

FIRST THINGS FIRST EXTERNAL EVALUATION
LONGITUDINAL CHILD STUDY OF ARIZONA
STATUS OF DESIGN, SAMPLING AND DATA COLLECTION
AND PROPOSALS FOR ANALYSIS

July 1, 2011

Prepared for

**The Arizona Early Childhood
Development and Health Board**

First Things First

Steve Lynn, Chairman
Rhian Allvin, Executive Director
Amy Kemp, Senior Director, Research
and Evaluation

Prepared by

**The First Things First External Evaluation University
Consortium and Leadership Team**

University of Arizona

Ronald W. Marx, Principal Investigator
David Yaden
Doug Taren
Chris Cutshaw
Raquel Bravo-Clouvet

Arizona State University

Nancy Perry
Beth Swadener
Billie Enz
Kathy Nakagawa

Northern Arizona University

Larry Gallagher
Jennifer Prior
Robert Horn

INTRODUCTION

This report updates earlier documents written in 2008 and 2009 in which the design, sampling and analysis plans of the Longitudinal Child Study of Arizona (formerly called the Longitudinal Child Cohort Study) were outlined prior to the actual collection of data. This current discussion of progress in the LCSA, whose central purpose is to assess the impact of First Things First (FTF) funded programs and services during early childhood and later upon children's health, readiness for and subsequent success in school, is written after a year's worth of data collection from over 6,000 children, ages 2 months - 66 months, across the state.

The university evaluation team is encouraged that during the past year, only slight modifications were made to the initial design which involved (a) extending data collection over the course of a full year following recommendations from the first FTTEE Advisory Board meeting in November of 2009, and (b) widening the age ranges of children from whom data was collected for primarily logistical reasons as families with more than one child (about 25% of the sample) preferred that each child in the family participate in the study.

Even though the severe economic downturn in Arizona has curtailed many services to families once provided by the state, and certain legislative initiatives possibly have resulted in limiting the enrollment of certain ethnic groups in the LCSA, as well as a much slower rollout of FTF programs over the past two years than anticipated, members of the FTTEE Advisory Board in their recent report reiterated that "it is important to recognize that the suggestions and feedback provided here and during the meeting were provided in the spirit *not of changing or modifying the studies*, but rather with an eye toward analysis and inferences that might be drawn from the completed work."

With these suggestions in mind, after a brief summary of the current design parameters and sampling progress as well as assessment progress, the following sections primarily focus upon additional analytic procedures which will be implemented to further illuminate the impact of FTF-funded programs upon children's cognitive growth, physical health and socio-emotional development.

DESIGN REQUIREMENTS

SCOPE OF WORK – JUNE 8, 2008

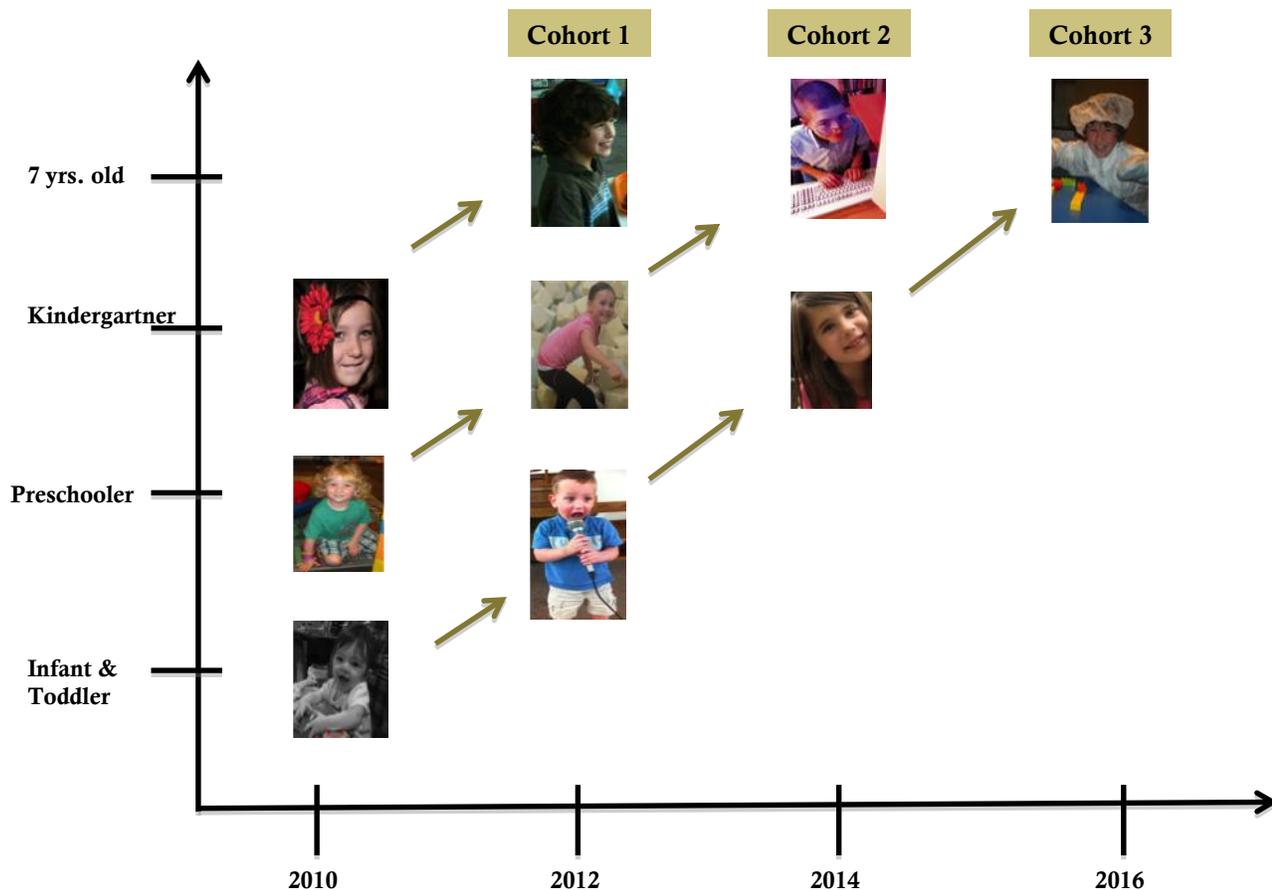
- Design a *longitudinal, matched comparison group study of child outcomes*, with the possibility of multiple counterfactual elder cohorts;
- Conduct longitudinal evaluation of services related to Regional and Statewide strategies undertaken to meet First Things First goals;
- Children *may not be assigned* to treatment or control groups;
- Must be a *representative* sample of the birth cohort population on all appropriate racial, ethnic, socio-economic and geographic variables;
- Methodology may include survey or observational scale of family knowledge & practice;
- Provide information to FTF that will be used to inform the legislature, the Governor, the Board of FTF, . . . and other stakeholders of progress towards achieving *system outcomes*.

