

## **Evaluation of Differential Response in Ohio: Challenges in Implementing a Randomized Control Trial**

**Madeleine Kimmich, D.S.W. and Linda Newton-Curtis**

Madeleine Kimmich, D.S.W., is a senior research fellow at Human Services Research Institute, where she also co-directs the office. Since 1987, Dr. Kimmich has led HSRI's Child & Family Services Team, with oversight responsibility for multiple projects in children's services arenas. She currently directs a multi-year evaluation of Ohio's Title IV-E Waiver Demonstration, and guides research on the impact of Differential Response in child welfare. In the past 35 years she has worked and published in the areas of quality assurance, performance measurement, systems reform, and program effectiveness related to children and families, adolescents, seniors, and people with developmental disabilities.

Linda Newton-Curtis is a research associate at Human Services Research Institute. She has worked in the field of Human Services evaluation for 10 years exploring issues such as the impact of the Adoption and Safe Families Act on children and families, the collaborative efforts between child welfare, drug and alcohol treatment agencies and the court on child welfare involved families as well as the effects of person-directed care for dementia patients. She is currently involved in the evaluation of Ohio's Title IV-E Waiver Demonstration and the Differential Response RCT. She holds a Master's degree in psychology from Portland State University.

### **Introduction**

In 2010, the Quality Improvement Center on Differential Response (QIC-DR), funded by the Children's Bureau, U.S. Department of Health and Human Services, began a three-site study of the impact of differential response (DR) on outcomes for children in the child welfare system. Using a randomized controlled trial methodology, this study seeks to replicate and extend previous research, and to contribute materially to the question of whether DR is an evidence-based practice (i.e., an effective intervention with certain child welfare populations). Ohio is one of three sites participating in the study, through a coalition of six counties.<sup>1</sup>

DR is being increasingly used across the nation as a method to intervene with families entering the child welfare system. It is designed to quickly engage families in services and offers an alternative to the traditional abuse/neglect investigation. DR represents a philosophical shift in the approach used by some child welfare agencies toward families with lower risk reports of abuse or neglect; punitive procedures and terminology associated with a typical investigation are reduced or negated, thus eliminating many of the initial barriers that caseworkers confront when first seeking to engage families. DR caseworkers interact closely with families in assessing family needs; appropriate services and supports may then be provided, without any formal allegation or determination of maltreatment being made. Workers are encouraged to partner with families, encouraging families to take the lead in describing their needs. The worker acts as a conduit through which families can connect with community agencies and resources that the *families* believe are necessary to support their well-being. For the current study, as each family

---

<sup>1</sup> Ten Ohio counties were also involved in an earlier study of DR (2007-2009). One of those original 10 counties is also a part of the current QIC-DR study.

enters the child welfare system, each of the six local agencies applies the common eligibility criteria to determine whether the case is eligible for DR; if so, the case is entered into a randomizer which then assigns it to the investigation track (IR) or to the alternative assessment track (DR).

Before proceeding to discuss the early experiences in Ohio, it is important to clarify the context in which this system reform occurs. Two factors bear careful consideration: First, the decision to employ a randomized control trial for the evaluation places increased procedural and data collection demands on the Ohio sites; and, second, Ohio enters this study already having substantial awareness of and commitment to the DR approach, so it does not offer a “clean slate” as is the case in the other QIC-DR study sites.

### *Randomized Control Trials*

In studying the effectiveness of DR as a system reform, the QIC-DR chose to mandate that its pilot sites utilize a randomized control trial (RCT) evaluation design. Because of its potential to clearly determine cause and effect, this methodology is strongly favored by many policymakers and researchers in the social services arena (Chelimsky, 1997; Wolff, 2000). The RCT design requires three conditions: precisely-defined intervention/protocols (both experimental and control), equivalent groups (tied to a meaningful target population), and equivalent trial environments. Evaluators need to be especially attentive to these conditions, assessing at the outset the degree to which they are met, and adjusting the scope of data collection as necessary to minimize the threats to validity of the research findings. Indeed, when studying the effectiveness of a complex social reform such as DR—characterized by variations in staffing arrangements and participant motivation, loosely-defined interventions, and interaction with broader social environments—evaluators are challenged to understand and control a wide array of factors within both the service delivery agencies as well as the larger child-serving community (Wolff; Audrey, Holliday, Parry-Langdon, & Campbell, 2006). Randomly assigning families to the IR and DR tracks is but the first step.

