



Information Summit on Research and Evaluation Final Report

March 17-18, 2009

The National Quality Improvement Center on Differential Response (QIC-DR) in Child Protective Services Information Summit on Research and Evaluation was held on March 17-18, 2009, in Sacramento, California. The discussions were very rich, providing important clarifications and insights regarding: evaluation design alternatives; components of the process, outcomes, and cost evaluations; site selection; and data collection and analysis tasks facing QIC-DR staff. The dissertation support program was also discussed within the larger context of the evaluation effort. This synthesis summarizes the main discussion points, as well as specific suggestions from summit participants on each of these issues.

The summit began with an overview of differential response and the QIC-DR project goals and activities by Caren Kaplan of American Humane. This provided summit participants with an understanding of the broader context for the specific concerns of this summit—the QIC-DR research and evaluation efforts. A summary of the points covered can be found in the synthesis of the Policy and Practice Information Summit.¹

This was followed with an overview by Gila Shusterman and her Walter R. McDonald & Associates, Inc. colleagues involved in the project's background research efforts to inform the evaluation, including: the comprehensive literature review and annotated bibliography; a synthesis of existing differential response evaluation research; a state survey of differential response programs; and stakeholder focus groups. Charley Wheeler provided an overview of the initial findings from the synthesis of previous evaluations of differential response. Summit participants were asked to review the bibliography for the literature review and evaluation synthesis, and were invited to provide additional citations. Several participants did, indeed, send additional references and documents following the summit.

Following these background presentations, QIC-DR team members conducted a series of intensive discussion sessions intended to explore key questions facing the project evaluation design team. Each session began with a brief presentation by team members to present the current thinking of the design team, followed by general discussion among invited participants and staff. The summarized outcome of those conversations follows.

¹Reports for all QIC-DR information summits may be found at <http://www.differentialresponseqic.org/>.

Alternatives for the Multi-Site Evaluation Design

From an evaluative perspective, the participants strongly urged that the programs funded by the QIC-DR have time to mature sufficiently to provide a “fair test” of differential response and its effects on child, family, and agency outcomes. They argued that among less mature programs, program effects will be less likely to be observed, and therefore the evaluation will be less likely to produce strong evidence to guide future policy and practice. There was skepticism that brand new programs could become fully mature even 24 months post-award, which the evaluation team estimated as the outside limit of time that could pass before starting the outcomes component of an evaluation. (NOTE: Members of the QIC-DR evaluation team agree that this is a critical concern, but believe it may be possible for new sites to develop relatively mature programs in less than 24 months with intensive training, technical assistance, and support.) Conversely, there was some concern raised that child welfare agencies with mature programs would be less willing to implement an experimental design with random assignment after years of assigning all eligible families to the non-investigation pathway in a differential response model.

Participants also highlighted the importance of sufficient funds to support a robust evaluation. New programs may need to devote a large portion of QIC-DR funding to readiness issues such as workforce development, service development, and delivery, leaving less to support the evaluation. New programs are also likely to require more intensive training, technical assistance, and support from the QIC-DR for proper implementation, perhaps leaving fewer resources to support the evaluation. Based on these concerns, participants developed three possible evaluation scenarios for the QIC-DR. These are briefly presented next.

Option A: Focus on Outcomes of Mature Programs

Core Characteristics:

- Focus only on existing, fully mature programs to provide the best and fairest opportunity to find significant effects of differential response.
- Require each program to conduct/participate in a robust evaluation design that includes process, outcomes (child, family, and agency), and cost evaluation. The maturity of the programs should allow for greater focus on outcome and cost portions of the evaluation.
- More than the customary 10-20% of the QIC-DR funds to each program would be allocated for evaluation. Perhaps as much as 50-60% of funds would be dedicated to evaluation. (The participants thought even 80% of funds might be needed.)
- A rough distribution of evaluation funds for each component of the evaluation might be 45% for process, 35% for outcomes, and 20% for cost.
- In order to achieve maximum comparability across programs, significant work would need to be done with all sites. Comparability will be more difficult to achieve with mature programs with an existing set of practices.

Advantages:

- This would be the strongest approach for producing strong evaluation results and developing a base of evidence that can be the foundation for future differential response initiatives, according to research summit participants.

Disadvantages:

- This approach would not shed much light on optimal strategies for the successful implementation of new differential response programs.
- It would limit the potential applicant pool to fewer than a dozen sites with mature programs that have not already been evaluated using a rigorous experimental design.

Option A was clearly the top choice among summit participants.

