Published in final edited form as:
Perspect Psychol Sci. 2013 July ; 8(4): . doi:10.1177/1745691613491579.

Advancing Science Through Collaborative Data Sharing and Synthesis:

NIMH Collaborative Data Synthesis for Adolescent Depression Trials Study Team including:

Tatiana Perrino[1], **George Howe**[2], **Anne Sperling**[3], **William Beardslee**[4,5], **Irwin Sandler**[6], **David Shern**[7], **Hilda Pantin**[1], **Sheila Kaupert**[1], **Nicole Cano**[1], **Gracelyn Cruden**[1], **Frank Bandiera**[8], and **C. Hendricks Brown**[1]

[1]University of Miami's Miller School of Medicine

[2]George Washington University

[3]National Institute of Mental Health, Bethesda, MD

[4]Children's Hospital, Boston, MA

[5]Harvard University

[6]Arizona State University

[7]Mental Health America, Alexandria, VA

[8]University of California, San Francisco

Abstract

The demand for researchers to share their data has increased dramatically in recent years. There is a need to replicate and confirm scientific findings to bolster confidence in many research areas. Data sharing also serves the critical function of allowing synthesis of findings across trials. As innovative statistical methods have helped resolve barriers to synthesis analyses, data sharing and synthesis can help answer research questions that cannot be answered by individual trials alone. However, the sharing of data among researchers remains challenging and infrequent. This article aims to (a) increase support for data sharing and synthesis collaborations among researchers to advance scientific knowledge and (b) provide a model for establishing these collaborations using the example of the ongoing National Institute of Mental Health's Collaborative Data Synthesis on Adolescent Depression Trials. This study brings together datasets from existing prevention and treatment trials in adolescent depression, as well as researchers and stakeholders, to answer questions about "for whom interventions work" and "by what pathways interventions have their effects." This is critical to improving interventions, including increasing knowledge about intervention efficacy among minority populations, or what we call "scientific equity." The collaborative model described is relevant to fields with research questions that can only be addressed by synthesizing individual-level data.

Corresponding Author: Tatiana Perrino, Department of Public Health Sciences, University of Miami Miller School of Medicine, 1425 NW 10th Avenue, #307F (R-669), Miami, FL 33136, tperrino@med.miami.edu.

Declaration of Conflicting Interests
The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Keywords

collaborative research; data sharing; data synthesis; scientific equity; integrative data analysis

---

Science advances through the replication of scientific results and through research that synthesizes, builds upon, and extends the findings from individual studies. It is not surprising that data sharing and research synthesis have received such widespread attention in the literature (cf. Cooper & Koenka, 2012; Curran & Hussong, 2009; Guttmacher, Nabel, & Collins, 2009; Kell, 2008; Nosek, Spies, & Motyl, 2012; Pashler & Wagenmakers, 2012; Savage & Vickers, 2009). Mounting concerns about a “crisis in confidence” in psychology and other fields have highlighted the need to confirm and critique existing findings through greater transparency in research and through data sharing (cf. November 2012 Special Section of *Perspectives in Psychological Science*). The scientific importance of data sharing among researchers extends beyond the replication of findings, as it can help answer new research questions that cannot be answered when researchers work independently with more limited data-sets. Combining datasets can increase sample sizes and available statistical power, as well as increase the diversity of samples, which allows more robust subgroup analyses. Through the National Institutes of Health’s (NIH’s) leadership, data-sharing efforts are increasingly being established at the start of projects. The NIH now supports several prospectively established data-sharing efforts that are making large datasets available to researchers for analyses, including data from genome-wide association studies and autism spectrum disorder research (National Database for Autism Research, 2011; NIH, 2007)—projects that are helping to accelerate science (National Cancer Institute, 2012).

These large-scale, prospectively planned and coordinated data-sharing efforts are critically important, but they represent major time and financial commitments. Opportunities also exist for sharing data from completed trials, which yield large, combined datasets much more cost effectively than do new trials started from scratch. Recognizing this, the NIH and the National Science Foundation (NSF) have established policies articulating expectations that funded researchers will share data gathered through their grants (NIH, 2003a; NSF, 2011), and they provide useful resources for researchers to address potential barriers to data sharing (NIH, 2009). However, available reports indicate that the sharing of existing data among investigators is infrequent (Savage & Vickers, 2009; Wicherts, Borsboom, Kats, & Molenaar, 2006). Additional approaches are needed to encourage widespread data sharing and synthesis.

The aim of this article is to encourage the establishment of collaborative efforts to promote data sharing and synthesis. We begin by describing recent advances in quantitative methods that facilitate synthesis analyses of combined datasets from multiple trials. We next summarize challenges to data sharing and synthesis, emphasizing the importance of collaborations and community based participatory research to address these challenges. Finally, using the example of the ongoing Collaborative Data Synthesis Study on Adolescent Depression Trials (CDSADT), a study funded by the National Institute of Mental Health, we describe a model for building collaborative data synthesis projects and discuss how these collaborative efforts can overcome barriers to data sharing and synthesis. Preliminary findings from the CDSADT study highlight the potential benefits of these collaborative data synthesis projects.

Methodological Advances for Data Synthesis

Meta-analysis is a synthesis tool that has played an important role in advancing scientific knowledge by integrating existing study results using summary data (see Chan & Arvey, 2012), and it has the advantage of being minimally reliant on data sharing. Yet, meta-analysis has noteworthy shortcomings and a limited ability to address scientific questions of moderation and mediation (Brown, Wang, & Sandler, 2008; MacKinnon & O'Rourke, 2012). Researchers across fields as diverse as developmental and social psychology, epidemiology and clinical trials to name a few, are proposing increasingly intricate models to explain complex psychological and health phenomena that include causal mechanisms leading to outcomes (i.e., mediators) as well as differences in for whom and under what conditions variables are related to outcomes (i.e., moderators; MacKinnon & Luecken, 2008; Rose, Holmbeck, Millstein Coakley, & Franks, 2004; Stice, Shaw, Bohon, Marti, & Rohde, 2009).

