

Discussion of Issues in Differential Response

Research on Social Work Practice
00(0) 1-15
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1049731513493674
rsw.sagepub.com



Ronald C. Hughes¹ and Judith S. Rycus¹

Abstract

In this article, the authors responded to nine commentaries by 17 contributors to their article, *Issues in Differential Response*. The authors found that a majority of the respondents agreed with the major conclusions of *Issues in Differential Response*. However, there were varying degrees of disagreement regarding the significance of some of the article's conclusions. The authors point out and discuss the considerable divergence in the respondents' definitions of differential response (DR), their assessment of DR reform's empirical support, and their assessment of its potential for progressive development as an evidence-informed model for child welfare practice. The authors conclude that research claims and public belief regarding DR reform's safety and effectiveness exceed its scientific support, and they make suggestions for improving model building and outcome research for DR reform.

Keywords

alternative response, child neglect, child abuse, child welfare, child welfare research, differential response, evidence-based practice, social welfare policy, outcome study

We would like to thank all of the respondents who contributed to this special issue of *Research on Social Work Practice (RSWP)* by reviewing and commenting on our article, "Issues in Differential Response." The 17 coauthors of the nine formal response papers clearly put a lot of time and thought into their submissions, and they have raised many important points that warrant further consideration. We also would like to thank Dr. Bruce Thyer, Editor of *RSWP*, for recognizing the centrality of differential response (DR) in the current child welfare landscape and for facilitating a broader conversation by dedicating an entire issue of the *RSWP* journal to the topic.

Our goal in producing our article was to promote dialog on the strengths, limitations, and developmental potentials of current DR research and practice as a means of supporting the refinement and sustainability of DR reform and to overcome barriers to its success. We believed then, and we still believe that significant unresolved issues remain that can potentially undermine DR's viability. We thought it important to raise these concerns despite the potential unpopularity of the message. The strong reactions, both pro and con, that we received from even a limited distribution of the article suggest that the issues we raise are of intense concern to many child welfare professionals around the country. This was even more apparent when we read the nine responses published in this journal, which range from laudatory—even celebratory—to scathingly critical and dismissive. We speculate that the disparities in these responses may mirror wide differences in perspectives and opinions regarding DR in the larger child welfare community. This, alone, is a powerful indication of how much dialog is still necessary if DR is to move forward in a cohesive and coherent manner, and we remain firmly convinced that the

issues we identified should be fully addressed, as we proceed further with DR reform.

We also found that while all the respondents presumed to be talking about the same thing, it was evident that we were often starting from very different premises and assumptions and had many differences in our definitions and understanding of key constructs. For example, a major theme throughout the responses was what exactly qualifies as a model, how a model differs from an approach, a method, a philosophy, or a practice, and which of these DR is or should be. Another issue is what constitutes an evidence-based practice or an evidence-supported treatment, whether DR can or should be evidence based, and on what criteria we should be assessing DR's progress. Even more basic was what specific characteristics and attributes differentiate an assessment from an investigation and how we define and assess risk and safety. It was remarkable to us that among this knowledgeable group of child welfare professionals and researchers, many of whom had been working on DR for many years, even a general definition of DR was not to be had. This fact, alone, should be a bellwether of concern. How is it possible to discuss, develop, implement, promote, measure, critique, and praise something when we cannot even agree on what it is? Perhaps most troubling was that, despite this widespread divergence in

¹Institute for Human Services, North American Resource Center for Child Welfare, Columbus, OH, USA

Corresponding Author:

Judith S. Rycus, Institute for Human Services, 1706 E. Broad Street, Columbus, OH 43203, USA.

Email: jsrycus@aol.com

key DR elements, some of the respondents seemed reluctant to acknowledge the threat that these divergent opinions posed to DR reform and thus to children and families.

On a more hopeful note, it was gratifying to read that some of the respondents had, on their own, identified the same issues we had identified in past DR research and were already attempting to remedy these problems in their present research efforts. Other respondents offered a variety of recommendations for ongoing research that could help answer some of the remaining questions and enhance DR as we move forward.

What follows are our thoughts about some of the comments in each of the nine responses and a discussion of some of the larger issues they introduce. We hope this can serve as a stimulus for further dialog and potentially as a tool to begin to solve some of the outstanding problems and issues in the development and implementation of DR reform.

We have reprinted, below, the five primary findings from our original article, as we often reference these in our responses.

Finding #1: DR programs do not adhere to a uniform, standardized practice model, nor are programs implemented consistently across sites.

Finding #2: Methodological problems in the DR research limit confidence in research findings and conclusions.

Finding #3: There is insufficient data to confirm the safety of children served in alternative tracks.

Finding #4: DR programs appear to prioritize allocating services and resources for families in alternative tracks.

Finding #5: DR literature misrepresents traditional child protective services (CPS) to enhance an alternative response model.

Ellett

Ellett provides an interesting assessment of both the historical context in which DR developed and the strengths and challenges DR faces as an evolving child welfare methodology. Her personal work experience in both direct child welfare practice, and in academia and research, informs her commentary, which addresses both research issues and larger programmatic concerns.

Relative to Finding #1, Ellett concurs that a clearly defined and uniform model for DR is needed, but she cautions against viewing DR as the solution to all CPS system problems. Referencing the variety of child welfare practices that have been “intermittently in vogue since child maltreatment was first recognized,” (p. 2), she claims “it is unlikely that a single practice model will be effective with all families . . . ,” even though, “. . . the public and particularly politicians want simple answers and a quick and cheap *silver bullet* solution to CPS” (p. 2). We believe that DR’s organizing principle—different service pathways for families with different service needs—is consistent with Ellett’s vision of a flexible and adaptable CPS service system. The implementation of DR in many states has challenged rigid policy and practice infrastructures that have historically characterized much of CPS. DR reform has championed

development of infrastructures that are more conducive to individualizing family assessment and service delivery. That said, we do agree with Ellett’s concern about DR being inappropriately viewed as a “silver bullet.” We firmly believe the promise of DR is still far ahead of the science, and the utility of DR in addressing many CPS practice issues has often been overstated.

Relative to our Finding #2, Ellett concurs with our conclusions about methodological flaws in the DR research literature. She contends that such methodological challenges are not unique to DR research but are comparable to the challenges encountered in most social science research conducted in direct practice settings. She states that this type of research is typically “. . . bounded by variables that cannot be controlled . . . ” and that a “host of variables” can affect CPS case outcomes (p. 2). This supports our contention that researchers working in applied CPS settings have a responsibility to recognize, acknowledge, and describe the potential effects of these uncontrollable variables when interpreting their research findings and conclusions. Failure to do so leaves ample room for users of such research to overinterpret the strength of the data, increasing the likelihood of viewing the intervention more favorably than is warranted, and potentially contributing to Ellett’s “silver bullet” syndrome.

Relative to Finding #3, Ellett agrees that attending to child safety in all CPS cases is essential. However, she follows by saying that “Hughes et al. attest that a full *investigation* is needed in all CPS reports to establish child safety or maltreatment” (p. 2). Some of the other respondents also interpreted our call for thorough fact-finding regarding child maltreatment dynamics for all families in both experimental (AR) and control (TR) tracks as an indication that we advocate forensic investigations for all families served by CPS agencies. This is not what we advocate, but it does point up confusion in the terminology and a potential lack of agreement about the respective purposes and attributes of CPS assessments and investigations.

A comprehensive family assessment is the primary means of gathering relevant information for treatment planning purposes. Family assessments in CPS cases must incorporate an appropriate level of maltreatment dynamics review (MDR)—the collection and assessment of information related to the potential existence and dynamics of child maltreatment that formed the basis for opening a CPS case. Family assessments are used to help families and caseworkers understand the conditions that increase risk to the children and the changes needed to ensure children’s safety and well-being. Family assessment focuses on why maltreatment occurred in a family, and families are helped to explore their strengths, protective capacities, and the resources available to mitigate risk. An understanding of what happened, how it happened, and who was involved is essential for treatment planning purposes. For this reason, family assessment is necessary in nearly all cases opened by CPS, as is some level of maltreatment dynamics review.

Maltreatment dynamics review in suspected criminal cases of child maltreatment must be done in consideration of criminal court rules of evidence. This makes the process adversarial and requires forensic interviewing and forensic investigation techniques. There is little controversy about the need for MDR

to be in the form of forensic investigation for this small percentage of CPS cases.

For CPS cases not subject to criminal court jurisprudence, maltreatment dynamics review need not be as adversarial. If court involvement is needed, these cases will be under the purview of juvenile or family courts, where the rules of evidence do not necessitate the same adversarial relationship between CPS staff and families. The mission and procedures in these courts are more conducive to family and worker collaboration. The majority of CPS cases will fall into this second category. Development and evaluation of an interviewing and assessment protocol that can promote the discussion of maltreatment dynamics in a supportive, collaborative, nonadversarial manner would be an important addition to enhance family-centered DR strategies.

