Chapter 2. Methods

This evidence report is based on a systematic review of the literature. Our EPC formed an Evidence Review Team consisting of pediatricians and EPC methodological staff to review the literature and perform data abstraction and analysis. The Evidence Review Team held several meetings and teleconferences with external technical experts representing the Social Security Administration (SSA), the American Academy of Pediatrics, and the Disability Law Center to refine key questions proposed by the SSA, and identify issues central to this report. A comprehensive search of the medical literature was conducted to identify the evidence available to address the questions. For this evidence report, we compiled evidence tables of study features and results, appraised the methodological quality and applicability of the studies, assessed the correlations of the predictors and outcomes, and summarized the results.

From the outset, the external technical experts and members of the Evidence Review Team found it necessary to distinguish between two definitions of disability, one used by the SSA for administrative decision-making and the other found in the medical literature.

SSA's statutory definition of disability in children includes specific "medically determinable impairments" which result in "marked and severe functional limitations", coupled with a temporal dimension: the impairment and functional limitations can be expected to result in death, or have lasted or can be expected to last for a continuous period of not less that 12 months (SSA, 1999). The medical literature, however, defines impairment and disability more broadly. Moreover, very few studies have examined the association between VLBW and SSA-defined disability, per se. As a result, this report will use the terms disability and impairment as defined by study authors. The definitions reflect those endorsed by the World Health Organization (WHO). Impairment is any loss or abnormality of psychological, physiological or anatomical structure or function; disability is any restriction or lack of ability to perform an activity in a manner or within the range considered normal for a human being. This evidence report also uses the assessments of functional limitations as defined by study authors.

Key Questions Addressed by the Evidence Report

For infants with birth weight < 1200 grams and for infants with birth weights between 1200 grams and 1500 grams:

1. What factors or combination of factors alone or in addition to birth weight will predict significant developmental disability in former premature infants?

2. Are such infants developmentally disability at 1 year, 2 years, or beyond?

To identify the functional or physical outcomes related to disability and the elements believed to predict them, we sought evidence that a specific factor(s) was significantly associated with (i.e. demonstrates a relationship with) a specific disability. For example, very low birth weight infants with bronchopulmonary dysplasia have lower receptive language scores; the degree of immaturity influences the risk of CP and neurodevelopmental disability in VLBW infants. The EPC Evidence Review Team looked for evidence of association of VLBW with six outcome conditions: cerebral palsy (CP) and neurological impairments; abnormal cognitive

development and mental retardation; speech/language delay, hearing loss, behavioral disorders, and learning disabilities; visual impairment (with or without other conditions); pulmonary impairment (with or without other conditions); and growth impairment.

Literature Search Strategies

Disability is not a specific medical condition that can be readily searched for. Thus we had to look at many studies with related concepts (i.e., medically definable impairments that are related to disability) to identify potentially relevant studies. The Evidence Review Team identified a comprehensive list of predictors and outcomes by organ system vulnerable to VLBW (Table 4). The predictors and outcomes formed the basis of MESH search terms.

A systematic literature search was performed for full journal articles of original data. We did not include review articles, editorials, letters or abstracts. English language studies were identified primarily through Medline searches conducted between October 2000 and February 2001. We performed an updated search in September 2001 and again in January 2002. Supplemental searches were also performed in ERIC, PsycInfo, HealthStar and Embase. Additional studies were identified from reference lists, review and primary articles, and from domain experts and reviewers.

The Medline literature searches were conducted to identify clinical studies published from 1966 through the search dates. Development of the search strategies was an iterative process that included input from domain experts. Each search strategy was designed to yield approximately 10,000 titles and the preliminary searches yielded multiple tens of thousands of titles. Samples of these searches were screened and keywords from potentially relevant articles were then used to refine future searches. All searches were limited to studies of humans that were published in English and included children (age < 18 years). Details of the literature searches are presented in Table 4 in Appendix A.

Search Results

The initial search for the literature review consisted of: 1) Medline search from 1966 through November 2001 (2,885 abstracts), 2) HealthStar search from 1975 to October 2000 (26 abstracts), 3) PsycInfo search from 1984 to September 2000 (333 abstracts), 4) ERIC search from 1966 to August 2000 (140). The total number of abstracts in the primary search was 3,384.

