

EARNING COMMUNITY TRUST IN DATA-DRIVEN INTERVENTIONS AT THE DUKE ENDOWMENT

RHETT MABRY

A MISSED STEP

In 2005, Hurricane Katrina hit the New Orleans area, causing more than 1,800 deaths and approximately \$160 billion in damages.¹ At its high-water mark, the category 5 storm left nearly 80 percent of the city and many of the surrounding parishes underwater.² More than 1 million people were immediately displaced; thirty days later, some 600,000 remained unable to return to their homes.³

In the aftermath, the Red Cross used a variety of techniques to help those in crisis locate loved ones living across the United States. The Red Cross's search system led to many successful connections, and leaders who work in the child welfare system began wondering if those same techniques might help locate relatives of children in foster care. Adoption placement agencies also saw potential for locating family members who might be unaware of a child's circumstances.

In North Carolina, leaders from the Department of Social Services, The Duke Endowment, and a prominent private state adoption placement agency met in Raleigh in 2008 to discuss using family search strategies to connect children with kin, whether to establish an ongoing relationship or possibly

entrust the child with a caring relative as a more permanent placement solution.

In short order, a plan was hatched. Nine North Carolina counties were selected for a randomized control trial to test if identifying more relatives of children in foster care might lead to increased permanent placements with kin. Early results were promising. Using internet and other search techniques, cases randomized into the treatment group identified, on average, almost ten times more relatives than control cases in which traditional methods were used. Even after excluding relatives who were not interested in or could not commit to a relationship, approximately four times as many family members in the treatment group were willing to commit to the child than in the control.

In the end, however, there was no difference between treatment and control cases in achieving permanent kinship placements.⁴ What happened? Truthfully, we cannot be sure, but we think we have identified what went wrong and what we might have done differently.

In our zeal to test this new, promising approach for locating family members, we failed to adequately engage caseworkers to capture their input before subjecting the approach to a rigorous trial. We failed to consider how front-line social workers, who typically are overburdened and carry caseloads above recommended standards, might respond to having additional family members (in the treatment group) to vet for possible placement options. More specifically, we did not account for the additional time the expanded options would require in determining the best placement for the child. Understandably, with nothing taken off their already full plates, demanding schedules limited the number of families with which social workers could work. Not surprisingly, then, the inability to capitalize on the increased family contacts resulted in no difference between the treatment and control groups in achieving a permanent placement.

Had we been more patient, we would have tested the approach on a smaller scale before rolling it out across nine counties. Doing so likely would have identified implementation challenges and surfaced ideas for freeing up necessary time. For instance, one possible solution would have been for staff who conducted the family searches to proceed with finalizing placements—which, in effect, could have served to extend caseworker capacity. Instead of vetting extra placement options on their own, caseworkers could have played a supervisory role.

ADDRESSING CHALLENGE

The unfortunate reality is that our experience is not unique. The United States spends billions of dollars each year to provide social supports for housing, food access, education, medical care, transportation, job training, and more. In many cases, this money is spent without knowing if the intended impact was achieved.⁵ For the past twenty years, The Duke Endowment has sought to increase the body of evidence for emerging promising practices, such as using family search techniques in foster care. We also have placed a priority on replicating what works. The list of evidence-based programs we support is long, including Nurse-Family Partnership (NFP), The Incredible Years, Triple P, Strengthening Families Program, Multi-systemic Therapy, and Trauma-Focused Cognitive Behavioral Therapy. For us to call these models “evidence-based,” each is required to have at least two randomized control trials documenting impact. While admittedly simplistic, this definition helped to guard against the tendency in social services to overuse and dilute the meaning of a catchy term like *evidence based*.

With the assurance that a model has demonstrated that it *can* have impact, our focus turned to effective implementation to ensure replication with fidelity. For instance, NFP has accumulated impressive short-term and longitudinal data supporting its impact across three randomized control trials dating to the 1970s. Spurred by this encouraging data and resulting cost savings, NFP has expanded dramatically during the past two decades, reaching more than 60,000 families across the country in 2019 alone.⁶ In most communities, public health departments serve as hosting agencies even though NFP’s more targeted and intensive home visiting approach does not always sync with the traditional public health orientation of providing lighter touch interactions and broad-reaching, community-wide strategies. Consequently, we suspect that, as NFP evolves, it will increasingly need to consider how it best integrates within existing community systems as opposed to operating as a standalone or adjunct program within health departments.

NFP leaders appear to agree with this direction and are taking steps to identify families who benefit most from the program and determine when home visiting services are (and are not) the most efficient use of resources. This may lead to accepting more targeted referrals and discharging families sooner based on progress assessments. NFP recently merged with Child

First, a home-based intervention that helps vulnerable children and families heal from the damaging effects of trauma, stress, and adversity, indicating its interest in integrating with complementary programs. Embedding services within a broader community tapestry may require adjustments to the model and will push funders enamored with strict model fidelity to think more adaptively.

