	52
<i>Example of Impact on Coding with Change of Classification System</i>	51

*Lists of Miscellaneous Outcomes or Measures by Dimension* 54

Impairment of Body Function 54

Impairment of Body Structure 55

Activity and Participation 55

Environmental Factors 57

**Appendix 3: Coding Levels of Evidence 58**

*Interpreting the Validity of Evidence* 58

Face Value Difference 59

Clinically Important Difference 59

Statistically Valid Difference 59

Causality or Attribution 61

*Levels of Evidence* 62

Types of Evidence 62

Evaluating External Evidence 62

Forming Hierarchies of Evidence 64

AACPDM Classification of Levels of Evidence 66

*Background* 66

*Classification* 69

*Operational Definitions* 71

Coding Studies (Revised July 2001) 75

**Appendix 4: Summary and Evidence Tables 78**

**References 81**

# Methodology

## for

### Developing Evidence Tables and Reviewing Treatment Outcomes Research

The AACPDMD has undertaken the development of “state of knowledge” documents, that is, summaries of the types of evidence that support use of interventions in developmental medicine. These are not “best practices” documents or practice guidelines. The Academy is neither endorsing nor disapproving of an intervention. Our goal is to present the evidence about interventions in an organized fashion. An evidence table is a convenient way to summarize evidence available for a specific condition and intervention. Evidence tables themselves do not specify how to treat a condition; they identify outcomes of treatment for which evidence is available. Such tables make it easier to visualize gaps in evidence, and can help identify new research that is needed.

This report describes a methodology for developing and presenting evidence tables that is based on a two-part conceptual framework. This framework makes it possible to 1) analyze treatment outcomes according to five dimensions of disablement and 2) differentiate between lesser versus stronger evidence about an intervention. (If you want background information about the Academy’s selection of this conceptual framework, read “Evaluating Research in Developmental Disabilities: a Conceptual Framework for Reviewing Treatment Outcomes “ which can also be found on this Web site.

#### **Limitations of this Methodology**

The development of this methodology originally grew out of the NCMRR disablement classification and Sackett’s levels of evidence classification. Neither of these classification systems was found to be complete or perfect for our purposes; both had to be adapted. See Appendices 2 and 3. When the WHO International Classification of Health, Functioning and Disability was adopted world-wide, this replaced the dimensions of disablement classification. Few operational definitions were available for either classification; these had to be developed. Moreover, reducing a virtually infinite variety of outcomes and research methodologies into a small number of categories to fit any classification system inevitably involves simplifications that carry with them some inaccuracy. Despite our efforts to operationally define the elements of each classification schema, inevitably some vagaries remain. In these cases, there is no option but to fall back on subjective judgments.

When a user of this methodology is uncertain about how to code an outcome or the level of evidence represented by a study, you are urged to use this as an opportunity to help the Treatment Outcomes Committee continue to refine this methodology. Please solicit the collective opinion of your colleagues by contacting the chairperson or designated individual of the Treatment Outcomes Committee to pose your question or uncertainty to the committee members. In this way, a group consensus can be reached and information continually added to this manual for the subsequent use of others. Thus, standardization of coding of evidence across reviews can be accomplished.

This first edition of the Methodology had initial field testing by the author and some members of the Treatment Outcomes Committee, but there is no doubt that parts may still be confusing or misleading to new users. Because it is a working document, however, and is in an on-line format, it can and will be continually updated as we use it and learn how it can be improved. Therefore, we request your suggestions to make successive “editions” clearer or easier to use.

Applying this approach requires two ingredients: some specific knowledge and a framework (a road map that organizes and directs that knowledge). This methodology provides the road map and tries to provide enough basic knowledge to follow it. If you need to develop further knowledge, especially in research concepts, we suggest that you supplement this material with additional reading from the references.

## **Developing the Evidence Tables and Review of Literature: Step-by-Step**

### **1. Define the intervention.**

Specifically state the intervention that will be evaluated to focus the literature search to relevant studies. In some cases, such as in Example 1 that follows, it is important to define precisely what the intervention includes and does not include.

### **2. Define the population for whom the intervention will be reviewed.**

Define the patient population as precisely as possible to focus the literature search to relevant studies. Also see Example 1.

#### **Example 1: *Specifying the intervention and the population addressed by a review.***

This review examines effects of intrathecal baclofen on individuals of any age with cerebral palsy. The intervention, *intrathecal baclofen*, includes baclofen administered by (1) single or multiple injections over the dorsal surface of the spinal cord) or (2) a subcutaneous implanted pump that delivers a continuous infusion into the lumbar cerebrospinal fluid. It excludes orally-administered baclofen or baclofen injected subcutaneously or into muscle.

This review is concerned with people with cerebral palsy. *Cerebral palsy* has traditionally been described as an evolving disorder of motor function secondary to a non-progressive pathology of the immature brain and is characterized by abnormalities of movement (i.e., spasticity, athetosis, chorea, dystonia, and ataxia). Two-thirds of individuals with cerebral palsy have spasticity, either alone or in combination with the other movement abnormalities. The evidence table includes studies (1) whose subjects were primarily individuals diagnosed with cerebral palsy with spasticity, alone or in combination with other types of abnormal movement or (2) that provided specific data about a subgroup of such individuals. This review excludes studies of spasticity of spinal origin (e.g., multiple sclerosis or spinal cord injury) or of cerebral origin due primarily to other causes such as traumatic brain injury.

### 3. Identify and record sources to be used in the search.

Decide whether the literature search will be limited to published literature or whether unpublished sources will be investigated. Unpublished data, such as results from dissertations, or clinical trials that have not yet been published, can also provide useful information. However, they are hard to track down and have not been subjected yet to peer review.

Determine which kinds of published articles will be examined and whether published abstracts from meetings will be acceptable. Citations from review articles and textbooks will be helpful to ascertain that all published studies are identified, but only original studies will be included in the review. Determine the bibliographic sources that will be used in the search:

- Bibliographic Indices: *Index Medicus*, *Excerpta Medica*, *Science Citation Index*, etc.
- Automated Data Bases: MEDLINE and other services in the National Library of Medicine's Medical Literature Retrieval System, and commercial data bases such as Ovid Medline and Healthstar
- Abstracting services
- Citations from bibliographies, textbooks, etc.
- Reports to granting agencies

Not all these sources have to be used; however, it is important to be explicit about the sources used and the process followed.

### 4. Keep a record of the search process and report it.

MEDLINE and similar databases have indexing problems that may fail to retrieve relevant studies. Refer to the chapter, "Searching for the Best Evidence", in *Evidence-based Medicine*, 1997, for thorough guidance on search strategies for MEDLINE and the Internet. Use the Clinical Queries feature in the Advanced PubMed search to target clinically relevant studies, but search MEDLINE in other ways, as well, to assure complete retrieval of studies.

- Identify key words or Medical Subject Headings (MeSH) used to search.
- State the inclusion and exclusion criteria applied to the search and to the subsequent selection process. Examples: time (all articles after 1966), language (English language only), publication type (randomized controlled trial, review, editorial).

See the following Example 3.

**Example 3: *Specifying types of research examined and search process used.***

The literature search was limited to published studies, full-text available in English, that are accessible through the United States university library system. Searching was conducted through the electronic, bibliographic MEDLINE data base of the National Library of Medicine (1966 to present). This data base has been searched in two ways to assure retrieval of all relevant studies. The Clinical Query feature of the PubMed search program used the exploded terms cerebral spasticity or cerebral palsy and intrathecal baclofen as Medical Subject Headings (MeSH) to identify clinically relevant studies about therapy. With the EndNote© search program, additional MeSH terms were used as follows: treatment outcomes, clinical trials, cohort studies, prospective studies, cross-sectional studies, clinical series, and case reports. Finally, the reference listings in relevant review articles and studies were examined.

Fifty-eight citations were identified. Based on review of their abstracts, forty-eight full-text articles were obtained for evaluation. Sixteen original studies or case reports were found; three were excluded because no results were specifically reported for any subjects with cerebral palsy. The evidence tables for this review of treatment outcomes are based on 14 studies.

**5. Select relevant studies.**

Document this process by keeping records of which studies were retrieved and which studies were excluded (along with the reasons why). Given the paucity of studies about many interventions that will be reviewed, generally speaking, all studies are of interest and will be included. When no empirical evidence is available, then articles that describe the intervention will be included provided they provide insight into the non-empirical evidence in support of the intervention, i.e., theory, expert opinion that is essentially testimonial evidence, or common sense. Studies of economic analysis of an intervention should also be obtained as these will provide information for the dimension of disablement called societal limitations or contextual factors.

**6. Extract and record data from each study.**

Appendix 1 contains the form to be used for consolidating information for each study. (Reviewers will provide a copy these completed forms to the Treatment Outcomes Committee when the review article is submitted.) Steps 7 (dimensions of disability) and 8 (interpreting the results) will complete information required in this form.

**7. Code each measured and anecdotal treatment outcome for the dimension of disability it represents.**

Appendix 2b provides background information, operational definitions, extensive explanations, and examples of outcomes (and their measures) for each of the five dimensions: impairment of body structure, impairment of body function, activity and participation, environmental factors, and other contextual factors. Appendix 2a contains the classification that was used for evidence reports developed before 2003, prior to the completion and adoption of the final WHO classification

## 8. Code each study for level of evidence it represents.

Appendix 3 provides background information about issues of internal validity, the classification for levels of evidence, and operational definitions and checklists for each level to differentiate further between lesser versus greater credibility of the evidence.

- Interpreting face value results: clinical importance
- Interpreting statistical validity: inferential statistics
- Interpreting uniformity of effect in subjects
- Attributing change to the intervention: levels of evidence
- Consistency of results across studies

## 9. Create summary tables of the studies.

See Appendix 4 for format and examples.

## 10. Create the evidence tables.

The table will be structured in two parts and allow analysis of the data for answering the research questions that will follow in the next step. The tables and examples sections are in Appendix 4.

## 11. Write the review article to answer the following questions.

1. What evidence exists about the effects of the intervention in the dimension of disability (pathophysiology, impairment, functional limitation/activity, disability/participation, or societal limitation/other contextual factors) in which it was expected to work?
2. What evidence exists about the effects of the intervention in the other dimensions of disablement?
3. What evidence exists for linkages of effects within and between these dimensions, i.e, which effects, if any, appear to carry over automatically?
4. What kinds and magnitude of medical complications have been documented?  
An understanding of the medical risks is necessary for assessing whether the benefits of an intervention outweigh its risks. The review will report the type and severity of complications that have been documented. It will not express a judgment about whether an intervention does more good than harm because a benefit/harm analysis can more appropriately be made by the reader who is considering the intervention for a particular individual.
5. What is the strength of the evidence?  
Internal validity concerns the level of credibility we can have in the face value results of an individual study. Is the evidence worthwhile, clinically speaking? Is it statistically valid? With what level of confidence can the observed results be attributed to the intervention? In addition, is the evidence consistently replicated in several studies (regardless of their level of evidence); consistently

positive, negative, or unchanged results across studies adds to the overall credibility of the efficacy of the intervention in producing that outcome.

How applicable is the evidence to other individuals? This question concerns the external validity or generalizability of the evidence to people beyond those that have been studied. This review will address the generalizability of the body of evidence rather than the generalizability of the results of an individual study. It will make statements about the total number of studies that have been conducted, the overall levels of evidence constituting the body of research, the total number of subjects that have been studied, the homogeneity or heterogeneity of the studied subjects, and to what extent results have been replicated across studies. The review will not make recommendations about people for whom the intervention may, or may not, be indicated or about the probability of generality of effect. Information about the subjects in the tables will allow the reader to decide whether generalization of effect is applicable to a particular individual for whom the intervention is being considered.

## **12. Submit the article to the Treatment Outcomes Committee for publication under the imprimatur of the AACPDM.**

Anyone wishing to develop a review article about an intervention is invited to do so. The Treatment Outcomes Committee is facilitating reviews based on a priority listing, but welcomes the review of any intervention you are willing to conduct using this format. To assure that an effort on the topic in which you are interested is not already underway by someone else, discuss your interest and intention with the Committee Chairperson (or designee).

See the Committee Procedures for Promoting, Reviewing and Disseminating Evidence Reports which is also on this web site.

Upon approval of the article by the review committee, the article will be posted on this Academy web site and will be submitted to *Developmental Medicine and Child Neurology* for publication in the name of the authors with the imprimatur of the Academy. The Committee may also organize the presentation of reviewed material at the annual meetings of the Academy or publish it in some other way.

This appendix contains a blank form that can be copied and used for recording the needed information from each study.

### Appendix 1: Study Summary Form

Reviewer's Name:

***Citation information:***

***Practice setting:*** (Place and/or type):

***Subjects:***

Number: In treatment condition:            In control condition:

Target Population:

Diagnoses and # in each Dx subgroup:

Ages:

***Specific intervention used:***

***Specific control used:***

***Complications:***

***Economic costs:***

***Period individuals were studied:***

***Research design and level of evidence for study:***

**Outcomes**

***Treatment compared to another condition (no treatment, placebo, or alternative treatment)***

<i>Outcome of Interest</i>	<i>Measure Used to Assess*</i>	<i>Dim</i>	<i>Result</i>	<i>Imp</i>	<i>Power</i>	<i>Stats</i>	<i>Lev</i>

Legend:

Dim      Dimension of disability (record after completing Step 7).

Results    Face value of difference (or change) between group means (median or mode) or between group pre-test and post-test in case series

Imp      Clinical importance or significance of the obtained difference. Write *yes* to indicate that author suggests that a face value difference was large enough to be clinically important (or if data suggests it to you); *no*, if not; ? if there is no way that you can judge this.

Power    Power calculations. Write *yes* if calculations were done prior to study / *yes* or *no* to indicate whether the study achieved the number of subjects needed according to the calculations.

Stats    Statistical validity of face value result. Write *yes* if valid at minimally acceptable level of  $p < .05$  or CI 95%. Then write the reported p value or CI. If the author only states that the result was statistically significant, write *SS*; if author stated only that the result is not statistically significant, write *NSS*.

Lev      Level of evidence of the study. Strength of confidence for attributing result to the intervention. (Record after completing Step 8).

- \* Include anecdotes; this will always be Level V evidence regardless of study in which they are reported

***Uniformity of effect within treated group***

<i>Outcome of Interest</i>	<i>Measure Used to Assess</i>	<i>Dim</i>	<i>Imprv</i>	<i>Wors</i>	<i>Unch</i>	<i>Stats</i>	<i>Lev</i>

Legend:

- Imprv Number and percent of group that improved
- Wors Number and percent of group that worsened
- Unch Number and percent of group that were unchanged

***Evaluating Level of Evidence***

The first step is to decide the highest level of evidence the ***research design*** could have provided using the Levels of Evidence classification, i.e., Level I, II, III, IV, or V. See Classification in Appendix 3. To do so, you must determine what research design was used for each study. Sometimes this is stated in the published report; when it is not, review the Methods section to figure out what design in the classification it is most like.

Next evaluate the ***actual conduct of the study*** using the following questions. Conduct of the study will be judged as Strong (score of 7 or 6), Moderate (score 5), or Weak (=4). Note, however, that Level V evidence is not empirical research; therefore, Level V evidence will not be coded for conduct of study. Indicate your answer with supporting notes below.

1. Were inclusion and exclusion criteria of the study population well described and followed?
2. Was the intervention well described and was there adherence to the intervention assignment? (For 2-group designs, was the control exposure also well described?)
3. Were the measures used clearly described? Were they valid and reliable for measuring the outcomes of interest?
4. Was the outcome assessor unaware of the intervention status of the participants (i.e., blind assessment)?
5. Did the authors conduct and report appropriate statistical evaluation including power calculations?
6. Were dropout/loss to follow-up reported and less than 20%? For 2-group designs, was dropout balanced?
7. Considering the study design, were appropriate methods for controlling confounding variables and limiting potential biases used?

## Appendix 2a: Coding Dimensions of Disablement (1999-2002) Using NCMRR/WHO Classification

(This was replaced January 2003 by Appendix 2b and is no longer in use for evidence reports begun in 2003.)

### Dimensions of Disablement

<i>Dimension</i>	<i>Description</i>
• Pathophysiology	Interruption or interference of normal physiology and developmental processes or structures
• Impairment	Loss or abnormality of body structure or physiologic body function
• Functional Limitation/Activity	Restriction of ability to perform functional activities
• Disability/Participation	Restricted participation in typical societal roles
• Societal Limitation/ Context Factors	Barriers to full participation imposed by societal attitudes, architectural barriers, and social policies; and other external factors

### ***Resolving the Terminology Confusion***

The AACPD Treatment Outcomes Committee found that experienced practitioners were not consistent with one another in coding research findings or treatment outcomes given only the descriptions provided by NCMRR as shown above. Why the confusion? The terms impairment, disability and handicap are often used interchangeably in everyday speech and writing. In a disablement classification system, however, these terms are given specific meanings that may differ from their everyday usage and are not interchangeable. Unfortunately, specific meanings and terms have not been consistently applied across the classification systems many of us have used. The article *Evaluating the research in developmental disabilities...* (also on this web site) discusses why the Nagi, WHO, IOM and NCMRR models each used somewhat different terms for conceptually similar reference levels: the level of molecules and cells versus the level of body parts or organs versus the level of activity performance. Regardless of the reasons, the inconsistent use of terms and meanings has created significant confusion. Adding to this problem, all four classifications were originally described in a relatively general manner; the basic attributes for each reference level were given, but the boundaries and measurement characteristics were not.

It was recognized that users of the classification must be guided by more specificity, i.e., by “operational definitions” which denote the basic attributes of each dimension and their measurement characteristics as well as their boundaries. For the AACPD to make comparisons between evidence tables produced by different authors, consistent use of terminology is imperative. The Treatment Outcomes Committee, therefore, took the following steps:

It established that the commonly used disablement models (World Health Organization or WHO, Institute of Medicine or IOM and NCMRR) shared common reference levels despite the different names that were given to the categories in each classification as shown in the following figure.

---

***Reference Levels Shared by Disablement Classification Systems***

---

Level of molecules, cells, tissues

Level of organs, organ systems, and the whole body

Level of activity performance

Level of social role participation

Level of external barriers to social role participation

---

The 1980 World Health Organization International Classification of Impairment, Disability and Handicap (ICIDH) and its 1997 revision draft (ICIDH-2), the 1991 U.S. Institute of Medicine model and its clarification in 1997, and the NCMRR model were combed for all comments and examples to gain understanding of the outcomes that were intended to be representative of these reference levels in each model.

The Treatment Outcomes Committee has relied heavily on the WHO revision work for two reasons: 1) ICIDH-2 has established operational definitions for three of the reference levels, i.e., the organ or body level, the activity level, and the societal role level and 2) the U.S. Public Health Service, IOM and NCMRR, along with Canadian and European agencies, are publicly supporting the ICIDH-2 revision process to reach an international agreement on and use of a single taxonomy. The AACPD, therefore, will want to have the option of converting to the use of the new WHO classification terminology once universal adoption is imminent or accomplished. WHO expects field testing of its main revision document to be completed in late 1999. WHO then plans on-going revision to provide operational definitions to 1) address the reference level of the molecular, cellular and tissue effects of a health condition and 2) the societal barriers reference level (i.e., which WHO currently refers to as “contextual factors”).