### *Ohio’s Involvement with DR*

Beginning in 2007, Ohio was originally the site of a DR pilot project sponsored by the Ohio State Supreme Court and the Ohio Department of Job and Family Services (ODJFS). The study encompassed 10 of Ohio’s 88 counties, including one of the most populous (see Table 1: Phase 1). Ohio has a county-administered child welfare system, operated through local Public Children Services Agencies (PCSA); this local decision-making role, coupled with significant funding generated at the local level, introduces substantial variation into local practice, in general and specifically in implementation of a systemic reform such as DR. The 2007-2009 study, also an RCT design, found favorable effects of DR when compared with IR, leading the state to commit to steadily expanding DR in subsequent years (Loman, Filonow, & Siegel, 2010). Indeed, ODJFS brought 10 more PCSAs into the DR arena in 2010 and 2011 (see Table 1: Phase 3), with the expectation that further rollout waves will occur. This means that all of Ohio’s PCSAs are aware of the DR approach and are somewhat familiar with the DR philosophy. The fact that, in 2009, six additional PCSAs (including one that participated in the original 10-county pilot project) came together to seek funding through the QIC-DR (Table 1: Phase 2) is but further testament to the widespread belief in Ohio that the future of child welfare lies in embracing DR.

Clearly, the varying levels of understanding of and commitment to DR predating the current study will need to be examined in the evaluation.

The current SOAR project (Six Ohio Counties Alternative Response) includes a mix of jurisdictions—large metro areas as well as tiny rural counties—and PCSAs with varying exposure to and experience in DR.

**Table 1. Schedule of DR Rollout Across Ohio**

<b>Table 1: Schedule of DR Rollout Across Ohio</b>		
	Number of Counties	Year
Phase 1: Original Study	10 counties	2007-2009
<b>Phase 2: QIC-RCT</b>	<b>6 counties</b>	<b>2010-2013</b>
Phase 3: Third Rollout	10 counties	2010/2011
Phase 4: Fourth Rollout	Up to 10 counties	Expected 2011/2012

This paper focuses on phase 2 of DR implementation in the six Ohio counties (highlighted in Table 1). The first section briefly describes the initial work in the SOAR counties to build the foundation for implementing and evaluating DR. The next section highlights four areas in which the counties and the evaluation team encountered particular challenges, and offers some remedies to the problems. The final section reflects on these experiences to identify key lessons learned, pinpointing actions which can be taken in future DR implementation efforts to help the process proceed more smoothly and effectively.

### **Laying the Foundation**

The SOAR counties approached the implementation of DR and its evaluation with excited energy and enthusiasm combined with a little trepidation. Implementing a new practice model in addition to being involved in a rigorously designed study was an enormous undertaking. The participation of one PCSA (Clark County) that had been part of the previous RCT study offered some reassurance to the SOAR team<sup>2</sup> that the mentoring and insights of the experienced member would help the team to avoid the most serious pitfalls of implementation and would guide them through the inevitable challenges of instilling a new practice philosophy and complying with a complex evaluation design.

In preparation for implementing DR and the accompanying evaluation activities, the SOAR sites engaged in four key activities: making changes in staffing arrangements, developing and documenting a standard case flow for the DR and IR tracks, training all staff, and establishing methods of ongoing support for the implementation process. To carry out these activities, several task-specific work teams were organized and operated over several months during the initial planning period of the study. The SOAR team and the evaluators participated in semi-monthly teleconferences or face-to-face meetings to establish a mutual support system and a common understanding of the goals of the study and the effort that would be required.

#### *Changes in Staffing Arrangements*

---

<sup>2</sup> The SOAR team consists of the manager and coordinator for DR in each of the six PCSAs plus the SOAR project manager.

One of the recommended changes for agencies adopting DR is to have caseworkers dedicated to DR and others dedicated to IR. The reasoning behind this recommendation is that, since DR and IR have different philosophies, it would be difficult for individual workers to bounce between DR and IR cases and still continue to preserve the respective philosophies. The larger SOAR counties have designated at least one whole unit, made up of several workers, to exclusively serve DR families. Each unit has its own DR-specific supervisor to provide advice and support to the workers. In the very small PCSAs, just one or two workers are designated as DR caseworkers. Their supervisors oversee both DR and IR caseworkers.