Given the study specifications from the original scope of work, the primary design employed was an ***Accelerated Longitudinal Design*** (also called a ***Cohort Sequential Design***). In this design, individual development can be studied over long intervals without the high costs associated with a single cohort since data is gathered during short intervals from multiple cohorts that overlap in certain age bands.

Thus, in the present study three cohorts (infants/toddlers, preschoolers and children beginning kindergarten), will be followed for five years with direct child data collection and family interviews currently planned for in 2010, 2012, and 2014. We aim to collect school records and test scores beginning at third grade and beyond for all cohorts. School achievement scores will be used to assess the long-term impact of high quality early childhood experiences upon later academic and socio-emotional growth.

The cohorts represent samples of children from a variety of child care arrangements (e.g., center, non-center, childcare home, non-maternal, relative, maternal care) and child care environments (e.g., public, profit, non-profit, tribal) across Arizona and include families and children participating in FTF-funded services and those that are not. The final sample to be assessed will total 7,486 children. The measures used for data collection assess development (including motor and physical development), height and weight, language and literacy, socioemotional development, executive functioning and math concepts (for preschoolers and kindergartners) and numerous facets of family life, parenting practices, service use, health and early childcare experiences. A diagram of the design and cohort comparisons appears on the next page.

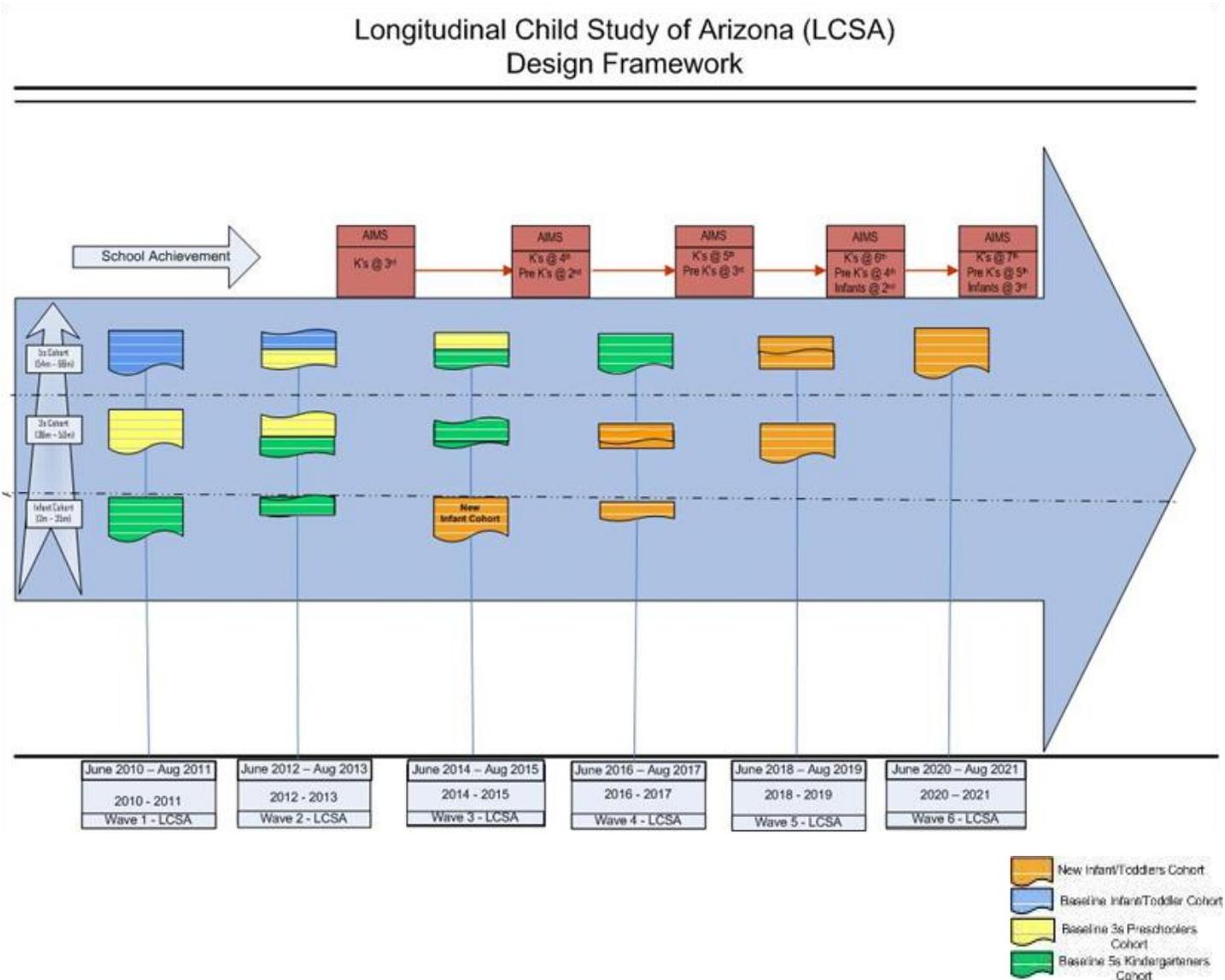
DIAGRAM OF COHORT SEQUENTIAL DESIGN - LCSA



In the cohort sequential design, children can be studied both longitudinally (same children over time at scheduled assessment points) and cross-sectionally (different children at different ages at the same time point). Biennial assessments allow for comparisons of historical cohorts (same age at different times) at preschool ages, kindergarten and seven years of age and should show enhanced child and family outcomes for cohorts having higher engagement levels with FTF-funded services.