The most significant misconception in DR reform is that strategies of case fact-finding in CPS can be as easily bifurcated as assigning tracks. In truth, safety and risk assessment, comprehensive family assessment, maltreatment dynamics review, and other essential aspects of case fact-finding and data collection cross DR track assignments in nuanced and case-specific ways. This is not always acknowledged by those persons who promote or undertake DR reform. For example, not all families need to be investigated, but some level of maltreatment dynamics review is necessary for treatment planning in all families served by CPS.

With respect to Finding #4, Ellett suggests that the provision of additional services and resources to the AR track may explain the higher satisfaction ratings reported in some DR research. She contends that CPS employees assigned to “more functional” families in the AR track often have “lower caseloads allowing them time to become more deeply engaged with families than those working in more traditional response practices” and that “Caseload issues alone may explain why CPS employees report higher levels of job satisfaction when engaged in DR activities” (p. 3). The outstanding question is, if TR families had increased access to services, and if caseworkers had lower caseloads, affording them more time to become deeply engaged with families, would the satisfaction ratings from workers and families in TR be comparable to those reported in AR?

We agree with Ellett’s conclusion that DR reform should be “situated within the context of large scale system change” (p. 2) and that DR needs a “sound theory base . . . that can be used as an impetus for future DR practice models and research” (p. 3). Ellett proposes that, “perhaps the time has come to devise a research-based CPS model that combines the best practice elements of both DR and traditional response as we know them . . .” (p. 3). This is consistent with our conclusion that many of the reform objectives suggested for alternative tracks, such as a heightened emphasis on family assessment, have equal utility for traditional tracks, and some objectives of traditional tracks, such as comprehensive case fact-finding regarding maltreatment dynamics, should also apply to AR tracks. We agree that a more holistic approach to child welfare reform is warranted, with equal attention paid to strengthening TR by integrating many of the family-friendly,

strengths-based interventions being implemented with families in the AR tracks.

Perry

Perry expresses agreement with the findings and issues raised in our article, and he poses several questions that are worthy of more discussion. He also offers recommendations for specific lines of research that could strengthen the empirical base of DR. We choose to address two of the issues he raises.

Regarding Finding #1, the lack of a standardized DR model, Perry says he finds “little to disagree with.” However, he proposes other potential explanations for the inconsistencies we found among implementing jurisdictions. He offers that these observed differences in programming may be “a by-product of variations in expectations for DR that are encapsulated with state legislation . . .” (p. 4) and that “. . . form and function may vary from state to state given variations in expectations by policy makers . . .” (p. 4). We agree that these are plausible explanations for some of the differences we found among implementing jurisdictions, and it is likely there are other possibilities as well. We would expect individual state legislatures and policy makers to adapt DR programming to fit their unique needs and circumstances. We also understand that even when programs have well-defined components and clear implementation guidance, ensuring fidelity and sustaining reforms over time remain a major challenge. Therefore, we would always expect to see some level of difference among and within jurisdictions, irrespective of the level of program standardization. However, taking any new program to scale before it has been fully defined, developed, and tested is of major concern to us. A goal of implementation science is to codify essential elements of an established program or model to prevent idiosyncratic divergence from the model’s design, thereby potentially undermining the program’s effectiveness.

Many of the respondents agreed with us that a lack of program and practice consistency reduces the generalizability of research findings, yet they did not see this as a serious concern. We view this as a problem because potential DR users look to prior research to provide verification of the program’s integrity and viability and to provide data to legislatures and policy makers to support pursuing DR reform. Inaccurate or overreaching claims in the research can create unrealistic expectations, especially when potential users may lack the research experience or the objective frame of reference needed to mitigate overzealous claims of program effectiveness. This is not a hypothetical concern. We have frequently heard DR advocates in a variety of venues communicate that prior DR research provides strong and generalized empirical backing that allows states to pursue DR reform with confidence. This issue is what prompted us to initially question the appropriateness of including promotional claims or marketing language in both outcome research and DR program literature.

This lack of a well-developed and articulated program model, combined with enthusiastic and overreaching promotion and

consumer susceptibility, results in the relatively inefficient, idiosyncratic, and disparate developmental reform efforts that characterize DR reform across the country. It would be better to provide a well-defined and empirically supported model that includes policies, practices, training, and tools that can be modified when necessary by individual jurisdictions, with the flexibility to further adjust the model when new evidence becomes available. However, maintaining an evidence-based focus in programming requires educating legislators and policy makers that making significant modifications to an evidence-based practice model can threaten their capacity to achieve their desired outcomes.

Perry raises questions regarding Finding #3, the safety of children served in AR tracks. He agrees with our assessment of the many potential failings of screening processes and the need for well-trained screeners. However, he questions whether there is sufficient evidence to support what he interprets as our claims of heightened safety risk of children served in alternative tracks and says, "Even if select investigatory tasks are 'discouraged' by alternative track DR models, this is not tantamount to or evidence that caseworkers typically . . . ignore evidence or engage in behaviors that would . . . 'prevent a thorough assessment of risk and safety from occurring in alternative tracks' . . ." (p. 2). Regarding our statement that " . . . it isn't possible to conclude that a DR model exists that can assure that children's safety is not compromised in alternative tracks . . ." (p. 3), he says, " . . . alternatively it could be asserted that it is not possible (with the evidence thus far) to conclude that a DR model exists where a child's safety is compromised in alternative tracks. Thus far, there is no evidence suggesting that one track has a more measured (and sustained) impact (across a variety of child welfare systems) on minimizing the likelihood of a future occurrence of maltreatment than another" (p. 3).

Perry is accurate in his summation of the issue, but we never claimed that children in AR tracks were unsafe, as he asserted. From our perspective, the safety of children in DR reform is still an open question, in spite of the considerable outcome research that claims otherwise. In our article, we provide a rationale for our concern about the potential for harm to children as a result of overzealous implementation of a reform whose safety has not yet been appropriately established, but we agree with Perry that such arguments must not be construed as proof that they are unsafe. That said, in any child protection agency, the absence of a standardized, empirically tested, and fully implemented system to identify and respond to risk in families, from screening until case closure, potentially increases the risk of harm to children, regardless of track assignment. Considering what is known about the dynamics of child maltreatment in families, and considering that this was the impetus a decade ago to develop empirically supported risk and safety assessment tools, we believe this caution is legitimate.

Perry also offers several recommendations for further research to refine our understanding of the issues around DR practice implementation, and we recommend these be considered when formulating an ongoing national research agenda.

Winokur and Gabel

Winokur and Gabel's response is organized into three sections. They first confirm their agreement with four of the five primary findings in our article. They then critique the methodology we used to arrive at our findings and conclusions. In the last section, they describe Colorado's DR program and their own ongoing multisite research project, and they explain how this work will help move both DR research and practice forward. A portion of the third section explains how they designed their research to remedy many of the methodological issues we had identified in our article as problematic in previous DR research.

Regarding Finding #1, Winokur and Gabel state, "From a national perspective, we agree that there is no defined or consistently implemented DR practice model" (p. 1). They suggest that this problem will be remedied in their collaborative multisite project in Colorado, Ohio, and Illinois, because the project directors from the three involved sites will be "developing an implementation guide," and because the "evaluation directors are conducting fidelity assessments" (p. 4). They also intend to use administrative and survey data to determine whether "practices and principles that defined the Colorado DR model were adhered to by the five counties during the research and demonstration project" (p. 2). We interpret these attempts to promote model building and fidelity in implementation as substantial agreement with our concerns about the negative consequences of trying to implement or evaluate DR when the program lacks uniform and clearly defined definitions, policies, procedures, standards and implementation strategies.

The authors spent considerable time articulating agreement with Finding #2 related to the lack of rigor in prior DR research, and the negative consequences of methodological problems on the validity of research conclusions. Early in their response, Winokur and Gabel claim, "Based on our understanding of the [DR research] literature, we too are concerned with insufficient control for differences between groups, the lack of standardized instruments to measure outcomes, and the challenge in generalizing findings because of wide variations in settings and practice models" (p. 1). They also agree with "the need to statistically control for pre-experimental differences between groups" (p. 1) and are "troubled that validated instruments used to measure such constructs as engagement and satisfaction are not widely available" (p. 1). In their description of their current cross-site evaluation project, they acknowledge that they " . . . have worked with the QIC-DR to address some of the limitations of past research on DR . . ." (p. 2). They describe their strategy to address how variables other than DR could potentially be responsible for observed outcomes and explain their methodology to statistically control for differences in caseworker characteristics as a means of controlling threats to validity from worker's self-selection into AR and TR tracks. They also echo our concern that inequities in training provided to caseworkers in AR and TR tracks may have contributed to lack of comparison group equivalence. They also reaffirm our concern about the lack of use of validated instruments, and they recommend "instrument development

and validation studies to better measure key DR outcomes including engagement, well-being, and safety” (p. 4). They support our concern about the lack of attention paid in previous DR research to the reasons families may not have responded to surveys, calling it a glaring “oversight” (p. 4) which, they explain, is addressed in their research through “nonresponse bias testing and weighting of the family and caseworker surveys” (p. 4).