The Treatment Outcomes Committee, therefore, investigated the equivalence of the NCMRR classification and the current ICIDH-2 draft. Equivalence between the NCMRR and the WHO classifications is imperative if eventual comparison of evidence tables produced under two classifications is to be possible. The ICIDH-2 uses a coding whereby the letters *i*, *s*, *a*, *p* and *c* will denote codes for Impairments of Function, Impairments of Structure, Activities, Participation and the list of Contextual factors, respectively. These equate to the NCMRR classification as follows:

**Reference Levels and Equivalence of NCMRR and WHO Classifications**

<i>Reference</i>	<i>Molecules, cells, tissues</i>	<i>Organs, organ systems, body parts</i>	<i>Activity performance, skills</i>	<i>Participation in social roles</i>	<i>External barriers to social participation</i>
NCMRR	Patho physiology	Impairment	Functional limitation	Disability	Societal limitation
ICIDH-2	Health condition	Impairment	Activity	Participation	Contextual factors
ICIDH-2 Coding		I code S code	A code	P code	

Finding that the two classifications are essentially equivalent, the Committee combined the NCMRR and ICIDH-2 operational definitions or examples and expanded their examples and will refer to the taxonomies together until the WHO terminology is adopted by the Academy. We have not proposed a hybrid NCMRR / WHO classification, but are explaining the NCMRR classification in terms of the WHO revision. The purpose is to facilitate a seamless transition to the use of revised WHO taxonomy when it becomes the universally accepted classification. Other commonalities found between these two classification systems have been maintained as follows:

- A glossary definition of each dimension tries to capture the essence of the concept referred to by the category heading, together with inclusion and exclusion criteria.
- Terminology regarding the dimensions should be thought of and denoted by "neutral" words as much as possible so that both positive and negative aspects of each dimension can be addressed. For example, use of the activity instead of activity limitation, body temperature instead of fever, and muscle force instead of strength or weakness, or muscle tone instead of hypertonus or hypotonus.
- The further specification of impairment as to impairment of function versus impairment of structure (as in the WHO revision) has been incorporated into our coding of studies using the NCMRR.
- The activities dimension in the ICIDH-2 (equivalent to functional limitations in NCMRR) is based solely on "activities" or skills of the person, which are actual performances. Thus, the unfruitful debates over "can do" versus "does do" or "might do" are avoided.
- Contextual factors in the ICIDH-2 (equivalent to societal limitations in NCMRR) are currently included as a factor of the participation dimension in the ICIDH-2, but are being coded separately in the field trials.

***Differentiating Between the Dimensions of the NCMRR / WHO Classification***

In each of the five dimensions that map the disablement process, there are three aspects that will be discussed:

1. **Manifestation** or how the disablement process acts in that dimension,
2. **Treatment outcome**, the outcome or result that is observed in association with an intervention, and

3. *Outcome measure*, the type of instrument or technique used to measure an outcome.

Seven health conditions (PKU, narcolepsy, cerebral palsy, reading disability, attention deficit disorder, panic disorder, and phocomelia) have been selected to serve as examples throughout the discussion of each of the dimensions in an attempt to address a wide variety of outcomes. These developmental disabilities are complex. Seldom, if ever, does a genetic abnormality or congenital defect produce a single effect such as mental retardation or abnormal muscle tone or a structural defect of the heart. They usually produce multiple primary anomalies as well as secondary effects. Consequently, a “primary condition” is a risk factor for one or more “secondary conditions. For the sake of simplicity, however, the seven health conditions used as examples will address only a single feature of the health condition.

At the end of this appendix, there are more outcomes of interest and outcome measures found in developmental disabilities research listed for each dimension. These lists do not include all outcome measures that may be extracted from studies, but are intended to be sufficiently representative to guide decision-making about other measures that may be encountered. These lists will grow over time to provide for consistent coding by all users.

### ***Pathophysiology (Reference Level: Molecules, Cells, Tissue)***

#### **Manifestation**

Pathophysiology is defined, for purposes of this classification system, as a health condition, i.e., an alteration or attribute of the health status of an individual. It is the interruption or interference of typical 1) physiological processes, 2) developmental processes, and/or 3) structures. (Loss of structure occurs at this level when some but not all cells or tissues of an organ or body part are affected. If the entire organ or body part is absent, then the loss is said to occur in the next dimension, impairment.) It may be a disease (acute or chronic), disorder, injury or trauma, or reflect other health-related states such as pregnancy, aging, stress, congenital anomaly, or genetic predisposition.

Pathophysiology includes the medical diagnosis, the etiology, and as much detail as is known about the underlying molecular, cellular and tissue disturbances. As such, this dimension also represents a description of the population for which an intervention has been evaluated in a particular study. The more specific researchers are in selecting subjects and in reporting the pathophysiology of their sample population, the more definitively we can understand for whom the findings of a intervention may apply. For example, cerebral palsy is a health condition, but there are many subgroups of cerebral palsy (i.e., spastic, athetoid, ataxic, ect.), with different etiologies, different sites of cellular damage, and varying physiologic dysfunction.

#### **Treatment Outcomes**

Interventions targeted at the pathophysiological dimension are those that attempt to restore the underlying genetic or cellular structures or metabolic processes to normal. Consequently, interventions here depend on specific understanding of a health condition at the molecular, cellular or tissue level. This understanding continues to evolve with the advent of new technologies (e.g., imaging and gene mapping) and the insights they provide.

## Outcome Measures

Techniques that quantify pathophysiological outcomes are those that measure anatomical changes of molecules, cells and tissues or how they function.

## Examples of Pathophysiology

### In PKU or Phenylketonuria

**Manifestation.** Inborn error of metabolism in which elevated levels of phenylalanine in the blood are toxic to the central nervous system producing mental retardation. **Treatment Outcomes.** Results of interventions that attempt to restore genetic structures or metabolic processes to normal by alternating the DNA structures or changing the biochemical processes they orchestrate or changing the blood phenylalanine levels. **Outcome Measures.** phenylalanine hydroxylase cDNA probe, measures of enzyme activity, measures of myelin sheathing, DGGE and SSCP analysis, plasma levels of phenylalanine.

### In Narcolepsy

**Manifestation.** Etiology and transmission unknown but strong association has been shown with histocompatibility antigen complex HLA-DR2, i.e., people with narcolepsy are HLA-DR2 positive. **Treatment Outcomes.** Results of interventions that attempt to change HLA-DR2 status. **Outcome Measures.** HLA-DR2 probe.

### In Cerebral Palsy or CP

**Manifestation.** Static encephalopathy with cell or tract damage in the brain. **Treatment Outcomes.** Results of interventions intended to repair the brain cell or tract damage, prevent the cell damage (e.g., blood transfusion for rh incompatibility, drugs to control intracranial hemorrhage) or restore normal function (e.g., neural firing). **Outcome Measures.** IVH grading.

### In Reading Disability

**Manifestation.** Unknown but suspected abnormality in medial geniculate nucleus in the brain resulting in too few neurons to “hear” staccato consonants; probably a multifactorial etiology. **Treatment Outcomes.** Change in number of neurons that fire when listening to these consonants. **Outcome Measures.** Imaging techniques, EMG.

### In Attention Deficit Disorder or ADD

**Manifestation.** Unknown; suspected multifactorial etiology. **Treatment Outcomes.** Unknown **Outcome Measures.** Unknown

## In Panic Disorder

*Manifestation.* Unknown. *Treatment Outcomes.* Unknown *Outcome Measures.* Unknown

## In Phocomelia

*Manifestation.* Arrested limb bud development; probable multifactorial etiology. *Treatment Outcomes.* Unknown *Outcome Measures.* Unknown

See additional list of miscellaneous examples about pathophysiology at the end of this appendix.

### ***Impairment (Reference Level: Organs, Organ Systems, or Body Parts)***

#### Manifestation

An impairment indicates a loss or abnormality of a body *part* (i.e., structure) or body *function* (i.e., physiological function of a body part or an organ system). Vision and hearing are examples of body function. Their structural correlates are the eye and its parts, the ear and its parts, the central nervous system and its parts. Functions include but are not limited to the “senses. Limitations in certain functions (inability to carry out a basic function of the body or body parts such as extension of arm, weight-bearing, symmetrical stance) are impairments. Patterns of neural firing of muscles is another example. The physiological functions include not only the physical functions of the central nervous system such as neural firing patterns, but also mental functions (i.e., cognitive and emotional functions).

Although “organ” has been the referenced term in the various models previously, the definition of “organ” has been unclear. The eye and ear are traditionally considered organs; however, it has been difficult to identify the boundaries of extremities and internal organs. Instead of an approach by “organ” which implies the existence of an entity or unit within the body, the ICDH revision replaces this term with “body structure”. “Body” refers to the human organism as a whole; hence, it includes the brain and its functions. Impairments are broader and more inclusive in scope than a disorder or a disease; thus, absence of a leg is a structural impairment while the cause, phocomelia (arrested limb bud development), is the health condition or pathophysiology.

Nouns are used to denote impairment of structure (i.e., eye and eye parts, brain, arm, foot, heart) and impairment of function (i.e., cognition; sensation; speech; vision; audition; respiration; and social, emotional and motor development). Abnormality here refers to a significant variation from established statistical norms and should be used only in this sense (i.e., as a deviation from a population mean within measured standard norms). Measures that yield developmental ages or other standard scores that compare these functions to the norm indicate impairment status (e.g., I.Q., gross motor age, 20/100 visual acuity, etc.).

Impairments are detectable or noticeable by others or the person by direct observation or by inference from indirect observation. They are absent or present; once an impairment is present, it may be scaled by its severity.

The effects of assistive devices occur in the functional limitation (i.e., activity level) of this classification because assistive devices have the potential to remove limitations on activities that can be performed. Without the assistive device, it is the person's activity that would continue to be limited.

Impairments may result in other impairments. For example, muscle spasticity can result in contractures which can restrict range of motion, impair movement, or both.

The presence of pain is an impairment and it may, or may not, be functionally limiting. The activities (e.g., walking, sleeping) that pain prevents or restricts are functional limitations. Endurance, too, is an aspect of impairment which may, or may not, be associated with a functional limitation. If some activity cannot be completed at all or without undue fatigue, the activity that is limited is coded in the functional limitation / activity dimension.

To be consistent with the ICDH-2, when coding manifestations, outcomes or outcome measures at the impairment level, specify whether the impairment is one of structure or of function.

### Treatment Outcomes

Interventions targeted at the impairment level are those that attempt to restore body structures or functions, as compared to normal. A kidney transplant is an intervention intended to restore a body structure. Gastrostomy feeding is intended to restore growth, a feature of body structure which has established statistical norms. A botox injection is an intervention to restore body function, i.e. to reduce spastic muscle tone. Infant stimulation programs are intended to promote progress in various developmental domains, i.e., social development, language development, cognitive development, etc.

### Outcome Measures

Techniques that quantify impairment outcomes are those that measure anatomical changes or features of organs, organ systems or body parts—or how they function—relative to population norms. Impairment is often expressed as a standard score, such as motor age, social quotient, I.Q., which reflect the degree of impairment in the various domains of child development compared to the norm. In addition, techniques that measure various aspects of physiologic function are measures of impairment. For example, speed or force of isolated movement (as in voluntary dorsiflexion of foot) or coordination of muscle activation are measures of impairment of physiologic or neurologic motor function. [In contrast, techniques that measure the speed with which a motor activity or skill is accomplished are measures in the dimension of functional limitation / activity. For example, timed samples of placing cubes in a box or walking a certain distance at school or within a certain time) are measures of functional limitation in activities.]

### In PKU or Phenylketonuria

**Manifestation.** Cognitive impairment, i.e., mental retardation of subnormality of intelligence.

**Treatment Outcomes.** Change in intellectual performance. **Outcome Measures.** standardized test of intelligence (e.g., WISC)

## In Narcolepsy

**Manifestation.** Excessive daytime somnolence and uncontrollable sleep attacks or episodes.

**Treatment Outcomes.**  
daytime sleep episodes.

**Outcome Measures.** Frequency count of

## In Cerebral Palsy or CP

**Manifestation.** Neuromuscular impairment (abnormal activation of muscles, i.e., motor dysfunction that varies depending on pathophysiology, e.g., athetoid type is a form that includes uncontrolled involuntary motion of the distal joints. **Treatment Outcomes.** Changes in factors that affect coordinated movement patterns such as muscle tone, posture and balance, reflex activity, equilibrium reactions **Outcome Measures.** Kinematic/dynamic and electronic recordings of EMG patterns, Ashworth Scale, goniometer reading of range of motion.

## In Reading Disability

**Manifestation.** Dysfunction of phonological processing due to auditory perceptual dysfunction or a specialized type of hearing problem. **Treatment Outcomes.** Change in acquisition of alphabetic code (application of phonological rules to print). **Outcomes Measures.** Decoding skill test.

## In Attention Deficit Disorder or ADD

**Manifestation.** Visual and auditory distractibility. **Treatment Outcomes.** Change in ability to sustain selective attention in spite of distracting sounds and sights. **Outcome Measures.** Standardized rating questionnaires about observed distractible behaviors at school and at home

## In Panic Disorder

**Manifestation.** Severe, uncontrolled, generalized anxiety. **Treatment Outcomes.** Change in intensity or focus of anxiety. **Outcome Measures.** Subjective rating scales of anxiety during day or when there is need to go out.

## In Phocomelia

**Manifestation.** Absence of lower legs and impaired gross motor function. **Treatment Outcomes.** Change in anatomical structure (not currently possible) and, therefore, function of lower legs **Outcome Measures.** Gross motor development age (e.g., Alpern-Boll Developmental Profile subscale score)

See additional listing of miscellaneous examples about impairment at the end of this appendix.

### ***Functional Limitation / Activity (Reference Level: Activity performance)***

#### **Manifestation**

Functional limitation is inability or restricted ability to perform a meaningful activity in the manner or within a range consistent with the purpose of a body part, organ or organ system. Activity is used in the broadest sense to capture any level of complexity, from simple skills to complex activities and behaviors. Activities include simple or basic physical skills of the person as a whole (grasping, moving a leg, or seeing), basic and complex mental functions (remembering past events or acquiring knowledge), collections of physical and mental activities at various levels of complexity (driving a car, a social skill such as interacting with persons in formal settings). The distinction between an activity (functional limitation) and impairment of function of an organ system or body part is made as follows. Memory, arm extension, and reciprocal leg movement are examples of impairments of physiologic or neurologic function (functions of cognition and of movement); recalling multiplication facts, throwing a ball, and walking are examples of *meaningful* activities or skills.

Activities are usually expressed as verbal nouns with an “ing” ending, i.e., walking, talking, eating, feeling, dressing, working, playing, reading, learning. Activities also include the skills that make up all human behavior such as gross motor activities, fine motor abilities, communication skills, coping, eating, dressing, basic learning skills (i.e., reading and writing), etc.

Functional or activity limitation is any difficulty in performance, accomplishment or completion of a skill or activity at the level of the person. Difficulty encompasses all of the ways in which the doing of an activity may be affected, i.e. not being able to do it at all, doing it too slowly or quickly, doing it awkwardly or otherwise not in the manner expected. It may range from a slight to severe deviation in terms of quality or quantity in doing the activity in a manner or to the extent that is expected. Thus, walking or some other activity may or may not be limited by the impairment of pain. Activity performance may be associated with endurance (an aspect of impairment) so that the activity cannot be done at all or done without undue fatigue.

The presence of assistive devices may remove limitations on activity in specific domains, whereas without the assistive device, the person’s activity (functional limitation) would be remain limited. Therefore, effects of assistive devices occur in this dimension of the classification system.

#### **Treatment Outcomes**

Changes in the quality or quantity of the performance of an activity or skill.

#### **Outcome Measures**

Any strategy or technique that measures the qualitative or quantifiable attributes of a meaningful activity. Speed or accuracy of an isolated arm movement (as in an laboratory setting) is a measure of impairment; speed or accuracy of writing or pouring water or throwing a ball are measures of activity or functional limitation.

<b>In PKU or Phenylketonuria</b>
----------------------------------

**Manifestation.** Limited scholastic achievement; inappropriate social interactions. **Treatment Outcomes.** Change in scholastic achievement or in inappropriate behaviors. **Outcome Measures.** Report card grades; frequency count of observed behaviors during social encounters.

### In Narcolepsy

**Manifestation.** Limited sustained reading and studying during the day. **Treatment Outcomes.** Change in quantity of reading or studying time. **Outcome Measures.** Count of minutes of reading and studying time before falling asleep.

### In Cerebral Palsy or CP

**Manifestation.** Limitations in the oral motor skills of eating, drinking, and swallowing. **Treatment Outcomes.** Change in quality, quantity or speed of eating and drinking. **Outcome Measures.** Frequency count of choking episodes during a meal; rating scale of loss of food and liquid during a meal; time count of minutes to complete a meal.

### In Reading Disability

**Manifestation.** Restricted reading. **Treatment Outcomes.** Change in recognition of words. **Outcome Measures.** Reading achievement tests, (i.e., word recognition subtest on Wide Range Achievement Test).

### In Attention Deficit Disorder or ADD

**Manifestation.** Limited persistence at assigned task. **Treatment Outcomes.** Change in number of seat assignments completed during allotted time or in number of chores at home completed without reminder. **Outcome Measures.** Count of completed assignments turned in at end of day or chores completed independently in specified time.

### In Panic Disorder

**Manifestation.** Restricted going out alone. **Treatment Outcomes.** Change in time spent out alone or number of times per week outing from residence is initiated. **Outcome Measures.** Time count of hours spent away from residence along and type of activity (i.e., grocery shopping, walking for exercising, going to medical appointments; frequency count of times leaving of residence was initiated during a week.

### In Phocomelia

**Manifestation.** Absence of lower legs and impaired gross motor function. **Treatment**  
**Outcomes.** Change in anatomical structure (not currently possible) and, therefore, function of lower legs **Outcome Measures.** Gross motor development age (e.g., Alpern-Boll Developmental Profile subscale score)

See additional listing of miscellaneous examples about functional limitation at the end of this appendix.

### ***Disability / Participation (Reference Level: Participation in Expected Social Roles)***

#### **Manifestation**

Disability is defined as inability or limitation in performing or participating in the expected roles within expected physical and social contexts of a culture or society. Participation consists of all areas or aspects of human life, including the full experience of being involved in a practice, custom or social behavior. Domains of participation—personal maintenance, mobility, exchange of information, social relationships, occupation, avocation, economic life, and family, community and civic life—are “social” in the sense that the character of these complex experiences are shaped by society. Social role and function refers to the complete lived experience of people with health conditions in the actual context of their lives and is designated by nouns such as involvement, engagement, independence and attainment.

Disability or participation restriction is relative to other people, i.e. is some diminishment of the degree or extent of participation compared to that expected of an individual without impairment—in that culture or society. The societal disadvantage in this dimension is created by internal features of the person, i.e., the health condition itself, as well as other internal factors such as age, race, gender, past and current experiences, personality, aptitudes, coping styles, lifestyle, habits, and social background. The disadvantage may take many forms including the creation of additional disablement such as anguish or mental illness.

#### **Treatment Outcomes**

Changes in the quality or quantity of participation in societal roles, i.e., worker, student, mother, father, boyfriend, citizen, friend, athlete, etc.

#### **Outcome Measures**

Any strategy or technique that measures the discordance between the observed participation of the person with an impairment and the expected participation of a person without an impairment.

**In PKU or Phenylketonuria**

**Manifestation.** Not able to attain high school diploma, participate in sophisticated social interactions, hold a job. **Treatment Outcomes.** Change in attainment of diploma, demonstration of social skills, getting and holding a job. **Outcome Measures.** Number of days employed at paying job.