These initial search strategies used the text words “low birth weight”, “preterm infant or newborn”, “disability”, “limitation”, “handicap”, “impairment”, “follow-up studies”, “longitudinal studies”, “cohort studies”, “case-control”, “randomized controlled trials”, “research design”, etc. The search strategies were updated in January 2001, using as MESH terms the list of predictors and outcomes by organ system vulnerable to VLBW. A secondary search in January 2002 for the year cohort 1985 through 2001, used MESH headings “Population: infant, premature” or “infant, low birth weight”; Study design: “follow-up” or “cohort”; Disease: predictor and outcome terms such as cerebral palsy, retinopathy of prematurity, etc.

Using such a very broad search strategy, the EPC identified 13,130 articles. The study designs included were cohort studies, case-control studies, and randomized controlled trials, which compared VLBW outcomes to normal birth weight outcomes.

Limitations of the Literature Search

The literature search protocols and study inclusion criteria were both designed to identify all possible correlations available in the literature. However, while literature searches were intended to be comprehensive, they may not have been exhaustive. As noted above, search strategies were limited to focus on studies likely to be relevant. Searches were limited to English language publications. Hand searches of journal were not performed, and review articles and textbook chapters were not systematically searched. Nonetheless, important studies known to the domain experts and studies found in reference lists were included in the review.

Study Selection

Results from Abstract Screening

By the second literature search in January of 2001, a total of 16,614 abstracts had been identified from four databases. We screened 16,164 abstracts that covered 13 categories: central nervous system (2,930); ophthalmology (398); audiology (80), pulmonary (1,833); nutrition and growth (2,533); medication (dexamethasome) (183); perinatal factors (875); illness acuity (56); infectious diseases (2,378); gastrointestinal (477); bone/osteomalacia (10); health care (466); and immune disorder (921). Approximately 1,693 articles were retrieved after screening of the abstracts.

Because we used search strategies with high sensitivity but low specificity to avoid missing potential relevant articles for this evidence report, the result was a large number of abstracts requiring review. Physician members of the Evidence Review Team screened each article against the inclusion criteria.

We focused the literature review primarily on premature infants born weighing less than 1,500 grams, including all subcategories of birth weights (e.g. less than 1,250 grams, less than 1,000 grams, and less than 750 grams). We also incorporated literature that included infants with birth weight less than 1,500 grams within a larger premature cohort and literature on infants whose prematurity was defined by gestational age, since many studies use gestational age and not birth weight criteria.

The Evidence Review Team reviewed all studies (retrospective or prospective) reporting impairments in infants or children who weighed 2000 grams or less, whose gestational age was 35 week or less, or whose birth weight or gestational age were below these thresholds. Longitudinal data for minimum of 6 months was preferred. Table C lists the inclusion and exclusion criteria.

Table C. Inclusion/exclusion criteria

Include	
Randomized controlled trials, case control studies, prospective cohort or retrospective cohort studies	• Only abstract data from primary studies • English language • Population (BW <= 2000gm and GA < 35 wk) • At least 10 subjects • At least one of the listed predictors (See Table 4) • At least one of the listed outcomes (See Table 4)
Exclude	• Articles if infants born before 1980 • Articles published before 1985

Revised Study Selection Criteria

The EPC categorized by sample size and birth year the articles that met the minimum inclusion criteria for LBW: articles with one or more relevant clinical outcomes; follow-up duration greater than of equal to 6 months; enrolled patients born after 1980, and study size greater than 10.

We then established a hierarchy of studies based on study size and birth year of the infants. Studies with birth years from 1990 onward were given preference, followed by studies with birth years between 1985 to 1989 and then studies before 1985. Within each birth year cohort, studies with more than 100 infants were selected first, followed by studies with 50 to 100 infants and less than 50 infants. Using this classification hierarchy, we worked through the most relevant (recent) and strongest (largest study size) studies in succession before older and smaller studies, until a complement of 178 articles was achieved.

Outcomes and Predictors Considered in this Evidence Report

The Evidence Review Team identified a list of predictors and outcomes (Table D). The universe of outcomes included those, which SSA recognizes as medically determinable physical or mental impairments that result from anatomical, physiological, or psychological abnormalities (SSA, 1999).