Also included in the current design (not shown in the diagram), are simultaneous comparison groups which have been defined to date as children in non-FTF funded centers/homes, parental care, and Family, Friends and Neighbor (FFN) care that are matched with children receiving FTF-funded services in similar arrangements. Some of these children, however, will be exposed to other FTF-funded services (e.g., home visiting, literacy programs, health screenings). Both of these groups will be followed in the years to come.

While the first 13-month data collection establishes an important baseline of child performance prior to First Things First-funded services being implemented at capacity throughout the state, the design will allow for comparisons of not only the impact of FTF's services, but also of the influence of other early childhood programs and services in the state, thus, providing perspective on how the developing early childhood system in the state is influencing children's and families' growth and well-being.



The LCSA sample has three cohorts: Infants/Toddlers (2m-35m), Preschoolers (36m-53m) and Kindergarteners (55m-66m). Data collection in Wave 1 (June 2010 - August 2011) constitutes the baseline measurement prior to substantial penetration of FTF services in the state. Data collection in Wave II will occur from June 2012 – August 2013. As a result of the wider age-band for infants/toddlers, some of the youngest infants in Wave I will still be in the infant/toddler cohort in 2012 while most will move to the preschool group. Similarly, some of younger preschoolers in Wave I will be in the upper range of the preschool band in 2012-13 (Wave II), while most preschoolers will be assessed as beginning kindergartners in 2012-2013.

In 2014 a new Infant/Toddler cohort is proposed to be constituted. Given that FTF-funded services will have been growing in capacity for five years, it is hypothesized that this group will show enhanced achievement over all other cohorts as they move forward in time. Finally, to estimate the contribution of high quality early childhood experiences, school achievement scores (AIMS) will be gathered from third grade on for each cohort. The first cohort to be measured at third grade (Spring 2013) will be kindergartners from Wave I.

DATA COLLECTION STATUS - LCSA**NUMBER OF CHILDREN ASSESSED STATEWIDE
AS OF JUNE 24, 2011**

	Infants & Toddlers	Preschoolers	Kindergartners	Total Collected	Goal	% of Goal
Child Assessment Total	2839	2134	1484	6457	7486	86%
Parent Interview Total	2831	2069	1455	6355	7486	86%

NUMBER OF FTF-FUNDED PARTICIPANTS***

Target FTF-funded Sample	= 4492	(60% of 7486)
FTF-funded as of May	= 1,882	
FTF-funded by end of August	= 2,258*	
15% from non-FTF	= 360**	
Total FTF sample estimate	= 4,490	

•Based upon 5,228 participants. This number is always lower than the total number reported from the field since there is a lag between when assessment packets are verified and entered into the database (ECOLE). All remaining children and families to be assessed through the end of August will be participants in FTF-funded programs.

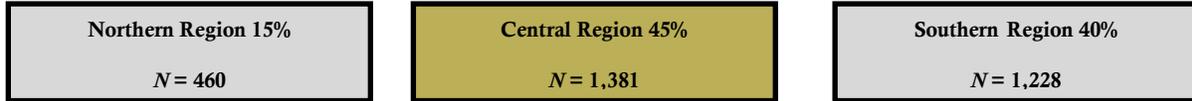
** Conservative estimate of percent of FTF-funded participants identified from the non-FTF sample across all categories. Actual percentage may be as high as 33% in which case the proportion of FTF participants would exceed 60%.

***See Appendix A for detailed percentages of FTF- and non-FTF-funded participants in each category of child care arrangement.

EXAMPLE OF TARGET NUMBERS AND CATEGORIES OF INTEREST FOR THE INFANT SAMPLE (Similar for preschool and kindergarten cohorts)

INFANT POPULATION* N= 92, 947

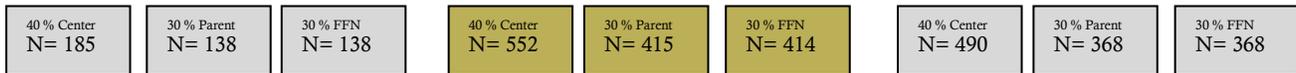
LCSA Total Sample Size N= 7,486
LCSA Infant Sample Size N= 3, 069 (41%)



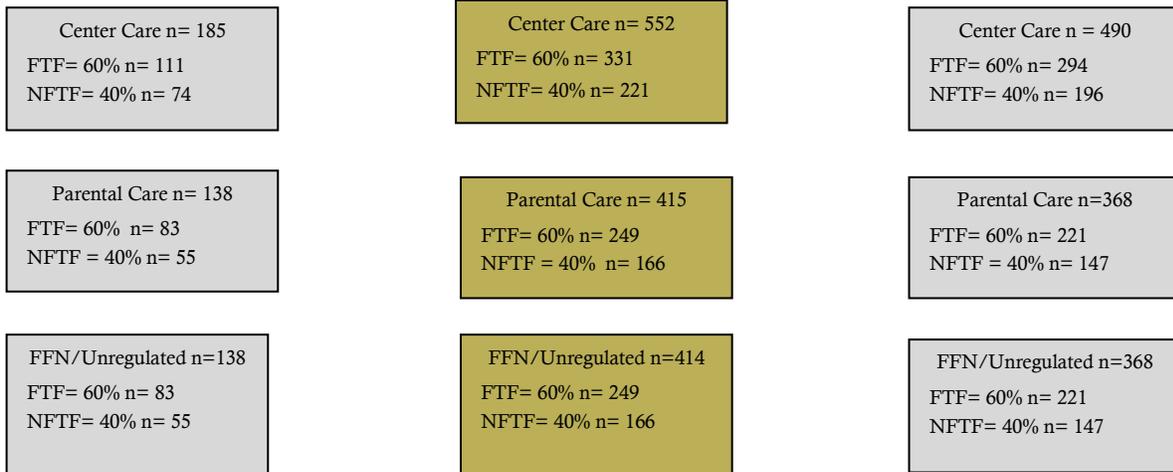
Stratification by Locale



Stratification by Type of Care



Stratification by FTF Funding Strategies



STRATEGIES FOR ANALYSIS

--Multi-level Linear Models --SEM
 --Propensity Scores --Weighting
 --Sensitivity Analysis --Imputation