### **In Narcolepsy**

**Manifestation.** Limited attainment of grade level curriculum. **Treatment Outcomes.** Change in amount of grade level curriculum completed. **Outcome Measures.** Number of school grades repeated, record of teacher concern about or amount of participation in extra help programs (e.g., tutoring, summer school).

### **In Cerebral Palsy or CP**

**Manifestation.** No involvement in school cafeteria (or restaurant dining) due to length of time required to eat and need for caregiver's attention to feed person. **Treatment Outcomes.** Change in participation in social dining. **Outcome Measures.** Timed samples of eating lunch in school cafeteria setting, count of lunches eaten within the school's allotted lunch period, count of family meals taken at a restaurant.

### **In Reading Disability**

**Manifestation.** Regular classroom participation limited by need for additional, daily reading instruction or assistance with assignments that involve reading. **Treatment Outcomes.** Change in number of hours of successful regular classroom participation. **Outcomes Measures.** Percentage of regular classroom assignments completed with at least 80% accuracy and without extra help.

### **In Attention Deficit Disorder or ADD**

**Manifestation.** Limited participation in group activities that require periods of focused attention and that have a high level of distractions (e.g., regular classrooms, lecture halls, Girl Scout meetings, Little League team practice and games). **Treatment Outcomes.** Change in ability to successfully participate in group activities. **Outcome Measures.** Count of number and type of group activities in which there is successful participation.

### **In Panic Disorder**

**Manifestation.** Self-imposed social isolation. **Treatment Outcomes.** Change in social isolation. **Outcome Measures.** Count of number and type of social engagements (i.e., visit to a friend's house, attending movie with someone, lunch with another person), count of number of friends seen at least once a month, days of attendance at school or a job.

## In Phocomelia

**Manifestation.** Sports, playground, and P.E. involvement limited to passive observation. **Treatment Outcomes.** Change of involvement (i.e., passive observer versus active player). **Outcome Measures.** Percentage of time in active play on playground, in sports activities, during PE classes.

See additional listing of miscellaneous examples about disability at the end of this appendix.

## **Societal Limitation / Contextual Factors (Reference Level: External Barriers to Participation and Other External Variables)**

### Manifestations

These are limitations that are imposed by society or the environment, as opposed to being limitations within the individual, that prevent fulfillment of roles. These barriers result from attitudes, architectural barriers and social policies that deny access to services and opportunities associated with full participation. In addition, these external factors may include the natural environment (i.e., weather or terrain), the human-made environment (i.e., tools, furnishings, the built-environment), other individuals, and costs (financial and other) associated with interventions.

### Treatment Outcomes

Changes in these contextual factors or barriers.

### Outcome Measures

Any strategy or technique that measures the qualitative or quantifiable attributes of these circumstances.

## In PKU or Phenylketonuria

**Manifestation.** Employment limitation due to employer's unwillingness to hire individuals with stigmata of mental retardation. **Treatment Outcomes.** Change in hiring policies. **Outcome Measures.** Number of employees with cognitive impairment on payroll.

## In Narcolepsy

**Manifestation.** Social restriction due to exclusion from play or teasing because of sleep attacks.

**Treatment Outcomes.** Change in exposure to teasing or other exclusion behaviors. **Outcome Measures.** Frequency count of teasing incidents or exclusion incidents on playground, in class or lunchroom.

## In Cerebral Palsy or CP

**Manifestation.** Social restriction due to rude behavior of diners at restaurant (e.g. staring at or commenting aloud about individual with impaired oral-motor involvement eating out in public).

**Treatment Outcomes.** Change in diners' behavior toward differences in eating behavior of others. **Outcome Measures.** Rating of restaurant diners' actual reactions to a diner with oral-motor involvement.

## In Reading Disability

**Manifestation.** Educational restriction due to specialized reading instruction or other tutorial help being available only in special education class. **Treatment Outcomes.** Change in service delivery of specialized reading instruction. **Outcome Measures.** Minutes per day of specialized reading instruction and tutoring delivered as part of regular reading curriculum.

## In Attention Deficit Disorder or ADD

**Manifestation.** Stigmatized as cognitively deficient so denied social acceptance or employment.

**Treatment Outcomes.** Change in societal understanding about competence of individuals with ADD. **Outcome Measures.** Questionnaire sampling understanding and attitudes about individuals with ADD.

## In Panic Disorder

**Manifestation.** Stigmatized as a mental patient, so denied social acceptance by neighbors and acquaintances. **Treatment Outcomes.** Change in social interaction initiated by neighbors and acquaintances. **Outcome Measures.** Count of verbal greetings initiated by other people, count of invitations received to participate in a social exchange.

## In Phocomelia

**Manifestation.** No ramps or elevators in church building precludes participation in spiritual life (e.g., taking communion, attending church school, joining in special services such as weddings,

funerals, christenings, etc.). **Treatment Outcomes.** Change in access to group spiritual activities. **Outcome Measures.** Count of number and type of church activities attended monthly.

See additional listing of miscellaneous examples about societal limitation at the end of this appendix.

### ***Sticky Wickets in Coding***

To develop evidence tables you will be identifying and coding the measures found in review of research studies. Each measure will be coded as to whether it represents an outcome at the level of pathophysiology, impairment, functional limitation, disability, or societal limitation.

Discrepancies may arise as you begin coding measures by this classification. In general, one should not accept measures at face value but must look for evidence that a particular measure is capable of measuring treatment outcomes at the intended level, regardless that its name or its common usage suggests that it does so. For example, the title of the Gross Motor Function Classification System for Cerebral Palsy may suggest to some users that this is a measure of impairment of function. Examination of the scale and the purpose for which the authors intend its use, however, will show it to be a measure that rates the activity of walking (or more specifically, locomoting) on a five-point scale—therefore, it is a measure of change in the functional limitation / activity dimension. The Pathophysiological Profile of Gait provides another example. According to its author, it describes the degree to which movement is impaired by the presence of neural activity in 1) spasticity (defined as velocity-dependent motor recruitment during muscle stretch, 2) paresis (defined as defective recruitment of motor units, 3) coactivation of antagonist muscles (loss of normal reciprocal inhibitory pattern of muscle activation) and by the presence of selected nonneural components, i.e., biomechanical mechanical properties. The data collected is kinematic tracings of electromyographic (EMG) activity to show the patterns of neural firing in muscle. In other words, the name of the instrument implies it will measure an outcome in the dimension of pathophysiology, but pathophysiology, in the context of this NCMRR/WHO classification includes outcomes of changes in cells or molecules. Impairment, in this context, means change in the function of a body part, organ, or organ system. Neural firing patterns in muscles during passive or voluntary movement describe changes in nervous system function. Despite its name, this is a measure of impairment.

Another more technical example reminds us to take care even in our understanding of some long-established measures. New methods of medical investigation are increasing not only our understanding of health conditions, but how we can, in turn, measure treatment outcomes with more sophistication than ever before. For years, the most commonly used clinical tool for measuring spasticity has been the Ashworth Scale of muscle tone. However, Lin et. al. (DMCN, 1994) point out that it is inappropriate to call this measure a test of spasticity. They explain that while the Ashworth Scale is sensitive to treatment that decreases spasticity, it is actually a measure of resistance to passive movement, which is caused by a variety of factors, only one of which is spasticity. Spasticity is a velocity-dependent resistance to passive stretch of muscles. Lin et. al. suggest that interventions that are purported to reduce spasticity should be specifically documented as producing a higher reflex velocity threshold and a lower reflex velocity gain. Hence, they have recently developed an electromyographic/goniometric recording system for documenting spasticity in terms of altered threshold responsiveness to stretch and increased gain of the system.

Clinicians interested in the use of a disablement classification have, heretofore, been operating without a standardized classification or operational definitions. Consequently, users may find that what a measure in a study is said to represent by the authors is different from the coding this classification would confer. A study recently reviewed by one of the Treatment Outcomes Committee members, for example, stated that there was “overall functional improvement” based on the outcomes of measures of “endurance, ease of transfers, and ADL’s.” Under the NCMRR classification schema as elaborated herein, however, endurance is coded as an impairment—not a functional limitation.

Even with the benefit of the operational definitions herein, confusion may still remain about the coding of some measures. Confusion is likely with regard to coding outcomes measured by normative developmental tests. For example, in early work on this classification, one member of the Treatment Outcomes Committee submitted that normed tests of gross motor, fine motor, intellectual, social development, etc. were examples of measures of functional limitation. Another argued that they represent measures of impairment. According to the operational definitions of this classification, normative comparisons are impairment outcomes. However, normative tests, intended as discriminative tests, have sometimes been used as if they were descriptive or evaluative tests. In other words, a normative test designed to reflect developmental status (i.e., a motor, social or intelligence quotient) may have been used as a measure of change in various activities or skills which are the items that make up the scale or subscale. While standardized tests of different domains of child development do include items about, for example, motor or social skills or activities, the tests include a variety of types of items and yield a motor age or social quotient. This informs about the degree of impairment of that domain of child development. It does not describe how well, for example, the result of the several items on walking versus those on sitting, etc.

The Treatment Outcomes Committee hopes that use of this classification will have the additional benefit of bringing clarity to our thinking and will help us in selecting appropriate measures in the future for what we intend to measure. If activities are the outcome of interest, an instrument that will describe specific activities needs to be used. Thus, the Gross Motor Function Scale will measure locomoting activity (despite its name) and the Denver Developmental Screening Test, Gross Motor Scale will measure motor age compared to the norm.

In the meanwhile, the coding of studies must be guided, not by what the author intended to measure or said was measured, but by what the outcome measure is capable of measuring. If you are not familiar with what an instrument actually measures (i.e., what the items are), contact the author.

In the event of discrepancies, users of this classification should rely on the operational definitions provided herein for coding rather than relying on statements in the article or the titles of measures. If you are still confused about how to code a measure after determining what it actually measures, please get the collective opinion of your colleagues by asking the Committee Chairperson to poll the committee or to organize a dialogue about the question. A group consensus, thus reached, can help you and can also be added to the growing list of measures at the end of this document for the subsequent consistent use by others.

Occasionally, an outcome measure is the single score that is an aggregate of several instruments that mix outcomes, usually in the impairment and functional limitations dimensions. For example, the Index of Motor Development in one study combines a measure of activities of daily living with

several measures of motor function and one of motor development (i.e., motor age) without providing results for the individual instruments. In these cases, a judgement must be made as to whether the score is primarily representative of impairment or functional limitations in activities.

Finally, investigators often report anecdotal outcomes and use wording such as “improvement in speech and mobility” without further description. In these cases, activities must have been observed that led the caregivers or investigators to make these comments. Since there are no measures, there are no comparisons to normed behavior as required in the impairment dimension. These anecdotal remarks, therefore, are regarded as activities and should be coded in the functional limitations / activities dimension.

### ***Lists of Miscellaneous Outcomes or Measures by Dimension***

To further help users of the classification determine in which dimension an outcome has been measured, the following lists will be updated as reviewers code outcomes to construct evidence tables using this classification. Different ways of stating an outcome may appear in the lists. Access to such lists will lead to increasing consistency of coding across users.

Note that death (death rates or survival rates) will be coded as a complication in the evidence tables rather than an outcome in one of these dimensions.

#### Pathophysiology

#### **Any Measure of the Following That Monitors Changes in Molecular, Cellular or Tissue Structures or Function**

Chromosome deletions, additions, translocations, or mosaicism, such as extra chromosome 21
Structure changes within genes such as the fibroblast factor receptor-3(FGFR3) on chromosome 4
Amino acid or organic acid inborn errors of metabolism
Malformations such as Arnold Chiari formation in brain or cleft palate
S/P decompression of posterior fossae
Cell damage from spinal cord disruption or brain damage
Intracranial hemorrhage or hydrocephalus
Myelination of nerve pathways
Abnormal presynaptic inhibition
Embryonic development of limb buds
Mineralization of bone
Structure of collagen
T-cells mass
Amino acid (or foreign matter such as lead) levels in urine or liver
Enzyme activity

## Impairment

### Any Measure of the Following That Monitors Changes in Structure or Function or Organs or Body Parts Compared to Established Norms

Musculoskeletal structures or isolated movements	muscle tone, strength or force, spasms, speed; range of motion, reflexes; bone strength, length, dislocations; absent limbs; kinematic and kinetic measures of abnormal function of limbs (co-contraction, slow speed of isolated movements like dorsiflexion of foot); movement patterns like dystonia, spasticity, chorea, athetosis; gait analysis (force plate measures, joint excursions, time in double limb support)
CNS	Arnold-Chiari formation of brain, neural firing patterns
Eye structures or vision	retina, lens (cataract), acuity, central vision, peripheral vision, rods, cones
Bowel and bladder	voiding, bladder capacity, bladder urgency, continence, constipation
Brain and functions	memory (verbal, visual, recognition, retrieval, long-term, short-term, active working); perception (figure-ground, constancy, position in space, visual-motor coordination, discrimination); arousal (somnolence, lethargy, activity level); sleep (disorders of maintaining sleep, sleep walking, night terrors, excessive sleep); attention (distractibility, concentration, perservation)
Emotion	mood swings, psychosis, depression, hallucination
Sensation	pain, tactile defensiveness, numb or insensate
Gastrointestinal system	gastrointestinal reflux, motility, absorption
Pulmonary	endurance, reactive airway, mucous production, cough reflex
Domains of development: motor, social, intellectual, language, etc.	tests or subtests that yield developmental quotients or scores that compare to a standardized norm for motor development, intellectual development, or social development (Bruininks-Osterestsky Test of Motor Proficiency, Peabody Developmental Motor Scales, Wechsler Intelligence Scale for Children, Vineland Scale of Social Maturity, Illinois Test of Psycholinguistic Abilities, Test of Language Development); motor development composite scores of ratings of extension, flexion, symmetry, weight bearing in various positions such as prone or sitting that could, but have not been, normed.
Growth	weight, length or height

## Functional Limitation / Activity

**Any Measure of the Following That Monitors Changes in Skill Performance or Activities of Organs or Body Parts**

### *Types of Activities*

Gross motor skills or purposeful activities	walking, sitting, standing, climbing, jumping, pushing, pulling, transferring, riding, running, driving, skiing, ease of being positioned
Fine motor	pinching, grasping, holding, writing, drawing, dressing, buttoning, combing, brushing, sewing, placing
Oral motor	eating, drinking, swallowing, sucking, chewing, licking
Communication	asking, telling, gesturing, speaking, smiling, laughing
Cognitive	thinking, learning, remembering, paying attention, listening, concentrating, reading, spelling, calculating
Social-emotional	feeling, interacting, socializing, coping, visiting

### *Types of Changes*

Presence of activity	Does not perform skill or activity Performs with help (amount of assistance) Performs with assistive technology
Quantitative change	Speed with which activity is accomplished
Qualitative change	Physical states: performed with fatigue, discomfort, energy cost Feeling states: anxious, confident, willing, persistent Coordination: smooth, jerky, awkward, fluency

### *Examples of Measures*

Timed speed of activity: walking 50', dressing for school, 10 buttons
Count of caregiver time required
Video recordings of sitting with support, with hands free, with hands to maintain balance, sitting tentatively and attending to task while sitting to evaluate change in sitting
Written log of means of communicating used during specified time period: communication device including computer, mouth stick to write, beeper switches, gestures, unintelligible or intelligible verbalizations
Timed samples of quality and types of communications: responding to question;, initiating a communication; communicating basic needs; communicating thoughts, feelings, questions
Count of choking episodes during meal
Rating scale: presence, absence, or reduction of restricted breathing
Count of hours of uninterrupted, restful sleeping through the night
Timed samples of time sitting on toilet before voiding begins (spastic bladder)
Gross Motor Function Classification for Cerebral Palsy
Pediatric Evaluation of Disability Inventory
Achievement tests of basic learning skills, e.g. California Achievement Test

Disability / Participation

**Any Measure of the Following That Monitors Changes in Participation in Social Roles**

### *Types of Participation*

<b>Participation as a</b>	<b>Participation in</b>
student	personal maintenance or self care
worker	exchange of information
neighbor	personal mobility
athlete	social relationships
friend	civic life
citizen	economic life
	occupation
	family life
	community life

*Examples of Changes that May Be Measured*

Job limitations: type of work, place of work, part or full time
Status of “problem kid” which excludes child from some class activities
Limited social eating (due to diet restrictions or oral-motor dysfunction)
School attendance
Variety and level of participation in school activities: academic classes, P.E., playground, lunchroom, after school activities
Variety and level of participation in sports: spectator, mascot, water boy, player, attends games, watches on T.V., collects baseball cards, keeps statistics, discusses sports
Church participation: Sunday School, Bible School, church services, special events
Scouting and other social or learning group participation
Amount of time spent with friends
Peer social activities: playing in neighborhood, sleeping over, going to movie
Family functioning: chores at home, lives away from family (assisted living, independent living)
Type and frequency of mobility: drives own vehicle, uses public transportation
Dating, co-habiting, marriage, parenting

## Societal Limitations

### Any Measure of the Following That Monitors Changes in Barriers to Participation Imposed by Society or Other External Variables

Private insurance and Medicaid funding policies regarding medical procedures and assistive technology
School policies regarding who can attend, how (transportation availability, parent volunteer work), and type of classroom programs and support available
Architectural barriers including braille in elevators, wheelchairs ramps or elevators
Community support, public policies and funding for low-income housing with support for independent living
Public access to TV and movies through close-captioning, telephones via TTY, and computers via accessibility options
State and federal laws requiring accommodations to people with disabilities in the workplace and school
Physical availability of a medical treatment, i.e., in community center, primary care hospital, tertiary treatment center
Financial availability of a medical treatment: costs, investigational status (not covered by insurance)
Funding to support policies of inclusion of people with disabilities in the workforce and in the educational system
Parenting behaviors
Home environment factors

## Appendix 2b: Coding Dimensions of Disablement (Beginning 2003)

### Using WHO 2001 Classification (Replaces Appendix 2a)

As planned, the AACPDMD coding system changed from use of the combined NCMRR/WHO (beta version of ICF) dimensions-of-disability classification to use of the WHO International Classification of Functioning, Disability and Health, published September 2001 (ICF) that has been accepted by 191 countries as the international standard to describe and measure health and disability. All evidence reports begun after January 2003 will use this classification. Reports that were in already in committee review will use the earlier classification shown in Appendix 2a.

**Background:** In the beta or field-test version of this ICF, WHO specified four dimensions: impairment, activity, participation, and contextual factors. These were consistent with the categories that NCMRR called impairment, functional limitations, disability, and societal limitations. In addition, the NCMRR classification included a fifth category that referenced cellular and molecular structure or function which was called Pathophysiology. Combining these two classifications, the AACPDMD dimensions of disability were 1) pathophysiology, 2) impairment, 3) functional limitation/activity, 4) disability/participation, and 5) societal limitation/context factors.

In the September 2001 final WHO version that resulted from various groups piloting the beta version, the categories (or dimensions) have been combined differently. The pathophysiology and impairment categories are collapsed together such that the new category is Impairment in two parts: 1) Impairment of Body Structures (molecular to whole body parts, i.e., arm, leg, heart) and 2) Impairment of Body Functions (again the whole range from molecular to organ systems). The functional limitation/activity and disability/participation categories were collapsed together to create a 3<sup>rd</sup> category which is called 3) Activity Limitations and Participation Restriction. The societal limitation/context factors category is now the final and 4<sup>th</sup> category and is called 4) Environmental Factors.