Winokur and Gabel did not spend much time discussing our Finding #4 but did point out that Colorado’s DR program did not make the mistake of disproportionately allocating services and resources to families in AR, since in Colorado “an increase in the amount of community-based services and financial support available through the grant is available to families in both the family assessment response (FAR) and investigation response (IR) tracks” (p. 2).

Regarding Finding #5, the authors implied that they agreed with our concern about the inappropriateness of marketing and promotion in DR research. They state that, “The QIC-DR has engaged an independent evaluation firm to conduct the cross-site evaluation” (we presume they are referring to themselves), and “As independent evaluators working with the QIC-DR on the evaluation [of DR], we are sensitive to these charges and have demonstrated our objectivity by stressing rigor and transparency in all aspects of the research enterprise” (p. 2).

Bauer (2001), Barkley et al. (2002), and Gambrill (2011, 2010) all argue that close collaboration between powerful promoters of ideas, and itinerate researchers with vetted ideologies, often creates a conflict of interest, resulting in the creation of knowledge monopolies and research cartels in which dissenting opinions and problematic issues are censured. Winokur and Gabel recognize that engagement of an independent evaluation firm to conduct the cross-site evaluation is a means to guard against these possibilities.

Winokur and Gabel did not comment on Finding #3, our concern about the lack of objective evidence supporting claims about the safety of children served in AR tracks. We find this puzzling, considering the importance of child safety to all DR implementation and research concerns, and the fact that the authors have focused in considerable detail on most other primary aspects of DR practice in their own research work.

In light of their fundamental agreement with our findings, we are confounded by the intensity of their criticisms of our methodology. Although we are gratified with their acknowledgment that we largely got it right, we would suggest that it was not just good luck or prescience but a thorough analysis of DR research reports and program literature that formed the basis of our findings. We find many of Winokur and Gabel’s criticisms to be exaggerated and unwarranted, particularly as they relate to our key informant survey. Some of their critique is based not on what was wrong with the survey but on what it *wasn’t* and *should have been*. In several places, they imply that the survey could have been more rigorous, exhaustive, and in depth. They challenged our decision to “not attempt any quantification of interview responses,” calling it a “red flag” (p. 2) and an open door to “cherry-picking responses or using one response as ‘evidence’”

(p. 2). They criticized the “mismatch between the length of the interview protocol and the time allotted for the interviews,” and they claimed it would have required “several hours, to properly complete” (p. 2).

Well . . . yes, we potentially could have done the key informant interviews in far greater depth, but their primary purpose was to assess whether and to what degree DR was being defined, organized, and implemented consistently among jurisdictions. It was not difficult in 1-hr interviews to identify wide differences in DR programs around the country. Interestingly, none of the nine respondents to our article contested the accuracy of this finding.

In contrast, we evaluated each of the 18 research reports independently and thoroughly, using well-defined and standardized evaluation criteria. Space limitations prevented us from appending the full analysis of each research study to our article, although we did provide a complete summary of the major issues found in each study in the appendices. The conclusions in our article were derived from these data, and we have myriad examples to support our concerns. Winokur and Gabel are simply wrong when they assert that we “simply did a literature review” (p. 2).

If the key informant survey had been the only or even the most important component of our study, it might have been worthy of Winokur and Gabel’s conspicuous and exhaustive focus of critique. Because the key informant survey data formed such a small part of our study, their singular focus on its methodological simplicity is curious.

We are confident that our findings are sound, and we are gratified that Winokur and Gabel agreed with their relevance and importance, even as they criticized the methodology we used to arrive at them. We do agree with the authors that further research in the depth and scope they recommend, including, at some point, a systematic review and meta-analysis, would certainly be a valuable addition in the ongoing evaluation of DR.

Baird, Park, and Lohrbach

Baird, Park, and Lohrbach represent the Children’s Research Center (CRC) of the National Council on Crime and Delinquency, and they provide a perspective based on two decades of CRC research on topics and tools to ensure the safety of children in public child protection and juvenile justice systems. Developers of the Structured Decision Making™ (SDM) assessment and case management model, they communicate their conviction that case decisions must be made based on data from vigorously tested and validated assessment tools and strategies. Considering their organizational history, it is not surprising that their comments reference the role of continuing research to support or refute some of DR’s fundamental underlying assumptions and to provide a sound empirical base to underpin DR’s future development. Their closing statement summarizes their general position related to DR: “In an era of evidence-based practice, no program should survive and flourish simply because it is viewed as a good idea” (p. 4).

The authors did not address two of our five findings in their response—Finding #1 related to the lack of a standardized model

for DR, or Finding #5, related to misrepresentation of the TR track to enhance the AR track. They did communicate their agreement with Finding #2, stating that, “claims of DR success go beyond what outcome data legitimately support” (p. 2). One of their primary criticisms of past DR research is that both funding and services were provided to families in the AR groups but were not provided to families in the TR groups. They contend that, “. . . instead of a level playing field on which the central premise of DR could be evaluated, DR programs were given resources not available to TR programs. Not only would these added services potentially affect outcomes, they would affect results of surveys of client and staff satisfaction as well” (p. 2), thereby making suspect claims related to the success of the DR program.

As might be expected, much of their commentary was related to Finding #3, the safety of children served in DR programs. Their primary issue echoed our concern about the challenges of accurately identifying which families belonged in which track, when standardized risk or safety assessment tools were not used. They stated, “The various methods used to determine which families are eligible for DR open the door for myriad unintended consequences” (p. 2), including the potential to overlook critical safety concerns in some families. They cited data from research they had conducted in California, demonstrating that 30.7% of the families identified as low risk had a safety concern that required an in-home safety plan; and in 2.1%, the safety concerns were at such a level that out-of-home placement was required. This provides further confirmation that vigilance regarding children’s safety is equally important for children served in AR tracks and that workers must ask about circumstances that prompted child protective services’ involvement.

We found Baird et al.’s discussion about Finding #4, the diversion of resources from higher risk to lower risk families, quite compelling. The authors report that in their risk assessment studies, researchers found little evidence that providing additional services to low-risk families reduced future maltreatment, and that reducing services to low-risk families had no detrimental impact. They also found little evidence that most low-risk families progressed to higher risk levels over time (Children’s Research Center, 1998; Johnson, Wagner, & Scharenbroch, 2007). By contrast, their research did demonstrate that targeting additional resources to high-risk families could significantly reduce rates of subsequent child abuse and neglect. They concluded that although “early intervention may prevent some future maltreatment, available research indicates using scarce resources to assist high-risk families has a far greater effect” (p. 2). These are important findings, because in our opinion, this is the most rigorous research available on this issue. We agree that this data should inform the refinement of a DR model in which planning for the use of CPS resources should focus on strategies that best meet the needs of families served in both AR and TR tracks rather than focusing solely or primarily on enhancing services to families served in AR. The authors recommend additional research to “determine if a change in approaches is warranted, or if DR

represents a misallocation of resources better allocated to high-risk cases” (p. 4).

Baird et al. propose additional research to evaluate the validity of two claims seen in the DR literature. The first is that as currently implemented, investigative approaches to child protection are “too adversarial” and “makes family engagement and service provision difficult” (p. 2), and further, that “. . . a less adversarial approach concentrating on needs rather than maltreatment allegations will more successfully engage families without negatively affecting child safety.” Baird et al. claim that this concept is “eminently testable” (p. 2). Because this premise is foundational in the philosophy that underlies the development of alternative tracks, we concur that developing a strong empirical base to support or refute this contention should be a paramount concern of policy makers, programmers, and funders, since so much of the current program philosophy and structure is based on this assumption.

We concur with the authors’ belief that DR reform offers rich opportunities to test practice hypotheses and intervention models and to determine which have the best potential to achieve our goals of child safety and family stability. As one desired outcome of our study, we had hoped to promote creation of a long-term, systematically implemented research agenda, whereby the unanswered questions that have plagued child protective services for decades could be strategically evaluated. This would be consistent with Ellett’s vision of a comprehensive and holistic model of child welfare interventions, offering a continuum of intervention strategies, all of which have been built, shaped, and refined based on evidence. Although such a large and coordinated research effort would be painstaking and probably very costly, we believe it would be a better use of available research dollars than each DR intervention site financing its own individual program evaluation. The added value to researchers building their projects collaboratively, using consistent and agreed-upon methodologies, and addressing common questions is that data could then be compared across studies and, ultimately, examined through a systematic review or meta-analysis (as was also recommended by Winokur and Gabel), thereby providing the most accurate representation possible of what really works in child protection. Baird et al.’s response is strongly in favor of building DR as an evidence-based or evidence-supported treatment, contrary to the position expressed by some of the other respondents who believe that evaluating DR as if it were an evidence-based or evidence-supported treatment is neither useful nor fair. We concur with Baird and colleagues on this point.