The challenges confronting an evaluation of such a multi-layered, complex system of integrated health care and education have been discussed at length in the literature as well as by FTFF's Advisory Board. These challenges include, but are not limited to, some of the following:

- Services and programs are unequally distributed across geographical areas influencing the level of access of certain participants;
- Use of services is mediated by individual, family, community, social and political factors which may result in the underrepresentation of some groups in the sample;
- Participation in services will vary in frequency, duration and intensity, and individual's participation is not independent of their need for services, thus complex models of exposure to FTF services will be needed (rather than "more services=better outcomes", for example);
- The degree to which individual services are implemented with fidelity will impact service quality, and the quality of FTF-funded services will likely impact child and parent outcomes; if data on implementation and quality are unavailable, efforts to assess aggregate services will be hindered.
- Participants may receive the same type of service from multiple providers that vary in how they provide that service making it difficult to define the precise nature of the services themselves, (e.g. home visiting);
- Lack of randomization of participants to services or the participant's ability to choose whether to participate in a given service makes it impossible to attribute changes in outcomes to a given service in and of itself.

All evaluation designs must consider the *unit of analysis* (the persons or things being studied). Thus, the evaluation of FTF provides a unique opportunity to look not only at children, but also at households and families, the schools children attend, service providers, and the regions in which they use services.

To evaluate how FTF impacts children, we must consider the hierarchical nature of these various units of analysis. However, examining data organized into hierarchies as if they were all at the same level leads to statistical errors and, more importantly, to interpretation errors. We believe that it is critical to analyze the data when possible using hierarchical statistical methods.

Multilevel linear modeling is a powerful statistical method used for "nested" evaluation designs. In this study, we will measure outcomes for infants who live within households, which are organized (nested) within regions and census locales. For our preschool and kindergarten cohorts, outcome variables will be analyzed for children within schools or centers which are, in turn, nested within census locales and regions of the state. This method makes it possible to examine separately the influence of the different levels.

STRATEGIES FOR ANALYSIS (CONT'D)

Intent-to-Treat Approach

Both First Things First and SURE START, a national program targeting families and children under five in the most distressed areas of the United Kingdom, are community-based initiatives (CBI) where each person in the community is a potential beneficiary of the multiple programs and services which are offered. However, in actual practice it is the case that some persons in need of services do not use them and when they do use them, the frequency, duration and intensity will vary.

While these participation factors may not seem surprising, they pose great challenges for evaluation because choices need to be made as to (a) what facets of the intervention can be evaluated, (b) what level of participation is needed to create a measurable impact, and (c) what factors lead persons to choose the particular service. Did they indeed have to seek it out or was participation mandatory? Was it by chance that they became aware it or were there recruitment efforts?

These varying types of circumstances, called *selection factors* have been found to confound interpretations of outcomes since the selection factors correlate with the indices of program performance. Thus, ***an intent-to-treat approach*** takes into consideration, both theoretically and statistically, the variation in participants' levels of engagement with service provision.

Propensity Score Matching

As pointed out by the evaluators of SURE START propensity scoring is a way to equate persons in "treated" vs. "untreated" (or alternatively treated) groups on multiple characteristics which they share in common, albeit in different degrees. Persons with similar propensity scores, then, can be considered roughly equivalent. In lieu of randomization, propensity scores allow for the comparison of groups to determine the impact of an intervention in which persons have participated.

The following are general examples of some of the 85 characteristics which SURE START used to calculate propensity scores to equate their groups. FTFEE believes that many similar ones can be derived from the *Parent Child Interview* which might be used in creating comparable groups. They include:

% of ethnicity distribution, % of single parent families, % of non working parents, % of long term unemployed, education level, health characteristics, income level, geographic area, type of occupation, % with no personal transportation, % who rent, % of home owners, % persons in household.

STRATEGIES FOR ANALYSIS (CONT'D)

Sensitivity Analyses and Structural Equation Modeling (SEM)

Complex interventions such as First Things First, Smart Start and SURE START fundamentally depend upon an underlying logic model in which assumptions are made about the kinds of interactions which occur between specified components in a system as well as what components might contribute more to the outcomes sought.

For example, First Things First's approach to families and children is described as *child- and family-centered, community-based, culturally responsive, collaborative, high quality*. In addition, several broadly defined components of the early childhood system have been specified such as *leadership and governance, standards of practice, public awareness, qualified and well-paid workforce, adequate and secure funding* as well as others.

Sensitivity Analyses and **SEM** are techniques of statistically modeling these components in various configurations to determine which ones contribute more to certain outcomes. While graphical models may depict all components as having equal importance in producing the desired outcomes for families and children, sensitivity analyses and SEM are helpful in suggesting the more nuanced and causal relationships in the system, depending upon the particular outcomes of interest.

Weighting and Imputation

Where samples are randomly drawn, individual and group characteristics in the population are assumed to be distributed evenly across the comparison groups of interest. However, in CBI studies services accessed and used differently, it is important to properly **weight** key demographic, individual and group characteristics so that interpretations and generalizations in reference to the larger population are accurate.

For LCSA, factors to be weighted will occur at several levels (e.g., geographic region, child care arrangement, and FTF service provision). Other factors may include persons over- or under-representation in the sample, individual and family characteristics and any other variable which is determined to have an influence upon the interpretation of outcomes.

Imputation techniques account for various types of missing data in the sample. Missing data can involve the absence of groups of interest, measures administered, or specific unanswered items on tests. Data also can either be determined as randomly missing or missing because of some factor related to study procedures or individual characteristics which relate to the purpose of the study itself.

In addition to making sure that the LCSA has as little missing data as possible through rigorous quality control in sampling, data collection and comprehensive retention and tracking strategies (see Appendix B for examples), only multiple imputation techniques will be used as necessary.