This classification has multiple uses and is being offered by WHO in a checklist form to use for taking a medical history of a patient; as operational definitions of a subject (or sample subject population) that both describe individuals and groups as well as allow determination of improvement as the result of an intervention. For our purposes, we are using it to code treatment outcomes from studies of interventions in order to aggregate otherwise disparate research results into categories to help us make sense of bodies of evidence.

**Dimensions of Disablement: International Classification of Functioning, Health and Disability (WHO, 2001)**

Impairment of Body Functions	Problems in the physiological functions of body systems (including psychological functions) as a significant deviation or loss at the molecular, cellular, organ, and organ systems level
Impairment of Body Structures	Problems in anatomical parts of the body from the molecular, cellular levels to organs, limbs and their components as a significant deviation or loss
Activity and Participation	Activity is the execution of a task or action by an individual. Activity limitations are difficulties an individual may have in executing activities. Participation is involvement in a life situation. Participation restrictions are problems an individual may have in involvement in life situations.
Environmental Factors	Environmental factors make up the physical, social and attitudinal environment in which people live and conduct their lives.

***Sticky Wickets in Coding***

To develop evidence tables you will be identifying and coding the measures found in review of research studies. Each measure will be coded as to whether it represents an outcome in the dimension of impairment of body structure, impairment of body function, activity and participation, environmental factors, or other contextual factors.

Discrepancies may arise as you begin coding measures by this classification. In general, one should not accept measures at face value but must look for evidence that a particular measure is capable of measuring treatment outcomes at the intended level, regardless that its name or its common usage suggests that it does so. For example, the title of the Gross Motor Function Classification System for Cerebral Palsy may suggest to some users that this is a measure of impairment of function. Examination of the scale and the purpose for which the authors intend its use, however, will show it to be a measure that rates the activity of walking (or more specifically, locomoting) on a five-point scale—therefore, it is a measure of change in the activity and participation dimension.

Another example reminds us to take care even in our understanding of some long-established measures. New methods of medical investigation are increasing not only our understanding of health conditions, but how we can, in turn, measure treatment outcomes with more sophistication than ever before. For years, the most commonly used clinical tool for measuring spasticity has been the Ashworth Scale of muscle tone. However, Lin et. al. (DMCN, 1994) point out that it is inappropriate to call this measure a test of spasticity. They explain that while the Ashworth Scale is sensitive to treatment that decreases spasticity, it is actually a measure of resistance to passive movement, which is caused by a variety of factors, only one of which is spasticity. Spasticity is a velocity-dependent resistance to passive stretch of muscles. Lin et. al. suggest that interventions that

are purported to reduce spasticity should be specifically documented as producing a higher reflex velocity threshold and a lower reflex velocity gain. Hence, they have recently developed an electromyographic/goniometric recording system for documenting spasticity in terms of altered threshold responsivity to stretch and increased gain of the system.

Clinicians interested in the use of a disablement classification have, heretofore, been operating with little in the way of standardized classification or operational definitions. Consequently, users may find that what authors of a study say a measure represents may be different from the coding this classification would confer. A study recently reviewed by one of the Treatment Outcomes Committee members, for example, stated that there was “overall functional improvement” based on the outcomes of measures of “endurance, ease of transfers, and ADL’s.” Under the classification schema as elaborated herein, however, endurance is coded as an impairment—not a functional limitation.

Even with the benefit of the operational definitions that are being provided in this appendix, confusion may still remain about the coding of some measures. Confusion is likely with regard to coding outcomes measured by normative developmental tests. For example, in early work on this classification, one member of the Treatment Outcomes Committee submitted that normed tests of gross motor, fine motor, intellectual, social development, etc. were examples of measures of functional limitation. Another argued that they represent measures of impairment. *According to the operational definitions of this classification*, normative comparisons are impairment outcomes. However, normative tests, intended as discriminative tests, have sometimes been used as if they were descriptive or evaluative tests. In other words, a normative test designed to reflect developmental status (i.e., a motor, social or intelligence quotient) may have been used as a measure of change in various activities or skills which are the items that make up the scale or subscale. While standardized tests of different domains of child development do include items about, for example, motor or social skills or activities, the tests include a variety of types of items and yield a motor age or social quotient. This informs about the degree of impairment of that domain of child development. It does not describe how well, for example, the result of the several items on walking versus those on sitting, etc. Whenever a standard score that represents variation from a norm is used to report the result of an intervention, that outcome (and measure) is an impairment outcome.

The Treatment Outcomes Committee hopes that use of this classification will have the additional benefit of bringing clarity to our thinking and will help us in selecting appropriate measures in the future for what we intend to measure. If activities are the outcome of interest, an instrument that will describe specific activities needs to be used. Thus, the Gross Motor Function Scale measures locomoting activity while the Denver Developmental Screening Test, Gross Motor Scale, measures motor age compared to the norm.

In the meanwhile, the coding of studies must be guided, not by what the author intended to measure or said was measured, but by what the outcome measure is capable of measuring. If you are not familiar with what an instrument actually measures (i.e., what the items are), contact the author.

In the event of discrepancies, users of this classification should rely on the operational definitions provided herein for coding rather than relying on statements in the article or the titles of measures. If this is insufficient, consult the WHO website for the ICF definitions. If you are still confused about how to code a measure after determining what it actually measures, please get the collective opinion

of your colleagues by asking the Committee Chairperson to poll the committee or to organize a dialogue about the question. A group consensus, thus reached, can help you and can also be added to the growing list of measures at the end of this document for the subsequent consistent use by others.

Occasionally, an outcome measure is the single score that is an aggregate of several instruments that mix outcomes, usually in the impairment and functional limitations dimensions. For example, the Index of Motor Development in one study combines a measure of activities of daily living with several measures of motor function and one of motor development (i.e., motor age) without providing results for the individual instruments. In these cases, a judgment must be made as to whether the score is primarily representative of impairment or functional limitations in activities.

Finally, investigators often report anecdotal outcomes and use wording such as “improvement in speech and mobility” without further description. In these cases, there is the implication that activities were observed that led the caregivers or investigators to make these comments. Since there are no measures, there are no comparisons to normed behavior as required in the impairment dimension. These anecdotal remarks, therefore, are regarded as activities and should be coded in the activities and participation dimension.

### ***Differentiating Manifestation of Dimension, Treatment Outcomes and Outcome Measures***

In each of the five dimensions that map the disablement process, there are three aspects that will be discussed:

4. ***Manifestation*** or how the disablement process acts in that dimension,
5. ***Treatment outcome***, the outcome or result that is observed in association with an intervention, and
6. ***Outcome measure***, the type of instrument or technique used to measure an outcome.

Seven health conditions (PKU, narcolepsy, cerebral palsy, reading disability, attention deficit disorder, panic disorder, and phocomelia) have been selected to serve as examples throughout the discussion of each of the dimensions in an attempt to address a wide variety of outcomes. These developmental disabilities are complex. Seldom, if ever, does a genetic abnormality or congenital defect produce a single effect such as mental retardation or abnormal muscle tone or a structural defect of the heart. They often produce multiple primary anomalies as well as multiple secondary effects. Consequently, a “primary condition” is a risk factor for one or more “secondary conditions. For the sake of simplicity, however, the seven health conditions used as examples will address only a single feature of the health condition.

At the end of this appendix, there are more outcomes of interest and outcome measures found in developmental disabilities research listed for each dimension. These lists do not include all outcome measures that may be extracted from studies, but are intended to be sufficiently representative to guide decision-making about other measures that may be encountered. These lists will grow over time to provide for consistent coding by all users.

## ***Impairment of Body Function (from molecular and cellular function to organs, organ systems function)***

### **Manifestation of Impairment of Body Function**

This indicates a loss or abnormality of body *function* (i.e., physiological function of molecules, cells, organs, organ systems, or body parts). Vision and hearing are examples of body function.

Functions include but are not limited to the “senses. Limitations in certain functions (inability to carry out a basic function of the body or body parts such as extension of arm, weight-bearing, symmetrical stance) are impairments. Patterns of neural firing of muscles are another example. The physiological functions include not only the physical functions of the central nervous system such as neural firing patterns, but also mental functions (i.e., cognitive and emotional functions).

Nouns are used to denote impairment of function (i.e., cognition; sensation; speech; vision; audition; respiration; and social, emotional and motor development). Impairments are detectable or noticeable by others or the person by direct observation or by inference from indirect observation.

They are absent or present; once an impairment is present, it may be scaled by its severity and refers to a significant variation from established statistical norms (i.e., as a deviation from a population mean within measured standard norms). Measures that yield developmental ages or other standard scores that compare these functions to the norm indicate impairment status (e.g., I.Q., gross motor age, 20/100 visual acuity, etc.). For example, speed or force of isolated movement (as in voluntary dorsiflexion of foot) or coordination of muscle activation are measures of impairment of physiologic or neurologic motor function that can be compared to population norms. [In contrast, techniques that measure the speed with which a motor activity or skill is accomplished are in the dimension of activity and participation. For example, timed samples of placing cubes in a box or walking a certain distance at school or within a certain time) are measures of functional limitation in activities.]

Impairment of one function may produce other impairments of function. For example, muscle spasticity can result in contractures which can restrict range of motion, impair movement, or both.

Pain is a manifestation of impairment and it may, or may not, functionally limit activity and participation. [The activities (e.g., walking, sleeping) that pain may prevent or restrict, however, are coded as activity and participation.] Endurance, too, is an aspect of impairment which may, or may not, be associated with a functional limitation in activity and participation. [ If some activity cannot be completed at all or without undue fatigue, the activity that is limited and is coded in the activity and participation dimension.]

Short List (each of the following is explained further in the full listing and discussion of body functions on the WHO web site but the short lists included herein will guide most, if not all, the coding decisions needed for AACPD Evidence Reports.)

## **NEUROMUSCULOSKELETAL AND MOVEMENT RELATED FUNCTIONS**

Mobility of joint

Muscle power

Muscle tone

Involuntary movements

#### **MENTAL FUNCTIONS**

Consciousness

Orientation (*time, place, person*)

Intellectual (*including retardation, dementia*)

Energy and drive functions

Sleep

Attention

Memory

Emotional functions

Perceptual functions

Higher level cognitive functions

Language

#### **SENSORY FUNCTIONS AND PAIN**

Seeing

Hearing

Vestibular (*including balance functions*)

Pain

#### **FUNCTIONS OF THE SKIN AND RELATED STRUCTURES**

#### **ANY OTHER BODY FUNCTIONS**

#### **VOICE AND SPEECH FUNCTIONS**

Voice

#### **FUNCTIONS OF THE CARDIOVASCULAR, HAEMATOLOGICAL, IMMUNOLOGICAL AND RESPIRATORY SYSTEMS**

Heart

Blood pressure

Haematological (*blood*)

Immunological (*allergies, hypersensitivity*)

Respiration (*breathing*)

#### **FUNCTIONS OF THE DIGESTIVE, METABOLIC AND ENDOCRINE SYSTEMS**

Digestive

Defecation

Weight maintenance

Endocrine glands (*hormonal changes*)

#### **GENITOURINARY AND REPRODUCTIVE FUNCTIONS**

Urination functions

Sexual functions

#### **Treatment Outcomes**

Interventions targeted at the impairment of body function dimension are those that attempt to restore to normal the metabolic processes, or processes of organs or organ systems. Consequently, interventions targeting this dimension depend on understanding of a health condition at the

molecular, cellular, tissue or organ level. Our understanding continues to evolve with the advent of new technologies (e.g., imaging and gene mapping) and the insights they provide. Examples: A botox injection is an intervention to stimulate Gaba<sub>b</sub> to restore body function, i.e. to reduce spastic muscle tone. Infant stimulation programs are intended to promote progress in various developmental domains and report standardized measures of social development, language development, cognitive development, etc.

### Outcome Measures

Techniques that quantify body function outcomes are those that measure changes in the function of molecules, cells, tissues and organs and include all measures that compare function to established norms, e.g., IQ scores, age equivalent scores.

### Examples of Impairment of Body Function vs. Treatment Outcomes vs. Outcome Measures

#### In PKU or Phenylketonuria

**Manifestation or impairment.** Inborn error of metabolism in which elevated levels of phenylalanine in the blood are toxic to the central nervous system producing mental retardation.

**Treatment Outcomes.** Results of interventions that attempt to restore metabolic processes to normal by changing the biochemical processes they orchestrate or changing the blood phenylalanine levels. **Outcome Measures.** phenylalanine hydroxylase cDNA probe, measures of enzyme activity, DGGE and SSCP analysis, plasma levels of phenylalanine.

**Manifestation.** Cognitive impairment, i.e., mental retardation or subnormality of intelligence.

**Treatment Outcomes.** Change in intellectual performance based on variation from established norms. **Outcome Measures.** Standardized test of intelligence (e.g., WISC)

#### In Narcolepsy

**Manifestation.** Excessive daytime somnolence and uncontrollable sleep attacks or episodes.

**Treatment Outcomes.** Change in somnolence and sleep attacks. **Outcome Measures.** Frequency count of daytime sleep episodes.

**Manifestation.** Etiology and transmission unknown but strong association has been shown with histocompatibility antigen complex HLA-DR2, i.e., people with narcolepsy are HLA-DR2 positive.

**Treatment Outcomes.** Results of interventions that attempt to change HLA-DR2 status.

**Outcome Measures.** HLA-DR2 probe.

## In Cerebral Palsy or CP

**Manifestation.** Neuromuscular impairment (abnormal activation of muscles, i.e., motor dysfunction that varies depending on pathophysiology, e.g., athetoid type is a form that includes uncontrolled involuntary motion of the distal joints. **Treatment Outcomes.** Changes in factors that affect coordinated movement patterns such as muscle tone, posture and balance, reflex activity, equilibrium reactions **Outcome Measures.** Kinematic/dynamic and electronic recordings of EMG patterns, Ashworth Scale, goniometer reading of range of motion.

## In Reading Disability

**Manifestation.** Dysfunction of phonological processing due to auditory perceptual dysfunction or a specialized type of hearing problem. **Treatment Outcomes.** Change in perception of certain sounds not previously distinguished. **Outcomes Measures.** Standardized phonological processing test that reports variation from established norms.

## In Attention Deficit Disorder or ADD

**Manifestation.** Visual and auditory distractibility. **Treatment Outcomes.** Change in ability to sustain selective attention in spite of distracting sounds and sights. **Outcome Measures.** Standardized rating questionnaires about observed distractible behaviors at school and at home that gauge variation from established norms.

## In Panic Disorder

**Manifestation.** Severe, uncontrolled, generalized anxiety. **Treatment Outcomes.** Change in intensity or focus of anxiety. **Outcome Measures.** Subjective ordinal rating scales of anxiety during day referenced to norms.

See additional list of miscellaneous examples about impairment of body function at the end of this appendix.

## ***Impairment of Body Structure (from cells to body parts)***

### Manifestation

This indicates a loss or abnormality of a body *part* (i.e., structure). Nouns are used to denote impairment of structure (i.e., cell, eye and eye parts, brain, arm, foot, heart). Abnormality here refers to a significant variation from established statistical norms and should be used only in this sense (i.e., as a deviation from a population mean within measured standard norms). Impairments

are detectable or noticeable by others or the person by direct observation or by inference from indirect observation. They are absent or present; once impairment is present, it may be scaled by its severity.

Short List (each of the following is explained further in the full listing and discussion of body structures on the WHO web site but the short lists included herein will guide most, if not all, the coding decisions needed for AACPD M Evidence Reports.)

**STRUCTURE OF THE NERVOUS SYSTEM**

Brain

Spinal cord and peripheral nerves

**STRUCTURE RELATED TO MOVEMENT**

Head and neck region

Shoulder region

Upper extremity (*arm, hand*)

Pelvis

Lower extremity (*leg, foot*)

Trunk

**SKIN AND RELATED STRUCTURES**

**ANY OTHER BODY STRUCTURES**

**THE EYE, EAR AND RELATED STRUCTURES**

**STRUCTURES INVOLVED IN VOICE AND SPEECH**

**STRUCTURE OF THE CARDIOVASCULAR,  
IMMUNOLOGICAL AND RESPIRATORY SYSTEMS**

Cardiovascular system

Respiratory system

**STRUCTURES RELATED TO THE DIGESTIVE,  
METABOLISM AND ENDOCRINE SYSTEMS**

**STRUCTURE RELATED TO GENITOURINARY AND  
REPRODUCTIVE SYSTEM**

Urinary system

Reproductive system

**Treatment Outcomes**

Interventions targeted at the impairment of body structure dimension are those that attempt to restore body structures, as compared to normal. Examples: A kidney transplant is an intervention intended to replace/restore a body structure. Gastrostomy feeding is intended to restore growth, a feature of body structure which has established statistical norms.

**Outcome Measures**

Techniques that quantify impairment outcomes are those that measure anatomical changes or features of organs, organ systems or body parts.

## Examples of Impairment of Body Structure vs. Treatment Outcomes vs. Outcome Measures

### In PKU or Phenylketonuria

**Manifestation.** Inborn error of metabolism in which elevated levels of phenylalanine in the blood are toxic to the central nervous system producing mental retardation. **Treatment Outcomes.** Results of interventions that attempt to restore genetic structures to normal by alternating the DNA structures or the cellular structure. **Outcome Measures.** measures of myelin sheathing, gene mapping.

**Outcome Measures.** HLA-DR2 probe.

### In Cerebral Palsy or CP

**Manifestation.** Static encephalopathy with cell or tract damage in the brain. **Treatment Outcomes.** Results of interventions intended to repair the brain cell or tract damage, prevent the cell damage (e.g., blood transfusion for rh incompatibility, drugs to control intracranial hemorrhage).

**Outcome Measures.** IVH grading.

### In Reading Disability

**Manifestation.** Unknown but suspected abnormality in medial geniculate nucleus in the brain resulting in too few neurons to “hear” staccato consonants; probably a multifactorial etiology.

**Treatment Outcomes.** Change in number of neurons that fire when listening to these consonants.

**Outcomes Measures.** Imaging techniques, EMG.

### In Phocomelia

**Manifestation.** Arrested limb bud development. **Treatment Outcomes.** Not currently known

**Outcome Measures.** None currently

### **Activity Limitations and Participation Restriction**

The activity and participation construct includes an individual’s capacity to execute a task or action at the highest probable level of functioning without assistance at a given moment in the current environment. Because the current environment brings in the societal context, participation can also

be understood as "involvement in a life situation" or "the lived experience" of people in the actual context in which they live.

Activities include simple or basic physical skills of the person as a whole (grasping, moving a leg, or seeing), basic and complex mental functions (remembering past events or acquiring knowledge), collections of physical and mental activities at various levels of complexity (driving a car, a social skill such as interacting with persons in formal settings). The distinction between an activity and impairment of function of an organ system or body part is subtle but will be distinguished as follows. Memory, arm extension, and reciprocal leg movement are examples of impairments of physiologic or neurologic function (functions of cognition and of movement); recalling multiplication facts, throwing a ball, and walking are examples of *meaningful* activities or skills.

Activities are usually expressed as verbs with an "ing" ending, i.e., walking, talking, eating, feeling, dressing, working, playing, reading, learning but they also include the skills that make up all human behavior such as gross motor activities, fine motor abilities, communication skills, coping, eating, dressing, basic learning skills (i.e., reading and writing), etc.

