August 9, 2010

MEMORANDUM

TO: Honorable Mary K. Wakefield
Administrator, Health Resources and Services Administration

Honorable David A. Hansell
Acting Assistant Secretary, Administration for Children and Families

RE: **Comments on Proposed Evidence Criteria in HHS Home Visitation Initiative**

The Coalition for Evidence-Based Policy – a nonprofit, nonpartisan organization – strongly supports the Department's proposed evidence criteria for the new HHS Home Visiting Program (*Federal Register*, July 23, 2010). Specifically, we believe the criteria in section 3.1 for rating the quality of impact studies as "high," "moderate," or "low" are consistent with the evidence standards articulated by respected scientific bodies including the National Academies,[1] Institute of Education Sciences,[2] U.S. Preventive Services Task Force,[3] and Food and Drug Administration,[4] as well as the scientific evidence on which study designs are most likely to yield valid estimates of program impact.[5] Furthermore, we believe they are an exceptionally clear and succinct statement of key principles – a refreshing departure from the complexity and jargon that characterize many discussions of evidence – and may therefore prove to have a major influence on policy and practice. We congratulate the Department on this important step, and urge it to retain both the clarity and rigor of the criteria as they are finalized and implemented.

We offer the following suggestions in the areas where the notice invites comment. Our suggestions are based on our broad experience in reviewing impact evaluations for Congress, the federal agencies, philanthropic foundations, and others, as part of the Top Tier Evidence initiative, the Congressionally-established Academic Competitiveness Council, and many other efforts. As a foundation-supported nonprofit, nonpartisan organization, we have no affiliation with any programs or program models, and no financial interest in any of these recommendations.

1. We strongly endorse HHS's plan (in 7.0) to competitively award a portion of funds based on strength of evidence and other factors, and suggest the following approach.

Consistent with the program's goals set forth in the statute, we recommend that the competitive process prioritize models that provide the highest confidence, based on evidence, of a sizable and sustained impact on important life outcomes. Specifically –

A. We suggest that the body of evidence for each model be rated on three main factors:

(i) Whether the evidence of impacts has <u>strong scientific validity</u>. To address this factor, reviewers would consider such items as:

- Whether the model was found to produce positive impacts in studies rated "high" under HHS's study rating criteria;
- Whether these high-rated studies were conducted in more than one implementation site (e.g., two or more studies, or a large, multi-site study);
- Whether the prevalence and/or replication of statistically-significant impacts in these studies makes it unlikely that they appeared by chance;

- Whether these studies evaluated the model in real-world community settings and conditions where it would normally be implemented; and
- Whether other well-conducted studies have produced any countervailing evidence regarding the model's impacts.

(ii) Whether this evidence shows sizable impacts on important, final outcomes. To address this factor, reviewers would consider such items as:

- The size of the impact estimates (e.g., a 5% reduction in child maltreatment injuries versus a 30% reduction); and
- Whether the impacts are on "final" outcomes that indicate true improvement in people's lives – not just intermediate outcomes that may or may not predict final outcomes.

Explanation: Examples of final outcomes might include verified incidents of child abuse and neglect, or increases in maternal employment and earnings. For some final outcomes, such as preventable deaths from SIDS or choking, even modest impacts may be of great policy importance.

Examples of intermediate outcomes include positive parenting skills or connection of families to community services. Whether impacts on such outcomes actually lead to true improvement in people's lives is often uncertain. HHS's Head Start Impact Study, for example, found a sizable impact on preschoolers' ability to identify letters and words (an intermediate outcome), but no significant impacts on actual reading ability or other educational outcomes at the end of first grade.[6] Similarly, MDRC's New Chance Demonstration for young mothers in poverty found a sizable impact on their receipt of a GED (an intermediate outcome), but no significant impacts on more final outcomes, including employment, earnings, welfare dependency, and reading skills.[7]

(iii) Whether the impacts were sustained long enough to constitute meaningful improvement in participants' lives.

Explanation: For some outcomes, such as the birth of a substance-exposed infant, even an immediate impact can represent a meaningful improvement in participants' lives. For other outcomes – such as negative parenting behaviors or children's mental development – impacts are likely to be meaningful to the extent they are sustained over time, and too often they do not endure, as some recent home visitation trials have shown.[8,9]

B. We suggest HHS's competitive criteria award the largest grants to models rated highly on all three factors, as their scale-up could greatly improve lives of thousands of at-risk children/families. The criteria might also include a second tier, awarding more modest-sized grants to models that score well on the three factors but fall short of the top ratings.

C. In addition, we suggest HHS report results of its evidence reviews (per section 4) in a format that enables readers to clearly identify the models rated highly on all three factors – along with the main reasons *why* these models met the three factors, summarized briefly in plain language (e.g., "two high-rated studies of model X found increases in annual family earnings of $1000-$1500 over a two-year period"). Such a format would enable grant applicants not only to identify and select models backed by the strongest evidence for their grant application, but also to understand the reasons why these models are given preference in the funding criteria. Such understanding may be critical to building acceptance of the program's evidence-based criteria among grant applicants and the larger policy community.