In our work to build evidence for promising practices, we have not always engaged in a sufficient formative evaluation to ensure proper systems integration. Systems are complicated and difficult to change. Introducing a new, well-researched or emerging practice for broader adoption requires far more than the naïve “plug and play” mentality we and other funders sometimes assume. Our work with the National Implementation Research Network in Chapel Hill, North Carolina, continues to inform our understanding of the importance of planning for and addressing implementation challenges before, during, and after introducing an innovation.

Relatedly, attempting to bring an evidence-based program to a much greater scale can have significant challenges. For starters, there is the question of whether there are enough high-quality or adequately trained service providers (social workers, nurses, clinicians, etc.) to deliver the intervention as designed. If not, service quality might be compromised. Scaling also requires logistical and technology enhancements. Consistent program delivery must include continuous staff training and recruitment, seamless telecommunications, and sufficient working capital to manage through reimbursement delays. Few challenges are insurmountable, but even smaller obstacles must be managed.

UNINTENDED CONSEQUENCES

At least two unintended consequences have emerged as foundations support replication of evidence-based practices. One is that the insistence on evidence and model fidelity may inadvertently stifle further innovation. Evidence-based solutions are coveted by funders and nonprofits alike, yet interventions that have documented impact from one or more randomized control trials are few. From 2002 and 2013, the U.S. Department of Education conducted some ninety randomized control trials, and nearly 90 percent produced weak or no effects.⁷ These results are consistent with those reported in other fields such as psychology⁸ and medicine.⁹ Once encouraging

data are confirmed, the instinct is to lock the program, cease further development, and proceed with high fidelity replication. This works fine for a while. Eventually, however, the world changes: Medicaid expands, telemedicine takes off, women stop smoking during pregnancy, smartphones start tracking vital health and exercise statistics. Interventions must constantly adapt to become more efficient and effective.

To be clear, the blame for this stagnation lies more with the funders, who insist on replicating “proven” models, than with the purveyors of the interventions, who would surely value research funding to continuously improve their approaches. The solution may be for funders to continue to demand and support strong implementation of well-documented and evaluated programs while also funding pockets of innovation in select communities. An example of testing new aspects of evidence-based programs in the context of a broader initiative is underway in Guilford County, North Carolina, where NFP, Family Connects, and Healthy Steps have agreed to meld their programs into a cohesive suite of services for young children and families. This collaboration, undergirded by an integrated data system, should allow NFP to serve highest-risk mothers and for other community providers to receive timely referrals for less intensive interventions. Another example is to capitalize on the popularity of telemedicine during the recent pandemic and use virtual connections to increase efficiencies and the number of contacts with families.

The second challenge with focusing on evidence-based programs is that it likely has funneled more resources to well-funded, established nonprofits at the expense of smaller grassroots organizations.¹⁰ Evidence-based models require not only evaluation expertise but also development staff to raise money for costly enhancements, along with policy advocates to tap into sustainable public funding. Without that capacity, many smaller organizations, which might have deep expertise in the issues faced by the communities they serve, are frequently passed over by foundations.

Fortunately, as philanthropy seeks more equitable solutions, foundations are realizing the importance of grassroots organizations, many of which are operated by leaders of color and located within or near communities that have been marginalized. Investing in these organizations for a sustained period would help them build and deliver interventions with documented results. The combination of earned community trust and data-driven interventions may prove potent for improving outcomes that have been difficult to change.

WHERE DO WE GO FROM HERE?

Given the challenges with building, testing, implementing, modifying, and sustaining evidence-based interventions, what might be a path forward? Abandoning support for those tried-and-true programs does not seem to be a good option. Perhaps there is a compromise.

For the past several years, in North Carolina, select churches in more than fifteen rural communities have agreed to offer The Duke Endowment's summer learning program for elementary-age students and help us test its effectiveness. Challenges along the way have required us to adopt a developmental approach to accumulating evidence. The initial goal is to refine the intervention by better understanding student profiles, pinpointing specific learning obstacles, creating an effective curriculum, developing teacher skill sets, and establishing student selection protocols while also collecting pre- and post-test data to discern directional impact. We hope options for summative evaluation will evolve over time as we amass a preponderance of evidence in support of the model's effectiveness as opposed to an all-or-none designation of the model as "evidence based." Boston-based Project Evident is helping us design a systematic approach for accumulating compelling evidence.

This iterative approach has broader applications beyond rural communities. The spate of place-based interventions emerging across the country will likely also need to adopt similar tactics. Place-based investments face considerable hurdles, not the least of which is measurement. This is particularly true in the early childhood field, which is the focus of many place-based efforts and for which there is a dearth of administrative data covering outcomes prior to school entry.