It is also important to note two other changes which occurred for workers involved in implementing DR. In each PCSA, workers interested in becoming DR caseworkers volunteered for the position, and then the manager responsible for DR made the selection from the pool of volunteers; this process often involved an in-depth interview with the candidates. The other crucial change regarded existing caseloads. Investigation workers maintained their cases as usual; in contrast, just prior to implementation, the caseloads of the workers assigned to DR were cleared. Both of these differences potentially influence the way the IR and DR workers handle their job responsibilities and the way in which they view the DR model and approach.

### *Understanding Agency Case Flow*

In order to ensure that DR practice was implemented similarly across the SOAR sites, the SOAR team, with the support of the evaluator, initially focused on creating a flow chart that documented in detail the course a family typically takes through the child welfare system. Included in this diagram were the different processes and procedures used in screening a report of abuse/neglect, as well as information about where in the case flow major decisions occurred, who was responsible for making each decision, and the criteria upon which each decision was based. Complicating this picture, of course, is the fact of county-level administration, meaning that even the small group of six SOAR counties had distinct differences, not only in sequencing of decisions and assignment of tasks, but also in the method of data entry into SACWIS (Statewide Automated Child Welfare Information System). SOAR representatives worked closely with the evaluation team over several months in order to explain precisely where those differences lay so that the research team could appropriately refine the evaluation process. Ultimately, the SOAR team was able to come to agreement on a generic case flow diagram. The evaluation team was then able to embed into the standard case flow the needed evaluation processes, especially the randomization step and subsequent data collection activities. Figure 1 provides a simplified view of this mapping.

*Pathway Assignment Tool:* A particularly critical step in the case flow process is determining a case's eligibility for DR. Initially, the SOAR sites varied in their judgments about the nature and the severity of abuse and family circumstances that were acceptable for DR services; these differences reflected both varying community norms as well as the perceived role of the PCSA in the larger county human services arena. While state rules clearly defined situations that required investigation, PCSAs had full discretion over a wide array of maltreatment reports. Consequently, while counties all agreed that they were unwilling to accept for DR services families they felt posed too great a risk, they varied somewhat in what they considered to be "too great a risk." To ensure the equivalence of the IR and DR populations, the evaluation team sought to bring more agreement to the discretionary decisions about DR

eligibility, and, where full agreement was not possible, to assure full documentation of the reasons for differing decisions. With additional data, the evaluation team would be able to examine the effects of the differing practices on the validity of the RCT. A work group formed to revise the pathway assignment tool, which had been originally developed for the 2007-2009 10-county pilot (while not required, this tool is still being used by some of the Phase 1 Ohio counties practicing DR). The work group's main objective was to build into the tool a new family characteristics section. In this way, characteristics that were acceptable to some counties while unacceptable to others could be closely tracked as counties accepted or rejected families for randomization. This tool enables evaluators to assess the impact of additional characteristics on family outcomes.

### *Training All Staff*

Training is an essential step in preparing for implementation of any new intervention. In the SOAR sites, a wide array of staff needed to be trained, both in the philosophy of DR and in the practice changes that define the alternative approach. All of the basic training in DR was provided by the QIC-DR, through staff and consultants of the American Humane Association. In addition, several key staff in each SOAR site visited other Ohio counties that had already implemented DR in order to observe the screening and casework activities; these visits were extremely valuable in alleviating anxiety and providing concrete answers to questions and concerns.

Equally important was evaluation training. Relevant information was shared with the SOAR team during semi-monthly teleconferences as well as through individual county conversations between the evaluators and key DR staff. The research team consistently sought local input to proposed evaluation tools and processes in order to ensure that burden would be minimized and accuracy enhanced. One month prior to implementation, all counties participated in a one-day in-depth evaluation training conducted by the evaluation team. The training included two sections: one for staff having screening responsibilities and another longer portion for supervisors and workers carrying caseloads of IR or DR cases. Both sessions explained the randomization process in detail, focusing on its critical importance to the evaluation as well as what it would mean in practical terms to agency workers. To make the training as accessible as possible to workers and supervisors, evaluators held multiple training sessions over a 2-week period, spread across locations in all the counties. Subsequently, training refreshers were offered through web-conferencing, affording smaller groups of staff the opportunity to ask questions that had arisen in their initial use of the evaluation tools.