Functional or activity limitation is any difficulty in performance, accomplishment or completion of a skill or activity at the level of the person. Difficulty encompasses all of the ways in which the doing of an activity may be affected, i.e. not being able to do it at all, doing it too slowly or quickly, doing it awkwardly or otherwise not in the manner expected. It may range from a slight to severe deviation in terms of quality or quantity in doing the activity in a manner or to the extent that is expected. Thus, walking or some other activity may or may not be limited by the impairment of pain. Activity performance may be associated with endurance (an aspect of impairment) so that the activity cannot be done at all or done without undue fatigue.

The presence of assistive devices may remove limitations on activity in specific domains, whereas without the assistive device, the person's activity would remain limited. Therefore, effects using assistive devices are coded in this dimension of the classification system.

Participation consists of all areas or aspects of human life, including the full experience of being involved in a practice, custom or social behavior. Domains of participation—personal maintenance, mobility, exchange of information, social relationships, occupation, avocation, economic life, and family, community and civic life—are "social" in the sense that the character of these complex experiences are shaped by society. Social role and function refers to the complete lived experience of people with health conditions in the actual context of their lives and is designated by nouns such as involvement, engagement, independence and attainment. Inability or limitation in performing or participating in the expected roles within expected physical and social contexts of a culture or society is coded in this dimension.

Participation restriction is relative to other people, i.e. is some diminishment of the degree or extent of participation compared to that expected of an individual without impairment—in that culture or society. The societal disadvantage in this dimension is created by internal features of the person, i.e., the health condition itself, as well as other internal factors such as age, race, gender, past and current experiences, personality, aptitudes, coping styles, lifestyle, habits, and social background. The disadvantage may take many forms including the creation of additional disablement such as anguish or mental illness.

## Outcome Measures

Short List (each of the following is explained further in the full listing and discussion of activity and participation on the WHO web site but the short lists included herein will guide most, if not all, the coding decisions needed for AACPD M Evidence Reports.)

### **MOBILITY**

Lifting and carrying objects

Fine hand use (*picking up, grasping*)

Walking

Moving around using equipment (*wheelchair, skates, etc.*)

Using transportation (*car, bus, train, plane, etc.*)

Driving (riding bicycle and *motorbike, driving car, etc.*)

### **SELF CARE**

Washing oneself (*bathing, drying, washing hands, etc*)

Caring for body parts (*brushing teeth, shaving, grooming, etc.*)

Toileting

Dressing

Eating

Drinking

Looking after one`s health

### **LEARNING AND APPLYING KNOWLEDGE**

Watching

Listening

Learning to read

Learning to write

Learning to calculate (*arithmetic*)

Solving problems

### **GENERAL TASKS AND DEMANDS**

Undertaking a single task

Undertaking multiple tasks

### **COMMUNICATION**

Communicating with -- receiving -- spoken messages

Communicating with -- receiving -- non-verbal messages

Speaking

Producing non-verbal messages

Conversation

### **DOMESTIC LIFE**

Acquisition of goods and services (*shopping, etc.*)

Preparation of meals (*cooking etc.*)

Doing housework (*cleaning house, washing dishes laundry, ironing, etc.*)

Assisting others

### **INTERPERSONAL INTERACTIONS AND RELATIONSHIPS**

Basic interpersonal interactions

Complex interpersonal interactions

Relating with strangers

Formal relationships

Informal social relationships

Family relationships

Intimate relationships

**MAJOR LIFE AREAS**

Informal education

School education

Higher education

Remunerative employment

Basic economic transactions

Economic self-sufficiency

**COMMUNITY, SOCIAL AND CIVIC LIFE**

Community Life

Recreation and leisure

Religion and spirituality

Human rights

Political life and citizenship

**ANY OTHER ACTIVITY AND PARTICIPATION**

## Outcome Measures

Any strategy or technique that measures the qualitative or quantifiable attributes of a meaningful activity. Speed or accuracy of an isolated arm movement (as in an laboratory setting) is a measure of impairment whereas speed or accuracy of writing or pouring water or throwing a ball are measures of activity or functional limitation.

Any strategy or technique that measures the discordance between the observed participation of the person with an impairment and the expected participation of a person without an impairment.

## Examples of Activity and Participation vs. Treatment Outcomes vs. Outcome Measures

### In PKU or Phenylketonuria

**Manifestation.** Limited scholastic achievement; inappropriate social interactions. **Treatment Outcomes.** Change in scholastic achievement or in inappropriate behaviors. **Outcome Measures.** Report card grades; frequency count of observed behaviors during social encounters.

**Manifestation.** Not able to attain high school diploma, participate in sophisticated social interactions, hold a job. **Treatment Outcomes.** Change in attainment of diploma, demonstration of social skills, getting and holding a job. **Outcome Measures.** Number of days employed at paying job.

## In Narcolepsy

**Manifestation.** Limited sustained reading and studying during the day. **Treatment Outcomes.** Change in quantity of reading or studying time. **Outcome Measures.** Count of minutes of reading and studying time before falling asleep.

**Manifestation.** Limited attainment of grade level curriculum. **Treatment Outcomes.** Change in amount of grade level curriculum completed. **Outcome Measures.** Number of school grades repeated, record of teacher concern about or amount of participation in extra help programs (e.g., tutoring, summer school).

## In Cerebral Palsy or CP

**Manifestation.** Limitations in the oral motor skills of eating, drinking, and swallowing. **Treatment Outcomes.** Change in quality, quantity or speed of eating and drinking. **Outcome Measures.** Frequency count of choking episodes during a meal; rating scale of loss of food and liquid during a meal; time count of minutes to complete a meal.

**Manifestation.** No involvement in school cafeteria (or restaurant dining) due to length of time required to eat and need for caregiver's attention to feed person. **Treatment Outcomes.** Change in participation in social dining. **Outcome Measures.** Timed samples of eating lunch in school cafeteria setting, count of lunches eaten within the school's allotted lunch period, count of family meals taken at a restaurant.

## In Reading Disability

**Manifestation.** Restricted reading. **Treatment Outcomes.** Change in recognition of words. **Outcomes Measures.** Reading achievement tests, (i.e., word recognition subtest on Wide Range Achievement Test).

**Manifestation.** Regular classroom participation limited by need for additional, daily reading instruction or assistance with assignments that involve reading. **Treatment Outcomes.** Change in number of hours of successful regular classroom participation. **Outcomes Measures.** Percentage of regular classroom assignments completed with at least 80% accuracy and without extra help.

## **In Attention Deficit Disorder or ADD**

**Manifestation.** Limited persistence at assigned task. **Treatment Outcomes.** Change in number of seat assignments completed during allotted time or in number of chores at home completed without reminder. **Outcome Measures.** Count of completed assignments turned in at end of day or chores completed independently in specified time.

**Manifestation.** Limited participation in group activities that require periods of focused attention and that have a high level of distractions (e.g., regular classrooms, lecture halls, Girl Scout meetings, Little League team practice and games). **Treatment Outcomes.** Change in ability to successfully participate in group activities. **Outcome Measures.** Count of number and type of group activities in which there is successful participation.

## **In Panic Disorder**

**Manifestation.** Restricted going out alone. **Treatment Outcomes.** Change in time spent out alone or number of times per week outing from residence is initiated. **Outcome Measures.** Time count of hours spent away from residence along and type of activity (i.e., grocery shopping, walking for exercising, going to medical appointments; frequency count of times leaving of residence was initiated during a week.

**Manifestation.** Self-imposed social isolation. **Treatment Outcomes.** Change in social isolation. **Outcome Measures.** Count of number and type of social engagements (i.e., visit to a friend's house, attending movie with someone, lunch with another person), count of number of friends seen at least once a month, days of attendance at school or a job.

## **In Phocomelia**

**Manifestation.** Sports, playground, and P.E. involvement limited to passive observation. **Treatment Outcomes.** Change of involvement (i.e., passive observer versus active player). **Outcome Measures.** Percentage of time in active play on playground, in sports activities, during PE classes.

## **Environmental Factors**

This dimension includes all aspects of the physical, social and attitudinal world and changes made in these environmental aspects. These are limitations that are imposed by society or the environment, as opposed to being limitations within the individual, that prevent fulfillment of roles. These barriers result from attitudes, architectural barriers and social policies that limit access to services and opportunities associated with full participation. In addition, these external factors may include the natural environment (i.e., weather or terrain), the human-made environment (i.e., tools, furnishings, the built-environment), other individuals, and costs (financial and other) associated with interventions.

Short List (each of the following is explained further in the full listing and discussion of environmental factors on the WHO web site but the short lists included herein will guide most, if not all, the coding decisions needed for AACPDME Evidence Reports.)

**PRODUCTS AND TECHNOLOGY (including all aspects of assistive technologies)**

For personal consumption (*food, medicines*)

For personal use in daily living

For personal indoor and outdoor mobility and transportation such as mobility devices

Products for communication

Design, construction and building products and technology of buildings for private or public use

**NATURAL ENVIRONMENT AND HUMAN MADE CHANGES TO ENVIRONMENT**

Climate

Light

Sound

**SUPPORT AND RELATIONSHIPS**

Immediate family

Friends

Acquaintances, peers, colleagues, neighbours and community members

People in position of authority

Personal care providers and personal assistants

Health professionals

Health related professionals

**ATTITUDES (including satisfaction with intervention)**

Individual attitudes of immediate family members

Individual attitudes of friends

Individual attitudes of personal care providers and personal assistants

Individual attitudes of health professionals

Individual attitudes of health related professionals

Societal attitudes

Social norms, practices and ideologies

**SERVICES, SYSTEMS AND POLICIES**

Housing services, systems and policies

Communication services, systems and policies

Transportation services, systems and policies

Legal services, systems and policies  
Social security, services, systems and policies  
General social support services, systems and policies  
Health services, systems and policies  
Education and training services, systems and policies  
Labour and employment services, systems and policies  
**ANY OTHER ENVIRONMENTAL FACTORS**

### Treatment Outcomes

Changes in these environmental factors or barriers.

### Outcome Measures

Any strategy or technique that measures the qualitative or quantifiable attributes of these factors.

Examples of Manifestation of Environmental Factors vs. Treatment Outcomes vs. Outcome Measures

#### **In PKU or Phenylketonuria**

**Manifestation.** Employment limitation due to employer's unwillingness to hire individuals with stigmata of mental retardation. **Treatment Outcomes.** Change in hiring policies. **Outcome Measures.** Number of employees with cognitive impairment on payroll.

#### **In Narcolepsy**

**Manifestation.** Social restriction due to exclusion from play or teasing because of sleep attacks. **Treatment Outcomes.** Change in exposure to teasing or other exclusion behaviors. **Outcome Measures.** Frequency count of teasing incidents or exclusion incidents on playground, in class or lunchroom.

#### **In Cerebral Palsy or CP**

**Manifestation.** Social restriction due to rude behavior of diners at restaurant (e.g. staring at or commenting aloud about individual with impaired oral-motor involvement eating out in public). **Treatment Outcomes.** Change in diners' behavior toward differences in eating behavior of others. **Outcome Measures.** Rating of restaurant diners' actual reactions to a diner with oral-motor involvement.

#### **In Reading Disability**

**Manifestation.** Educational restriction due to specialized reading instruction or other tutorial help being available only in special education class. **Treatment Outcomes.** Change in service delivery

of specialized reading instruction. *Outcomes Measures*. Minutes per day of specialized reading instruction and tutoring delivered as part of regular reading curriculum.

### **In Attention Deficit Disorder or ADD**

*Manifestation*. Stigmatized as cognitively deficient so denied social acceptance or employment. *Treatment Outcomes*. Change in societal understanding about competence of individuals with ADD. *Outcome Measures*. Questionnaire sampling understanding and attitudes about individuals with ADD.

### **In Panic Disorder**

*Manifestation*. Stigmatized as a mental patient, so denied social acceptance by neighbors and acquaintances. *Treatment Outcomes*. Change in social interaction initiated by neighbors and acquaintances. *Outcome Measures*. Count of verbal greetings initiated by other people, count of invitations received to participate in a social exchange.

### **In Phocomelia**

*Manifestation*. No ramps or elevators in church building precludes participation in spiritual life (e.g., taking communion, attending church school, joining in special services such as weddings, funerals, christenings, etc.). *Treatment Outcomes*. Change in access to group spiritual activities. *Outcome Measures*. Count of number and type of church activities attended monthly.

### **Example of Impact on Coding with Change from Earlier NCMRR/WHO Classification to ICF (WHO 2001)**

Following is a modified Summary of outcomes, measures and results table from the NDT study which was originally coded using the NCMRR/WHO dimensions. The table below shows the dimensions of disablement coded in the NCMRR/WHO classification and in the WHO final classification in both classifications. This demonstrates a relatively seamless transition can be made from the earlier classification to the WHO classification.

I:S = Impairment of Body Structures

I:F = Impairment of Body Functions

A&P = Activity Limitations and Participation Restriction  
Env = Environmental Factors

Study	Outcome of Interest	NCMRR/WH O beta dimensions	ICF dimen- sions	Measure	Level of Evidence
1973 Wright <sup>12</sup> ∅	Automatic reflexes	I	I:F	Rated observation	II
	ROM (2 movements)	I	I:F	Not specified	II
	Gross motor activities	FL/A	A&P	Rated observation	II
1975 Carlsen <sup>13</sup>	Motor age	I	I:F	Bayley Motor Scale	II
	Gross motor age	I	I:F	DDST, Motor Scale	II
	Fine motor age	I	I:F	DDST, F. Motor Scale	II
	Social age	I	I:F	DDST, Social Scale	II
	Language age	I	I:F	DDST, Lang. Scale	II
1976 Scherzer <sup>14</sup>	Physiologic function	I	I:F	Motor Dev. Evaluation	II
	Social activities	FL/A	A&P	Questionnaire	II
	Home management	SL/C	Env	Questionnaire	II
1981 Sommerfeld <sup>15</sup>	Dev. reflexes	I	I:F	Wilson DR Test	II
	Gross motor age	I	I:F	Gross Motor	II
	ROM (6 movements)	I	I:F	ROM Scale	II
1983 DeGangi <sup>16</sup>	Positioning/activities	FL/A	A&P	Rated observations	II
1985 Laskas <sup>17</sup>	ROM (dorsiflexion)	I	I:F	Biofeedback instrument	III
	ROM (heel strike)	I	I:F	Biofeedback instrument	III
1987 Herndon <sup>18</sup>	ROM (hip flexion)	I	I:F	Goniometer	IV
	ROM (hip abduct.)	I	I:F	Goniometer	IV
	ROM (knee)	I	I:F	Goniometer	IV
	ROM (dorsiflexion)	I	I:F	Goniometer	IV
	Rising from chair	FL/A	A&P	Video analysis	IV
	Walking	FL/A	A&P	Video analysis	IV
	Turning	FL/A	A&P	Video analysis: walking	IV
	Trunk rotation	I	I:F	Video analysis: walking	IV
	Trunk rotation	I	I:F	Video analysis: sitting	IV
	Postural alignment	I	I:F	Video analysis: sitting	IV
	Weight shift	I	I:F	Video analysis: sitting	IV
	Assuming position	FL/A	A&P	Video analysis: sitting	IV
1988 Palmer <sup>19</sup> +	<u>Part I</u>	~		~	~
	Motor age	I	I:F	Bayley Motor Scale	I
	Motor milestones	FL/A	A&P	Attainment defined skills	I
	Walking attainment	FL/A	A&P	Observation/defined skill	I
	Tone/spasticity/reflexes	I	I:F	Neurological exam	I
	Mental age	I	I:F	Bayley Mental Scale	I
	Social age	I	I:F	Vineland Social Maturity Sc.	I
	<u>Part II</u>	~		~	~
	Motor age	I	I:F	Bayley Motor Scale	I
	Spasticity	I	I:F	Neurological exam	I
	LE reflexes	I	I:F	Neurological exam	I
	Joint limitation/ROM	I	I:F	Bracing recommended	I
	Contractures/ROM	I	I:F	Surgery recommended	I
	Motor milestones	FL/A	A&P	Attainment defined skills	I
	Age/independent walking	FL/A	A&P	Observation/defined skill	I
	Mental age	I	I:F	Bayley Mental Scale	I
	Social age	I	I:F	Vineland Social Maturity Sc.	I
	1989 Hanzlik <sup>20</sup>	Infant compliance	I	I:F	Video analysis: DMIB
Infant responsiveness		I	I:F	Video analysis: DMIB	II
Independent play		FL/A	A&P	Video analysis: DMIB	II
Maternal directness		SL/C	Env	Video analysis: DMIB	II
Maternal initiation		SL/C	Env	Video analysis: DMIB	II
Maternal responsiveness		SL/C	Env	Video analysis: DMIB	II
Adaptive seating provision		SL/C	Env	Video analysis: DMIB	II
Maternal holding		SL/C	Env	Video analysis: DMIB	II
Face to face contact		SL/C	Env	Video analysis: DMIB	II

	Physical contact				
--	------------------	--	--	--	--

(Table V. continued)

1990 Palmer <sup>21</sup>	Infant activity	I	I:F	CITQ	I
	Infant rhythmicity	I	I:F	CITQ	I
	Infant adaptability	I	I:F	CITQ	I
	Infant approach	I	I:F	CITQ	I
	Infant threshold	I	I:F	CITQ	I
	Infant intensity	I	I:F	CITQ	I
	Infant mood	I	I:F	CITQ	I
	Infant distractibility	I	I:F	CITQ	I
	Infant persistence	I	I:F	CITQ	I
	Maternal acceptance	SL/C	Env	RMCRE	I
	Maternal overprotection	SL/C	Env	RMCRE	I
	Maternal overindulgence	SL/C	Env	RMCRE	I
	Maternal rejection	SL/C	Env	RMCRE	I
	Maternal responsiveness	SL/C	Env	HOME	I
	Maternal involvement	SL/C	Env	HOME	I
	Restriction avoidance	SL/C	Env	HOME	I
	Environment organization.	SL/C	Env	HOME	I
	Play materials	SL/C	Env	HOME	I
Variety of stimulation					
1990 Lilly <sup>22</sup>	Physiologic function	I	I:F	Rate of movements	II
1990 Embrey <sup>23</sup>	ROM (knee flexion)	I	I:F	Goniometer ; video	II
1990 Kluzik <sup>24</sup>	UE movement time	I	I:F	Video; kinematics	IV
	UE movement unit	I	I:F	Video; kinematics	IV
	% of reach in 1 <sup>st</sup> unit	I	I:F	Video; kinematics	IV
	UE associated reactions	I	I:F	Video; kinematics	IV
1991 Law <sup>25</sup> ÷	Fine motor age	I	I:F	Peabody FM Scale	I
	Physiologic hand function	I	I:F	QUEST	I
1994 De Gangi 26	Qualitative movement	I	I:F	Video analysis; checklist	V
1994 Bower <sup>27</sup>	Gross motor skills	FL/A	A&P	Goal setting/GMFM	III
	Parent satisfaction	SL/C	Env	Questionnaire	III
1996 Fetters <sup>28</sup>	UE movement time	I	I:F	Kinematic analysis: reach	III
	UE movement unit	I	I:F	Kinematic analysis: reach	III
	UE reaction time	I	I:F	Kinematic analysis: reach	III
	UE displacement	I	I:F	Kinematic analysis: reach	III
1997 Jonsdottir <sup>29</sup>	Postural alignment	I	I:F	PAS	III
	Postural alignment	I	I:F	Kinematic analysis	III
1997 Law <sup>30</sup>	Fine motor age	I	I:F	Peabody Fine Motor Scale	I
	Physiologic UE function	I	I:F	QUEST	I
	Hand activities	FL/A	A&P	COPM	I
	Parent satisfaction	SL/C	Env	Rating scale	I
1999 Trahan <sup>31</sup>	Gross motor activities	FL/A	A&P	GMFM	IV
2000 Adams <sup>32</sup>	Gait: stride length	I	I:F	Pedographs	IV
	Gait: step length	I	I:F	Pedographs	IV
	Gait: cadence	I	I:F	Pedographs	IV
	Gait: velocity	I	I:F	Pedographs	IV
	Gait: foot angle	I	I:F	Pedographs	IV
	Gait: base of support	I	I:F	Pedographs	IV

### ***Lists of Miscellaneous Outcomes or Measures by Dimension***

To further help users of the classification determine in which dimension an outcome has been measured, the following lists will be updated as reviewers code outcomes to construct evidence tables using this classification. Different ways of stating an outcome may appear in the lists. Access to such lists will lead to increasing consistency of coding across users.