Through its use of study-level (e.g., group means) rather than individual-level data, meta-analysis, in particular meta-regression, concentrates on overall impact and pays limited attention to within-trial mediation or moderation. These study-level analyses can lead to an ecological fallacy or the potentially erroneous assumption that an association between effect sizes and study-level variables found in meta-analytic regression is identical to the association between individual-level variables (Berlin, Santanna, Schmid, Szczech, & Feldman, 2002). In fact, study-level analyses have much less power and might produce effects different from those found when data are analyzed at the individual level (Dagne, Brown, Howe, & Kellam, 2013.) Another concern with meta-analysis is that nonsignificant findings are less likely to be published (see Nosek et al., 2012). Thus, meta-analysis that relies only on published findings tends to overestimate the strength of these moderation and mediation effect sizes (Brown et al., 2011; MacKinnon, 2008).

Recent advances in quantitative methods, such as integrative data analysis (IDA; Curran & Hussong, 2009), provide a promising alternative to meta-analysis as a form of data synthesis; however, the sharing of existing individual-level data is required. IDA is the "… analysis of multiple data sets that have been pooled into one" (Curran & Hussong, 2009, p. 81), which increases the available statistical power to answer new and important research questions that cannot be answered adequately by single trials or by independent study teams. IDA—sometimes called individual patient level meta-analysis—extends the idea of meta-analysis as it is inherently multilevel due to the integration of study-level and participant-level data. Multilevel modeling is then used to directly address questions of measure comparability, evaluation of change, and moderation and mediation at both the between-trial and within-trial levels. IDA has been used to synthesize findings from longitudinal observational studies (Bauer & Hussong, 2009; Curran & Hussong, 2009; Higgins, Whitehead, Turner, Omar, & Thompson, 2001; Hofer & Piccinin, 2009), but it is only now being applied to sets of intervention trials. IDA methods can substantially increase power to detect moderation and mediation and increase our confidence in null findings. Pooling samples and conducting IDA can provide other potential advantages for researchers, such as increased confidence in generalization, given that combined samples are more heterogeneous than any single trial. By combining data from multiple trials that assess participants of different ages, IDA techniques have also been successfully used to examine developmental processes across multiple stages of life (Curran & Hussong, 2009). IDA has also shown potential for calibrating results across age in randomized trials (Brown et al., 2011).

Barriers to Data Sharing and Synthesis

The development of novel statistical methods has further strengthened the case for more open sharing of data and for collaborative efforts that bring together the expertise of multiple investigators in a given area of research. Unfortunately, there are several barriers to such efforts, including the tendency of researchers to work independently and in competition with one another, a pattern reinforced by academic and research cultures that incentivize investigators to test specific and novel hypotheses from their own theoretical models rather than to develop a broader understanding of shared and unique findings across studies. We have identified three broad issues that present challenges to sharing and synthesizing existing data: data ownership, data protection, and data interpretation.

Data ownership is a complex topic, and although it is common to speak about "owning data," it may be more accurate, in the realm of grant-generated data, to say that what investigators and their institutions really own are rights to access data. The institution that supported the study—which is typically the actual recipient of the grant from the funding agency—has access rights. Individual research participants may also have rights to their own data. The public, whose tax dollars funded the study, also has rights to the aggregate data (Guttmacher et al., 2009). The investigators who collected the data may have more of a sense of "psychological ownership" than others and they have a vested interest in exploring the data and preventing its misrepresentation.

In terms of data protection, investigators, their institutions, and institutional review boards (IRBs) have a responsibility to protect the privacy rights and welfare of study participants as outlined in the study protocol and informed consent process. However, explicitly restrictive consent form language may make data sharing impossible. Investigators with NIH grants, especially those whose grant terms and conditions detail a data-sharing plan, have a responsibility to ensure that the informed consent language is consistent with data sharing and avoids excessively restrictive terms (e.g., specifically guaranteeing that data will not be shared with others; NIH, 2004). For recent studies, excessively restrictive consent language should be unusual because many funding institutions, professional organizations, and journals have policies that require data sharing.

As data are gathered, another barrier to successful data sharing is the need to develop a common understanding of each dataset. Existing documentation may not be sufficiently complete for outsiders to fully grasp the intricacies of the data or to permit further analyses or proper interpretation of the findings. Individual investigators and their teams are the experts in their own trials and should be key informants in interpreting findings. Yet in typical data sharing, their involvement ends once they share their data, which is a considerable disadvantage as far as the interpretation of findings.

In addition to the challenges involved in data sharing, there are also challenges to data synthesis. Despite the methodological advances that have been made, several analytic challenges remain. These include harmonizing different measures across different trials, accounting for differences in samples, and accounting for differences in the delivery of the interventions themselves. To resolve these challenges, statisticians and methodologists need the substantive and practical expertise of the individual investigators who developed, implemented, and evaluated the trials, as they can communicate about issues they experienced in the delivery of their interventions and about the assessment and analysis of their data.