SUMMARY AND CONCLUSIONS

FTF is a bold, innovative, comprehensive and multidimensional program that needs and deserves an evaluation equal to its vision, sweep, and complexity. The evaluation developed by the First Things First External Evaluation employs multiple methods and measures to assess multiple outcomes across the state over time. The strength of its initial design to accomplish these goals is expressed by its Advisory Board's members at their recent meeting in Tucson (see full report in Appendix C):

The Advisory Group was deeply impressed by the scope of work that has taken place since the initial meeting of the group at the beginning of the project. Substantial progress has been made on the complexities related to hiring and training staff, identifying , piloting and finalizing data protocols, negotiating cooperative agreements with agencies and providers, securing access to participants, collecting data, developing data handling and storage capacities, and conducting preliminary data analysis. Given the number of "moving parts" in this project, we commend the research group on progress to date. (June 16-17, 2011)

The conceptual framework behind this evaluation is based on a logic model of hypothesized causal links in a chain that runs from the creation of FTF; to the Regional Partnership Councils; to the communities and programs they support, create, and integrate into the early care and education system; to the children, teachers and other service providers, and families they serve. Our evaluation to date has gathered substantial amounts of important baseline data prior to the complete rollout of FTF into the state and performed preliminary quantitative and qualitative analyses which are the beginning of further understanding of the current state of children's readiness for school, their families' experiences of accessing a variety of education and health services and the intensity, frequency and duration of those services as a crucial indicator of how outcomes in health, education and early development are influenced.

As many of our Advisory Board members noted, the evaluation of Arizona's First Things First early childhood initiative across the state presents challenging complexities. However, by the application of the analytic techniques briefly described in this paper as well as the integration of information from the sample studied by the Family and Community Case studies, FTTEE feels that the combination of all of these levels and types of analysis makes for an evaluation that is more comprehensive and multidimensional than those being done by other states, which will make Arizona a leader and a model nationwide for not only creating a high quality, coordinated early childhood system but also for understanding how to evaluate one.

REFERENCES

1. Angrist, J., Imbens, G., Rubin, D., (Jun., 1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, Vol. 91, No. 434, pp. 444-455
2. Begg, C., (June 2000). Ruminations on the Intent-to-Treat Principle (Commentary). *Controlled Clinical Trials*, Vol. 21, Issue 3, pp. 241-243
3. D'Agostino Jr, R., Rubin, D., (Sep., 2000). Estimating and Using Propensity Scores with partially Missing Data *Journal of the American Statistical Association*, Vol. 95, No. 451, pp. 749-759
4. Holland, P., (Dec., 1989). Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training: Comment. *Journal of the American Statistical Association*, Vol. 84, No. 408, pp. 875-877
5. Lachin, J., (June 2000). Statistical Considerations in the Intent-to-Treat Principle. *Controlled Clinical Trials*, Volume 21, Issue 3, pp. 167-189
6. Little, T.D., Schnabel, K.U., Baumert, J., (Eds.). (2000). *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Mahwah, NJ: Lawrence Erlbaum.
7. Melhuish, E., Belsky, J., Leyland, A.H., Barnes, J., The National Evaluation of Sure Start Research Team. (Nov 2008). **Effects of fully-established Sure Start Local Programmes on 3-year-old children and their families living in England: a quasi-experimental observational study**. *The Lancet*, Vol. 372, Issue 9650, pp.1641-1647
8. Melhuish, E., Belsky, J., Anning, A., Ball, M., Barnes, J., Romaniuk, H., Leyland, A., The NESS Research Team, (June 2007). Variation in community intervention programmes and consequences for children and families: the example of Sure Start Local Programmes. *Journal of Child Psychology and Psychiatry*, Vol. 48, Issue 6, pp. 543-551
9. Melhuish, E., Belsky, J., Barnes, J., (Nov 2010). Sure Start and its Evaluation in England. *Encyclopedia on Early Childhood Development*. Centre of Excellence for Early Childhood Development.
10. Ribisl, K.M., Walton, M.A., Mowbray, C.T., Luke, D.A., Davidson II, W.S., Bootsmiller, B.J., (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: review and recommendations. *Evaluation and Program Planning*, Vol. 19, No. 1, pp. 1-25.
11. Shadish, W.R., (2002), Revisiting field experiments: Field notes for the future., *Psychological Methods*, Vol. 7, Issue 1, pp. 3-18
12. Shadish, W.R., and Cook, T. D., Comment-Design Rules: More Steps toward a Complete Theory of Quasi-Experimentation. *Statistical Science*, Vol. 14, No. 3 (Aug., 1999), pp. 294-300
13. Shadish, W.R., Cook, T. D., Campbell, D.T., (Eds.). (2002). *Experimental and Quasi-Experimental Designs for a Generalized Causal Inference*. Boston, MA: Houghton Mifflin Co.
14. Sheiner, L. B., and Rubin, D.B., (1995), Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics*, Vol. 57, pp. 6-15
15. Sutherland, C., McCroskey, J., Halfon, N., (2000), The Challenges of Measuring the Impact of First 5. [Policy Brief]. UCLA Center for Healthier Children, Families and Communities. California Policy Research Center. University of California. Number 20.