### *Providing Ongoing Support*

The final preparation activity was to institute ways to provide ongoing support to staff, in terms of both DR practice and evaluation responsibilities. Staff in Clark County reached out to the new DR sites during SOAR team meetings and at the individual request of any participant, providing mini-trainings on certain aspects of practice (e.g., how to use the pathway assignment tool) and addressing any questions or confusions that arose in the course of everyday DR practice. In addition, SOAR representatives have made personal connections with staff in other DR counties to whom they turn for advice, and many of the SOAR sites attend quarterly meetings for all Ohio DR counties.

A similar ongoing relationship exists for evaluation issues. The evaluation team has offered refresher trainings for those who need further clarification, follow-up webinar or teleconference trainings for those who missed the original in-person trainings, and subsequent in-person trainings tailored specifically for the respective individuals or counties and their unique issues. Each county also has electronic access to an evaluation manual. Perhaps most helpful, a local evaluation team member located at one of the SOAR sites is available to provide hands-on practical evaluation advice and training as necessary, including emergency consultation.

## **Reflections on Early Challenges Faced in Implementing the RCT**

*Random assignment is truly random and will not immediately generate a constant flow of cases into each track.*

Funding for the SOAR project began in February 2010, and will continue until June 2013. The random assignment of cases formally began in December 2010; a 3-month pilot period preceded the start to enable county staff to become familiar with the new intervention model and to iron out any wrinkles in the random assignment and data collection processes. During this pilot period of implementation, the six counties encountered varying obstacles related to both the new practice and the evaluation, and worked together to fashion creative solutions.

This section highlights four areas that proved particularly challenging to the SOAR counties, both during the pilot period and in the first few months of the demonstration: random assignment, technology, workload, and exposure to the intervention. The discussion provides insights to inform future decisions by both agencies and evaluators who are considering a human services RCT. Implementation of DR itself presents challenges (as described in other articles in this journal); overlaying a new practice with the demands of an RCT can be formidable. The following sections describe each issue and its challenges, and offer some solutions or at least ways to reduce the impact of the challenges on practice and evaluation.

### *Random Assignment*

The use of random assignment of cases to an innovative intervention or to the traditional casework approach immediately raises concerns among most child welfare workers. In the first instance, they may be reticent to see the families they support being denied the best intervention that is available, regardless of whether scientific evidence has proven its effectiveness for similar families (Little, Kogan, Bullock, & van der Laan, 2004). In addition, caseworkers value their roles as gatekeeper and advocate, using their discretion to decide which interventions are most appropriate for a family. To relinquish that role in the interests of evaluation, they must come to believe that the knowledge gained from the study will bring even greater benefit to families now and in the future. Without this perception, caseworkers will face a continuing ethical dilemma that may ultimately lead to compromises in the random assignment process.

Separate from the ethical issue is a complex practical issue: Random assignment is truly random and will not immediately generate a constant flow of cases into each track. The six SOAR counties and the evaluation team developed county-specific ratios for the number of cases *eligible* for DR that would be *randomized* to the DR track versus the IR track. The ratios varied greatly between counties and ranged from 49% of all study-eligible cases tracked to DR in one county to just 18% of cases in another. The ratios were initiated to reflect a balance between

evaluation requirements and caseworker resources in counties of various sizes. Nonetheless, such ratios do not ensure that cases will be evenly spaced. The ratios apply over a long period of time and do not guarantee that new DR workers will immediately have a full caseload. The ebb and flow of cases is not unknown to child welfare; all agencies face periods of time in which reports of abuse and neglect are heavier than at other times, and intake workers are accustomed to adjusting and sharing the strain. However, DR is a new option in the intake phase exercised only by certain workers. Where IR intake workers equally share the times of high caseloads and low caseloads, random assignment may serve to screen DR workers from becoming as overwhelmed as traditional IR sometimes become; similarly, when random assignment does not immediately generate enough DR cases to fill a caseload, IR workers may begin to perceive DR workers as having lighter workloads. In the interests of maintaining good relationships between DR and IR workers, and in order to support good casework practice, the SOAR counties have adopted a variety of solutions:

- One county had a swing worker who would take both types of cases in order to somewhat reduce the pressures on workers in one track or the other;
- Some counties had the DR workers take other related types of cases—dependency or families in need of services (FINS)—when the DR caseload was too low;
- Some counties requested that randomizer ratios be increased or decreased over time. The challenge was to not change the ratio so dramatically that it simply shifted the overload to the other track.