Precisely because of these barriers to data sharing and synthesis, as well as the fact that many individuals and groups have an interest in and expertise regarding the data, it makes sense to discuss data sharing and synthesis efforts in terms of partnerships and collaborations. *Collaborative data synthesis*, the framework advanced in this article, provides a useful model for building research collaborations across multiple research teams for data sharing and synthesis. This model, its strengths, and its challenges are described through the case example of the CDSADT study. Although this article describes issues that have arisen in this specific study, many of these issues are highly relevant and applicable to others who wish to develop a collaborative data synthesis study in another research area.

The CDSADT Study

The CDSADT study is an ongoing collaborative data synthesis project that has developed partnerships among researchers testing interventions to prevent or treat adolescent depression and has established collaborations with groups who have an interest in the data and study results, such as community leaders, practitioners, and advocates. This study aims to develop and refine statistical models to synthesize research findings in randomized trials with longitudinal outcomes, to apply these methods across multiple randomized trials of depression interventions in adolescents using individual-level data, and to develop scientific guidelines for conducting the next generation of intervention trials. The long-term goal of this work is to inform the adolescent depression field about what works and for whom, as well as the underlying mechanisms by which interventions work.

Randomized controlled trials have established that there are efficacious interventions to prevent and treat adolescent depression (Beardslee, Gladstone, Wright, & Cooper, 2003; Brent et al., 1997; Clarke et al., 2001; Compas et al., 2009; Dishion, Kavanagh, Schneiger, Nelson, & Kaufman, 2002; Garber et al., 2009; Gillham et al., 2007; March et al., 2004; Prado et al., 2007; Rossello & Bernal, 1999; Sandler et al., 2003; Spoth, Trudeau, Guyll, Shin, & Redmond, 2009; Young, Mufson, & Davies, 2006). Meta-analyses and descriptive reviews of research across some of these trials have been published (Horowitz & Garber, 2006; Merry, Hetrick, Cox, Brudevold-Iverson, & McDowell, 2011; Stice et al., 2009; Substance Abuse and Mental Health Services Administration, 2011; National Research Council [NRC]/Institute of Medicine [IOM], 2009a; National Research Council [NRC]/Institute of Medicine [IOM], 2009b). These individual trials vary substantially on important factors, including the populations studied and the types of interventions employed. Existing evidence suggests that these interventions are not equally efficacious for all adolescents, including youth from different racial and ethnic backgrounds, and the mechanisms by which these interventions work are not completely understood (Cardemil, Reivich, & Seligman, 2002; Horowitz & Garber, 2006). As a result, research questions about moderators and mediators of intervention effects have become increasingly important. By combining data across multiple studies, synthesis analyses can increase sample sizes of select subgroups of youth who are often underrepresented in research studies, such as those from ethnic minority backgrounds. These analyses have the potential to answer pressing research questions about how to better address health disparities or inequities. As such, synthesis work can help promote “scientific equity,” which we define as equality in the amount of scientific knowledge that is produced to understand causes and solutions to health disparities, which can ultimately lead to improvements in the efficacy and reach of interventions.

Strategies for Building a Collaborative Data Synthesis Project

The steps involved in establishing a collaborative data synthesis study are illustrated in Figure 1, the Collaborative Data Sharing and Synthesis Model.

Identifying a research question

An initial step in collaborative data synthesis is specifying an important research question in the field that can be addressed through a data sharing and synthesis study. As noted, many substantive areas within and outside of psychology must contend with the fact that individual studies are underpowered to properly test complex causal models that can explain behavioral, psychological, and health phenomena. There are also clear scientific benefits to combining the expertise of leading investigators in a field to answer critical questions that can elude individual research teams and can slow a field's research progress (Brown et al., 2011).

In the CDSADT study, the research questions were identified through carefully reading the existing literature on adolescent depression interventions and through discussions with leading substantive and methodological experts about what was needed to better address the problem of adolescent depression. For this study, having a clearer understanding of the differential effects of interventions across subgroups and of the mechanisms by which interventions work has the potential to improve interventions and to reach more adolescents in need of prevention or treatment. In any substantive area, identifying a research question that investigators and community stakeholders need answered also acts as an important incentive to motivate collaborators to participate in collaborative synthesis.

Identifying stakeholders

The formative work of collaborative data sharing and synthesis also involves the identification of "stakeholders"—individuals who have an interest or a stake in the data (*data stakeholders*) or who have a stake in the findings of the data synthesis study (*stakeholders in the synthesis findings*). In the CDSADT study, the stakeholders are the CDSADT study's facilitating team, individual study investigators and their institutions' IRBs, community advocates, adolescent depression healthcare providers, and adolescents at risk for depression and their families. In this study, the facilitating team is the group of CDSADT study investigators and consultants who originally proposed the project through the grant application and who are responsible for developing and refining the methods, recruiting and engaging the collaborators and other stakeholders, gathering and organizing the deidentified datasets, and conducting the analyses per the grant's narrative—all in close collaboration with the other stakeholders. The individual study investigators are the investigators who share the datasets from their intervention trials with the CDSADT study for the purpose of synthesis analyses. These investigators have a clear stake in their data, but also have a stake in the findings of the CDSADT study. The synthesis analyses have the potential to advance their own research and interventions by clarifying which groups benefit from different interventions and how these interventions operate. Thus far, the facilitating team has identified 42 intervention trials for adolescents through a literature search (23 prevention and 19 treatment trials) and has contacted the investigators who conducted these trials. This project is currently in progress, and the initial focus has been on building collaborations among the investigators of the 23 prevention trials. Both mental health advocates and mental health service providers have been identified as important stakeholders in the synthesis study's findings because of their need to apply the findings to shape practice and to advocate for the next generation of programs to prevent and treat depression. The interest of community members is to promote research that improves interventions for adolescents who have depression or who are at risk for depression. All of these groups are considered to be research partners in the collaborative data synthesis project and are expected to share knowledge and build upon each other's expertise to improve the research and utility of findings.