APPENDIX A

TARGET TOTAL SAMPLE = 7486 ACTUAL SAMPLE RECEIVED = 5525

Infants:

TARGET INFANT SAMPLE/TOTAL SAMPLE		3069/7486 = 41%		% of Infants Collected/Total Sample		2424/7486 = 32%		% of Infants Collected/Total Infants		2424/3069 = 80%	
		40% Regulated care = 1227/3069 60%		Unregulated care = 1842/3069							
CENTERS/HOMES (40%)	n = 1227	% of FTF Ctr/home	PARENTAL CARE (30%)	n = 921	% of FTF Parent	UNREGULATED CARE (30%)	n = 921	% of FTF Unreg	TOTAL	Infant Target	
FTF (60%)	736		FTF* (60%)	553	nFTF (40%)	368	FTF* (60%)	553	368		
	385			401				76			
	736	52.3%		368	72.5%			153	3069	2424	79.0%
	491	21.4		1128							

↑ FTF* service in non QF ctr Unknown FTF funding at this time ↑ Unknown FTF funding at this time

Threes:

TARGET THREE SAMPLE/TOTAL SAMPLE		2546/7486 = 34%		% of Threes Collected/Total Sample		1805/7486 = 24%		% of Threes Collected/Total Threes		1805/2546 = 71%	
		60% Regulated care = 1528/2546 40%		Unregulated care = 1018/2546							
CENTERS/HOMES (60%)	n = 1528	% of FTF Ctr/home	PARENTAL CARE (20%)	n = 509	% of FTF Parent	UNREGULATED CARE (20%)	n = 509	% of FTF Unreg	TOTAL	Three Target	
FTF (60%)	917		FTF (60%)	305	nFTF (40%)	204	FTF (60%)	305	204		
	448			113				14			
	917	48.9%		204	37.0%			102	2546	1805	70.9%
	611	73		632							

↑ FTF* service in non QF ctr Unknown FTF funding at this time ↑ Unknown FTF funding at this time

Fives:

TARGET FIVE SAMPLE/TOTAL SAMPLE		1871/7486 = 25%		% of Fives Collected/Total Sample		1296/7486 = 17%		% of Fives Collected/Total Fives		1296/1871 = 69%	
		60% Regulated care = 1123/1871 40%		Unregulated care = 748/1871							
CENTERS/HOMES (60%)	n = 1123	% of FTF Ctr/home	PARENTAL CARE (20%)	n = 374	% of FTF Parent	UNREGULATED CARE (20%)	n = 374	% of FTF Unreg	TOTAL	Five Target	
FTF (60%)	674		FTF (60%)	224	nFTF (40%)	150	FTF (60%)	224	150		
	291			57				4			
	674	43.2%		150	25.4%			76	1871	1296	69.3%
	449	70		448							

↑ FTF* service in non QF ctr Unknown FTF funding at this time ↑ Unknown FTF funding at this time

APPENDIX B

TABLE 2
COMPREHENSIVE LISTING OF RETENTION AND TRACKING TECHNIQUES FOR PANEL STUDIES

Information collected from participant	
Demographics of participant	<ul style="list-style-type: none"> ● First and last name, middle initial (or name) and all aliases, nicknames ● Intentions to change name ● Social Security number ● Medicaid/Medicare number ● Date of birth and place (city, town, state, hospital) ● Home address(es), mailing address(es) and phone number(s) ● Current and previous occupation, work address(es) and phone number(s) ● Veteran status—if applicable, claim number and dates of service ● Student status, name and address of school, school district ● Driver's license number ● Participant's moving plans in the next year
Demographics of relatives (parents, spouse, significant other, stepparents, siblings, children) ^a	<p>Obtain the following information on at least 2 people:</p> <ul style="list-style-type: none"> ● Full names (maiden or birth names for women if married and changed name) ● Addresses and phone numbers ● Date of birth
Demographics of collaterals ^a	<ul style="list-style-type: none"> ● Name, address, phone number of significant others/friends ● Name, address, phone number of representative payee, state welfare workers, and religious contacts, if applicable ● Name, phone number of landlord, if applicable
Demographics of professionals ^a	<p>Identifying information for:</p> <ul style="list-style-type: none"> ● Community Mental Health Caseworker or Primary Therapist ● Department of Social Services Caseworker ● Parole/probation officer—note times and dates of incarceration ● Medical doctor and/or clinics/hospitals utilized ● Names of shelter workers or shelters the participant frequents
Relatives, correspondents, and professionals	<ul style="list-style-type: none"> ● Contact these people to check the accuracy of information (optional) ● If the participant is institutionalized, verify contacts before their release and discuss with them any conflicting or inaccurate information
Retention and tracking procedures	
Participant	<ul style="list-style-type: none"> ● Request that participant calls project office if any tracking information changes or give them prepaid change of address cards. ● Call or write 2–3 weeks after initial contact (or midway between first follow-up) since the “trail is fresh” and contact again at least 2 weeks before follow-up data point ● After locating the person, go over all current tracking information and add any new information (keep copies of all old information) ● Schedule interview appointments for difficult-to-locate participants on the same day of contact or within 1 week, at a time most convenient for the participant ● Give participant an interview card with the logo featuring: the name, address, and phone number of the study; a description of the incentive for completing an interview (if offered); time and place of next interview; and a reminder to call the office if any location information changes ● Offer to help defray the costs that participants incur for getting to the interview (cab or bus fair, babysitting costs, etc.) ● Send participants cards on their birthdays and any other appropriate occasions
Relatives/correspondents	<ul style="list-style-type: none"> ● If you cannot locate the actual participant, call or write to relatives or correspondents ● If the relative does not know the person's whereabouts ask them if you can call them again in a week or so because people tend to resume contact eventually, also send the collateral an interview card to forward to the participant if they see him or her ● Contact other participants who were in treatment, jail, etc. at the same time (see McCoy & Nurco, 1991; Nurco et al., 1977 for a method to do this and still maintain confidentiality)
Public records	<ul style="list-style-type: none"> ● Telephone—check phone directory, call directory assistance, use criss-cross directory for participant and collaterals ● Mail—contact post-office for change of address information for participants and collaterals, use Postal Service forwarding and record updates, certified mail, registered mail, stamped return envelope ● Government records—check police and prison records, check parole and probation information—check names and addresses of visitors to jails or prisons. Contact marriage license bureaus; city, county, and state tax rolls; drivers license bureaus; social security; welfare; FBI; population statistics. Contact State Department of Public Health or Vital Statistics to inquire if the participant is deceased ● Contact agencies: Alumni offices, local utility companies, high school records, professional organizations, treatment centers, credit agencies, psychiatric institutions, VA hospital