Drake

Drake contends that DR “is not a specific intervention,” that DR programs show such “massive variability” that DR cannot be evaluated because “there is no such thing as DR . . . ” (p. 2), and that empirically determined conclusions about DR programs “. . . might not tell us much about a different DR program in a different state” (p. 2). He contends that the “single similarity between all DR programs is that there are always at least two

‘tracks’” (p. 1), but then states that it is not possible to generalize even what “tracks” means, as the differences include, “. . . when the ‘track assignment’ is made, on what criteria it is made, by whom it is made, (and) how each track is staffed . . . ” (p. 1). He also states that the alternative track is no more “voluntary” than traditional practice (p. 2), even though this is a frequently used argument in describing the uniqueness of alternative tracks. Drake summarizes that given this state of affairs, it is not possible to determine whether “DR is safe” (p. 2) because the program is so “chimeric” in nature that there really is nothing to measure. He states that DR is best defined as a “policy orientation” (p. 1) but indicates that even this minimal definition of DR is of limited utility, since the policy being implemented in a number of different states and counties is only “loosely similar” (p. 1). We think it is safe to say that Drake agrees with Finding #1 that DR programs are not implemented consistently across sites and do not adhere to uniform standards. Yet, Drake states that these conclusions in our article are unnecessarily harsh and trivially true. Drake evidently believes that everyone knows and accepts this fact, that DR proponents have not communicated otherwise, and that states are not promoting DR as a distinct, substantive, and empirically strong reform. But, in fact, DR has been strongly promoted as all of the above, and thus we undertook our study to gather data to determine what DR is and is not. We certainly did not attempt to be harsh in our commentary but understand that reality can sometimes be so.

In his discussion of research to assess child safety, Drake contends that we suggested that the DR (experimental) and the TR (control) subjects should have been “*randomly selected from among all reports*” (p. 3), and he takes “strong issue” with the “ethics” of this recommendation (p. 3). We take equally “strong issue” with Drake’s incorrect assumption regarding our discussion of random assignment and random selection in research methodology design. It is clear from a thorough reading of our discussion that we made no such suggestion. We pointed out that there is a difference between random selection and random assignment and that readers should be aware of these differences when interpreting the validity of intergroup data regarding recurrence rates. Given Drake’s professed concern with “harsh” tone in article critiques, we are puzzled with his implication of a disregard of ethics on our part.

In several places in his response, Drake defends existing DR research in its evaluation of the safety of children in DR programs. He first states, “I believe that we have a convincing body of data demonstrating that DR does not put children at higher risk” (p. 3). Later he claims, “. . . in my view . . . the existing research is sufficiently strong in quality and quantity to show that child safety is not being degraded” (p. 5). However, one paragraph later, Drake acknowledges that a portion of the same research he found so “convincing” should be replicated, “. . . preferably by an independent research team . . . ” because of faulty design, in which the same researchers provided additional money for services to families in the DR track (experimental variable) but not in the TR track (control variable), which likely contaminated their findings. For this reason and others that we made clear in our article, we cannot share Drake’s “belief” in

a “convincing body of data demonstrating that DR does not put children at higher risk” (p. 3).

Regarding Finding #5, Drake agrees with our contention that DR literature often criticizes and misrepresents traditional CPS, but he states that this is not unique to DR. He claims that CPS are “. . . almost universally believed to be . . . offensively intrusive and unhelpful by clients and even other professionals, when this is not, in fact, the case” (p. 5). He then questions whether we were “implying intentional falsehood” (p. 5) by pointing out this phenomena in the DR research and the program literature. Drake states that since this misrepresentation of CPS is so endemic, why would we point it out so dramatically relative to DR? We do not believe that just because misrepresentation of CPS is common that it is inappropriate to point it out in the DR research and literature. We have no opinions regarding anyone’s motives, or lack thereof, for promulgating this misrepresentation. Our intent was to educate consumers of the DR literature to the existence of this dynamic so they could take it into account in their analysis of DR policy and programming.

Drake states that “something like a traditional ‘investigation’ should be a necessary part of DR practice” (p. 4), because a “comprehensive risk assessment” is a “necessary part of either a DR track or traditional services” (p. 4). He then suggests that more study may be needed to determine whether understanding past events in a family is important in DR cases. Drake apparently shares our concern about DR’s assertion that fact-finding regarding maltreatment dynamics may be unnecessary, and he apparently also agrees that failure to do appropriate fact-finding to inform risk assessment potentially compromises child safety. We refer Drake (and readers) to the large body of research on actuarial risk assessment that identifies prior maltreatment as the single most highly correlated factor with future maltreatment.

Drake finds fault with our criticism of DR researchers who fail to acknowledge that high-risk cases are more likely to recidivate than low-risk cases, and, therefore, a finding of similar recidivism rates between higher risk TR cases and lower risk AR cases should not be interpreted as a positive finding. Drake cites research that has found similar recidivism rates between substantiated and unsubstantiated cases and states that, therefore, we should not expect to see different base rates of recidivism for higher and lower risk cases. We would first point out that other researchers have found higher recidivism rates for substantiated versus unsubstantiated cases (Fuller & Nieto, 2009). But, that is beside the point, as we are talking about high-risk and low-risk cases, not substantiated and unsubstantiated cases. If Drake’s unlikely premise were true, it would be fatal to all the DR outcome research we analyzed. The assumption that low-risk and high-risk cases would tend to recidivate at lower and higher rates, respectively, was a universal and basic premise underlying all DR outcome research.

Drake implies that our criticism of DR’s lack of progress toward becoming an evidence-based model is unwarranted. He claims DR “is not a specific intervention” (p. 1), and, therefore, it is foolish to think it will ever be an evidence-based program. He states that the “variability between DR programs is not likely

to change,” because of “obdurate differences between locales . . . union contracts, geographic considerations,” and other barriers to standardized program development and implementation (p. 2). Drake further points out that program research, like that which we reviewed for our study, does not go through the same peer-review process as academic research, and, apparently, we should not expect the same level of research quality in evaluations of DR programs. Drake also cautions that while DR research may, in fact, have some of the problems and issues we report, it is no worse than in other areas of child welfare research, and he suggests we should use a better “sense of proportion” (p. 2) and a less “negative tone” (p. 2) in our review—sort of like grading DR programming on a sliding scale rather than doing an objective measurement using empirical criteria.

All things considered, we found Drake’s response more notable for its intensity than its relevance. His Panglossian apology for the existing barriers that prevent DR from progressive empirical development toward an evidence-based model of reform is, in our opinion, not only epistemologically short sighted but also discounts the intellectual and moral commitment of child welfare agencies in this country, and the potential for leadership to ensure that child welfare practice is evidence based. We believe much more can and should be done to realize the “promise” of DR’s “promising practice” status.

Fluke, Merkel-Holguin, and Schene

Fluke, Merkel-Holguin, and Schene have been involved in the evolution of DR as staff members of the American Humane Association and the National Quality Improvement Center on Differential Response (QIC-DR), funded by the The United States Department of Health and Human Services (USDHHS)—Administration on Children, Youth and Families (ACYF) Children’s Bureau. The QIC-DR is currently located at the Kempe Center at the University of Colorado, and two of the authors are still affiliated with the Center. We reviewed several publications by these authors during our study.

Fluke et al. define DR as a “method to restructure the CPS system to have multiple ways to respond to accepted (or screened in) reports of child maltreatment” (p. 1) rather than, as some states claim, a program or a practice. Fluke et al. join other respondents in arguing that DR lacks the strong empirical support that is typical of evidence-based programs, but Fluke et al. contend that DR was never intended to be a standardized intervention, and they question why we made the effort to document DR’s lack of empirical support.

First, as we have documented throughout our article, there are many who do claim that DR is or should be an evidence-based practice, program, or model. The California Evidence-Based Clearinghouse for Child Welfare (CEBC4CW) defines the highest level of evidence-based practice as that which is “well supported by research” (<http://www.cebc4cw.org/ratings>). We raised the question of DR’s empirical support in our article because of the ubiquitous claims in the DR literature, by DR advocates, and subsequently by states promoting it, that DR is “well supported by research.” Although it may be no

surprise to Fluke et al. that DR is not an evidence-based program, there are many in the field of child welfare for whom this will be informative.

Second, we did not measure claims of strong empirical support in the DR literature against some high standard of empirical legitimacy. Our intent was simply to identify the empirical support, or lack of it, in response to pervasive claims in the DR literature that DR was a practice that had strong empirical support. It is interesting that Fluke et al. would be the third respondent to claim that we inappropriately characterized DR as a model, because such a belief is unsupported. It would be a clever “straw man” indeed, if so constructed that your challenger tears it down for you.