None of these practices was perfect; DR workers could quickly become overloaded if a large number of cases happened to be randomized to DR soon after the DR workers had picked up other cases. But any of the solutions was preferable to having the site coordinators arbitrarily choose not to use the randomizer—an understandable reaction by a supervisor but a serious threat to the evaluation. Overall, the randomizer proved a little frustrating at study start because of the unpredictability it caused in caseload numbers for staff.

### *Technology*

Technology has proven to be both a boon and a challenge—helpful when it works correctly, anxiety-inducing and irritating when it does not. The RCT study relies on three separate electronic data systems: the randomizer, the SOAR web-based data collection system (named SOARDS), and the state administrative data system, SACWIS.

*Randomizer:* Just as the idea of random assignment was difficult for staff to accept, so too was the concept of an electronic process to make that random assignment. The randomizer is seen as a “black box” wherein a decision is made without input from the worker/supervisor who is accustomed to using professional discretion to decide how a family should be engaged by the child welfare system personnel. The technical process is very simple: Once a case is screened as eligible for DR, the screener logs into the randomizer website; a new screen then pops up with a box in which the study-eligible family’s intake number can be entered. The staffer clicks the “submit” button, and the randomization result appears on the screen: assigned to DR or IR, and

selected for a survey or no survey.<sup>3</sup> However, this process initially presented two problems to the SOAR sites. The first problem was the separation of the randomizer from SACWIS. It meant that staff had the extra step of going to an outside web address while working within SACWIS to obtain the correct case ID number. Workers were initially anxious about being able to log in back and forth between systems; this soon proved to be a relatively smooth process. The second challenge was that the randomizer sometimes was out of service and was not available to the staff. Because PCSAs have strict state rules dictating the allowable time between receiving a report of abuse and a worker initiating the DR or IR process, a non-working randomizer threatens to compromise the casework process. This problem also concerned evaluators: If the unreliability of the randomizer reduced sites' commitment to always using the random assignment process, it would undermine the integrity of the RCT. This problem was resolved early in the study by moving the randomizer to a different type of server. However, while the problem was being addressed, the evaluation team established a Plan B. Whenever the randomizer was not available, the PCSA staff person would use the old-fashioned coin toss to randomize families. This was not ideal but at least maintained a random assignment of families; the decision was then communicated to the appropriate casework staff. At the same time, the evaluators required a notation to be made in SOARDS for any cases in which a coin toss had been used in order to reduce the likelihood of workers making arbitrary decisions to assign cases to one track or the other.

Several benefits come from use of an electronic randomizer. It has enabled the evaluation team to relatively easily control and change DR/IR ratios within each county when caseloads become too unequal. It allows counties to painlessly assign families to each track without having to put substantial additional effort into their work. Finally, it provides an electronic record of the cases that have been entered into the system, along with the historical record of the date the family was randomized and the random assignment result. This gives the evaluators an opportunity to double-check randomizer track assignments when and if data entry errors occur at later points in the process.

*SOARDS*: The initial reason for creating a web-based data system for the SOAR cases was to provide a centralized tracking system to ease the burden county coordinators face when tracking families. In the 2007-2009 DR pilot, Clark County had developed its own spreadsheets to accomplish this task; a centralized, standardized web-based system seemed to be an easier approach, and it has ultimately come to serve an even larger set of purposes. SOARDS is a stand-alone web-based data collection system utilized to varying degrees by PCSA screening staff, DR coordinators, and casework supervisors. It contains a search tool so that counties can check whether a family with a new abuse/neglect report has been previously entered into the randomizer. It includes tabs for the pathway assignment tool and a family characteristics tool, both of which need to be completed for each family. It has a resource page with downloadable copies of surveys, consent forms, the evaluation manual, and team contact information. And it has a tracking sheet for randomization results—track assignment and survey selection—as well as for dates of survey distribution and completion. In addition, it offers sites the option of running a series of canned reports so they can track their own performance over time, including survey response rates. At the same time, it is invaluable for keeping the evaluation team alert to

---

<sup>3</sup> The randomization is conducted at two levels. The first is to DR or IR. Each case is then randomized again to determine whether it is part of the subgroup receiving a survey (SACWIS data will be used for all cases).

potential errors and omissions in data; in this way, evaluators can urge counties to keep up with data cleaning as the study progresses.