Using a community based participatory research framework

The barriers to collaborations and the sharing of existing data (i.e., data ownership, protection, and interpretation) are difficult to overcome in the absence of strong working partnerships with the stakeholders. In the CDSADT study, the research team has moved away from an approach in which the facilitating team collects and analyzes data independently. Instead, a participatory method involving collaborative data synthesis has become the goal. In this approach, data stakeholders have been invited to participate actively in virtually all phases of the research, including the articulation of research questions, interpretation and publication of results, and dissemination of findings. The building of strong partnerships with stakeholders throughout the research process is conceptually similar to the community-based participatory research process in which academic researchers join with community stakeholders to collaborate as active research partners (Israel, Eng, Schulz, Parker, & Satcher, 2005).

Building and maintaining partnerships

The work of developing partnerships involves individual discussions with stakeholders to involve them actively in the collaborative data synthesis study by establishing shared goals for the overall project and goals for the stakeholders themselves. In the CDSADT study, common goals emerging from these discussions have covered substantive issues as well as methodological or process issues, all of which relate to improving adolescent depression interventions. Among the goals that emerged are the development of new quantitative methods required to analyze aggregated data, the production and dissemination of new findings related to the aggregated set of studies that emerge from data synthesis, and the establishment of priority and best-practice recommendations for the adolescent depression field to provide guidance in the next generation of studies and in practice improvement.

The CDSADT study's facilitating team has supported the aspirations and goals of individual investigators in mining their own data and ensured that these investigators receive recognition for that work. The potential that this collaborative study holds for the advancement of the field, in particular the answering of research questions that cannot be answered through individual efforts, has generated significant enthusiasm and participation among the stakeholders. Although data stakeholders are not responsible for analyzing data, they provide input about specific features of their trials, such as details about measurement of constructs and implementation of interventions, and they help develop hypotheses for the synthesis study. They also review and interpret analyses and have a role in disseminating findings so that they can be incorporated into the next generation of trials on adolescent depression. As stakeholders have come to consider themselves an integral part of this project, open discussions have helped bridge differences and overcome challenges to collaboration.

Thus far, verbal agreement to collaborate in the CDSADT study has been obtained from the investigators of 22 of the 23 prevention trials identified. These investigators have joined quarterly conference calls and shared their ideas about the most pressing research questions for the field and outcomes they would like to see from this collaboration. At the time of this writing, the study has received deidentified datasets for 14 of these prevention trials. A conference symposium conducted at a national research meeting brought together leading investigators and stakeholders to share their findings and to identify research gaps that need to be addressed by future research (Howe & Perrino, 2012). Investigators for the other trials are in the process of obtaining internal approvals to share data from their IRBs, institutions, or other collaborators. Two investigators have agreed to share their datasets following the completion of their data collection and the publication of initial findings. In total, data for approximately 6,100 individual participants across the different studies have been gathered, allowing the CDSADT study's facilitating team to begin the development, refinement, and application of new statistical methods to these data. It has also allowed the study to conduct initial integrative data analysis using the data that have been shared with the CDSADT study, as described later in this article. Community advocates and representatives have been involved to identify research questions that have important implications for mental health services.

Securing funding to support collaborative data synthesis

All of the steps in the process of developing these research collaborations for collaborative data sharing and synthesis require time and resources. The funding provided by NIMH has been a critical resource for the CDSADT study as it has allowed the facilitating team to contact and engage collaborators and stakeholders, address issues of data rights and data protection through drafting data sharing plans, gather and manage data, develop and refine statistical methodologies, conduct analyses, and work with the collaborators to interpret findings and delineate new research questions. Although collaborative data synthesis is more cost effective and less time intensive than conducting a new study with a sufficiently large and diverse sample to address pressing research questions, it does have financial and time investments associated with it. However, there is a strong interest from funding agencies to capitalize on existing data. Indeed, funding for secondary data analyses and synthesis is available through various grant funding mechanisms.

Addressing Barriers and Incorporating Incentives to Collaborative Data Synthesis

The CDSADT study's formative work of collaborative data synthesis identified barriers and incentives to data sharing on the part of the different stakeholders that can provide valuable information for others seeking to build their own collaborative data synthesis projects.

Individual investigators

Though individual study investigators have expressed a strong interest in the scientific benefits of data sharing through CDSADT and the opportunity to better understand the effects of these interventions, some expressed reservations about data sharing—they felt that it would be time-consuming, that their data might not be correctly interpreted, and in particular that they wished to protect their intellectual investments and ensure that the CDSADT team did not publish findings that the individual study team was planning to publish.

To protect the individual teams' intellectual investments, the facilitating team worked to clearly differentiate the CDSADT study's research questions from the individual investigators' research questions, committing to exclusively investigate research questions across aggregated datasets as opposed to within individual study datasets. Thus, the facilitating team has supported individual investigators in continuing to analyze their data and receiving recognition for their work and has emphasized how the synthesis work can produce findings that will ultimately inform and advance the investigators' work.[1] Furthermore, all stakeholders have been invited to quarterly group telephone calls, which are used to provide updates on the analytic progress, facilitate continued reciprocal feedback between stakeholders and the facilitating team, and provide an opportunity to monitor investigators' commitment to the project. The use of data sharing agreements—delineating how areas of possible contention will be addressed—has helped to alleviate concerns and articulate benefits to individual research groups. These plans have also ensured that the process is transparent and allows for input at key junctures.