This lack of administrative data rules out the most straightforward designs for impact evaluation. Instead, evaluators are exploring multi-pronged designs and strategies for collecting a variety of data that produce a "basket" of converging evidence. For instance, primary or original data collection—family-by-family surveys and interviews—may be used to capture information on children's social and emotional development. A convincing evaluation surely requires a credible counterfactual, so data collection will need to occur in both the treatment community and matched comparison sites. Kindergarten readiness assessments, currently a hodgepodge of tools administered with varying levels of rigor and credi-

bility, will need standardization. Individual program-level data will be tracked to assess contributory effects and directional trends. These data on outcomes will need to be complemented by documenting implementation details—numbers served, satisfaction with services, referrals made, funding redirected, reimbursement revenue opened, and policy changes enacted. The latter output measures should help shed light on the progress of system changes underway.

The difficulty of data collection for place-based investments, especially those targeting early childhood outcomes, applies to other large-scale interventions as well. Implementation costs are high, formative and summative evaluations are expensive, and integrated data systems are complicated to build and frequently fail to achieve sufficient buy-in. It is worth considering a collaborative investment between philanthropy and government to establish systems and administrative processes for routinely capturing data where gaps exist. Such a system would decrease both the expense and risk of evaluating large-scale, community-wide interventions. Accessible data also would be useful in calculating associated cost savings, which are important drivers for changing policies and practices.

An important caution about adopting a “preponderance of evidence” approach to determining likely effectiveness of a program or initiative is that it may give comfort to those who are fundamentally uninterested in measuring outcomes. Abandoning clear-cut definitions for what constitutes sufficient evidence—such as having at least two RCTs with statistically significant findings—may be seen as an invitation to ignore data altogether. Doing so would be a mistake. Assuring impact requires developing a discipline (by funders and practitioners) of capturing qualitative and quantitative data in carefully planned feedback loops and using that data to improve our approaches. Measuring outcomes is an important part of such a discipline. When the most rigorous methods (“the perfect”) are not available, the best data available (“the good”) will have to suffice. Either way is preferable to ignoring measurement and data altogether.

Philanthropy, too, often mirrors government in this regard. The complexities that accompany rigorous evaluation should not thwart efforts to use data. Rather, they should spur us to continue to seek new solutions and approaches. No outcome data, regardless of evaluation design or rigor, is absolute, and no findings, no matter how weak or strong, are fixed indefinitely.

I am confident that advances in technology and evaluation approaches will continue to drive improvements in the social sector. The first human coronavirus—the cause of the common cold—was identified in 1965. Since then, seven coronaviruses have been known to sicken us, and scientists had little success developing vaccines until the recent breakthroughs in response to SARS and COVID-19. That success built on decades of work studying the mechanisms of viral transmission, messenger RNA transcription, and therapeutics.¹¹ Just as a combination of meticulous laboratory observation, epidemiological studies, and clinical trials surely drove our vaccine success, research in the social sciences must follow a similar painstaking path, iteratively building on prior findings with an appropriate mix of methods. In addressing society’s complex challenges, let us be encouraged and committed to that journey.

NOTES

1. See Britannica website, www.britannica.com/event/Hurricane-Katrina.
2. Ibid.
3. Allison Plyer, “Facts for Features: Katrina Impact,” The Data Center, August 26, 2016, www.datacenterresearch.org/data-resources/katrina/facts-for-impact/#:~:text=The%20storm%20displaced%20more%20than,housed%20at%20least%20114%2C000%20households.
4. Child Trends conducted the evaluation.
5. See the Coalition for Evidence Based Policy website, <http://coalition4evidence.org/>.
6. Annual Report 2019—Nurse-Family Partnership.
7. See the Coalition for Evidence Based Policy website, <http://coalition4evidence.org/wp-content/uploads/2013/06/IES-Commissioned-RCTs-positive-vs-weak-or-null-findings-7-2013.pdf>.
8. The “replication crisis” in psychology was triggered by a report that just 36 percent of studies published in top psychology journals could be replicated. Andria Woodell, “Leaning into the Replication Crisis: Why you Should Consider Conducting Replication Research,” APA, March 2020, www.apa.org/ed/precollege/psn/2020/03/replication-crisis, and Ed Yong, “Psychology’s Replication Crisis is Real,” *The Atlantic*, November 2018, www.theatlantic.com/science/archive/2018/11/psychologys-replication-crisis-real/576223/.
9. Reviews have found that 50 to 80 percent of positive results in initial (“phase II”) clinical studies are overturned in subsequent, more definitive RCTs (“phase III”). John P. A. Ioannidis, “Contradicted and Initially Stronger

Effects in Highly Cited Clinical Research,” *Journal of the American Medical Association* 294, no. 2 (July 13, 2005), pp. 218–28.

10. Cheryl Dorsey, Jeff Bradach, and Peter Kim, “Racial Equity and Philanthropy: Disparities in Funding for Leaders of Color Leave Impact on the Table,” The Bridgespan Group, June 4, 2020, p. 15.

11. See https://www.cdc.gov/coronavirus/2019-ncov/vaccines/different-vaccines/how-they-work.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fvaccines%2Fdifferent-vaccines%2Fmrna.html.