Note that death (death rates or survival rates) will be coded as a complication in the evidence tables rather than an outcome in one of these dimensions.

#### **Impairment of Body Function**

##### **Any Measure of the Following That Monitors Changes in Molecular, Cellular, Tissue, Organ or Organ System Function**

Musculoskeletal organ system: muscle tone, strength or force, spasms, speed; range of motion, reflexes; bone strength, length, dislocations; absent limbs; kinematic and kinetic measures of abnormal function of limbs (co-contraction, slow speed of isolated movements like dorsiflexion of foot); movement patterns like dystonia, spasticity, chorea, athetosis; gait analysis (force plate measures, joint excursions, time in double limb support)

CNS: patterns of neural firing

Cell: Amino acid or organic acid inborn errors of metabolism

Bowel and bladder: voiding, urgency, continence, constipation

Eye: central vision, peripheral vision,

Brain: Intracranial hemorrhage or hydrocephalus

Brain: memory (verbal, visual, recognition, retrieval, long-term, short-term, active working); perception (figure-ground, constancy, position in space, visual-motor coordination, discrimination); arousal (somnolence, lethargy, activity level); sleep (disorders of maintaining sleep, sleep walking, night terrors, excessive sleep); attention (distractibility, concentration, perservation)

Nerve: Abnormal presynaptic inhibition

Emotion: mood swings, psychosis, depression, hallucination

Sensation: pain, tactile defensiveness, numb or insensate

Gastrointestinal system: gastrointestinal reflux, motility, absorption

Pulmonary system: endurance, reactive airway, mucous production, cough reflex

Cell: Amino acid (or foreign matter such as lead) levels in urine or liver

Cell: Enzyme activity

Motor, social, cognitive, language development: tests or subtests that yield developmental quotients or scores that compare to a standardized norm for motor development, intellectual development, or social development (Bruininks-Osterestsky Test of Motor Proficiency, Peabody Developmental Motor Scales, Wechsler Intelligence Scale for Children, Vineland Scale of Social Maturity, Illinois Test of Psycholinguistic Abilities, Test of Language Development); motor development composite scores of ratings of extension, flexion, symmetry, weight bearing in various positions such as prone or sitting that could, but have not been, normed.

## Impairment of Body Structure

**Any Measure of the Following That Monitors Changes in Structure of Cells, Tissues, Organs, Organ Systems or Body Parts Compared to Established Norms**

Brain	Arnold Chiari formation
Face	Cleft palate
Spine	S/P decompression of posterior fossei
Cell	Cell damage from spinal cord disruption or brain damage
Nerve	Myelination of nerve pathways
Cell	Embryonic development of limb buds
Chromosome	Chromosome deletions, additions, translocations, or mosaicism, such as extra chromosome 21
Chromosome	Structure changes within genes such as the fibroblast factor receptor-3(FGFR3) on chromosome 4
Bone	Mineralization of bone
Collagen	Structure of collagen
Cell	T-cells mass
Eye structures or vision	retina, lens (cataract), acuite, rods, cones
Bowel and bladder	Capacity
Growth	weight, length or height

## Activity Limitations and Participation Restriction

**Any Measure of the Following That Monitors Changes in Skill Performance or Activities and Participation in Social Roles**

### *Types of Activities*

Gross motor skills or purposeful activities	walking, sitting, standing, climbing, jumping, pushing, pulling, transferring, riding, running, driving, skiing, ease of being positioned
Fine motor	pinching, grasping, holding, writing, drawing, dressing, buttoning, combing, brushing, sewing, placing
Oral motor	eating, drinking, swallowing, sucking, chewing, licking
Communication	asking, telling, gesturing, speaking, smiling, laughing
Cognitive	thinking, learning, remembering, paying attention, listening, concentrating, reading, spelling, calculating
Social-emotional	feeling, interacting, socializing, coping, visiting

### *Types of Changes in Activity*

Presence of activity	Does not perform skill or activity Performs with help (amount of assistance) Performs with assistive technology
Quantitative change	Speed with which activity is accomplished
Qualitative change	Physical states: performed with fatigue, discomfort, energy cost Feeling states: anxious, confident, willing, persistent Coordination: smooth, jerky, awkward, fluency

### *Examples of Measures of Activity*

Timed speed of activity: walking 50', dressing for school, 10 buttons
Count of caregiver time required
Video recordings of sitting with support, with hands free, with hands to maintain balance, sitting tentatively and attending to task while sitting to evaluate change in sitting
Written log of means of communicating used during specified time period: communication device including computer, mouth stick to write, beeper switches, gestures, unintelligible or intelligible verbalizations
Timed samples of quality and types of communications: responding to question;, initiating a communication; communicating basic needs; communicating thoughts, feelings, questions
Count of choking episodes during meal
Rating scale: presence, absence, or reduction of restricted breathing
Count of hours of uninterrupted, restful sleeping through the night
Timed samples of time sitting on toilet before voiding begins (spastic bladder)
Gross Motor Function Classification for Cerebral Palsy
Pediatric Evaluation of Disability Inventory
Achievement tests of basic learning skills, e.g. California Achievement Test

### *Types of Participation*

<b>Participation as a</b>	<b>Participation in</b>
student	personal maintenance or self care
worker	exchange of information
neighbor	personal mobility
athlete	social relationships
friend	civic life
citizen	economic life
	occupation
	family life
	community life

***Examples of Participation Changes that May Be Measured***

Job limitations: type of work, place of work, part or full time
Status of “problem kid” which excludes child from some class activities
Limited social eating (due to diet restrictions or oral-motor dysfunction)
School attendance
Variety and level of participation in school activities: academic classes, P.E., playground, lunchroom, after school activities
Variety and level of participation in sports: spectator, mascot, water boy, player, attends games, watches on T.V., collects baseball cards, keeps statistics, discusses sports
Church participation: Sunday School, Bible School, church services, special events
Scouting and other social or learning group participation
Amount of time spent with friends
Peer social activities: playing in neighborhood, sleeping over, going to movie
Family functioning: chores at home, lives away from family (assisted living, independent living)
Type and frequency of mobility: drives own vehicle, uses public transportation
Dating, co-habiting, marriage, parenting

**Environmental Factors**

**Any Measure of the Following That Monitors Changes in Barriers to Participation Imposed by Society or Other External Variables**

Private insurance and Medicaid funding policies regarding medical procedures and assistive technology
School policies regarding who can attend, how (transportation availability, parent volunteer work), and type of classroom programs and support available
Architectural barriers including braille in elevators, wheelchairs ramps or elevators
Community support, public policies and funding for low-income housing with support for independent living
Public access to TV and movies through close-captioning, telephones via TTY, and computers via accessibility options
State and federal laws requiring accommodations to people with disabilities in the workplace and school
Physical availability of a medical treatment, i.e., in community center, primary care hospital, tertiary treatment center
Financial availability of a medical treatment: costs, investigational status (not covered by insurance)
Funding to support policies of inclusion of people with disabilities in the workforce and in the educational system
Parenting behaviors or beliefs
Home environment factors

### Appendix 3: Coding Levels of Evidence

(Abstracted from writings by DL Sackett, et.al.; from *A Manual for Assessing Health Practices & Designing Practice Policies* by DM Eddy; from *Selected Issues in Program Evaluation: Arguments for the Individual* by OR White; and from *communiqués* with Sackett and White. Updated with information from Zaza, et al. on procedures to reduce bias in systematic reviews)

#### **Interpreting the Validity of Evidence**

A common tendency when interpreting the results of a study is to search for the validity of evidence in the form of statistically significant results, often assuming that if such significance is found, the intervention should be recommended. Conversely, if such significance is not found, the intervention should not be recommended. This is erroneous thinking. The presence of statistical significance only establishes that there is an acceptably low level of *probability* (traditionally acceptable if less than 5%) that an observed difference among the groups may have occurred by chance. Statistical significance does not specify whether an observed difference may be desirable, clinically useful, or even attributed to the intervention. It is an important step, but only one of several in the search for validity of the results of a study.

The appropriate search for validity of the evidence from a particular study requires three steps of interpretation. A fourth step in the search for validity of evidence requires *replication* of the result by other studies.

1. The first step takes the observed results of a study at face value: the presence or the absence of a difference among groups. Seldom is there no absolute difference; usually there is at least some small difference found among the groups studied. An important question at this point is whether the magnitude of this difference is large enough to represent a clinically useful or important difference.
2. Where there is a face value difference, the second step applies statistical calculations to that difference to ascertain the likelihood that it may not have been a true difference but only occurred by chance. This step affects our level of confidence about the validity of the face value difference, i.e., our certainty or uncertainty that the observed difference was a true difference.
3. The third step identifies any bias (or threat to validity) in the study that may be responsible for the result. In other words, some bias may have existed in the subjects that constituted the treatment or control groups or in the circumstances in which they were studied that led to the obtained result. This step affects our level of confidence about the extent to which the result can be attributed, at least in part, to some aspect of the groups being studied or how they were studied instead of to the intervention.
4. The appropriate search for evidence about an intervention has yet a fourth step. It is only when other studies have replicated a particular result that one can conclude with real confidence that there is evidence either for—or against—the efficacy of an intervention.

## Face Value Difference

The face value effect of an intervention can be measured in several ways. Common measures are absolute or actual changes or absolute or actual differences in the outcome measures with and without the intervention, ratios of the outcome measures with and without the intervention, relative or percent changes in the outcome measures, odds ratios, and effect sizes.

## Clinically Important Difference

Determining the clinical importance of a face value result requires knowledge of the measure used and what incremental differences may signify. Clinical importance may be evaluated by a reader who is familiar with the measure, by a reader who is given enough information in the article about the measure to recognize a meaningful change, and/or by authors' direct statements about whether the magnitude of change was sufficient to be regarded as clinically significant or important. [Clinical importance in Sackett's definition also includes a benefit/risk assessment which takes complications into account. This will be handled later in the review.]

## Statistically Valid Difference

This is where inferential statistics come in.

Statistical Significance for Hypothesis Testing. The fundamental purpose of most inferential statistics is simply to establish the existence of a change or difference per se. Univariate analyses (t-test, F-test or ANOVA or analysis of variance), covariate analysis (analysis of covariance), and multivariate analyses are used to infer the statistical validity of a difference. They yield a value (called P) that represents the probability of the difference being valid (did not occur by chance); by convention a minimally acceptable level is  $P < 0.05$ .

These statistical calculations are used to test the hypothesis that an intervention had no effect; this is called the null hypothesis. If the calculations show a sufficiently low probability ( $P < 0.05$ ) that the observed face value difference occurred by chance, the null hypothesis is rejected, and it is said that the result is statistically significant. If the probability is not sufficiently low, then it is said that the result is not statistically significant—meaning that there was not a demonstrated effect in this study. While this *implies* that the intervention, in general, is not effective, it is an implication which is not always appropriate because the probability of detecting a difference depends on the power of the particular study. Power is the result of two conditions: the number of subjects combined with size of effect of the intervention. Thus, it is valid that the size of the effect was not large enough to be statistically significant, given the number of subjects investigated. On the other hand, a similar study with a larger number of subjects may detect that size of an effect. Thus, in the absence of statistical significance, the result of a individual study is best regarded as inconclusive. When the study is replicated using a larger number of subjects and statistical significance is still not demonstrated, then it becomes appropriate to conclude that an intervention is not able to produce a particular effect.

Power calculations. The power of a study is the probability that the study, given its particular sample size and design, could detect a true effect of a given size, if it were present. Calculating the power requires deciding the size of the effect that will be meaningful (clinically important or significant) and specifying the level of statistical significance that will be acceptable. Power calculations are an important complement to calculations of statistical significance. Power calculations can be done to determine the number of subjects needed for the study. Power

calculations can also be done, in retrospect, to estimate the number of subjects that would have produced statistical significance, given the size of effect that was observed in the study. Although investigators are strongly encouraged to report the powers of their studies, such reports are still relatively uncommon.

Confidence intervals. The CI gives a measure of the precision (or uncertainty) of study results for making inferences about the population of all such subjects with a narrow CI representing greater certainty. The CI is based on the idea that the same study carried out on different samples of similar subjects would not yield identical results but would be spread around the true but unknown result. The CI estimates this sampling variation. The estimation approach aims to quantify the effect of interest and also to quantify the uncertainty of this effect. Most often this is a range of values either side of the estimate in which we can be 95% sure that the true value lies. The convention of using the value of 95% as a minimum is arbitrary, just as is that of taking  $P < 0.05$  as being significant. The two values that define the interval are called confidence limits. Thus, the confidence limits (e.g., 0.5 and 0.73) are the two end points of a confidence interval, and the confidence interval is the entire range of numbers between the confidence limits. Technically, a particular interval (e.g., 95% CI) is the interval that, if the study were repeated a very large number of times, would contain the point estimate that is calculated for each repetition, 95% of the time. Despite the considerably different philosophical approaches, CI's and significance tests are closely related mathematically. Thus, P value of  $P < 0.05$  will correspond to a 95% CI which also excludes the value indicating equality of the groups (i.e., zero). In other words, if the CI includes zero, there will not be a statistically significant P value. The uncertainty expressed by a CI is associated with a correspondingly wider CI and this is affected by sample size (with the CI being wider in smaller samples, all else being equal). The increasing use of CIs in medical research papers over the last decade has benefited a more correct understanding of the external evidence used in the practice of evidence based health care, but CIs are still not reported in all studies.

**Example** (from Eddy): Imagine a randomized controlled trial of an intervention with 50 people each in the treated and control group. Imagine the outcome was that 12 people in the treated and 20 people in the control group died.

#### Statistical significance

Under a null hypothesis (meaning that the intervention had no effect), the P value (using the two-tailed test) of this experiment is  $P=0.086$ , and the result (difference) is not statistically significant at the traditional  $P<0.05$  level. That is, even if the treatment actually has no effect, the probability of observing a difference of either an increase or decrease in death rate as great as 8 of 50 is about 8.6%.

#### Power calculations

The power of a study this size to detect (in the sense of being statistically significant at the  $P<0.05$  level) a decrease in death rates as large as 10% is 0.18. That is, if the true effect of the treatment were to decrease the probability of death by 5 of 50, a randomized controlled trial with 50 people in each group would have an 18% chance of yielding a statistically significant result at the  $P<0.05$  level.

#### Confidence intervals

The observed difference between the treated and control group (point estimate) for the effect is -0.16 or -16%. The 95% confidence intervals are +2.4% to -33% which includes the value zero.

These inferential statistical models are tools that address only issues related to the likelihood that a true difference was present. They do not allow inferences to be made about the likelihood that “no true difference existed”.

### Causality or Attribution

Can the treatment outcome be attributed to the intervention? This is where levels of evidence come in and the remainder of this appendix is related to this topic. After establishing the clinical importance of a face value difference and determining the statistical validity of that face value difference, the study must be examined for possible biases such as some aspect of the groups studied or how they were studied as possible explanations for the obtained difference. The effect of any of such biases on the study’s results must be estimated, and a decision must be made about how the study should be regarded. We have chosen to use the weighted approach called “levels of evidence.” In any weighted approach, the assessor subjectively assigns a “weight” to the study, with the intention of modifying its influence, relative to other pieces of evidence. The levels-of-evidence method gives greater weight to research designs that are less subject to bias and error. It is relatively easy to apply this method because it does not require the formal training and extensive analytic time needed for more explicit and sophisticated methods of accounting for threats to validity. It is an approach that is better suited to reviews of evidence that will be done by volunteer clinicians, as in the AACPD reviews.

## **Levels of Evidence**

### **Types of Evidence**

It is important to know whether health care decisions (especially those involving high risk to people) can be based on the results of rigorously controlled investigations, or to know that they can rest only on the results of uncontrolled clinical observations—or on even lesser evidence. Potentially useful interventions in health care originate with a common sense idea, or with expectations that are based on knowledge of the pathophysiology of a medical condition and the mechanism of action of an intervention. Sometimes they are based on beliefs arising from unsubstantiated theories of physiology or from analogies to other conditions. These forms of evidence may constitute the sole basis for an intervention that is being offered. Usually, however, these types of evidence serve only in a preliminary way, until the intervention has been submitted to systematic observation and evaluation. While the foregoing types of evidence can have a compelling logic, the actual effectiveness of an intervention can only be established by external evidence derived through empirical research.

### **Evaluating External Evidence**

There are several approaches to treatment evaluation including group methods, single subject methods, outcomes research methods, and qualitative methods. Though there are these different approaches to evaluating external evidence, a long tradition of group research has led to the almost unquestioning acceptance that results of group studies are superior to other methods. Therefore, a brief primer on research approaches will follow that pays particular attention to single subject research. The single subject approach gets attention here because 1) medical researchers who are more accustomed to group methods generally lack training about this approach, 2) the approach is well suited to studying low incidence, highly heterogeneous populations such as those commonly found in developmental disabilities, and 3) studies using this approach are increasingly found in the DD literature. Some background about the single subject approach is also important for understanding the “adapted” levels of evidence classification that will be used in coding studies for the AACPD systematics reviews. This classification has been adapted from one by Sackett et al. that originally included only group study designs (1980) and still relies primarily on such designs (1998).

In group research (more appropriately called between-subjects methods), one group of individuals is treated differently from another group (or groups), and inferential statistics are commonly used to evaluate the obtained group differences in outcomes. The various groups are assumed to be equivalent at the onset of experimentation, an assumption which is only reasonable with random assignment of subjects to groups or matching of groups on important dimensions, and the measured difference between the mean outcome of the groups is attributed to their differential treatment. The more common term, “group research”, will be used.

In single subject research (more appropriately called within-subjects methods), the same person is exposed to both the treatment and the control condition(s), thus acting as his or her own control. In contrast to most group designs, where the outcome of interest is measured once (or at most a few times) in each of a large number of individuals, within-subjects designs typically involve a large number of measures for each condition in a relatively small number of subjects. Attributing the

measured difference to the intervention depends on comparing stability of outcomes measured repeatedly during each condition and on shifts in the obtained differences being consistently coincident with the shifts between the treatment and control conditions. Visual analysis is usually aided by trend and level analyses although inferential statistics can be used to evaluate the obtained differences. Research designs based on this approach are commonly called single subject designs. This is an unfortunate designation for it leads to the common misunderstanding that such studies are only about single subjects. A valid study can be conducted with a single individual, but it is rarely done. Most often, small groups are studied; but very large groups can also be investigated using within-subjects methods. Because the terms "single subject research" and "single subject designs" are so widely used, these will be used hereafter.

While the common practices of these two approaches (i.e., number of subjects and intensity of study of the individuals), may differ, these common practices are not necessary differences. What differentiates the two approaches is the type of variation each measures. Group research measures the variation of results at a group level and is limited to this. Single subject research measures the variation of results for an individual, but when multiple individuals are studied, single subject research can also measure the consistency of variation for the group.