Table D. List of predictors and outcomes

PREDICTORS	OUTCOMES
Various BW categories (such as…	**All outcomes listed below:**
< 1500 gm	Neurodevelopmental/behavioral outcomes,
< 1250 gm	Vision, Hearing, Pulmonary, Growth, etc.
< 1 kg	
< 750 gm	
Various Gestational Age categories (such as	**All outcomes listed below**:
< 35 weeks	Neurodevelopmental/behavioral outcomes,
<32 weeks	Vision, Hearing, Pulmonary, Growth, etc.
<30 weeks	
<28 weeks	
<26 weeks	
Central Nervous System – CNS	
Intracranial /intraventricular hemorrhage	Motor delay/ Cerebral palsy
Periventricular leukomalacia	Cognitive delay Mental retardation
Seizures	Behavioral
Hypoxic-Ischemic encephalopathy (HIE)	School performance
Ventriculomegaly/ ventricular dilatation	Learning disabilities
Primary Outcomes feeding / swallowing	
Prolonged apnea	Hearing disorders /Deafness
Intrauterine substance abuse (opiates, cocaine, ethanol)	Visual impairment/Blindness
Low Apgar Score	Speech / Language / Communication disorder
	Feeding / Swallowing disorders
	Post hemorrhagic hydrocephalus (PHH)
	Ventricular peritoneal (VP) shunt & other disabilities
	Neurodevelopmental Impression: Normal, Abnormal, Suspect
	Neurological examination: Normal, Abnormal, Suspect
Ophthalmology	
ROP (retinopathy of prematurity)	Visual outcome
HIE	
Audiology	
Aminoglycosides	Hearing outcomes
Furosemide	
Hearing Screen Failure	
Cardiovascular diseases	
Cor pulmonale	
Pulmonary	
Chronic Lung Disease	Asthma
Bronchopulmonary dysplasia	Pulmonary function
	Tracheostomy
	Upper airway
	Chronic Lung Disease
	Reactive airway
	Exercise tolerance

PREDICTORS	OUTCOMES
Gastrointestinal	
Short gut	Short gut
Necrotizing “enterocolitis” – NEC	Cholestasis / Cirrhosis
Total parenteral nutrition – TPN	Gastrostomies / GERD / Fundoplication
Cholestasis	
Nutrition / Growth	
	Weight
	Height
Bone	
Osteomalacia	Osteomalacia
Infectious Diseases	
RSV	
Meningitis	
Sepsis	
Congenital/Acquired Immune Disorders	
Hospital / Health Care Resource Utilization	
	Rehospitalization (for any reason)
	Costs
	Physical/occupational therapy
	Orthopedic
Illness Acuity	
SNAP	
CRIB	
Medications	
Dexamethasone	
Perinatal factors	
Drug hx	
Antenatal steroid use	
Chorioamnionitis	
Chorionic villous sampling	
Diabetes	
Pre-eclampsia	

Data Abstraction Process

EPC staff developed draft data abstraction forms, which were refined through an iterative process with the methodologic and domain expert members of the Evidence Review Team (Appendix B). Information abstracted for assessment included the study population characteristics (i.e. such information as age, height, weight and gestational age), inclusion and exclusion information, study design, study funding source, the results and conclusions of the

study. In addition, data for quality assessment of individual studies were systematically abstracted, including data for rating the internal validity and applicability of the study.

Pediatrician domain experts performed all the data abstraction. The EPC staff trained abstractors. As part of the training each Team Member abstracted three studies in duplicate with the Team Leader and meetings were held to discuss discrepancies. After training, all remaining studies were abstracted by one pediatrician. Abstracted data were verified by a member of the EPC staff when the data was transferred to evidence and summary tables.

Reporting the Evidence

We summarized the evidence we found for the LBW condition in two complementary forms. The evidence tables contain detailed information about the study characteristics, population and disease characteristics, patient demographics, treatment comparisons, and outcome measures. We used this information to derive an evidence-grade to provide an indication of "quality" for each of the studies used to address the key questions. This evidence-grading scheme captures dimensions of a study that are important for the proper interpretation of the evidence: internal validity, applicability, magnitude of treatment effect (for treatment studies), and the size of the study. This evidence-grading scheme is used as part of the reporting of the results.

A narrative and a tabular summary of the strength and quality of the evidence of each study are provided for each outcome condition. For some conditions, studies are grouped first by study sample disease type. For some conditions, studies are ordered first by methodological quality (best to worst), then by study size (largest to smallest). For other conditions, studies are grouped first by methodological quality, then by applicability of study sample.