Despite all its advantages, SOARDS nonetheless represents another level of complexity for county coordinators. Having the two stand-alone systems—the randomizer and SOARDS—can lead to data discrepancies between systems as intake numbers are copied from SOARDS into the randomizer, and randomizer results are copied into SOARDS. Tracking and logging data inconsistencies are laborious and can be hard to resolve. Progress is slowly being made, but the tasks remain burdensome at times to both county staff and the evaluation team.

*SACWIS:* The state administrative data system is an extremely large, complex system that has recently undergone a series of changes in response to the expansion of DR throughout Ohio. SACWIS will provide much of the data needed to evaluate outcomes. At the time of the study’s rollout, several gaps still remained between what was in the system and what the evaluation would require. As a safeguard, the evaluation team chose to capture additional information in SOARDS. However, this means that the PCSAs now have three unconnected data systems all housing critical evaluation data. The evaluation team knows that data entry discrepancies will emerge among the three systems and these will take some time and additional information to unravel. By the end of data collection for this study, SACWIS will contain information to remedy many of these issues, making it easier for Ohio counties that are implementing DR in the future.

### *Workload*

Workload issues arise whenever a new casework practice or approach is introduced because it takes some time for staff to adjust to the changes. However, the impact of the shift and its duration are very much influenced by the magnitude of the change being sought. In other words, a fundamental change in child welfare philosophy (as in the move to DR) may be harder for workers to embrace and may require a longer transition period. Thus, the stress associated with workload pressures needs to be alleviated promptly rather than being allowed to build up and threaten the longer-term stability of the reform—and, indeed, the viability of the RCT.

The RCT design requires that data is collected on both the experimental and control side. The evaluation team is collecting survey data for a randomly chosen subset of the study families. For each family selected to complete a survey, the caseworker is also asked to complete a survey.

The RCT design introduces workload concerns in three main ways. First, because the RCT is a “live experiment,” the evaluation requires some types of data that are not available in state administrative data systems but which are essential to the local evaluation as well as to the cross-site study (see later article in this journal). Asking workers and supervisors to complete multiple surveys and to enter information into several different data systems is extra work and there is no easy way to reduce the burden. However, letting staff know early on that they will be required to do extra work on all cases in the study, not just those selected for DR, is essential. This was a problem in some SOAR sites, causing extra resistance to data collection responsibilities. Training and ongoing support to staff in their data collection tracking tasks has somewhat relieved this stress. Nonetheless, it is also important to note that this extra work is purely due to the evaluation: States that choose to adopt DR will not necessarily have an accompanying evaluation.

A second source of workload pressure for caseworkers is the fluctuation in case assignments. As described briefly above, the random assignment process generates an uneven flow of cases to DR and IR staff. While the numerical inequities can be somewhat ameliorated, there may be an accompanying negative undercurrent that is more difficult to address: Workers on the IR side may begin to feel less valued, interpreting the attention to keeping DR caseloads manageable as a message that their own work is not as important. This feeling may be aggravated by the enthusiasm on the DR side, since these workers have self-selected for the DR track.

The PCSAs sought to minimize this situation through orientation and training messaging, openly explaining the value of both types of practice for families and for evaluation. The evaluation will assess the magnitude of this attitudinal factor through a general survey of caseworkers' experiences and perspectives of DR; these data may help agencies to fashion an appropriate response (e.g., bringing workers together to give positive feedback to both groups, or giving each group the opportunity to fully vent their anxieties and resentments). It has also been helpful to share with all staff the *actual* caseload differences, to replace impressions with facts, and to allow a more direct discussion of how best to remedy any genuine inequity.

The third source of workload burden comes from the need to carefully track what happens to cases as they proceed through the randomizer to be assigned to DR or IR staff. Each case is randomized at two levels: Not only is the case randomly assigned to the IR or DR track, but it is also randomly selected to be surveyed. A surveyed case requires the worker to do two things when the case closes: distribute a paper survey to the family and complete a case report form asking for details about the case. While workers receive formal notification of these random assignments at the time they receive the case, it can be difficult for them to keep on top of the extra tasks that come with a case's selection for surveying. Some counties initially left these tasks to workers or line supervisors, but subsequently have realized they need clerical support to keep up with reminders, and have thus brought on additional support staff.