Single subject research strategies offer an alternative to group research; some researchers would argue that they represent the approach of choice in many situations. While this may be a controversial issue, there are situations when many would agree that the single subject approach is the method of choice. One of these situations is the study of interventions in populations so heterogeneous in nature that any summative statements of groups as a whole might be terribly misleading. Another situation is the study of low-density populations in which it is not feasible to muster even the smallest of group sizes that would be acceptable for a reasonable group study. An increasing number of single subject designs are surfacing in the developmental disabilities research literature for the very reason that one or both of these situations exist in the study of interventions for people who have chronic, sometimes severe, and often complex disabilities present from early childhood.

In treatment evaluation, single subject strategies can prove to be as powerful and persuasive as group strategies, and they are more likely to inform and less likely to mislead in two ways: what happened to the individual subjects and how the outcome came about. In other words,

1. In a group design, the act of averaging the individual scores to describe a single, overall group effect obscures the individual effects. One is misled if one interprets the group effect to be representative of each individual, i.e. that all the individuals' scores clustered around the mean score. One cannot know whether the scores clustered—or whether there was wide variance with some people doing much better, and some doing much worse, but all averaging somewhere in between. Researchers do attempt to describe the variance in their outcomes in a group study by using statistical strategies such as analysis of variance, but less often do they specifically indicate how many subjects improved or worsened, and less often still do they describe the characteristics of those who varied. In single subject research, effect is described and reported at the individual level so that the number of subjects who benefited from the intervention (or not)—and their specific characteristics—is always apparent to any reader.
2. Single subject designs tap into the process by which an intervention works because they measure frequently over time. In other words, these studies can detect whether an intervention

needs time to work or whether effects recede over time. A study that measures each individual only once (as in most group designs) may fail to detect change simply because of the timing of that measurement. Alternatively, it may reflect change that existed only on that one day of measurement because of a reason totally extraneous to the intervention, e.g., if the subject was sick with a bad cold, it may not reflect change that would have been noted on other days.

While most of the same factors can threaten the credibility of findings from a group or single subject study, the group methods usually seek to control for those threats by distributing potentially confounding factors evenly among the various groups. The uneven (albeit unknown and unintentional) distribution of these factors in one group provides one of the most common threats to validity (called biased subject selection) in group studies. Single subject methods obviate this particular threat to internal validity because, with no groups, there is no potential for unknown bias in one of the groups. In addition, single subject methods allow for the direct observation and analysis of other threats. Thus, single subject methods can produce strong credibility (internal validity) that the observed changes can be attributed to the intervention, or conversely, that the intervention was not efficacious (not able to bring about the desired result).

### Forming Hierarchies of Evidence

Sackett, (1980) first proposed a hierarchy that differentiated three major divisions of group studies: experimental, observational and before-and-after studies. He reasoned that the first division that can be made is between experimental and observational studies. In experimental studies, the investigator controls the manner in which participants are allocated to the different groups, while in observational studies treatment allocation is a haphazard mixture of many unknown factors, such as patient and clinician preference. A hierarchy also exists within observational studies. Those studies which are planned in advance and undertaken prospectively are also less likely to be biased than studies which are undertaken retrospectively. Cohort studies, in which groups receiving the different interventions are evaluated concurrently, are, therefore, regarded as more likely valid than studies which make comparisons with "historical" controls. This reflects a tendency for there to be many more differences between two groups separated in time, than there are between two groups at the same point in time. In prospective studies, data collection is likely to be more uniformly reliable and complete; and, in direct contrast to retrospective studies, it is impossible for the selection of the participants in prospective studies to be influenced by their outcomes. Case-control studies, being retrospective are, therefore, prone to many extra biases and fall below concurrent cohort studies in the hierarchy. A third division includes before-and-after studies, where the same subjects are studied before and after an intervention with no additional comparator or "control" group. Here it is often very hard to conclude that differences seen are attributable to the intervention. For example, a reduction in cardiovascular mortality rates over a decade within a country may be due to the interplay of several economic, social and medical factors, which may not all be easily identifiable. However, there are some circumstances (such as the all-or-none study described later) when large differences seen in before-and-after studies prove quite convincing evidence of the effectiveness of an intervention. On the basis of these considerations of bias, the standard group study designs can be graded into a hierarchy of decreasing strength of evidence.

Recently, Sackett et al. (1998) made four changes to improve the original hierarchy. 1) Studies that were not well executed (i.e., "poor quality") are down-graded to the next level to reflect the

possibility of more bias, and thus, less credibility for the results of that study. 2) A special case of the before-and-after research design which they call an "All or none case series" was added. 3) A design from "outcomes research" methodology (i.e., analytic survey design) was added; this was the first design outside the standard group methods. 4) Non-empirical types of evidence were added at a fifth level. They placed these additions in the classification based on similarity of threats to validity as compared to the original group designs.

Owen R. White at the University of Washington collaborated with Charlene Butler to apply the concept of levels of evidence to single subject methods. Using similar reasoning to Sackett's about relative vulnerability to threats of internal validity, the main types of single subject designs were placed on the CEBM hierarchy. Whether group or single subject methods are used, studies can be placed in a hierarchy according to basic features such as randomization, manipulation of exposure, comparison, concurrent observation, and prospective versus retrospective conduct that define the research design and which reduce or at least account for biases or threats to validity. Some of these features are specific to a particular bias, but others are general in nature, controlling for several biases simultaneously (e.g., random allocation). The design features that control or account for internal validity manifest somewhat differently when used in group versus single subject methods, as shown in the following table, but the same features are present as follows in the table.

<i>Design Features</i>	<i>In Group Methods</i>	<i>In Single Subject Methods</i>
<b>Comparison to determine differential outcomes</b>	Intervention vs. control group(s)	Intervention phase vs. control phase(s) within a subject
<b>Similarity of comparison (initially and throughout the study) except for intervention</b>	Of intervention group compared to control group(s)	Of circumstances of a subject during intervention and control phase(s)
<b>Random allocation</b>	Of subjects to groups	Of conditions to a subject at various times
<b>Concurrency</b>	Treatment and controls groups are investigated concurrently	Treatment and control phases are investigated concurrently or in close temporal proximity
<b>Manipulation of exposure</b>	Treating one group vs. withholding it from another	Treating a subject at some times vs. withholding treatment at other times
<b>Ascertainment of exposure (compliance with control vs. treatment condition)</b>	That each person actually experienced their assigned treatment or control condition—and only that condition	That each assigned treatment or control phase was experienced by a subject—and only that condition—during the specified times
<b>Loss to follow-up</b>	Loss of subjects from groups	Loss of data during phases
<b>Sufficiently large sample to accurately determine the variation of effect of intervention and non-intervention</b>	Data representing the status of different people in treatment vs. control groups which yield a group estimate of variance	Data representing the status points of a subject during all phases which yield an individual estimate of variance
<b>Statistical evaluation of the presence of a change or difference</b>	Inferential statistics to aid evaluation of plotted data	Level and trend analysis of plotted data; sometimes inferential statistics

## AACPDM Classification of Levels of Evidence

### *Background*

The classification that the AACPD reviews will use is based on the work of Sackett, first with the Canadian Task Force on the Periodic Health Examination. First published in 1980, it was for many years referred to as "Sackett's levels of evidence and grades of recommendation". The classification was republished with little change in 1993, but more recently has evolved further and changed under the auspices of the NHS Research and Development Centre for Evidence Based Medicine (CEBM) in Oxford, England. The current version by Sackett (Director of CEBM) and his colleagues was posted on the CEBM web site on the Internet at [www.cebm.jr2.ox.ac.uk](http://www.cebm.jr2.ox.ac.uk). in late 1998.

Both the original and later version of this classification have some limitations. Both versions are a hierarchy of evidence that is based on research design types. Reducing a variety of research studies to fit a small number of categories in a classification system inevitably involves some over-simplifications. More difficult is that neither iteration of the classification has operational definitions that allow for its consistent use by others. Despite these limitations, however, the AACPDm feels that "levels of evidence" is a useful concept and that the classification developed by Sackett and his colleagues is the most thoroughly tested. Such a concept moves us beyond the uncritical acceptance of any study as being "true" just because it is "research", or beyond the too-critical rejection of research as being inconclusive because even the best studies are never perfect. Levels of evidence provides a relatively simple means by which to judge whether interventions are based on studies that approach some "gold standard" that is more likely to inform and less likely to mislead or whether they rest on less conclusive evidence.

There are further limitations of the CEBM hierarchy, however, that did not allow its "as is" use for the AACPDm review process. Those reasons follow:

1. Sackett's levels of evidence and grades of recommendation and its CEBM successor, grew out of an intent to develop practice guidelines, as follows. Sackett et al. identified three components of critical appraisal: 1) deciding whether the research study is internally valid, 2) deciding whether it is clinically important (the magnitude of the positive effect outweighs the risk of adverse effects), and 3) deciding whether it is applicable to other people. They wanted their levels of evidence designation for a study to reflect an composite of all three aspects of critical appraisal.

After appraising available studies to find evidence appropriate for a very specific guideline, they select the study or studies that reflect(s) the highest level of evidence to support recommendations in the guideline. For example, their guideline for treating giant cell arteritis is "Give steroids (A) immediately (D) at doses of at least 20 mg (C) daily (A)." This means that there is grade A evidence for use of steroids, grade D evidence for starting them immediately, etc. Their grades of recommendation (A, B, C, D) are extrapolated from the levels of evidence (I-V).

In contrast, AACPDm's interest in levels of evidence grew out of the intent to describe the credibility of the evidence that currently supports interventions that are in use. This descriptive process is more appropriately served by addressing the three components of critical appraisal, not as a composite, but separately, as can be seen by the next points. Moreover, the AACPDm reviews will not be making practice recommendations, thus the grades of recommendation portion of the classification is not needed.

2. The CEBM classification is basically concerned with the classification of group research. Despite the fact that Sackett et al. (1997) recommend an N-of-1 randomized controlled trial as the appropriate means by which to sort out whether a trial of therapy is efficacious in a particular person, in the CEBM application all within-subjects studies are ranked at the 4th level of evidence. Thus, unimpeachable evidence of a therapy's efficacy via an N-of-1 randomized controlled trial in their current classification is grouped in their classification alongside case series designs which, by virtue of the lack of a comparison with similar others who were not exposed to the intervention, can do no more than hint at causality. This is the also the result of their classification schema which combined the three components of critical appraisal. Since a single

subject study rarely has a large number of subjects (although it can) and since the other two components of appraisal (generalizability and ability to identify infrequent but potentially serious side effects) depend on a large number of subjects having been studied, they decided to downgrade single subject study designs. The only exception to single subject studies being placed at Level IV is when a single subject design has been used to study a larger number of subjects and been subjected to group (as well as individual) analysis using inferential statistics. The CEBM reviewers regard this as a transformation to group research, call it a multiple cross-over trial, and place it at either Level I or II.

3. The CEBM classification prioritizes group studies because their practice guidelines address the most common disease conditions in adults. These conditions affect large numbers of people, so large numbers of subjects are available. Group research is common, outcomes research is increasingly common, and single subject research is scarce or non-existent in the literature they review.

The AACPDm, however, is appraising the strength of evidence for interventions in low-frequency, heterogeneous populations. For example, a group study of an intervention for cerebral palsy (a relatively low-frequency condition) often includes all types and distributions of cerebral palsy; this heterogeneity within the condition of cerebral palsy produces important prognostic variability, regardless of intervention. Moreover, withholding a therapy in order to form a control group presents ethical as well practical problems in these populations. Single subject designs represent feasible, alternative approaches that are increasingly being encountered in the developmental disabilities literature.

4. The CEBM classification departs from some other weighted systems by mixing a requirement of inferential statistics (i.e., a narrow confidence interval with a randomized controlled trial) for evidence to qualify as Level I evidence). This is in contrast to, for example, the US Preventive Services Task Force's Grade of Evidence classification which defines Grade I evidence to be "evidence obtained from at least one properly randomized controlled trial" with no statement regarding specific outcomes of inferential statistical calculations). It is not the requirement, however, so much as the inconsistency of that requirement throughout the CEBM classification. At no other level is a narrow confidence interval or other inferential statistical finding a requirement. This inconsistency of requirement for Level I versus Levels II-V can be handled in practice guideline development because Sackett et al. are reporting only selected results from a study.

The AACPDm reviews, however, must characterize all the results in a study, some of which may be statistically significant or have narrow confidence intervals, some of which may not. An inconsistent requirement related to inferential statistics is problematic in this descriptive endeavor. Either all—or none—of the levels need to require it.

Given these reasons, and upon the recommendation of Sackett, we have adapted this CEBM tool to our purposes. The specific adaptations to the CEBM classification include: 1) limiting the use of the classification to an appraisal of internal validity, 2) changing the decision levels about single subject designs because they are placed in the classification on the basis of internal validity alone, 3) deleting the items in the classification that list systematic reviews since only original research studies will be included in the AACPDm reviews, 4) deleting the grades of recommendation because no recommendations are being made, 5) expanding operational definitions

to accomplish systematic reviews, 6) and deleting the requirement of narrow confidence intervals for Level I evidence in order to resolve the inconsistency in the classification related to confidence intervals and to allow the “level” to be assigned to a whole study and simplify the descriptive process. [ Update May 2001: After use of this methodology for two evidence reports and based on the input from a large group of reviewers, our readers want more detailed information about the level of evidence assigned, objecting to the CEBM convention of reducing the level of the study by one level if it were poorly conducted. Accepting the recommendation of a subcommittee charged with addressing this issue, the committee approved the following change in September 2000 and authors (also committee members) piloted its use in two reports in progress. The AACPDm level of evidence will be composed of two parts. See Coding Studies section below.]

Other aspects of critical appraisal (generalizability and clinical importance) will be addressed individually in the AACPDm review process, not as a part of a composite level of evidence. Instead, only the internal validity of studies will be signified by the level of evidence designation.

### *Classification*

The following table shows the AACPDm classification of levels of evidence of internal validity. In descending order, the designs are decreasingly able to demonstrate that the intervention—and not something else—was responsible for the observed outcome. Level I evidence is the most definitive for establishing causality; Level IV can only hint at it; Level V only suggests the possibility. See table next page.

<i>Level</i>	<i>Non-empirical</i>	<i>Group Research</i>	<i>Outcomes Research</i>	<i>Single Subject Research</i>
<b>I</b>		<b>Randomized controlled trial All or none case series</b>		<b>N-of-1 randomized controlled trials</b>
<b>II</b>		<b>Nonrandomized controlled trial Cohort study with concurrent control group</b>	<b>Outcomes research analytic survey</b>	<b>Multiple phases (treatment/no treatment) design Alternating treatments Multiple baseline across =3 subjects ABA design</b>
<b>III</b>		<b>Case-control study Cohort study with historical control group</b>		
<b>IV</b>		<b>Before and after case series without control group</b>		<b>AB design</b>
<b>V</b>	<b>Descriptive (after) case series or reports Anecdotes Expert opinion without explicit critical appraisal (or testimony) Theory based on physiology, bench, animal research Common sense/ first principles</b>			

Experiments are necessary when our expectation of an outcome is not certain. As a rule of thumb, Level I designs are well controlled experiments that must also include random allocation and manipulation of the intervention. Level II designs do not include randomization but are otherwise well controlled experiments or comparison studies. Level III designs are comparison studies, but one (or both) of the comparisons is retrospective. Levels I-III all include some control group or condition, the purpose of which is to establish an expectancy about the outcome in people in the absence of intervention. Level IV designs have no comparison group or condition. If there is a firm base of expectancy (i.e., all people with a certain condition previously died within 1 year of diagnosis), then a control is not needed to demonstrate convincing evidence for an intervention. For most conditions, however, we lack good descriptive information (natural history of conditions) or

baseline information, so an expectancy about outcome is necessary to establish convincing evidence about an intervention. Level V evidence is non-empirical evidence—or Level IV case series with high potential for bias reducing the information more to the equivalency of case reports at Level V. Level V evidence is not empirical research; it can only hint at possible relationships between intervention and outcome.

### *Operational Definitions*

Because there is no standard taxonomy for discussion of research concepts and designs, even within the field of medicine, the following definitions are given for the terms used in the classification. Some variations of these designs are included; others may be encountered in the literature. These definitions derive from the writings of Sackett et al., Eddy, and White, or from personal communications between them and Charlene Butler.

Control group vs. index group. A control group is a group of healthy or non-disabled individuals who do not receive the intervention. An index group is a group of non-disabled people who, like the disabled group, are exposed to treatment and no treatment conditions or phases during the study. This is most likely to be encountered in drug studies. Though the publication may call the non-disabled group, a control group, it is really an index group. It is not providing a control against which to compare the disabled group's response to the drug. That is provided by the no treatment phases of exposure to the disabled group (i.e., by the group crossing over between treatment and no treatment). The non-disabled group simply provides additional information for general interpretation of the outcome observed in the disabled group. Therefore, in coding the study, disregard the non-disabled group; it does not add to the internal validity of the outcome of treatment/no treatment observed in the disabled group. Footnote that there is an index group in the table but otherwise ignore an index group for coding the level of evidence of a study by its research design.

### Randomized controlled trials.

The distinguishing feature of the randomized controlled trial (RCT), in group research, is that people are randomly allocated to a group that is offered the intervention or a group that is offered nothing, "usual care", a placebo, or some other intervention. When groups are relatively small, strategies in addition to random allocation may be used to increase the similarity of the groups. Strategies such as matching and stratification by age or disability, for example, will precede random allocation; statistical strategies such as analysis of covariance will follow completion of data collection.

Variations of the RCT include the **randomized cross-over trial**. This is a RCT with the added feature of internal comparison of people against themselves as well as external comparison of the groups. People are randomly allocated to an intervention and control group and receive the intervention or control condition for a specified period of time, as in the RCT. Then the groups cross-over, i.e., the group that received the treatment initially now receives the control condition and vice versa. People in both groups have the medical condition of interest. The term cross-over is more commonly used in medical writing and usually refers to a group data.

Drug trials are often called a **double-blind, placebo-controlled trial** in people with the medical condition of interest. If they are randomly allocated to the treatment and placebo phases, then this is

an RCT. It may or may not have the word random in the title but it will be stated in the methods section. (There may or may not also be an index group. See above.)

#### All or none case series.

Not all certain knowledge is obtained through controlled experiments. Experiments are necessary only when our expectation of an outcome is not certain. If an outcome is essentially certain, such as a uniformly fatal disease, a case series in which one or more people survive after receiving an intervention provides very convincing evidence. Criteria for the all or none case series are met when all people died before the treatment became available, but some now survive on it; or when some people died before the treatment became available, but none now die on it.

#### N-of-1 randomized controlled trials.

In single subject research, treatment versus control conditions are manipulated within a single person; the order of these exposures is randomly allocated. There are several variations of the N-of-1 RCT, sometimes called a **randomized cross-over trial**; these include the **blind cross-over trial** or **double blind cross-over trial**. The difference between this and a group crossover is that there are repeated measures in multiple phases. A person frequently undergoes pairs of periods in which one period applies an experimental treatment (B) and the other applies a placebo (C) or baseline (A)—in other words, an **ABABA type of design** or **ABCBCBA** or variation. The order of these periods within each pair is randomly selected so that the conduct of the trial may be, for example, ABBAAB. Treatment outcomes are monitored to document the effect of the condition currently being applied. These phases are *repeatedly* measured until the person being treated and the investigator are convinced that the treatment period is clearly different, or clearly not different. In a blind trial, the person making the outcome assessments is blind to the treatment condition; in a double blind trial, both the subject and the assessor are unaware of the treatment condition. Although this method can also provide a group comparison when more than one subject has been studied, the focus of the published report is the individual comparisons. Alternatively, when multiple N-of-1 randomized controlled trials conducted under the same protocol have been summed and a group comparison is provided, this is called a **multiple cross-over trial**.