2. We suggest two additional key principles to include in the study-rating criteria in section 3.1 (criteria which we believe are otherwise excellent):

A. "Outcome Measures." This criterion would address whether the study's outcomes were measured in ways that might introduce bias or raise other concerns about validity. For example, a study of a model that teaches parents that spanking and sharp verbal reprimands are inappropriate forms of discipline probably should not rely exclusively on parent self-reports to measure parenting behavior. The reason is that the parents in the treatment group may under-report spanking/sharp reprimands relative to the control group if they know such methods are frowned upon by program officials and researchers.

Similarly, in a study that measures parenting skills through observations of parent-child interactions that leave room for subjective judgment, the observers should probably be unaffiliated with the program provider and/or blinded as to which parents are in the treatment versus control group. Such steps are important to ensure that observer bias does not influence their outcome measurements.

B. "Analysis." This criterion would address whether the analysis of study impacts was done correctly – for example, whether its tests for statistical significance took into account key features of the study design, such as whether groups or individuals were randomly assigned, and whether the sample was sorted into groups (i.e., "stratified") prior to randomization. Inappropriate analysis methods often lead to false findings that impacts are statistically significant.

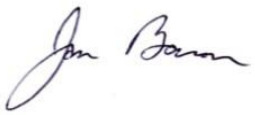
3. We suggest HHS establish a streamlined appeals process, managed by an independent expert, to correct any errors in the evidence ratings and to improve the criteria over time.

The reason we suggest this is that any review process, no matter how well designed, sometimes makes errors. And since the evidence ratings will have important funding consequences, it may be essential to have an effective, streamlined mechanism for study authors, program providers, and others to bring possible errors to HHS's attention so that they can be independently assessed and – if verified – promptly corrected.

Such an appeals process would also enable HHS to identify any problems with the evidence criteria themselves, and make appropriate refinements over time. As an illustrative example, if HHS's review criteria were to prioritize models showing "favorable impacts sustained at least one year after program completion" (as suggested hypothetically in Section 7.0 of the HHS notice, and as some other evidence review initiatives currently provide), a model with strong evidence of important impacts *during* the intervention – such as prevention of infant deaths from SIDS – could make their case that the criteria are too restrictive. HHS could consider such input and then make any needed adjustments.

In conclusion, we are highly supportive of HHS's proposed evidence criteria and funding plans for the home visiting initiative, and appreciate this opportunity to provide input on key elements of the proposal.

Sincerely,



Jon Baron
President, Coalition for Evidence-Based Policy

References

[1] *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities* (National Academies Press, 2009, recommendation 12-4, p. 371, linked here).

[2] U.S. Department of Education, "Scientifically-Based Evaluation Methods: Notice of Final Priority," *Federal Register*, vol. 70, no. 15, January 25, 2005, pp. 3586-3589. U.S. Education Department, Institute of Education Sciences, *What Works Clearinghouse Procedures and Standards Handbook, Version 2.0*, December 2008, linked here.

[3] U.S. Preventive Services Task Force, "Current Methods of the U.S. Preventive Services Task Force: A Review of the Process," *American Journal of Preventive Medicine*, vol. 20, no. 3 (supplement), April 2001, pp. 21-35.

[4] The Food and Drug Administration's standard for assessing the effectiveness of pharmaceutical drugs and medical devices, at 21 C.F.R. §314.12.

[5] Howard S. Bloom, Charles Michalopoulos, and Carolyn J. Hill, "Using Experiments to Assess Nonexperimental Comparison-Groups Methods for Measuring Program Effects," in *Learning More From Social Experiments: Evolving Analytic Approaches*, Russell Sage Foundation, 2005, pp. 173-235. Thomas D. Cook, William R. Shadish, and Vivian C. Wong, "Three Conditions Under Which Experiments and Observational Studies Often Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons," *Journal of Policy Analysis and Management*, vol. 27, no. 4, pp. 724-50. Steve Glazerman, Dan M. Levy, and David Myers, "Nonexperimental versus Experimental Estimates of Earnings Impact," *The American Annals of Political and Social Science*, vol. 589, September 2003, pp. 63-93.

[6] U.S. Department of Health and Human Services, Administration for Children and Families, *Head Start Impact Study Final Report*, January 2010.

[7] Janet C. Quint, Johannes M. Bos, and Denise F. Polit, New Chance: *Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children*, MDRC, January 1997.

[8] For diminishing impacts on negative parenting behaviors, see: Dumont, Kimberly, Susan Mitchell-Herzfeld, Rosa Greene, Eunju Lee, Ann Lowenfels, Monica Rodriguez, and Vajeera Dorabawila, "Healthy Families New York (HFNY) randomized trial: Effects on early child abuse and neglect," *Child Abuse and Neglect*, 2008, vol. 32, no. 3, pp. 295-315. Dumont, Kimberly A, et. al, "Effects of Healthy Families New York on Maternal Behaviors: Observational Assessments of Positive and Negative Parenting," New York State Office of Children & Family Services Working Paper, July 2008. Although this trial found that impacts on negative parenting behavior for the full sample had dissipated by the three-year follow-up, it did find impacts at three years on positive parenting behaviors.

[9] For diminishing impacts on child cognitive development, see: Landsverk J, Carrilio T, Connelly CD, et. al., *Healthy Families San Diego Clinical Trial: Technical Report*, Child and Adolescent Services Research Center, San Diego Children's Hospital and Health Center, 2002.