Additionally, we are uncertain how the term “method” differs substantially from terms such as “program” or “practice.” All of these terms are nonspecific and fairly synonymous and equally subject to dissembling rhetoric. Focusing on the labels begs the question of what empirical evidence exists to justify DR reform, whatever its descriptive eponym. We are OK calling it a method rather than a program or practice, forestalling claims that strong empirical support was ever a DR developmental or research objective, as long as it is clear that such strong empirical support does not exist. We therefore accept the authors’ demotion of DR from a “practice” that is well supported by research to a “method” that is not expected to have such empirical support. If these characterizations had been more accurately communicated to state governments, child welfare organizations, and the public with half the intensity of the mischaracterizations and overestimations of empirical support for DR, our article would not have been necessary.

Later in their response, Fluke et al. seem to reverse their characterization of DR as a method by saying, “In our experience, DR has propelled or created the need for the implementation of *more uniform* (our italics) micro-child welfare practices, including enhanced screening protocols and screening decisions, safety organized practice, team decisions, and the use of family meetings . . .” (p. 2). This appears to affirm the need and efficacy of model building, and for incorporating more “uniform” practices into the day-to-day work of CPS. Interestingly, the authors identify the very same practices that we have noted throughout our article as being essential for effective CPS practice, particularly to ensure child safety in AR tracks. We fully concur that the absence of uniformity in these particular practices poses the greatest potential danger to children. Whether DR is a model, itself, or a template for a model that includes a compilation of integrated “micro” evidence-based practices, is a moot point. We interpret Fluke et al. as ultimately concurring that model building is necessary for DR to continue progressing toward its potential.

Fluke et al. cite federal government latitude as allowing variability in DR implementation across states. Their claim of inconsistency in maltreatment investigation in states, and the lack of consistent standards and programs in traditional CPS across the country, may or may not be an accurate account of the state of affairs. But even if true, it does not justify a similar inconsistency, lax oversight, or lack of standardization of DR.

We should be advocates of evidence-based practice in all aspects of child welfare practice. Claiming that DR is no more troubling than other areas of practice, or that the federal government's oversight allows this, provides little confidence in the practice, comfort for those served, or justification for continuing marginal practices because they have a lot of company.

Fluke et al. state their agreement with the common AR contention that case fact-finding regarding maltreatment dynamics sets up "barriers to successful engagement" (p. 3) and may not be necessary. Eliminating case fact-finding because it can set up potential barriers to engagement begs the question of whether case fact-finding regarding maltreatment dynamics is essential for CPS cases. The essential philosophical question of DR reform—one which is not often addressed openly and transparently—is whether it is a necessary responsibility of child protective services to ask the relevant questions to fully assess risk and ensure children's safety. We disagree with the respondents who claim that asking these targeted questions will set up insurmountable barriers to engagement by social work practitioners. We believe case fact-finding is essential to meeting CPS' responsibility for child safety, and our position is that the DR research on child safety has not proven otherwise. Moreover, even if research found that eliminating fact-finding about maltreatment dynamics in AR tracks did not compromise safety when compared to traditional practice investigations, we would be reluctant to assume that, therefore, everyone is "safe." It would be just as likely that nobody is. If DR researchers really believe that TR does not ensure children's safety, as several respondents have claimed, then what is the point of a finding that children in AR tracks are no less safe than children in TR tracks?

It is also important not to confuse our recommendation for thorough case fact-finding regarding maltreatment dynamics in AR cases with the issue of whether substantiation is necessary in AR tracks. Identifying the fact of maltreatment and the identity of a perpetrator may be important information for safety planning, case planning, or intervention, but whether this information is used for formal substantiation and state registry is a moral and political issue, and there are strong rational arguments for limiting its use. However, we are promoting fact-finding, not fault finding. Drake, another respondent who questioned the need for assessing maltreatment dynamics for families served in AR tracks, completed research to identify factors that influenced recidivism in unsubstantiated cases. Researchers have determined that factors related to the perpetrator, such as relationship to the victim, age of the perpetrator, and financial resources available to the family affected recidivism rates (Child Welfare Information Gateway, 2003). This is important information for safety and case planning. How would Fluke, Drake, and others obtain this essential information without identifying maltreatment dynamics, including the identity of the perpetrator?

Many of the issues seen in families served by CPS agencies may be uncomfortable or challenging for families to address. Alcohol and drug abuse is estimated to be an issue in a majority of the families we serve. These concerns often greatly increase risk to children. Should the issue be ignored because it may be

uncomfortable for families to discuss? Effective completion of safety plans and case plans and provision of the most relevant service interventions are all dependent on this information. With proper social work intervention, addressing these issues can be a means to engage families in exploring their own needs, strengths, and protective capacities. We believe it is condescending, dishonest, and disrespectful to assume that most of the families we serve are unable to understand, accept, and address issues that threaten their children's safety and their family's well-being, in spite of the discomfort inherent in such discussion. We also believe such patronizing approaches discount the social work profession's capacity and success in engaging families with difficult histories.

Fluke et al. contend that whether thorough case fact-finding "enhances or diminishes the likelihood of successful engagement remains an open question from a research perspective" (p. 3). We agree.

Fluke et al. state that "contrary to Hughes et al.'s suggestion, CPS systems typically have clear policy and procedures for assessing risk and safety in both TR and AR cases." We are aware that many jurisdictions have policies requiring risk and safety assessment. What we found, however, is that only about half of our key informants indicated they used standardized protocols to do these assessments and even fewer reported use of empirically supported tools to make these assessments. For all the reasons discussed in our policy white paper, *Issues in Risk Assessment in Child Protective Services*, (Rycus & Hughes, 2003) risk and safety ratings based on professional judgment or using consensus-based instruments are significantly lacking in their capacity to identify risk levels with any statistical reliability or validity. This is a serious issue that can and should be addressed.

Fluke et al. state, "the notion that DR or any other systems level reform should not be implemented until the highest standard of research is conducted is unrealistic." It is not clear from this response whether Fluke et al. are suggesting that we believe DR reform should wait until better research has been done regarding safety and effectiveness. This is not our position, as we have indicated both in our article and in our response to Vaughan-Eden and Vandervort. Rather, we are saying that 10 years of opportunity to do better and more coordinated research regarding safety and effectiveness has been missed, and it is past time to correct this.

In their concluding section, Fluke et al. state that they "... do not necessarily share a common perspective regarding the efficacy of DR or the status of the research base" (p. 4). We appreciate their candor. Their statement further validates our observation about large discrepancies in professionals' beliefs and opinions related to DR and underscores the importance of continuing to negotiate and build consensus among DR leaders and advocates on fundamental issues. This is essential if we are to avoid continuing polarization and resulting inertia. We hope the dialog begun in this issue of RSWP can be a constructive step for such a process.

Fluke et al. also make several recommendations for areas in need of research to improve DR reform. These recommendations are well conceived, and we hope the authors are in a position to

facilitate these research initiatives. They conclude their response by citing the California Evidence-Based Clearinghouse for Child Welfare's designation of DR as a promising practice, but they also declare that DR's progress toward an evidence-based practice will be "daunting," and they express ambivalence regarding whether attempting development of an evidence-based program model would be desirable. We hope that with further consideration, they may provide leadership in moving DR toward an evidence-based model of practice.

Vaughan-Eden and Vandervort

Vaughan-Eden and Vandervort raise many questions about the fundamental integrity of DR in their response. Among their concerns is the lack of "objective empirical evidence supporting [its] efficacy" and literature that is "deeply flawed methodologically" (p. 1). In their opening paragraph, the authors state, "... in an era of evidence-based practice, why has a program with so little empirical support been so widely and aggressively utilized by the nation's child welfare systems?" (p. 1).

In an attempt to answer their own question, the authors suggest that advocates have come to regard DR as a "panacea" (p. 1), similar to the dynamics surrounding the family preservation movement 20 years ago. The authors agree that family preservation is an effective practice model for some families, but they suggest it was "utilized far beyond what research and reason would suggest was appropriate" (p. 2), and "... many children were seriously harmed ... and many died" (p. 2) as a result. Vaughan-Eden and Vandervort state that marketing a program model, such as DR, without sufficient research evidence to support its claims can "place children in harm's way because such programs have an unrealistic belief in their own effectiveness—an unrealistic belief that has been fed and continues to be fed by ... over marketing ... " (p. 3).

The authors support their thesis by describing the legal and political environments in which previous child welfare reform efforts have been implemented, and they suggest that DR reform reflects "history repeating itself" (p. 2). They place DR within a framework of society's historic ambivalence about the role of public child protection, particularly its competing priorities of protecting children from maltreatment by their parents or caregivers, and of supporting parents' rights to raise their children in a manner consistent with their values and culture, without societal interference.