Another variation of the N-of-1 RCT is the **alternating treatments design** in which the subject is exposed to the treatment condition and control condition(s) in close temporal proximity. For example, a subject is assessed during a 20 minute exposure to a control condition followed by a 20 minute exposure to the treatment condition; these exposures are determined by random allocation. Yet another variation is the **multiple baseline across subjects design**; several subjects are assessed for differing periods of exposure to the non-treatment condition (called baseline) and then assessed during treatment exposure. The order in which subjects change from the control condition to the treatment condition is established through random allocation.

#### Non-randomized controlled trials.

This is like a RCT with the one exception that subjects are not randomly allocated to groups. Instead, the groups are selectively allocated or convened on the basis of factors such as convenience (for example, a comparison of people in two regional hospitals), availability (children who attend the investigator's clinic) or by the voluntary behavior of the subjects. A variation is the **cross-over design** without randomization. Most **double-blind placebo-controlled trials** use selective allocation rather than random to determine the exposure to treatment/placebo. Similarly, the absence of random allocation of treatment and control conditions to a single subject makes the

**alternating treatments design** and **multiple baseline across subjects design** the single subject research equivalent of this experimental design, as does the **ABABA design**. In the latter design, multiple opportunities are available to observe change between treatment and control phases.

Cohort study with concurrent control group.

This study is essentially the same as the non-randomized controlled trial.

Outcomes research analytic survey.

This is not a direct study of people. Instead, “groups” are created from retrospective review of information obtained from a database. This is sometimes called a correlational study. Depending on the criteria used for inclusion in a database, the observations might be controlled and calibrated. The database can be on a large scale (e.g., national surveillance registry for spina bifida) or small scale (e.g., the spina bifida clinic of a hospital). In outcomes research, the investigator *starts with an intervention of interest*, sorting all the people in the database in rows, by whether they received the intervention or not, and then sorting them in columns, by whether they had the outcome of interest or not. As an example, the investigator uses a database that was created by a hospital's neuromuscular clinic to follow up its patients with cerebral palsy. Data can be extracted allowing the following question to be answered: "Are children with spastic diplegia cerebral palsy who received selective posterior rhizotomy (SPR) surgery more likely to walk unaided at age 10 than children who did not? The investigator organizes the numbers obtained as follows:

	<b>Outcome of interest</b> (unaided walking)	
	Present	Absent
<b>Exposed to intervention</b> (SPR)		
Yes	a	b
No	c	d

The numbers in cells a,b,c, and d are analyzed for differences in rates of outcomes. A major value of this design is that it demonstrates whether the outcomes we might expect to observe (or have observed) in controlled experiments are also being observed in the real world of clinical care.

Case-control study.

Like the outcomes research analytic survey, this, too, is a correlational study of "groups" created by retrospective review of information for analysis. In this correlational analysis, not all people who receive an intervention are tracked, only those with a particular outcome. Similarity of people with an outcome of interest (cases) to people without the outcome (controls) can only be established retrospectively; and their previous exposure to an intervention of interest can only be verified retrospectively. In the most common form of this design (called a **2x2 case-control study**), the investigator starts with identifying a group of people who already have an outcome of interest (called cases) and another group of similar people who do not have the outcome of interest (called controls). The investigator then retrospectively examines the histories of both groups to determine current or previous exposure to the intervention. This design is useful for establishing the relationship between low-frequency outcomes that may be important (i.e., serious adverse outcomes) and an intervention . Using the same example as that used in defining outcomes research (i.e., spastic diplegia, selective posterior rhizotomy, and unaided walking), the matrix that is created for the analysis looks exactly the same; the numbers in the cells differ, of course.

There is a variation of this design sometimes called a **matched case-control study** which begins by identifying the cases; then for each case a specified number of controls (typically two to five) that

match the case with respect to several important characteristics (e.g., gender, age) are identified. Another variation is used when the intervention can occur in degrees or categories of intensity such as the dose of a drug or exposure to varying degrees of asbestos; this is called a **2xk case-control study** with the k referring to the number of degrees or categories of intensity of the intervention.

ABA design. In this study, baseline (A) is established for the outcome of interest through multiple measures made over a period of time. A treatment period (B) follows and the level or trend of the outcome is established. Finally, the treatment is withdrawn with multiple measurements made again (A) to observe whether the outcome reverses. Two opportunities to observe change between treatment and control phases are available.

Cohort study with historical control group.

This is a cohort study that compares two groups of people, but only one group is currently studied by the investigator. Rates of outcomes for the studied group are compared with those of a control group of people who were studied at an earlier time when, or in a place where, treatment policy (or availability) differed from that being investigated. Alternatively, the rates are compared with rates of outcomes published for a similar group who received a different intervention (literature control). Because the comparison studies were not conducted under the same protocol, it is almost certain that the people and the way their outcomes were measured is not the same, posing significant threats to the validity of results under this research design.

Before and after case series without control group.

A case series typically consists of a single group of people who receive an intervention and are followed for a time to observe their outcomes. The outcome is measured before and after the intervention, but any rate of change is not compared directly with the rates that occurred in people who were not receiving the intervention but were otherwise comparable. In the absence of a firm base of expectancy or a control group to establish an expectancy, a rate of change in a single group has little credibility. The observed rate of change may have occurred for some reason other than the intervention or may have even happened without the intervention. The single subject research equivalent of this group design is the **AB design**. In this study, the investigator makes repeated measures during a baseline phase followed by measures during an intervention phase. Only one opportunity to observe change between the treatment and control phases is available.

Descriptive case series or case reports.

These describe small collections of cases that usually involve a careful review of records, i.e., are usually retrospectively done. They only describe the person or group after the intervention; they do not include *data* about the person or group's condition prior to the intervention. Their main value is in documenting the occurrence of events that otherwise are known to be exceedingly rare.

Anecdotes.

Anecdotes are uncontrolled observations reported in a study. They do not establish causality, but they are useful to document that an outcome has, at least, been observed to occur in association with the intervention. Anecdotal reports of outcomes may also be useful for the formation of hypotheses for subsequent study. Anecdotal evidence includes comments reported by the author with or without any quantification. If the information is formalized and/or quantified (e.g., subjects' responses in an interview or questionnaire at the conclusion of a study) but not compared to responses to similar questions elicited prior to treatment, this is regarded as anecdotal evidence. All data reported from interviews will be regarded as anecdotal unless the interview also included a

rating scale that allows quantification of the interviewees' responses and allows comparison of interviewees' perception prior to and after the intervention.

Expert opinion without explicit critical appraisal (testimonial evidence).

This is defined as a statement of belief by an individual or a group about the effect of an intervention on an outcome without description of supporting evidence or rationale.

Theory based on physiology, bench or animal research.

Evidence about an intervention is said to be theoretical, if 1) no empirical observations exist about the effect of the intervention on outcomes, but 2) there is an appeal to a set of beliefs, based on knowledge of the pathophysiology of the disease and the mechanism of action of the intervention. This can be based on basic science (bench) research, animal research, or an analogy (i.e., screening for breast cancer may rest on the evidence (or belief) that screening for another type of cancer is effective in reducing mortality).

Common sense or first principles.

Occasionally there will be virtually unanimous agreement about the merits of an intervention, despite the fact that there is no external evidence pertaining to the intervention and no obvious biological theory that directly supports it. An example of a first principles or common sense intervention (and one that went wrong) is offered by Sackett. Controlling blood pressure in people with aortic dissection is based on the principle that lowering blood pressure will reduce ventricular output and so decrease the risk of further extension of the dissection. First principles suggest that prophylactic treatment with anti-arrhythmics that reduce these arrhythmias must also reduce mortality. The CAST study randomized people with these post-MI arrhythmias to flecainide/encainide or placebo, but found that more people on an anti-arrhythmic died than those on placebo. In developmental disabilities, an example of a common sense intervention (untested) is the multidisciplinary approach to care of children with complex developmental disabilities.

## Coding Studies

[ Update May 2001: After use of this methodology for two evidence reports and based on the input from a large group of reviewers, objections about the CEBM convention of reducing the level of the study by one level if it were poorly conducted led to coding the Level of evidence as follows. A subcommittee charged with addressing this issue in May 2000 made a recommendation that was approved by the whole committee in September 2000 and piloted in the following several months in two evidence reports in progress being developed by committee member/authors as a pilot of the new coding. . This being successful, the level of evidence will subsequently be coded in two parts as follows.]

The first step is to decide the highest level of evidence the *research design* could have provided using the Levels of Evidence classification, i.e., Level I, II, III, IV, or V. To do so, you must determine what research design was used for each study. Sometimes this is stated in the published report; when it is not, review the Methods section to figure out what design in the classification it is most like. Refer to the operational definitions of studies shown above for guidance.

Next evaluate the *actual conduct of the study* using the following questions. Conduct of the study will be judged as Strong (score of 7 or 6), Moderate (score 5), or Weak (=4). Note,

however, that Level V evidence is not empirical research; therefore, Level V evidence will not be coded for conduct of study.

8. Were inclusion and exclusion criteria of the study population well described and followed?
9. Was the intervention well described and was there adherence to the intervention assignment? (For 2-group designs, was the control exposure also well described?)
10. Were the measures used clearly described? Were they valid and reliable for measuring the outcomes of interest?
11. Was the outcome assessor unaware of the intervention status of the participants (i.e., blind assessment)?
12. Did the authors conduct and report appropriate statistical evaluation including power calculations?
13. Were dropout/loss to follow-up reported and less than 20%? For 2-group designs, was dropout balanced?
14. Considering the study design, were appropriate methods for controlling confounding variables and limiting potential biases used?

The level of evidence will be reported as a combination of the research design designation and the conduct of the study designation, i.e., I-S or I-W. I-S indicates that the outcome is from one of the strongest types of research design (Level I) that was well (or strongly) conducted (S). II-W indicates an outcome from a Level II type of design that was poorly conducted (so is weak Level II evidence).

When coding studies, make a copy of the questions for each study and make notes on it so you can recall, when asked later, what facts led you to answer yes or no to each question in order that you may support your coding. The committee may require these worksheets.

Assessment of factors to determine how well controlled a study was tends to be a highly subjective enterprise. Standards for what constitutes a well controlled study and standards for comprehensive reporting of various types of studies are needed, but have only recently been recommended for randomized controlled trials (only) in the CONSORT statement in JAMA, 1996. Thus, this assessment itself is subject to bias. This bias can be partially overcome by using a checklist in order that factors can be considered systematically. It can be further overcome by agreement of two or more authors independently coding the studies. Agreement by the members of the Treatment Outcomes Committee will reduce this bias still further. In the absence of independent agreement, the Committee and authors can reach a consensus ; in this way, we will all be honing our skills for making judgments about the levels of evidence and will develop consistency of judgment.

The checklist poses questions about some of the most important issues in comprehensive conduct of a study. Answering these questions with confidence will depend on the comprehensive reporting of the study as well as its comprehensive conduct. Unfortunately, judgment about the latter will depend on the former. Unless there is information in the report to answer a specific question in the affirmative, one cannot assume that certain threats to validity and/or strategies to reduce or account for them were considered and/or included in the conduct of the study. On the other hand, length restriction for publication and lack of standards for reporting mean that all

important details are seldom included in an article. Because of this, one will have to rely, to some extent, on the *apparent* thoroughness of the study. Finally, this coding will be based on the assumption that the investigator employed appropriate statistical evaluation of the results and interpreted the evaluation correctly. In other words, the conclusions of the authors about the results of a study reached through statistical evaluation will be accepted on their face value unless they know otherwise. Authors' remarks about clinical significance or meaningful effects will be similarly accepted. Thus, this coding is an analysis directed by a relatively rough set of guidelines which yields a rule-of-thumb sort of appraisal rather than a sophisticated analysis guided by discrete rules for decision.

## Appendix 4: The Evidence Tables

Study the evidence tables that appear in the published evidence tables as guides for displaying your data, relying on the format of tables from later articles because their tables represent improvements over earlier tables. Each evidence report will include:

Table I: Dimensions of Disablement

Table II: Levels of Evidence

Table III: Summary of Studies: Interventions and Participants

Table IV: Summary of Studies: Research Methods

Table V: Summary of Studies: Outcomes, Measures and Results [possibly in two sections (see below)].

Table VI: Evidence Table: Outcomes of (name of intervention) for (name of medical condition of population) [also possibly in two sections, (see below)].

Reviews that contain studies with data about medical complications will include an additional table.

Table VII: Evidence Table: Adverse Effects from Medical Complication

The contents of each table follows:

### ***Table: Summary of Studies: Interventions and participants***

- Study citation: first author, date of publication, and reference (in superscript)
- Specifics about intervention
- Control intervention, if any, and specifics
- Population sampled by study; add relevant detail
- Total number of subjects in study
- If not all the participants had the medical condition being reviewed, but included data for those that did, add a column to report the number with the condition of interest for whom data was available. (See ITB report for example.)
- Ages (range)

### ***Table: Summary of Studies: Research methods***

- Research design used; sometimes studies are conducted in two parts
- Duration individuals were studied
- Number of subjects in treated group. Use only the number who had the condition of interest IF this is reported in the summary of participants table.
- Number of subjects in control group
- Specific analysis of results available for subjects of interest (only) in this review; if yes, subsequent results section will contain outcomes for only this subgroup

### ***Table: Summary of Studies: Outcomes, Measures and Results***

This table may need to contain two sections because the type of results reported may be different: Section 1. Results of treatment compared to another condition including no treatment, placebo, and alternative treatments. Studies that reported a result, usually a mean score, that reflects the

difference in outcomes 1) between a treated condition and a control condition, 2) before versus after treatment in a group, 3) between baseline and treated periods in individuals are reported in one type of table. (See the NDT report.)

Section 2. Results of treatment that show the uniformity of effect of a treatment within a group. Studies that reported the number of participants in a case series whose outcome was improved on whatever was measured, versus the number who worsened, versus the number who were unchanged. (See the ITB report.)

Some studies report both types of results requiring the evidence report to do so. See ITB report for example of both in the same report.

**Section 1. Results of treatment compared to another condition (specify what comparisons were, i.e., no treatment, placebo, and/or alternative treatments)**

- Study: first author, year, and reference in superscript
- Outcome of interest. Anecdotal reports on outcomes *for which no formal measures are available* will be included as Level V (non-empirical) outcomes. Anecdotal outcomes can, in no way, establish causality, but when empirical data about an intervention is limited, anecdotal reports are useful to document that an outcome has, at least, been observed to occur in association with an intervention.
- Dimension of disablement in which outcome has effect
- Measure used
- Result: improvement, worsening or no change in face value results
- Indication that result was of sufficient magnitude to be clinically important
- Indication that result was statistically valid: record P value and/or CIs, if provided; if article only states that the outcome was not statistically significant, record NNS

**Section 2. Uniformity of results within treated groups**

- Study: first author, year, and reference in superscript
- Dimension of disablement
- Outcome of interest
- Measure used
- Number of participants in study who improved on the measure (+)
- Number of participants who got worse (-)
- Number of participants who were unchanged ( $\pm$ )

**Table: Evidence Table: Outcomes of (name the intervention) for (name the medical condition)**

Arranging the measured and anecdotal outcomes in a matrix or evidence table, based on dimensions of disablement, reveals the state of knowledge (i.e., the evidence) that has been documented or anecdotally reported and the gaps in our knowledge base. Outcomes may have to be aggregated in 2 sections in this table.

**Section 1. Outcomes of treatment compared with another condition (specify)**

**Section 2. Uniformity of outcomes within treated groups**

**Table: Evidence Table: Adverse Effects from Medical Complications**

Be as descriptive as possible within a table.

**Example: Intrathecal baclofen for spasticity of cerebral origin.**

<i>Study</i>	<i>Type of Complication</i>	<i>Cases</i>
Albright, 1991 N=23	Lethargy	1
	Disorientation and agitation	2
Albright, 1993 N=37	Urinary hesitancy	4
	Pedal edema	1
	Infections requiring pump removal and antibiotics	4
	Operations to correct catheter kinking, occlusion, breaks	5
Gilmartin, 1995 N=51	Seizures with bolus or CIBI in 31 subjects with no history of seizures	0
	Seizures in 19 subjects with history of epilepsy +current use of anti-epileptic drugs	9
Almeida, 1997 N=1	Operation to correct catheter break	1
Armstrong, 1997 S=19	Sedation	1
	Bradycardia	2
	Hypotension	2
	Apnea	1
	Respiratory depression	1
	Meningitis	2
	CSF fistula persistent at catheter insertion site	1
	Operations to correct pump and catheter system	10
Seizure activity during IB	3	
Gerszten, 1997 N=24	Infection, pump related	1
	Catheter related issues	3
Gerszten, 1998 N=48	Operations to correct catheter breaks or dislocations	11

## References

- Eddy DM. (1992) *A Manual for Assessing Health Practices and Designing Practice Policies: The Explicit Approach*. Philadelphia, PA: American College of Physicians.
- National Institutes of Health. (1993) *Research Plan for the National Center for Medical Rehabilitation Research*. NIH Publication no. 93-3509. Washington, DC: U.S. Department of Health and Human Services.
- American Academy of Pediatrics. (1994) *Methodology and Examples for Practice Parameter Development: A Step-By-Step Approach*. Elk Grove Village, IL: American Academy of Pediatrics.
- Pew Health Professions Commission. (1995) *Critical Challenges: Revitalizing the Health Professions for the 21st Century*. San Francisco, CA: UCSF Center for the Health Professions.
- Pope AM, Tarlov AR. (Eds.) (1991) *Disability in America: Toward a National Agenda for Prevention*. Washington, DC: National Academy Press.
- Sackett DL. (1993) Rules of evidence and clinical recommendations for the management of patients [see comments]. *Can J Cardiol* **9**:487-9.
- Sackett DL, Richardson WS, Rosenberg W, Haynes RB. (1997) *Evidence-based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingstone.
- White OR. (1984) 'Selected issues in program evaluation: Arguments for the individual'. (Eds.) *Advances in Special Education*. JAI Press,
- World Health Organization. (1980) *International Classification of Impairments, Disabilities and Handicaps*. Geneva, Switzerland: World Health Organization.
- World Health Organization. (1992-1994) *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* familiarly abbreviated as *ICD-10*. Geneva, Switzerland: World Health Organization.
- World Health Organization. (1997) *ICIDH-2: International Classification of Impairments, Activities and Participation*. Geneva, Switzerland
- World Health Organization. (2001) *International Classification of Health, Functioning and Disability*, Geneva, Switzerland (Internet [www.who.org](http://www.who.org))
- Seltser R. (1992) *Public Health Service Task Force on Improving Medical Criteria for Disability Determination: Summary Report*. Washington, DC: Public Health Service.
- Nagi S. (1991) 'Disability concepts revisited: Implication for prevention'. In Pope A, Tarlov A (Eds.) *Disability in America: Toward a National Agenda for Prevention*. Washington, DC: National Academy Press, 309-327.
- Zaza, S. et.al. (Task Force on Community Preventive Services). (2000) Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. *American Journal of Preventive Medicine*; 18 (1S): 44-74.