### *Exposure to the Intervention*

To clearly assess the impact of participation in the DR pathway on child and family outcomes, the evaluation team must ensure that all families included in the study (i.e., all those sent to the randomizer to be assigned to DR or IR) have not had any past experience with DR. The evaluation is testing whether an initial "dose" of DR leads to better outcomes than those that typically occur for families served in the traditional approach. If any family in the current study had been previously involved in DR, that experience could be expected to affect how the family fares during the current case episode (the case episode begins with the initial report of abuse/neglect and ends when the case is closed by the PCSA). If families in either the DR or IR track have such prior exposure to DR, they are *different* than other families in the study and thus constitute a potential threat to the validity of the random assignment process.

Given Ohio's earlier involvement in a DR pilot study, the risk is real that current PCSA families may have already been served through DR. This risk stems from two sources: First, Clark County is one of the SOAR sites and was one of the earlier pilot sites; families served a few years ago during the pilot study could very well emerge again with a report of abuse or neglect that makes them eligible for the current DR study. Second, families often move from one county to another in the state, so families served in any of the 10 counties in the 2007-2009 pilot study might reappear in any of the six SOAR counties.

There is no way to completely eliminate this risk, but it can be reduced. The first step taken by the evaluation team was to pre-populate SOARDS with the names and ID numbers of all cases served in the DR track in Clark County during the previous study. Therefore, by completing a quick search in the SOARDS system, counties could be assured that the family was not already a part of the current study or one of Clark's previous family study participants. Second, also to reduce the likelihood that cases will reappear twice during the current study, the randomizer retained identifying information on all cases entered since the study began. This ensured that a case would not end up randomized a second time—a confusing and possibly harmful situation for a family, and a problem for the evaluation as well.

These steps taken to control prior exposure to DR address the most likely sources of sample contamination, but there remain other potential problem areas. There is the danger that cases from 2007-2009 sites may slip by and be entered into the SOAR randomizer. And, as Ohio continues to expand its use of DR (10 counties were added in 2010 and 10 more will be added in 2011), more and more families will experience the DR approach. However, beginning in late 2010, all cases served through DR have been and will be flagged in SACWIS. While this does not prevent a case from being entered into the SOAR randomizer, it does provide the evaluation team with a way to statistically control for prior DR exposure.

## **Lessons Learned**

The evaluation of the SOAR project using an RCT faces challenging but not insurmountable conditions. Careful and thorough preparation is essential, especially in terms of building relationships with the study sites, communicating well throughout the study, and having a pilot period for testing. The research must be seen by all parties as a *collaboration* and not as something that is being done to the agency; therefore, at the heart of implementing and sustaining a successful RCT is developing a shared vision between community agencies and researchers (Senge, 1990). Agency representatives and researchers must work closely, shoulder to shoulder, to come to a common understanding of the value of the research, and to establish the common path that will achieve the vision. Leaders will then be more willing to advocate for the internal structural and procedural changes necessary to ensure full implementation of DR and success of the RCT study, and to support the research process when worker motivation and enthusiasm for the study begin to waver.

### *Relationships and Communication*

The top priority in preparing to field an RCT is to establish a solid partnership with the study sites and mechanisms for ongoing communication. In the current study, the evaluation team maintains frequent contact with the SOAR counties through individual phone conversations, emails, and monthly all-site meetings; key to these interactions is listening, respecting, and being responsive to the counties' needs for support. Equally important is that staff in all the sites feel the freedom and security to call the researchers whenever they feel confused, anxious, or have a specific question. Most people want to do the best job they can, but enthusiasm may be quickly lost if external support is not available. Having an on-site evaluator who is readily accessible to troubleshoot as needed has been invaluable. This person understands the unique context of each site, can balance this context against the needs of the evaluation, and

is always available to take a question or visit the individual site to provide extra support or training.

Regular feedback to the counties has been very important. This takes the form of monthly data reports showing, for example, the number of family participants in the study to date and the number of surveys received from the selected families. The reports are delivered electronically to each site and then given further explanation at the monthly all-site call. This call provides an open forum in which the SOAR team and the evaluators work together to identify any problems that may be occurring and partner to find solutions to those problems. For example, researchers were surprised to discover the workload toll that tracking families through the study pipeline was taking in one of the larger counties. Adjustments were made to the process so that it could be sustained throughout the duration of the study. Communication of study progress in the public arena of the monthly conference call has been positive in that counties that are relatively successful in their RCT efforts can provide tips to other counties.