Minimizing Attrition

19

TABLE 2—continued

Information collected from participant	
Neighborhood environment	<ul style="list-style-type: none"> ● Contact next-door neighbors when the participant has moved ● Go to their home and workplace ● Talk to landlord, neighbors, and walk around neighborhood and ask anyone “hanging out” ● Go to corner restaurants, convenience stores
Preventing refusals	
Potential dropouts	<ul style="list-style-type: none"> ● Have same interviewer track participant over time to build rapport or switch interviewers to see if participant likes the new one better ● Provide immediate reinforcement for attending appointments ● Provide snacks and beverages during interviews or treat participant if conducting interview at a restaurant or coffee shop ● If participant has a history of missing interviews, send appointment reminder card or remind by telephone; mention incentives ● Be nonjudgmental and open
Relatives, correspondents, and professionals	<ul style="list-style-type: none"> ● Stress that all information will be held strictly confidential ● Discuss incentives, if any, for participation ● Remind the participant of the importance of the study and their participation ● Have study director personally call or visit participant ● Ask if you can call them back after a few days so they can think about their decision to drop out of the study ● Do not coerce participant
Relatives, correspondents, and professionals	<ul style="list-style-type: none"> ● From the beginning, have the participant inform contacts that researchers may contact them ● Describe, mail, or show in-person the “Release of Information” form that the participant has signed mentioning that person’s name

General notes. Not all of the information featured in this table typically needs to be collected, nor do all of the tracking procedures need to be attempted for all populations. The purpose of this list is to be as comprehensive as possible.

Specific notes. *You must have a signed “Release of Information” for each of these individuals.

References: Bale et al. (1984), Crider et al. (1971), Clarridge et al. (1977), Eckland (1968), Freedman et al. (1980), Gregory et al. (1992), Hastrup and Jepsen (1984), LaPorte et al. (1981), Moos and Bliss (1978), Nurco and Lerner (1971), Nurco et al. (1975), Nurco et al. (1977), Pirie et al. (1989), Sobell and Sobell (1973), and Vannicelli et al. (1976).

From Ribisl, K.M., Walton, M.A., Mowbray, C.T., Luke, D.A., Davidson II, W.S., Bootsmiller, B.J., (1996). Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: review and recommendations. *Evaluation and Program Planning*, Vol. 19, No. 1, pp. 1-25.

APPENDIX C

First Things First External Evaluation Advisory Meeting

June 16-17, 2011

Summary of Advisory Board Feedback and Recommendations

Maryann Santos de Barona¹

Diane Bricker

Robert Rueda

Catherine Snow

Kyle Snow

Jack McArdle (Consultant)

This document represents the feedback and recommendations of the First Things First External Evaluation Advisory Board and Consultant. The group met with project staff and directors over a two-day period (June 16-17, 2011). Prior to the meeting, members were sent the *First Things First External Evaluation Arizona Kindergarten Readiness Study Technical Report* for review. During the meeting, overviews, summaries and progress reports were given on the three simultaneous studies that encompass the larger project: the *Arizona Kindergarten Readiness Study* (AKRS), the *Longitudinal Child Study of Arizona* (LCSA), and the *Family Community Case Study* (FCCS). The Advisory team dedicated time during both days of the

¹ Additional members of the Advisory Committee who were not in attendance at the meeting include W. Steven Barnett, Donna M. Bryant, Patricia Edwards, Joan La France, and Eugene W. Thompson. This document does not include their input. Catherine Snow had to leave before the final Advisory Committee private meeting, so contributed less content to this report than other Committee members.

meeting for consultation and discussion as a group, and were therefore able to develop a set of recommendations with a high degree of agreement. Rather than focus on the detailed issues and questions pertaining to each study, we focus on the “big picture” issues and suggestions that appear to us to be critical to consider for the long term success of the project. We present these recommendations in the following paragraphs.

First, the Advisory Group recognizes that this is an extremely large and complex undertaking. While most evaluation projects of this type are conducted by private “for profit” groups, the current project involves a consortium of three state universities and thirty one Regional Partnership Councils (RPC’s) which are under local control. This unique collaboration introduces complexities that offer both unique challenges as well as unique opportunities. In projects of this scope there are always multiple constituencies and competing interests and viewpoints. Not surprisingly, some of the competing viewpoints were expressed during the Advisory meeting. This is neither unexpected nor necessarily negative however, it is important to recognize the existence of varying perspectives and the opportunities and challenges that such variations may bring to the project. In this context, the Advisory Group was deeply impressed by the scope of work that has taken place since the initial meeting of the group at the beginning of the project. Substantial progress has been made on the complexities related to hiring and training staff, identifying , piloting and finalizing data protocols, negotiating cooperative agreements with agencies and providers, securing access to participants, collecting data, developing data handling and storage capacities, and conducting preliminary data analysis. Given the number of “moving parts” in this project, we commend the research group on progress to date.

Second, during the two-day meeting, it became clear that the larger social, political, and economic context in the state has undergone substantial change in even the short time that the project has been in operation. A partial list of the most important and visible of these changes include the economic downturn resulting in increased job loss, decrease in funding for programs and services, increased levels of poverty and associated risk factors. In addition, the state has been affected by the relatively negative attitudes and legislative initiatives related to immigrants in the state and the (fortunately unsuccessful) challenge to continued funding of First Things First at the ballot box. These circumstances both increase the number and severity of risk factors which impact early childhood development, and decrease the willingness of families to access services and to participate in this evaluation. While these factors are beyond the control of the evaluation project staff, they do form an important backdrop to the conduct of the work and the interpretation of the findings. We believe it is important to capture these factors, if even at the descriptive level. It is possible that work in the *Family Community Case Study* may be most suited to document these factors, especially as they are experienced by participating families and agencies, and we emphasize that they should be an important part of the research record.

In addition to these general observations, there were issues related to specific aspects of the evaluation in the areas of design, sampling and data analysis. These are discussed next.

Design issues

In the current configuration, the studies taken together are descriptive and correlational, with cross-sectional and quasi-experimental components. Without randomized assignment, these designs are not optimally suited to demonstrating causal effects. If it is determined by the research team that causal inferences will be requested at a later time, the Advisory Group suggests that it might be useful to show on a smaller scale that intervention makes a difference. It might be possible to do a smaller study or set of studies that use specific indicators of success, such as “not retained in Kindergarten,” “not referred to special education,” etc. While ethical and policy concerns preclude random assignment to treatment and non-treatment groups, there were some ideas generated at the meeting regarding embedded studies that could be done by differentiating cases where participants get services in what amounts to a random design, i.e. based on random factors such as area of residence, from those where participants actively sought services. We strongly recommend that it would be worthwhile to explore options in this area.