Option B: Focus on Implementation of New Programs

Core Characteristics:

- Primary focus is on new, innovative differential response programs rather than established programs.
- Evaluation design would focus primarily on process/implementation. Child and family outcomes would be limited to specific indicators such as being re-reported to Child Protective Services (CPS) within a period of time. A cost evaluation focusing on cost benefits would have a low priority under this scenario due to high costs, though a cost analysis of how funds are expended could be conducted.
- It is expected that most of the site funds (80% or more) would be needed to support implementation of the new programs, leaving about 20% for evaluation activities.

Advantages

- Allows for detailed documentation of implementation from which one may be able to derive best practices for new programs.

Disadvantages

- According to summit participants, even allowing for a 2-year implementation period, programs may not be sufficiently mature to assess differential response program effects on outcomes.
- Even if the programs are fully up to speed in 24 months, the modest amount of funds that would be left for evaluation work may put substantial limits on the outcomes that could be measured. It also puts a substantial time limit on the period when outcome data can be collected due to the extended implementation period.

This option was considered by summit invitees to be the least desirable option, producing limited evidence to guide future policy decisions for differential response.

Option C: Hybrid Option

Core Characteristics:

- A mix of Options A and B, with two mature programs and two new programs, with the Request for Applications (RFA) clarifying that two types of grants would be funded.

Advantages

- Allows for a robust and broad evaluation design for the two mature sites, while also generating valuable information on implementation strategies for new programs. It would require two substantially different evaluation designs, one for mature programs and one for new programs.

Disadvantages

- It may affect the use of QIC-DR resources, since it would require two designs and also two different technical support approaches.

This option was offered by the group as a worthwhile strategy for the evaluation, particularly if Option A proved to be unfeasible for any reason.

SAMPLE SIZE

Summit participants discussed sampling issues at several points during the summit, because so many of the design issues have implications for sample size requirements. Sample sizes must be large enough to support the desired analyses with adequate statistical power to yield meaningful results. It was agreed that levels of comparability will not be great enough to justify combining data across sites, so each site must be capable of supporting desired analyses on its own. Sample size requirements should be determined by the QIC-DR staff and included in the RFP. Final sample size requirements ought to take account of the following:

- **Outcomes:** Power analyses setting minimal sample sizes are specific to the outcomes used to test the model. Staff will need to identify a desired power level (e.g., .8, .9), and produce power analyses for key outcomes. Results can vary substantially from outcome to outcome.
- **Subgroup Analyses:** It must be able to support separate analyses of high priority subgroups (e.g., first time/two-plus-time report families; risk level; accepted/refused services; age group; income level; race/ethnicity).
- **Data Availability:** Need to determine the percentage of families for whom no outcome data are likely to be collected (through attrition or non-response) and adjust sample size requirements accordingly.

It was recommended that each proposed site offer intake flow tables with associated numbers of intakes and responses for their proposed sites to make sure that each site can support the minimum sample size.

A LOCAL VERSUS CENTRALIZED EVALUATOR

The current design for the evaluation calls for strong local evaluators, with QIC-DR staff responsible for the cross-site portion of the evaluation, as well as providing technical assistance and coordination across the sites. Summit participants expressed concern that the local evaluator strategy runs a risk of data quality problems for some sites, which would reduce the power of the multi-site strategy. They recommended that consideration be given to a single, centralized evaluator if Option A is implemented. Similarly, if Option C is selected, the evaluation for mature sites might benefit from centralized control for at least some aspects of the evaluations. If local evaluators are used, and this is valued from an empowerment perspective, the selection of the sites should include a process to assure the quality and

capabilities of the evaluators proposed. It should be emphasized that this recommendation came from the invited researchers, and was not suggested by the QIC evaluation team.

LOGIC MODEL

Participants recommended that the QIC-DR staff develop a more detailed logic model that will identify the essential program features and expected outcomes that must define the core of each site's program. The current logic model is a start, but is too general. This will provide crucial guidance to those responding to the RFA, and should promote cross-site consistency to support the multi-site evaluation.

PROCESS EVALUATION

Harry Day engaged the participants in a discussion of process evaluation. The purpose of the process evaluation is twofold: to encourage and assess fidelity to the model (both cross-site and site-specific components); and to provide the information that would be needed for future sites to replicate programs that proved to be effective. The QIC-DR evaluation team expects to provide feedback from early assessments to sites during the first year of the program so that sites have the opportunity to improve implementation fidelity. Following the first year, the formal process evaluation will begin.