Vaughan-Eden and Vandervort describe a metaphorical "swinging pendulum," which they believe has characterized the child protection field since its inception. They explain how CPS is initially driven by societal demands to intervene to ensure the safety of maltreated and at-risk children, even when against parents' wishes; then, in response to perceived violations of family rights, society demands that CPS be less intrusive, more engaging, and more collaborative in addressing each family's needs. Then, when the use of protective authority has been inappropriately set aside in favor of engaging and partnering with families, and children are seriously

hurt or die from abuse or neglect, our laws, policies, and direct practice approaches are re-created once again to focus primarily on children's safety.

In our opinion, the swinging pendulum is a real phenomenon in child welfare practice, resulting in part from a lack of understanding among policy makers and advocates of the equally compelling commitments to both children and their families that must characterize all child protection work. This dual responsibility cannot be resolved by overemphasizing one or the other horns of this dilemma, which Vaughan-Eden and Vandervort state characterizes DR reform. The solution is to design and implement practice models that ensure concurrent attention to both child safety and family support, which requires primary reliance on social work methods to engage and empower families, but which preserves the use of protective authority to ensure children's safety when working collaboratively and voluntarily with families cannot.

The basic philosophy of DR reform, in contrast to its marketing, rhetoric, and implementation history, appears to acknowledge the necessity of balancing these service objectives. DR was founded on the premise that different families have different needs and require different interventions and that the least intrusive intervention that can achieve case objectives should be applied in any individual family. We have no issue with this guiding principle of DR. A less intrusive approach to working with families when children are not at high risk of serious harm is entirely consistent with the values of a family-centered approach to child protection. Further, the types of family-empowering service interventions promoted by DR reform are appropriate for most of the families served by CPS agencies, including many families served in TR tracks, and their widespread dissemination is in the best interests of all families served by the system. It is within this context that we disagree with Vaughan-Eden and Vandervort's concluding recommendation, that, "Policy makers and child welfare staff on the front lines should impose a moratorium on. [DR's] use until the program is better defined and its utility is rigorously and honestly studied. Failure to do so runs too high a risk of violating the fundamental commitment of child welfare practice: keeping children safe" (p. 4). DR reform has been quite successful in engaging child welfare organizations to re-engineer their infrastructures to respond more individually to families with different needs, thus overcoming long-standing inertia and opening the door to more constructive family-centered practice reform over time. Our hope is that a practice model will emerge that effectively supports the integration of practice strategies to ensure children's safety while concurrently supporting and sustaining their families.

Ensuring the safety of children served in AR tracks could be largely addressed by following the lead of DR states like Ohio and Minnesota that require the use of empirically supported decision-making protocols, such as Structured Decision MakingTM, with *every family* served by the agency. This ensures that contributors to risk are monitored and that safety planning is implemented whenever called for, regardless of track assignment, and throughout the life of every case. To

improve the validity of track assignments, we still contend that the child welfare field should adopt a standardized screening protocol to help screeners elicit accurate and relevant information on risk factors and safety concerns from the very first contact with a referent. These tools are available—they need to be piloted, integrated, and more widely adopted. To assist in fact-finding during family assessments, we would recommend a standardized interview and assessment protocol (with training to use it properly) to enable caseworkers to involve families in collaborative, respectful, and supportive discussions about risk factors and the protective capacities available to mitigate them. Perhaps if workers were better armed with skills to engage families in challenging discussions in a family-friendly manner, they would not feel the need to avoid fact-finding in attempts to preserve a family's trust. Identifying, developing, piloting, and evaluating such tools across a wide spectrum of DR users could greatly enhance the model building we believe to be so essential to DR's effectiveness, while simultaneously reducing the possibility of harm to children as DR is being further developed and refined.

That said, we strongly concur with Vaughan-Eden and Vandervort that DR must be a well-researched, evidence-based intervention. We also agree with these authors that providing extensive training and coaching to staff and supervisors in both AR and TR tracks is essential for any child protection reform to work. Having devoted our professional careers to child welfare education and training, we fully agree that the knowledge and skills needed to successfully negotiate the dual responsibilities of child safety and family preservation are not easily attained. Effective practice requires very skilled professionals, and our current child welfare systems are not always set up to develop their staff to this level of practice competence. This reopens long-standing workforce development concerns, particularly the lack of skilled social work professionals interested in making a career commitment to the child welfare field. However, there is no simple infrastructure fix to the inherent complexity of CPS practice.

Loman and Siegel

Loman and Siegel state that one "troubling matter" is "the authors' lack of clarity of what DR is" (p. 1). We plead guilty. Many others in the field of child welfare, including many of the other respondents to our article, share our guilt. We are distressed that after more than a decade of DR reform and research, there is so little consensus about what DR really is. It may be as Drake suggests, in spite of Loman and Siegel's contention otherwise, that "there is no such thing as "DR" (Drake, p. 2), and we are stuck with "massive variability" (Drake, p. 2) in programs that claim the eponym, differential response.

Loman and Siegel claim that we make "sweeping statements" . . . and "provide so little documentation to support their pronouncements." This is followed by a selected example of our sweeping statements: Considerable DR literature discounts the need for case fact-finding in the AR track regarding child maltreatment dynamics. In fact, this sweeping statement

is well documented in our article and is not disputed by many DR advocates who indicate that such fact-finding interferes with family engagement and is unnecessary to ensure children's safety or to effectively serve families. Loman and Siegel respond to this sweeping statement by saying that we failed to provide evidence that adequate case fact-finding was being done in traditional response investigations. What? This is an example of a debate tactic referred to as the *Tu quoque* (you too) fallacy, in which one defends an error in one's own reasoning by attempting to show that the opponent has made a similar, though unrelated error, rather than addressing the issue under discussion. Because poor case fact-finding may exist in traditional child welfare practice, Loman and Siegel appear to imply that we have no right or responsibility to question DR's claims regarding child well-being and safety. We believe that all aspects of child welfare practice could greatly benefit from legitimate empirical outcome research, but just because ineffective practice may exist in one area does not justify our ignoring it in another.

Loman and Siegel contend that we do not understand safety and risk and are apparently ignorant of Ohio's utilization of empirically supported risk and safety assessments in all CPS cases. In fact, the Institute for Human Services (IHS) was one of the early proponents of actuarial risk assessment. We have authored many publications on the topic (Hughes & Rycus, 2007; Rycus & Hughes, 2004, 2008, 2012), including a policy white paper entitled, *Issues in Risk Assessment* (Rycus & Hughes, 2003), which are among the most widely cited guides to the utilization of risk assessment in child welfare. In our opinion, it is Loman and Siegel who show the same fundamental misunderstanding of risk and safety assessment that our policy paper, *Issues in Risk Assessment*, addressed a decade ago. Regardless, our DR article was not an evaluation of Ohio's DR program.

Loman and Siegel either totally misunderstand our claim regarding bias and promotion in child welfare outcome research, or they misconstrue the issue. They state, "we are not marketing ourselves in any way." Whether or not they, or anyone else, market themselves or their services was never an issue or contention in our article. Our concern was the marketing of DR in what is supposed to be objective and disinterested outcome research. We documented in the appendices of our article many examples of this kind of promotion in the DR research we reviewed for our article. We are not, nor have we ever been, concerned with any researcher's motives with respect to this widespread and endemic problem in outcome research. Bias in research has many possible sources, including advocacy, ideology, cognitive style, intellectual capacity, politics, values, and self-interest. It can be conscious and intentional or unconscious and unintentional when a researcher is earnestly striving for accuracy (MacCoun, 1998). Our intent was to alert consumers of the research, including states and counties considering adoption of DR, to be more aware of marketing and promotion in research findings and to understand how it is fatal to the goals and objectives of such research. Loman and Siegel make this into another *ad hominem* attack on IHS by stating that they

“do not have a marketing communications coordinator on staff, as does the company of the authors” (p. 4). Although we have no issue with marketing and communications, in our 35-year history we have never had staff with marketing and advertising responsibility, much less a marketing and communications director. However, this does provide yet another example of how inadequate research and biased interpretation can be so misinforming to readers. Moreover, our concern was with allowing promotional statements into outcome research, not with marketing in general. We stand by our conclusions that this has been a significant problem in DR research.

Loman and Siegel state we have a “fundamental ignorance” of their outcome studies because we did a “not-very-close reading” of the research reports we analyzed (p. 4). This is not the case. Each of the 18 research studies included in our review was thoroughly assessed using standardized criteria. Because of the sheer scope and length of the research reports, totaling several thousand pages, it was not possible to provide a detailed commentary in our article of our analysis of each of the studies. We did, however, include a lengthy appendix with a more detailed description of the basis for our conclusions regarding each of the research reports. This was done so that interested researchers could obtain and analyze these same reports and compare their conclusions to ours.