The correct interpretation of data has been another concern of investigators. This has included the possibility that the facilitating team might not fully understand assessment measures, interventions, and populations being studied and might misinterpret the effects of

---

[1] A unique feature of behavioral interventions, in comparison with pharmacological interventions as identified in the CDSADT study, is that investigators of behavioral interventions have been willing to share the parts of their datasets that have been sufficiently analyzed, delaying full data sharing until all of their primary analyses are completed.

the intervention. Synthesis findings are checked carefully against existing publications, and individual investigative teams are contacted when clarification is needed. In addition, concerns about scientific validity have been addressed during the regular stakeholder conference calls when the facilitating team has an opportunity to further understand the data, studies, and stakeholders and ensure the proper interpretation of data analyses.

Providing the opportunity for investigators to remain involved in the CDSADT study has helped alleviate many of the expressed concerns and has significantly enhanced the quality of the study. As experts in their own studies and in the field of adolescent depression, individual investigators possess essential information and experience to guide and interpret the analyses, thus enriching the research findings and laying the groundwork for disseminating results. The investigators function informally as a scientific advisory board by raising questions important to the field as a whole. The inclusion of collaborators and stakeholders as coauthors on publications has been important in incentivizing participation and building collaborations.

Investigators’ practical concerns have also included limited time availability to prepare the data for sharing. The CDSADT facilitating team has helped by taking responsibility for some of these tasks, such as preparing responses to local IRB concerns and reviewing existing data documentation, rather than relying solely on the limited resources of individual trial investigators. When data preparation tasks have been especially burdensome and local resources unavailable, financial support has been provided to help prepare datasets for transfer, a special concern for older datasets.

IRBs

Universities and health institutions, as well as their IRBs, are responsible for ensuring that human subjects’ rights and welfare are protected, and data sharing requires an examination of how these rights will be protected (especially the right to confidentiality). For the CDSADT study, approval for the synthesis study was first obtained from the University of Miami’s IRB to permit the collection and analysis of deidentified data from the individual trials. Subsequently, individual investigators were asked to contact their own local IRBs to seek guidance on any specific local requirements prior to sharing data. This procedure ensured that a local IRB’s interpretation of protection requirements was honored by the CDSADT study in the use of data from studies approved by that institution. Variability has been found across IRBs in terms of requirements. Some IRBs determined that, given the involvement of the University of Miami IRB and the fact that shared data would be deidentified, there was no need for additional local IRB review. Other IRBs had very specific questions about the study and required formal submissions to their boards regarding subject protections in the form of official memos or protocol amendments. Common concerns included whether data would be shared with third parties, how data would be deidentified, what types of data security would be in place, and what would be done with the data after the synthesis analyses were complete.

Communications with local IRBs similarly involved a strong participatory process. It was important to establish that the protection of the rights and welfare of study participants was a shared commitment of the CDSADT study, and the facilitating team worked with both trials’ investigators and IRB staff to understand each IRB’s concerns. Openness to such concerns was important in gaining IRB participation and approval. The CDSADT study’s facilitating team followed the recommendations set forth by NIH in their “NIH Data Sharing Policy and Implementation Guidance” document, which provides extremely useful information on planning and implementing data sharing efforts (NIH, 2003b), including guidance on writing data sharing plans (NIH, 2009). To date, all local IRBs contacted have approved the sharing of deidentified data with the CDSADT study. Certainly, there is a need to harmonize IRB regulations and consent form practices with the data sharing expectations of the NIH and other funding agencies, as well as those of certain journals and professional organizations. As research collaborations and data sharing efforts become more common, and as IRBs and investigators work through these data sharing agreements, procedures should become more consistent.

Practitioners, community advocates, and community members

In any substantive area, there are stakeholders who can apply synthesis findings to address areas that interest or concern them. In the CDSADT study, mental health practitioners, advocates, and community members have a clear interest in researchers' collaborative efforts to share and synthesize data. Sharing and synthesizing data across studies can help advance professional understanding about what interventions work, and for whom, in terms of the prevention and treatment of adolescent depression. Meanwhile, advocacy groups are a portal to access community members, assist in brokering community participation in studies, disseminate findings, and build political will for service system improvement. Community practitioners and advocates hold pivotal roles in translating scientific findings into practice for the benefit of the general public.

The CDSADT study's facilitating team contacted a leading mental health advocacy group to learn about their own questions for research and to seek their collaboration and that of community members as study partners. The advocacy group's representative has become a part of the CDSADT study's advisory board and provides valuable input throughout the study from the perspective of advocacy groups and community members. In particular, he advocates for closing the gap between research and practice; for ensuring that the research being conducted is relevant, accessible, and practical to individuals with or at risk for mental illness; and for adequate dissemination to those who would benefit from the findings, including national and state policy audiences charged with health system oversight and improvement. This is particularly critical given the current efforts to implement both the 2008 Mental Health Parity and Addictions Equity Act (MHPAEA, 2008) and the 2010 Patient Protection and Affordable Care Act (PPACA, 2010), both of which have critical components related to treatment and preventive services.

Given their direct experiences coping with depression and its treatment, adolescents with depression and their families have perhaps the greatest stake in the outcomes of synthesis studies such as the CDSADT study. The ability to pool and analyze data from multiple research studies is important to help clarify findings about the safety and efficacy of interventions. As a result, the CDSADT study team also worked with a national, nonprofit organization that is dedicated to helping families identify, address, and cope with depressive disorders. Discussions with this organization's director have accentuated challenges that adolescents and their families encounter while seeking treatment, their need for greater clarity in understanding which treatments are most effective for which youth, and the need for a more systematic approach to treatment.