Evidence Tables

EPC staff constructed evidence tables for each the outcomes within the six conditions of interest. These tables are presented in Chapter 7 of this evidence report:

Table E. Evidence tables created in this report

Table number	Table name
Evidence Table 1	Studies evaluating association of LBW and multiple outcomes: CNS, Eye, Lung, Growth, etc
Evidence Table 2	Randomized Control Trials for LBW infants and multiple outcomes: CNS, Eye, Lung, Growth, etc
Evidence Table 3	Studies evaluating association of LBW and cerebral palsy and neurological outcomes
Evidence Table 4	Studies evaluating treatment effects of LBW and cerebral palsy and neurological outcomes

Table number	Table name
Evidence Table 5A, B	Studies evaluating association of LBW with speech/language and hearing loss
Evidence Table 6A, B	Studies evaluating association of LBW with behavioral disorders and school performance
Evidence Table 7	Studies of association of LBW to ophthalmic outcomes
Evidence Table 8	Studies of treatment effects for LBW infants with retinopathy of prematurity (ROP)
Evidence Table 9	Studies evaluating association of LBW and pulmonary outcomes
Evidence Table 10	Studies evaluating association of LBW and bone and growth outcomes
Evidence Table 11	Randomized Controlled Trials in LBW Neonates for growth outcomes
Evidence Table 12	Studies evaluating association of LBW and nutritional outcomes
Evidence Table 13	Studies evaluating association of LBW and other outcomes

Summarizing the Evidence of Individual Studies

Grading of evidence can be useful in appreciating the overall "quality" of a group of studies addressing a question. Over two-dozen scales have been proposed to evaluate the quality of randomized controlled trials (Moher, Jadad, Nichol, et al., 1995). While it may be desirable to have a simple evidence grading system using a single quantity, the "quality" of evidence is multi-dimensional and a single metric cannot fully capture information needed to interpret a clinical study (Ioannidis and Lau, 1998). A recent empirical study applied 25 quality scales to one meta-analysis and found that different quality scales could result in different conclusions hence quality scales are inconsistent among themselves (Juni, Witschi, Bloch, et al., 1999). Another empirical study demonstrated the greater usefulness of assessing studies according to specific study design features (Lijmer, Mol, Heisterkamp, et al., 1999).

Methodological Quality

Methodological quality, often referred to as internal validity, addresses the design, conduct, and reporting of the clinical trial. Some of the items belonging to this entity have been widely used in various "quality" scales and usually include items such as concealment of random allocation, treatment blinding, and handling of dropouts. In this evidence report, we define a three category internal validity scale: A (least bias), B (susceptible to some bias), C (likely to have large bias).

Study Quality

A **Least bias**. Prospective study that is clearly reported, uses explicit and appropriate eligibility criteria, uses appropriate definitions of predictors and outcomes that are properly measured or estimated, uses appropriate statistical and analytical methods, and is free of obvious bias. Retrospective studies, irrespective of other aspects of quality, cannot be in category A. Study size should not be a factor for quality.

B **Susceptible to some bias**. Prospective or retrospective study that does not meet qualifications of category A but deficiencies are unlikely to cause major bias.

C **Likely to have a large bias**. Major deficiencies that cannot exclude possibility of significant bias. Insufficiently reported information.

Applicability

Applicability, also known as generalizability or external validity, addresses the issue of whether the evidence from the study population is sufficiently broad as to be able to generalize to the population at large. Individual studies are often unable to achieve broad applicability due to restricted study population characteristics and a small number of study subjects (Lau, Ioannidis, Schmid, 1997).

For questions where all studies within a given table evaluate children with the same (or similar) diseases, a designation for applicability was assigned to each article, according to the following three-level scale:

🧍🧍🧍 Sample is representative of the whole population of babies with prematurity and low birth weight condition relevant to the topic question (eg, whole population of preterm babies and infants with BW 1200-2000g). This implies a reasonable sample size, a diverse group of infants with the condition, and inclusion/exclusion criteria that will capture the whole group.

🧍🧍 A relevant sub-group or subgroups of very low birth weight and prematurity, (only those with a specific, though common, condition eg: BW 1200-1500 grams).

🧍 A very narrow group of subjects who are a limited sample of very low birth weight and prematurity (only those with a relatively rare condition, or a non-representative demographic group e.g. crack babies).

Study Size

The study (sample) size is used as a measure of the weight of the evidence. In general, large studies provide more precise estimates of prevalence and associations. In addition, large studies are more likely to have wide applicability, depending on eligibility criteria. However, large study size does not guarantee applicability.

Results

The type of results available is determined by each study's design, the purpose of the study, and the question(s) being asked. Therefore, the results presented vary across summary tables. Summary tables for some question present either association or percentage of subjects with impairment. Other questions include separate columns for the strength of the association of the predictors with the outcome of interest. For appropriate topics, associations are described with the following arrows:

- ⬆ Statistically significant positive association found between the predictor and the outcome. Significantly increased the risk of MR/CP/Growth/BPD/POP (any disability) associated with VLBW or GA.

- ⬇ Statistically significant negative association found. Significantly decreased the risk of MR/CP/Growth/BPD/POP (any disability) associated with VLBW or GA.

- ⇔ No association.