Participants offered the following suggestions for a strong process evaluation:

- Develop a detailed implementation fidelity checklist. This should be included in the RFA as guidance, and also used to guide the process evaluation once implementation is under way.
- Sites must be able to fully implement the design within the timeframe set. This may affect design, and will also affect site choice.
- Components for the process evaluation of DR should include:
 - Worker style, to capture the different approaches under alternative response and traditional investigation (a Barton-Holmes measure like the one developed by Washington State was suggested to capture this).
 - Worker attitudes toward available services for the families in their communities based on the Texas worker decision-making survey, in order to determine the degree of worker confidence in the level and types of services at their disposal.
 - Mix of available services and the degree of match to family needs (may need to use a panel of independent raters for this).
 - Worker and family engagement (good engagement measures exist).

OUTCOMES EVALUATION

Brett Brown presented four categories of outcome measures for discussion: child, family, agency, and community outcomes. Summit participants were given a list of many potential outcome measures that have been used in child welfare research and evaluation and were asked to identify the most relevant and feasible measures given the purpose and scope of the evaluation.

Overall Discussion Points

- The evaluation must track outcomes for all families, not just those who receive services. If one cannot follow those who do not get services, or who end services before the follow-up, it will undermine the evaluation. Child and family outcomes would be essentially limited to re-report.
- Outcomes chosen should have a realistic chance of being affected within the time period chosen for follow-up, somewhere between 6 and 12 months.
- Some Child and Family Service Review (CFSR) outcome measures may not be appropriate for this evaluation (e.g., some of the foster care-related outcomes). Consider carefully which CFSR measures are adopted for the evaluation.
- Other well-being measures will be needed not for outcomes, but to help assess the appropriateness of the services that are available or given at a particular site.

Child Outcomes

Participants expressed differing views concerning which child outcome measures should be followed. Some suggested minimizing the focus on child outcomes in favor of more family outcome measures, which may have a better chance of being affected by differential response interventions. The following outcomes were identified by one or more participants as important for the evaluation.

- Re-report
- Safety assessment post-services
- Removal (including length of time)
- Number of placements
- Behavior problems scale (shorter than Child Behavior Checklist)
- Emotional security with parent
- Individualized Education Plan placement
- Health service receipt
- School attendance and performance

Family Outcomes

Summit participants supported the collection of service-related and material hardship family outcome measures. Some expressed skepticism over the utility of such personal well-being measures as mental health, stress, and sense of belonging, though others were comfortable with using such measures.

- Economic hardship (e.g. adequate housing, number of moves in the last 6 months, adequate food, joblessness)
- Parental stress
- Social supports
- Parental mental health (e.g. Short Form 12 Health Survey)
- Sense of belonging to the community.
- Self re-report

- Parent satisfaction with services
- Engagement in service planning
- Service receipt levels
- Appropriateness of the services to needs of the family.
- Speed of service receipt

Agency Outcomes

Improvements in agency functioning are often identified in the literature as an important goal for differential response. Measures that were identified by summit participants as important included:

- Job satisfaction
- Turnover (need to control for the economy)
- Worker perceptions of differential response effectiveness
- Workload levels

Community Outcomes

While changing perceptions and attitudes within the community towards CPS can be identified as an important goal for differential response, summit participants did not consider these as crucial for purposes of the evaluation. These community outcomes were perceived as less likely to change in the limited time period covered by the evaluation.

Cost Evaluation

A brief presentation of the cost evaluation methodology and findings in Missouri was provided by Tony Loman, and summit participants then had an opportunity to provide comments and suggestions. These comments included the following:

- A full cost-benefits study would be extremely expensive to produce and would be impractical for this evaluation. A cost analysis could be produced that goes beyond CPS costs by using family surveys that ask families to provide information on income, dollars received via cash welfare, and extent of services from other non-CPS agencies. This would be important, as community partnerships are an essential philosophical feature of differential response. Alternatively, one could do a restricted cost analysis based on report recurrence and removal placement costs, comparing experimental and control groups. This would be useful but less informative.
- One could consider collecting cost information from a subsample of families if funds were not available to collect data on everyone, though this would limit the statistical power of what was collected.
- Cost savings should not be advertised as a likely outcome of differential response. While we might be able to show an experimental-control differential, this does not necessarily translate to lower agency costs.
- Conducting a cost study is in part dependent on the amount of evaluation funds made available to local sites. If evaluation funds are limited (e.g., to 10-20%) the most important elements are the

process analysis and a limited outcome study. If funds are adequate then it will be possible to conduct a cost study.