Preliminary Evidence of the Promise of Collaborative Data Synthesis From CDSADT

As the CDSADT study's collaborations continue and data are combined, the work of developing and refining more effective statistical methods and more comprehensive theoretical models on adolescent depression has been moving forward. Initial analyses have examined moderators and mediators of intervention effects across trials using subsets of data that have already been shared (Howe & Perrino, 2012). From a methodological perspective, the goal of these preliminary analyses has been to test IDA's power to detect these effects when individual trials are underpowered. The preliminary analyses described below indicate that IDA can be instrumental in detecting moderation and mediation effects.

Analysis #1: Prevention studies

Together with the original study investigators, the CDSADT team combined data from three separate trials of a family-based preventive intervention to examine the intervention's effects on adolescent internalizing symptoms or depressive and anxiety symptoms. This intervention had already been found to be efficacious in preventing adolescent drug use and sexual risk behaviors, but its effects on internalizing symptoms had not yet been examined. Our collaborators were interested in examining whether levels of parent–adolescent communication at baseline influenced the effects of this intervention on adolescent internalizing symptoms. Family functioning and parent–adolescent communication were among the key targets of this intervention, given that positive parent–adolescent relationships have been important in the prevention of depression and other adolescent problems (NRC/IOM, 2009b). The hypothesis was that baseline levels of family communication would moderate the intervention's effects or, more specifically, that the effect of the intervention on internalizing symptoms would be greater for adolescents in families with poorer baseline communication than for those in families with better communication.

When analyses were conducted separately for the three trials, there were no significant moderation effects. However, when the three studies' datasets were combined and analyzed together using IDA, a significant interaction effect was found (Perrino et al., 2012). As predicted, the family-based intervention was more effective in reducing adolescent depressive symptoms in families with poor communication. Moderated mediation analyses further showed that program-induced changes in parent–adolescent communication mediated the intervention's effects on internalizing symptoms (with stronger effects for those with poorer baseline communication). This finding has been important to the individual study's investigators because it helps identify mechanisms by which their intervention works. Along with exploring research questions important to the original research team, these preliminary synthesis analyses permitted the facilitating team to resolve methodological issues and demonstrate the value of integrative data analyses to the larger collaborative group.

Analysis #2: Treatment studies

Recently, Gibbons, Hur, Brown, Davis, and Mann (2012) reported on an IDA of four adolescent trials of fluoxetine for the treatment of depression—trials that are part of the CDSADT study. Significant main effects of these medications in reducing depression were found for adolescents when the data were combined. Depression scores decreased 16 units on placebo compared to 21.6 units on fluoxetine—a 30% greater improvement. Analyses also examined whether baseline depression levels moderated the effects of fluoxetine on outcomes. Both low- and high-baseline depressed patients benefitted from fluoxetine equally, with the slopes of their depressive symptoms virtually constant across baseline depressive symptoms (Gibbons et al., 2012). Moderation by baseline analyses in the separate trials had too little power to come to any conclusion. Unexpectedly for youth, reductions in depressive symptoms did not mediate the effects of fluoxetine on suicidal ideation, but mediation was found for adults when their trial data were combined. Analysis #2 illustrates an additional advantage of IDA beyond providing greater statistical power to detect effects, which involves increasing our confidence in null findings. For both the moderation and mediation analyses for youth listed above, the confidence intervals were much smaller for the combined analysis than they were for those for the separate trials. These examples, plus theoretical results in Brown et al. (2011), illustrate that IDA can often be used to make much more precise statements about mediators or moderators than can be achieved with single trials.

Conclusion

In a time of expanding data processing capabilities, increased calls for public access to research findings, and increased scrutiny into the outcomes of research investments, scientists, treatment professionals, and advocates have begun to rally around the cause of rapid, widespread data sharing. The critical importance of data sharing has been addressed across disciplines (e.g., Hernan & Wilcox, 2009; Kell, 2008; Piwowar, Becich, Bilofsky, Crowley, & caBIG Data Sharing and Intellectual Capital Workspace, 2008) and the rise of data processing solutions and methodological advances have transformed the potential for research collaborations and synthesis work across different substantive fields. The most direct way to ensure collaboration in data sharing and synthesis is to integrate it into a study as a central research objective during the inception of the study with early and explicit commitment from investigators to collaborate and share data. This approach permits proper planning that can avoid later challenges, addressing issues of participant confidentiality, informed consent, selection of common assessment instruments, and comprehensive data documentation. However, rich opportunities also exist in sharing datasets that have already been collected and combining the expertise of researchers and experts in a particular substantive field for the purposes of synthesis studies. Collaborative data sharing and synthesis that utilizes a community-based participatory research approach provides a model for establishing and maintaining effective working relationships among stakeholders that can overcome barriers to data sharing and synthesis and help answer key research questions that cannot be answered by individual investigators working on their own.

Acknowledgments

We thank Julie Totten, President and Founder of Families for Depression Awareness, for her comments on sections of this article.

Funding

This work was supported by National Institute of Mental Health Grant No. MH040859- “Collaborative Data Synthesis on Adolescent Depression Trials” (C. Hendricks Brown, PI) and National Institutes of Health Grant No. UL1TR000460- “Miami Clinical and Translational Science Institute” (J. Szapocznik, PI).