We have selected one example here to illustrate the type of exhaustive analysis we completed of these reports. We chose Loman and Siegel’s Minnesota study (Institute of Applied Research, 2004), because it influenced subsequent implementation and evaluation efforts conducted in other jurisdictions, because it is generally considered one of the stronger evaluations because of its use of random assignment to experimental and comparison groups, and because nearly all of the source data provided below can be found within a few pages of the Minnesota report, allowing readers to easily review the material and draw their own conclusions.

In our analysis of this study, we encountered inconsistencies in the data presented, problems with the explanations the researchers adopted to explain differences between experimental and comparison groups, and, consequently, the conclusions drawn from researchers’ analysis. In nearly every instance, these methodological errors produced results favoring the alternative response group.

An initial problem was the lack of comparability of the experimental and control samples. The researchers randomly assigned families to comparison groups from a sample of cases deemed eligible for the AR track. After comparison group assignments had been made, it was discovered that 170 cases placed in the experimental group were not actually eligible for AR placement, and they were removed from the analysis. Given the fact that ineligible cases were found in the experimental group, it should be assumed that a similar number of ineligible cases could be present in the comparison (TR) group as well. They should have been identified and removed, but there is no indication in the report that this was done. Although 170 cases represent a small number of families in the experimental group, the reported differences in

recurrence rates for the two groups were also small—about 3%. This factor alone could have accounted for a significant proportion of the reported improvement in outcomes for the AR group.

More importantly, the TR (control) group appears to have had a substantially higher risk profile than the AR group. The authors did not present breakdowns by risk levels, but they did provide a chart indicating how cases in each group scored on each risk factor, and these data clearly illustrate that the average case in the TR control group presented a greater risk than the average family assigned to the experimental AR group. Therefore, it could be expected that more control than experimental group cases would be reported for maltreatment in the future. The authors initially intended to use risk levels as a control when reporting results, but the substantial difference in risk profiles created a conundrum. Their answer was as follows: (1) conclude that cases in the AR group were at lower risk because they were being served by AR workers and (2) to use prior involvement in child protective services, rather than risk, as a control. Concluding that the AR approach, itself, lowers risk is very problematic, and using prior involvement with CPS as a control produced results that cannot be reconciled with other data presented.

As noted, families in the control (TR) group scored higher on almost every risk factor on the Minnesota risk assessment scale. The largest differences were on factors related to cooperation. The researchers focused on these items and concluded that the AR approach actually lowered the risk represented by experimental group families because these families’ scores on several items reflected higher levels of cooperation and motivation. There is nothing in their analysis that supports this conjecture. In the AR environment, it is possible that families might appear more cooperative and motivated, but this may have little impact on their risk of future maltreatment. The conclusion that AR, by itself, lowers risk ignores the fact that AR cases also scored lower on almost every case history item. It also ignores the fact that the difference in recurrence rates between groups was relatively minor and could be explained well by differences in scores on other factors, not to mention the removal of 170 “noneligible” families from the experimental group. This part of the research, while not supported by any real analysis, is central to the conclusions drawn about the effectiveness of DR, and it is the main point used to support the claim that AR, irrespective of extra services, lowers risk and produces better outcomes.

We find the inconsistencies in the reporting of results even more disturbing. On page 122, the researchers state that recurrence rates for the AR (experimental) and TR (control) groups were 27.2% and 30.3%, respectively. They state on page 123 that the overall recurrence rates for cases with no prior CPS involvement were 26.2%, while families with prior involvement had a recurrence rate of 48.5%. They then use prior CPS involvement as a control and compared rates observed for the experimental and comparison groups in a survival analysis chart. Given that cases with prior involvement had a much higher recurrence rate, recurrence rates for both the AR and

TR groups should have been below the 27.2% and 30.3% overall figures. Instead, the researchers report recurrence rates that were between 5% and 7% higher for both the groups (Figure 10.1 on page 124). When combined, the results presented in figure 10.1 were nearly 10% higher than the overall rate given for families with no prior involvement. The researchers provided no explanation for these serious inconsistencies. Further, on page 123, the researchers report that the percentage of AR and TR cases with prior CPS involvement was 8.2% and 10.5%, respectively, —well below the recurrence rates reported in figure 10.1, again without explanation. Because of these inconsistencies, it is difficult to have much confidence in any of the data presented in this section of the report.

The researchers sought to analyze data from subsamples of families from both the experimental and the control groups who received case management services. The researchers noted that (1) a far higher percentage of cases were opened for services in the AR group and (2) lower risk cases in AR were more likely to receive such services and that this led to a slightly lower rate of recurrence for this group. This raises two important questions. If TR workers had been provided the same level of resources that AR workers were provided, would the TR workers have opened a comparable number of cases, and would providing an equal number of services to the families in the TR group have produced better results? We are not claiming that these issues could have been effectively addressed in this study, given its design, but the researchers did not raise or discuss these important possibilities. Rather, all inferences from this data, some of which were highlighted in italics in the report, claimed that AR, again, proved superior. A more objective review of the data would have at least entertained these questions and put qualifications on the interpretation of the findings.

The cost analysis exhibited many problems, including the lack of equivalence between comparison groups and large variance in the follow-up periods. The differences noted in the number of children placed in out-of-home care (clearly the most expensive outcome) could be attributed solely to the higher risk profile of the comparison group and the 170 “ineligible” cases removed from the AR group without assessing for the presence of similar ineligible cases in the TR group. These problems could also explain why results were not consistent even within Minnesota. In Ramsey County, TR appeared far more cost effective, while AR was reported more cost effective elsewhere. These inconsistencies in reported outcome data, the impact of these data on costs, large differences in follow-up periods, and questions regarding sample comparability suggest that not much confidence should be placed in the cost analysis provided in this report.

Throughout this report, we found that in spite of errors in data, the assumptions and inferences drawn from the analysis of the data favored AR. This was true of other studies as well. This is what led to our concerns regarding potential evaluator bias in favor of the AR track in DR research. Loman and Siegel’s response to our article in this issue did little to change our perspective.

In their response, Loman and Siegel claim that our article was more of an opinion-editorial paper than a scientific study. We can accept this characterization if readers recognize the

following. First, as we stated clearly in our article, our intent was not to corroborate or disconfirm any hypothesis, but rather to show that DR outcome research has not confirmed its hypotheses regarding claims of child safety and program effectiveness. Second, our opinions are based upon both an exhaustive analysis of 18 research study reports, such as the example above, and a well-documented and detailed literature review. And third, we support DR reform, and have stated this clearly in our article.

We believe the promise of reform can best be achieved by adhering to a strategic plan of model development and implementation based on objective, transparent, and well-designed empirical support.

Samuels and Brown

Samuels and Brown are affiliated to the ACYF-USDHHS. Their response includes a discussion of the many ways that “ACYF strongly encourages the use of evidence-based and evidence-informed practices in the programs it supports” (p. 2). In 2008, ACYF provided grant funding to establish the QIC-DR. Their response cites a number of other initiatives supported by ACYF to increase the scope of available evidence to inform child welfare practice.

In the first paragraph of their response, Samuels and Brown cite a “growing body of evaluation research” that found that “children in the alternative response track are as safe as or safer than similar children receiving traditional investigation” (p. 1). As stated in our article, we disagree. We do not share their confidence in the body of evaluation research that they reference. We believe our article presents compelling concerns about the validity of this research and therefore concluded that the safety and effectiveness of DR have not yet been demonstrated. The majority of the respondents to our article share this concern.

Samuels and Brown state that, “. . . families are more satisfied with alternative response and receive more services” (p. 1). We accept that families served in AR tracks would report being more satisfied, since by design, they received more services than did families in TR tracks, and they were not required to participate in the potentially more stressful process of case fact-finding about prior maltreatment. Logic would suggest they would be more satisfied. Ten years of repeated DR research with numerous satisfaction surveys documenting this obvious outcome is a waste of time, resources, and effort. Further, it is unclear why Samuels and Brown would cite the fact that AR families received more services as a positive AR research finding. AR families were provided more services than TR families by design. We question this approach, both in serving families and in conducting outcome research. It is certainly not an informative outcome. As stated in our article, since higher risk families in TR tracks, who have an equal or greater need for these same services, did not receive them, we do not think a trend of favoring families in one track over another bodes well for DR’s long-term success.

We commend ACYF for implementing their 5-year multi-site randomized controlled trial evaluation of DR programs. Based on Winokur and Gabel's response to our article, this appears to address some of the methodological problems found in previous research studies (as described in our Finding #2) including the studies that Samuels and Brown are referencing.