Data Collection Strategies

Summit participants had various suggestions throughout the two days about strategies for collecting data for the purpose of such an evaluation. These included the following:

- It is critical that data, including outcome data, be collected on all families that are part of the study. Many families in both the alternative response and investigative tracks will cease to interact with CPS staff altogether once their assessments/investigations are complete. Therefore, it is necessary that families be followed after they have left CPS services, and a means developed for gathering service- and outcome-related data.
- Participants were clear that survey data would need to be collected from the families. To maximize response, the following suggestions were offered:
 - Have the CPS caseworker collect detailed contact information at the time of the initial interview including the names and phone numbers of several relatives or friends who will know where they are should they move. Standard administrative data is not sufficient for this task, as there is often too much missing information.
 - Maximize response rates by offering a payment in return for families' cooperation. Be careful to get Institutional Review Board approval on the amount offered, as a too-generous payment may be seen as coercive. Done properly, it will substantially increase response rates with this population.

Because surveys are expensive, participants offered several suggestions for reducing costs if necessary:

- Survey random subsamples of the families in the study, but ensure that sample sizes are adequate to support analysis with the desired statistical power.
- Consider a combination of mail and in-person responses to reduce costs.
- A respondent burden table should be produced to estimate and control the level of respondent burden.
- For sites with advanced and integrated data warehouse capabilities, administrative data may be a rich source of service receipt data. Most sites will not have this capability, but it would be a valuable resource for any site that had this capability.
- Summit participants stressed the need to follow all families once they have left child welfare services. A large proportion of families in both tracks (alternative response and investigative) will cease contact with child welfare after receiving non-investigation or an investigation. In many places it is likely that less than 20% will have ongoing contact through in-home services or foster care services with the child welfare department. Thus, the issue of including surveys to track families was raised. There are many challenges and difficulties with such surveys, including highly likely low response rates; therefore, the costs to support all aspects of the evaluation would likely need to be increased.

Site Visits

Participants recommended that QIC-DR staff perform pre-award site visits to verify that the initially qualified sites are fully capable of meeting the requirements of the RFA, as the requirements will be difficult for many sites to meet, and their true capabilities may be difficult to discern from their proposals. Alternatively, in-depth telephone interviews should be conducted to verify their capabilities as much as possible.

Dissertations

In the discussion that followed John Fluke's presentation of the dissertation support efforts, the following points were made by participants:

- Consider funding a methodological topic for one of the dissertations — for example, the development and testing of initial screening protocols. If possible, pay the student directly rather than going through the university, to avoid overhead charges.
- Consider funding a dissertation that would help the QIC evaluation staff build a “fidelity index” for differential response implementation. It was noted that David Cordrey at Vanderbilt has students who would be interested in this.
- Fathers would be an important topic to cover as part of dissertation research.

Appendix A: Participant List

Donald Baumann, Ph.D.
Texas Dept. of Protective and Regulatory
Services

Brett Drake, Ph.D.
Washington University St. Louis

Jill Duerr Berrick, M.S.W., Ph.D.
Center for Child and Youth Policy
University of California-Berkeley

Diana English, Ph.D.
Casey Family Programs

Todd Franke, Ph.D.
Dept. of Social Welfare
University of California-Los Angeles

Rob Geen
The Annie E. Casey Foundation

Ruth Huebner, Ph.D.
Kentucky Cabinet for Health and Family
Services

Jacquelyn McCroskey, D.S.W.
School of Social Work
University of Southern California

Peter Pecora, Ph.D.
Casey Family Programs and University of
Washington

Susan Wells, Ph.D.
Psychology and School of Social Work
University of British Columbia/Okanagan

Project Staff

Brett Brown
Walter R. McDonald & Associates, Inc.

Karen Davis-Brown
Walter R. McDonald & Associates, Inc.

Harry Day
Walter R. McDonald & Associates, Inc.

Tony Loman
Institute of Applied Research

Mary Jo Ortiz
Walter R. McDonald & Associates, Inc.

Walter R. McDonald
Walter R. McDonald & Associates, Inc.

Tessa Robinette
Walter R. McDonald & Associates, Inc.

Ted Scott-Femenella
Walter R. McDonald & Associates, Inc.

Gila Shusterman
Walter R. McDonald & Associates, Inc.

Charles Wheeler
Walter R. McDonald & Associates, Inc.

¹Reports for all QIC-DR information summits may be found at <http://www.differentialresponseqc.org/>.