References

Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. Psychological Methods. 2009; 14:101–125. [PubMed: 19485624]

Beardslee WR, Gladstone TRG, Wright EJ, Cooper AB. A family-based approach to the prevention of depressive symptoms in children at risk: Evidence of parental and child change. Pediatrics. 2003; 112:e119–e131. [PubMed: 12897317]

Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI. Individual patient- versus group-level data metaregressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. Statistics in Medicine. 2002; 21:371–387. [PubMed: 11813224]

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, Johnson BA. A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Archives of General Psychiatry. 1997; 54:877–885. [PubMed: 9294380]

Brown CH, Sloboda Z, Faggiano F, Teasdale B, Keller F, Burkhart G. Prevention Science and Methodology Group. Methods for synthesizing findings on moderation effects across multiple randomized trials. Prevention Science: The Official Journal of the Society for Prevention Research. 2011 Advance online publication.

Brown CH, Wang W, Sandler I. Examining how context changes intervention impact: The use of effect sizes in multilevel meta-analysis. Child Development Perspectives. 2008; 2:198–205. [PubMed: 20585469]

Cardemil EV, Reivich KJ, Seligman MEP. The prevention of depressive symptoms in low-income minority middle-school students. Prevention & Treatment. 2002; 5:Article 8. Retrieved from http://www.clarku.edu/faculty/ecardemil/pdf/The%20Prevention%20of%20Depressive%20Symptoms%20in%20Low-Income%20Minority%20children%202002.pdf.

Chan ME, Arvey RD. Meta-analysis and the development of knowledge. Perspectives on Psychological Science. 2012; 7:79–92.

Clarke GN, Hornbrook M, Lynch F, Polen M, Gale J, Beardslee W, Seeley J. A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. Archives of General Psychiatry. 2001; 58:1127–1134. [PubMed: 11735841]

Compas BE, Forehand R, Keller G, Champion JE, Rakow A, Reeslund KL, Cole DA. Randomized controlled trial of a family cognitive-behavioral preventive intervention for children of depressed parents. Journal of Consulting and Clinical Psychology. 2009; 77:1007–1020. [PubMed: 19968378]

Cooper H, Koenka AC. The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. American Psychologist. 2012 Advance online publication.

Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. Psychological Methods. 2009; 14:81–100. [PubMed: 19485623]

Dagne, G.; Brown, CH.; Howe, G.; Kellam, S. Moderated network meta-analysis with individual participant data. Miami, FL: University of Miami Miller School of Medicine; 2013. Unpublished technical report

Dishion TJ, Kavanagh K, Schneiger A, Nelson S, Kaufman NK. Preventing early adolescent substance use: A family-centered strategy for the public middle school. Prevention Science: The Official Journal of the Society for Prevention Research. 2002; 3:191–201. [PubMed: 12387554]

Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TRG, Iyengar S. Prevention of depression in at-risk adolescents: A randomized controlled trial. Journal of the American Medical Association. 2009; 301:2215–2224. [PubMed: 19491183]

Gibbons RD, Hur K, Brown CH, Davis JM, Mann JJ. Benefits from antidepressants: Synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. Archives of General Psychiatry. 2012 Advance online publication.

Gillham JE, Reivich KJ, Freres DR, Chaplin TM, Shatte AJ, Samuels B, Seligman MEP. School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn resiliency program. Journal of Consulting and Clinical Psychology. 2007; 75:9–19. [PubMed: 17295559]

Guttmacher AE, Nabel EG, Collins FS. Why data-sharing policies matter. Proceedings of the National Academy of Sciences, USA. 2009; 106:16894.10.1073/pnas.0910378106

Hernan MA, Wilcox AJ. Epidemiology, data sharing, and the challenge of scientific replication. Epidemiology. 2009; 20:167–168. [PubMed: 19234410]

Higgins JP, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. Statistics in Medicine. 2001; 20:2219–2241. [PubMed: 11468761]

Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Psychological Methods. 2009; 14:150–164. [PubMed: 19485626]

Horowitz JL, Garber J. The prevention of depressive symptoms in children and adolescents: A meta-analytic review. Journal of Consulting and Clinical Psychology. 2006; 74:401–415. [PubMed: 16822098]

Howe, GH.; Perrino, T. Synthesizing mediation and moderation findings across prevention trials for adolescent depression. Poster Symposium presented at the 2012 Society for Prevention Research Conference; Washington, DC. 2012 May.

Israel, B.; Eng, E.; Schulz, AJ.; Parker, EA.; Satcher, D. Methods in community-based participatory research for health. San Francisco, CA: Jossey-Bass; 2005.

Kell DB. Progress being made on standards for use in data sharing. Nature. 2008; 456:29. [PubMed: 18987714]

MacKinnon, DP. Introduction to statistical mediation analysis. Mahwah, NJ: Erlbaum; 2008.

MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. Health Psychology. 2008; 27(Suppl 2):S99. [PubMed: 18377161]

MacKinnon, DP.; O'Rourke, HP. A new approach to increasing statistical power based on the mediation model. 2012. Manuscript submitted for publication

March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J. Treatment for Adolescents With Depression Study Team. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for adolescents with depression study (TADS) randomized controlled trial. Journal of the American Medical Association. 2004; 292:807–820. [PubMed: 15315995]

Mental Health Parity and Addiction Equity Act. Paul Wellston and Pete Domenici Mental Health Parity and Addiction Equity Act. 2008. Retrieved from https://www.cms.gov/Regulations-and-Guidance/Health-Insurance-Reform/HealthInsReformforConsume/downloads/MHPAEA.pdf

Merry SN, Hetrick SE, Cox GR, Brudevold-Iverson T, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database of Systematic Reviews. 2011:Article CD003380.10.1002/14651858.CD003380.pub3

National Cancer Institute. Advancing scientific progress through genomic data sharing and access. 2012. Retrieved from http://epi.grants.cancer.gov/datasharing/

National Database for Autism Research. Data sharing. 2011 Jul 12. Retrieved from http://ndar.nih.gov/ndarpublicweb/sharing.go

National Institutes of Health. Final NIH statement on sharing research data. 2003a Feb 26. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

National Institutes of Health. NIH data sharing policy & implementation guidance. 2003b Mar 5. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

National Institutes of Health. Frequently asked questions: Data sharing. 2004 Feb 16. Retrieved from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#922

National Institutes of Health. Policy for sharing of data obtained in NIH supported or conducted Genome-Wide Association Studies (GWAS). 2007 Aug 28. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html

National Institutes of Health. NIH data sharing policy. 2009. Retrieved from http://grants.nih.gov/grants/policy/data_sharing

National Research Council/Institute of Medicine. Depression in parents, parenting, and children: Opportunities to improve identification, treatment, and prevention. Washington, DC: Committee on Depression, Parenting Practices, and the Healthy Development of Children; 2009a.