We also hope the project staff will consider whether there are lessons (concerning evaluation design, making causal inferences, and/or communicating to the public about limitations on the kinds of inferences that can be made from the study) to be learned from the national evaluation of the British Sure Start program, which has many of the features (distributed implementation, diverse programs supported, local control, and so on) of First Things First External Evaluation Project, though on a national rather than a state-wide scale. Information about the Sure Start evaluation, which has progressed to the point of producing many articles and at least one book, is available at <http://www.ness.bbk.ac.uk/>.

Sampling issues

The project has done a notable job in recruiting the sample for the evaluation. However, there is an actual threat and a potential threat related to sampling that merits consideration. First, while some cooperation has been garnered from Native American participants for some studies, representative numbers of Native Americans are not included in the sample. This is understandable, given the negative history and abuses which Native Americans have experienced in the past associated with research endeavors. This is complicated by the sovereign nation status for Native American tribal lands. Nevertheless, the absence of this population in the data set could potentially result in findings that present a more positive picture of children’s status across the state than is accurate. We recognize the difficulty in this endeavor, and commend the project on their efforts thus far. However, if there are strategies which might be emulated from projects which have had relative success using innovative ways to secure cooperation, these alternative strategies should be explored. Second, given the political, economic, and social conditions in the state, the issue of participation and attrition may be especially problematic for populations such as low-income Latino immigrants who are undocumented and again offer a picture of child status that is unwarranted. Such conditions should continue to be monitored.

An additional issue for consideration is that the control groups may not be “clean” – that is, there may be families who get needs met from services outside the scope of FTF. It may be important to assess the full array of services families have access to, since this is an important part of the statewide picture. One

possible solution would be to look at specific services such as developmental screening “across the board” without regard to who pays or whether it is FTF-funded.

Data Analysis

During the course of the meeting, it became evident to the Advisory Group that there is a need to demonstrate “what works.” However, it also became clear that more work is needed to define in operational terms the “what,” that is, the content and activities that produce desired change. While a simple input-output model cannot begin to capture the complexity of the project being evaluated, it may be helpful to think in such terms initially. For example, “inputs” appear to vary substantially across the RPC’s. Therefore, it may be necessary to think in terms of “dosage effects” and to create a dosage scale of “FTF exposure” or something similar that would allow inferences about the variety of services to which participants are exposed. In addition, it seems that additional work is needed to operationally define the “outputs” or the expected and desired outcomes. For example the FTF Brochure (*Ready for School. Set for Life. First Things First – Creating the Model Early Childhood System*) references the goal that “All Arizona children by the time they are 5 years old have a solid foundation for success in school and in life...”. However, as desirable as this goal is, it is not clear that there is unambiguous agreement about how success in school might be defined and measured. The initial conceptual models that were developed in early documents may be of assistance in helping to operationalize these constructs. Some of the primary measures used in the study have normative data with already available calibrations of bands of performance and levels of risk that might be used as outcome measures. These already existing bands might assist in making the outcomes easily understandable and accessible to readers without requiring statistical data that might not be easily understandable to all consumers of the work.

An additional point with respect to data analysis is that during the meeting it was indicated that data analysis would commence as soon as the entire data set is complete. We strongly recommend that data analysis begin immediately. It is possible to create models and estimate preliminary effect sizes and patterns, for example, even with the data collected so far. While later estimates will be more precise, we believe that the foundation for later work can begin now. It would also be important to try to model the characteristics of people who have dropped out *based on data collected to date*, and to model the characteristics of people who only contributed partial data *based on data so far*. In more specific terms, we recommend that data analysis commence with the first 50% of cases or if that goal is not already reached, then with the data collected thus far. We recognize that this might entail shifting some project resources; however, we believe that the effort is worthwhile, and will facilitate the ultimate analyses of the full data set.

An additional issue is related to assessing the characteristics of those who drop out of the study. The purpose is not just to use this information for descriptive reporting, but rather to use these data to correct the sample and get a more accurate estimate of program effects. Thus we strongly recommend using selection effects in the ongoing data analysis.

Finally, given logistical, time, and cost constraints, it may be useful to reassess the essential priorities of the work, to determine which are the most important. It would then be possible to use project resources to address these specifically, and permit doctoral and postdoctoral students to address “desired but not essential” data collection on a more leisurely timetable.

General Feedback

The Advisory Group wants to reaffirm both the values and direction of the current constellation of studies. Each is unique and contributes to the overall picture. As an example, the Longitudinal study (LCCS) and the Community study (FCCS) are valuable because the implementation of FTF is in the developmental stages and it will be important to capture that growth and developmental process. If there is a way to link statewide data systems in pursuing the general goals of the project in this regard, it would represent a substantial resource to the project and others interested in this work². Moreover, longitudinal work is important because the literature on early intervention suggests that, while the attenuation of intervention effects after program termination is a continuing issue in the field, nonetheless longer-term follow-up studies find impacts of early childhood interventions in domains like increased rates of marriage, later childbearing, and reduced criminality. If some of the possible effects of exposure to FTF services are distal and long term, these will not be captured by proximal short-term measures but could ultimately be tracked through well-maintained integrated data systems. It is important to recognize that the suggestions and feedback provided here and during the meeting were provided in the spirit not of changing or modifying the studies, but rather with an eye toward analysis and inferences that might be drawn from the completed work.

The Advisory Group would like to emphasize that the evaluation consortium represents an extremely valuable training and research opportunity above and beyond the contractual obligations. The final product will also represent a valuable data set of general interest to the field. It is with this goal in mind that these suggestions and recommendations are presented to the evaluation team.

² A model for such an effort would be the Youth Data Archive founded by the Gardner Center at Stanford University, vide http://gardnercenter.stanford.edu/current_initiatives/youth_archive.html