Several aspects of the evaluation require particular attention, through training and ongoing support:

- Explaining the how and why of random assignment in a very down-to-earth way so that workers see their roles as integral to the success of the study;
- Adopting user-friendly technology, explaining why each piece is needed, and being open to making modifications to increase worker comfort with the processes;
- Ensuring that data collection responsibilities are laid out in detail from the start, especially the purpose of each task in terms of the overall integrity of the study, and emphasizing that the process will get easier over time, as staff become accustomed to the tools and the data entry routine.

Overall, it is essential for the evaluation team to be attentive to the level of stress workers experience during the lengthy transition to the new practice, and to do whatever is feasible to acknowledge the contribution staff are making and to try to relieve some of their burden.

### *Piloting of the RCT*

An important complement to nurturing relationships and communication between the evaluation team and the sites is having a pilot period before the study formally begins. This “no-fault” time period offers researchers and agency participants the opportunity to conduct a trial run. During this time, theory becomes practice and hands-on learning occurs—for everyone. It is a time when mistakes can be made without fear of reprimand or criticism, additional trainings or supports can be put in place, and final adjustments to procedures and responsibilities can be made.

Piloting the RCT is a necessary and critical undertaking in order to identify and address situations that might threaten the internal validity of the study. A 2-month pilot period was originally scheduled for the current study. However, it quickly became evident that 2 months was too short a time frame, and the piloting period was extended by an additional month. Two reasons drove the extension decision: During the first 2 months of the study, only a very small number of the study participants had enough time to pass completely through the study pipeline, so not all aspects of the study process were adequately tested; additionally, the pipeline had not had time to reach full capacity on the intervention or control side of the study. It was not possible

to fully understand how well the procedures for tracking the assignments and collecting data about exiting families were working in practice. Researchers have recently wondered if even the extended period has been long enough.

Even with the best training, when so many different people are involved in the smooth operation of an RCT, there will be misunderstandings. The notion of an RCT is not necessarily an intuitive one for workers who are trained to deal with family problems rather than research design. One of the clear benefits of the pilot period has been to uncover where those misunderstandings lay. The pilot period provided a much-needed window during which follow-up individualized trainings could be provided. These have been welcomed by the agency partners, and ad hoc refresher trainings will continue to be provided as necessary throughout the study.

## **Conclusion**

The study described above is a real world experiment located in a variety of child welfare settings with families that have complex issues. The intervention itself is a shift in philosophy as well as practice, making it more challenging to evaluate. Many factors interact to influence the PCSAs, the workers, the families, and the communities. This first year of implementation has been an intense learning experience for all those involved: It has included learning new language, new practices, and new rules, and it has been an opportunity to establish strong relationships between project staff and evaluators. The enormous progress made by the SOAR sites has established a solid foundation for practice and evaluation, and promises to support the long-term examination of the impact of DR on child and family outcomes. The study team anticipates producing findings in 2013 which will clearly address issues of child safety, subsequent maltreatment, and differences in rates of placement into foster care as they relate to comparable families that have received the traditional approach versus DR.

## **References**

- Audrey, S., Holliday, J., Parry-Langdon, N., & Campbell, R. (2006). Meeting the challenges of implementing process evaluation within randomized controlled trials: The example of ASSIST (A Stop Smoking in Schools Trial). *Health Education Research, Theory and Practice, 21*(3), 366-377.
- Chelimsky, E. (1997). Thoughts for a new evaluation society. *Evaluation, 3*, 97-118.
- Little, M., Kogan, J., Bullock, R., & van der Laan, P. (2004). ISSP: An experiment in multi-systemic responses to persistent young offenders known to children's services. *British Journal of Criminology, 44*, 225-240.
- Loman, L. A., Filonow, C. S., & Siegel, G. (2010). *Ohio alternative response evaluation: Final report*. St. Louis: MO: Institute of Applied Research.
- Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Wolff, N. (2000). Using randomized controlled trials to evaluate socially complex services: Problems, challenges and recommendations. *Journal of Mental Health Policy and Economics*, 3, 97-109.

Figure 1: Agency Case Flow and Evaluation Mapping