We are confused regarding Samuels and Brown's position on whether DR should be an evidence-based practice. The authors first explain that, "The QIC-DR was established with the specific goal of advancing DR along the continuum from 'promising' toward 'evidence-based practice' ..." (p. 3). They continue, "... we believe that differential response is a promising practice, and we expect that results from the QIC-DR evaluation ... will significantly enhance the evidence base." However, in the middle of this commentary, Samuels and Brown state, "We agree that there is no single, clearly articulated and testable model of DR, but are less troubled by this fact than Hughes and colleagues. We are comfortable with treating DR as an approach capable of producing a variety of testable and replicable models that share a minimum core set of characteristics ..." (p. 3). Therefore, it is difficult to tell whether Samuels and Brown see DR as "a promising practice" (p. 1) or "a variety of testable and replicable models that share a minimum core set of characteristics" (p. 3). We have seen this shift of focus frequently in communications about DR, including in several of the responses in this issue. DR is initially labeled a promising practice with strong empirical support until the strength of the evidence or the consistency of implementation—both essential elements of evidence-based practice—are called into question. Then, DR is described as an approach, a policy orientation, a method, a philosophy—anything but a practice—ergo, a general approach that encompasses a heterogeneous array of differing practice interventions that are exempt from critique using criteria of evidence-based programs.

Samuels and Brown appear, at times, to say they support the development of an overarching model of DR reform when they note the CEBC4CW designation of DR as a promising practice, and when they look to Winokur and Gabel's multisite research to enhance DR's development toward an evidence-based model. In their response, Winokur and Gabel articulate a goal of their research to provide impetus to elevating DR to an evidence-based model by addressing the methodological shortcomings of previous research and by including important elements necessary for the development of an evidence-based program model. On the other hand, Samuels and Brown appear, at times, to say that they do not support development of a single DR model, but rather they would support any number of models, as long as they possess a "minimum core set of characteristics" (p. 3). If this is the case, then any discussion of the possibility of the CEBC4CW, or similar organization, vetting DR practice as evidence based, would be irrelevant.

The CEBC4CW (www.cebc4cw.org) is one of many organizations that were designed to assess and communicate the existing level of empirical support of child welfare programs, practices, and models and to track progress over time in their

development toward an "evidence-based" program. The categorization of a program or practice as a promising practice communicates the Clearinghouse's conclusions that the practice holds promise in evolving toward the level necessary to identify it as evidence based. There are several criteria necessary to earn a designation of promising practice from the Clearinghouse, including that the practice must have a manual or other tool that specifies the components of the practice and describes how to administer it; and the overall weight of evidence must support the benefits of the practice (California Evidence-Based Clearinghouse For Child Welfare, 2013). All these are prerequisites for a consistent and well-researched practice that can be replicated with confidence. Fluke et al., Samuels and Brown, and several other of the respondents contend that it is either impossible or undesirable to support the development of a consistent, uniform DR practice model, and that individual states and jurisdictions do have (and, we read, *should* have) the latitude to develop their own practice models as long as they have at least two tracks. As a result, each state initiative would have to be vetted independently with solid outcome research. Such an approach would ensure that the promise of the current promising practice designation is a hollow one. More importantly, although practice integrity could potentially be addressed to varying degrees of validity with multiple outcome evaluations, we would miss the opportunity to develop an evidence-based model of child welfare reform that could be reliably replicated, comprehensively evaluated, and easily improved upon. If a uniform model was developed, states could use their limited research resources in a more effective and efficient way, and findings could be shared by all jurisdictions with confidence in applicability and safe and effective replication.

Moreover, if each jurisdiction studied is essentially implementing its own unique version of a program, it undermines opportunities to collaborate across jurisdictions in an emendatory and progressive developmental process that can ultimately lead to the best model possible - a model that will likely continue to evolve and improve as new research data becomes available. The same lack of consistency also undermines our ability to aggregate data across multiple studies and compare the data in a meta-analysis to obtain a more reliable and valid assessment of a program's outcomes and effectiveness. This has considerable significance for program planners who depend on existing research to justify implementation of complicated and potentially costly initiatives in their own jurisdictions.

If we want a best-practice model with proven effectiveness and general applicability, we must first do the heavy lifting to define, standardize, and consistently implement the program before we can begin to research its efficacy or its effectiveness. It is unclear to us whether Samuels and Brown support this approach.

In our opinion, ACYF has demonstrated a commitment to evidence-based practice that is unprecedented and robust and is evident in both recent policy initiatives and practice directives. Samuels and Brown clearly believe it is important to seek

strong empirical support for services provided to CPS families, and they describe current ACYF initiatives intended to identify evidence-informed practices to promote children's well-being, especially in the social and emotional domains most affected by trauma. We commend this commitment to principles of evidence-based practice and believe it can provide the foundation for significant practice improvement in services to children and families. Our primary disagreement with Samuels and Brown's response is their support of the idiosyncratic development of DR reform in different states, with only minimal requirements for consistency, as the optimal way of moving forward in DR reform. For reasons we have made clear, we believe this position represents a missed opportunity for the federal government to champion model development—the highest commitment to evidence-based practice—for an important child welfare reform.

Conclusion

Child welfare reform efforts are being implemented throughout the country under the banner of differential response, but there is no agreement on a definition and little consistency in policies, practices, or program components. At the same time, much of the outcome research and program literature communicates that DR has, in fact, been clearly articulated, that its safety and effectiveness have been demonstrated by research, and that it can be consistently and safely implemented.

The basic premise of DR is that families in the child welfare system have a wide range of needs, strengths, capacities, and problems, which necessitate an equally wide range of potential intervention strategies to address family needs and achieve goals of child safety, permanence, and well-being. This is highly complicated work. This is a highly complicated work, and child welfare organizations have much to do to achieve these goals for all children being served. Although DR shows considerable promise, it has yet to demonstrate outcomes that can justify the large expenditure of time effort and resources over the past 10 years on DR research and program implementation. It is past time to remedy this. We are strongly supportive of the family-centered practice principles that underlie DR reform, and we hope that the articles in this special issue of *RSWP* can generate constructive and objective dialog among key child welfare leaders and advocates, and that it will result in strategies to shape ongoing research and development so the promise of DR can be realized for the children and families served in our public child welfare systems.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Barkley, R. A., Cook, E. H., Diamond, A., Zametkin, A., Thapar, A., Teeter, A., . . . Willcutt, E. (2002). International consensus statement on ADHD. *Clinical Child and Family Psychology Review*, 5, 89–111.
- Bauer, H. (2001). *Fatal attractions: The troubles with science*. New York, NY: Paraview Press.
- California Evidence-Based Clearinghouse For Child Welfare. (2013). Retrieved from. <http://www.cebc4cw.org/ratings/scientific-rating-scale/>
- Child Welfare Information Gateway. (2003). *Reducing re-referral in unsubstantiated child protective services cases: Research to practice*. Washington, DC: U.S. Department of Health and Human Services
- Children's Research Center. (1998). *The urban caucus CPS decision making system: Revalidation of the risk instruments and an analysis of the effectiveness of child protective services in three counties*. Madison, WI: Author.
- Fuller, T., & Nieto, M. (2009). Substantiation and maltreatment reporting: A propensity score analysis. *Child Maltreatment*, 14, 27–37.
- Gambrill, E. (2010). Evidence-informed practice: Antidote to propaganda in the helping professions? *Research on Social Work Practice*, 20, 302–320.
- Gambrill, E. (2011). Ethical aspects of outcome studies in social, behavioral and educational interventions. *Research on Social Work Practice*, 21, 654–663.
- Hughes, R. C., & Rycus, J. S. (2007). Issues in risk assessment in child protective services. *Journal of Public Child Welfare*, 1, 85–116.
- Institute of Applied Research. (2004). *Minnesota alternative response evaluation: Final Report*. St. Louis, MO: Author. Retrieved from <http://www.iarstl.org/papers.htm#ancC5Loman>
- Johnson, K., Wagner, D., & Scharenbroch, C. (2007). *Risk assessment validation: A prospective study*. California Department of Social Services, Children and Family Services Division. Madison, WI: Children's Research Center.
- Maccoun, R. J. (1998). Biases in the interpretation and use of research results. *Annual Review of Psychology*, 49, 259–287.
- Rycus, J. S., & Hughes, R. C. (2003). *Issues in risk assessment in child protective services: A policy white paper*. Columbus, OH: North American Resource Center for Child Welfare.
- Rycus, J. S., & Hughes, R. C. (2004). Issues in risk assessment in child protective services. *APSAC Advisor* (16). Oklahoma City, OK: The American Professional Society on the Abuse of Children.
- Rycus, J. S., & Hughes, R. C. (2008). Assessing risk throughout the life of a child welfare case. In D. Lindsay & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 201–213). New York, NY: Oxford University Press.
- Rycus, J. S., & Hughes, R. C. (2012). Assessing risk in child protective services. In K. Frysztacki & A. Iliz (Eds.), *Between America and Poland: Opole sociological studies* (pp. 97–110). Cracow, Poland: Nomos Publishing.