National Research Council/Institute of Medicine. Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities. Washington, DC: Committee on the Prevention of Mental Disorders and Substance Abuse Among Children; 2009b.

National Science Foundation. Other post award requirements and considerations. 2011 Jan. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

Nosek BA, Spies JR, Motyl M. Scientific Utopia II: Restructuring incentives and practices to promote truth over publishability. Perspectives on Psychological Science. 2012; 7:615–631.

Pashler H, Wagenmakers EJ. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? Perspectives on Psychological Science. 2012; 7:528–530.

Patient Protection and Affordable Care Act. (2010). Pub. L. No. 111–148, §2702, 124 Stat. 119. Retrieved from http://www.healthcare.gov/law/full

Perrino T, Pantin H, Prado G, Huang S, Brincks A, Howe G, Brown CH. Preventing internalizing symptoms among Hispanic adolescents: A synthesis across Familias Unidas Trials. 2012 Manuscript submitted for publication.

Piwowar HA, Becich MJ, Bilofsky H, Crowley RS. caBIG Data Sharing and Intellectual Capital Workspace. Towards a data sharing culture: Recommendations for leadership from academic health centers. PLoS Medicine. 2008; 5(9):e183. [PubMed: 18767901]

Prado G, Pantin H, Briones E, Schwartz SJ, Feaster D, Huang S, Szapocznik J. A randomized controlled trial of a parent-centered intervention in preventing substance use and HIV risk behaviors in Hispanic adolescents. Journal of Consulting and Clinical Psychology. 2007; 75:914–926. [PubMed: 18085908]

Rose BM, Holmbeck GN, Millstein Coakley R, Franks EA. Mediator and moderator effects in developmental and behavioral pediatric research. Journal of Developmental & Behavioral Pediatrics. 2004; 25:58–67. [PubMed: 14767360]

Rossello J, Bernal G. The efficacy of cognitive-behavioral and interpersonal treatments for depression in Puerto Rican adolescents. Journal of Consulting and Clinical Psychology. 1999; 67:734–745. [PubMed: 10535240]

Sandler IN, Ayers TS, Wolchik SA, Tein J, Kwok O, Haine RA, Griffin WA. The Family Bereavement Program: Efficacy evaluation of a theory-based prevention program for parentally bereaved children and adolescents. Journal of Consulting and Clinical Psychology. 2003; 71:587–600. [PubMed: 12795581]

Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. PLoS ONE. 2009; 4:e7078. [PubMed: 19763261]

Spoth R, Trudeau L, Guyll M, Shin C, Redmond C. Universal intervention effects on substance use among young adults mediated by delayed adolescent substance initiation. Journal of Consulting and Clinical Psychology. 2009; 77:620–632. [PubMed: 19634956]

Stice E, Shaw H, Bohon C, Marti CN, Rohde P. A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. Journal of Consulting and Clinical Psychology. 2009; 77:486–503. [PubMed: 19485590]

Substance Abuse and Mental Health Services Administration. SAMHSA's National Registry of Evidence-Based Programs and Practices. 2011 Jul 11. Available from http://nrepp.samhsa.gov/

Wicherts JM, Borsboom D, Kats J, Molenaar D. The poor availability of psychological research data for reanalysis. American Psychologist. 2006; 61:726–728. [PubMed: 17032082]

Young JF, Mufson L, Davies M. Efficacy of interpersonal psychotherapy-adolescent skills training: An indicated preventive intervention for depression. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2006; 47:1254–1262.

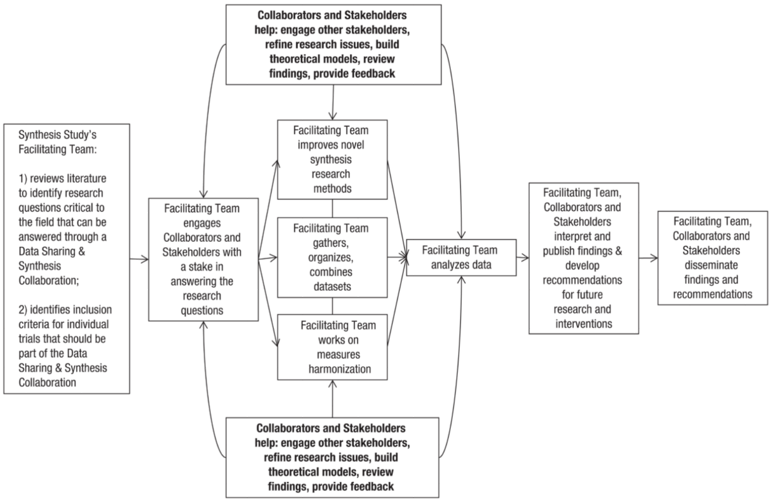

Fig. 1.

Collaborative data sharing and synthesis model.