Qualitative/Quantitative Synergies in a Random-Assignment Program Evaluation

To appear in T. Weisner (Ed.). Mixed methods in the study of child and family life (in press)

Christina M. Gibson

Greg J. Duncan

Northwestern University

Abstract:

This paper describes the productive synergy that resulted from a mixed qualitative/quantitative approach to evaluating program impacts of New Hope, an anti-poverty program developed and implemented in Milwaukee, Wisconsin in the mid-1990s. An overarching conclusion is that the combination of the two methodologies can indeed enrich evaluation efforts. The assessment of New Hope was enhanced by decisions to randomly sample qualitative cases from the larger population of New Hope participants and to train graduate student research assistants to both conduct the qualitative interviews and analyze the quantitative data. The mixed method approach was helpful for understanding program impacts estimated in the quantitative data and for identifying subgroups for which program impacts were the strongest. The n=43 qualitative data were helpful for identifying some but not all kinds of experimental impacts. Case studies from the qualitative data provided important context for the evaluation report and were used to generate measures for a follow-up survey.

Acknowledgment:

This paper summarizes an evaluation conducted by MDRC under a contract with the New Hope Project, Inc. and in collaboration with the MacArthur Network on Successful Pathways Through Middle Childhood. It was supported by the John D. and Catherine T. MacArthur Foundation, the Helen Bader Foundation, the Ford Foundation, the State of Wisconsin Department of Workforce Development, the William T. Grant Foundation, the Annie E. Casey Foundation, and the U.S. Department of Health and Human Services. Hans Bos, Aletha Huston, Robert Granger, Greg Duncan and Vonnie McLoyd co-directed the quantitative study of child and family portion of the evaluation. Tom Weisner and Lucinda Bernheimer co-directed the qualitative study. We are grateful to the MacArthur Foundation and the National Institute of Child Health and Human Development (1 R01 HD360-38) for supporting the child and family portion of this research.

Qualitative/Quantitative Synergies in a Random-Assignment Program Evaluation

I. Introduction

"I could not be where I am without New Hope" was how Maria described the impact on her life of the New Hope anti-poverty program. She is a bright, articulate mother of two small children living in Milwaukee, Wisconsin. Prior to New Hope, she had wanted to work full-time, but had been unable to afford childcare. Through its provision of a child care subsidy, one of its four financial benefits, New Hope allowed Maria to secure and sustain full-time employment. She quit her part-time job as a cashier in a drug store and launched a career in sales.

What lessons about New Hope can be drawn from qualitative narratives such as Maria's? She was but one of over 1,300 participants in the program, and there is little reason to suspect that her experience was representative of the larger New Hope population. An ethnographic evaluation of the project, limited in case size, cannot identify statistically significant results, nor can it provide an understanding of the likely effects of a large-scale implementation of New Hope. Yet quantitative measures from surveys and administrative data sources not only miss thick descriptions that are provided by the qualitative data, but may also fail to identify key circumstances that are critical for understanding program impacts. Perhaps program evaluations need both narratives and numbers to be complete.

The evaluation of New Hope offered both of these elements (Bos et al., 1999). Survey and administrative data about employment, family functioning and child well-being were gathered from nearly all participants between the point of random assignment and twenty-four months later. Forty-six families, including Maria and her two sons, were selected at random to participate in a longitudinal ethnographic study. In this paper, we demonstrate how the combination of these methodologies greatly enhanced the evaluation of New Hope.

This productive synergy between qualitative and quantitative methodologies was facilitated by two key elements of the evaluation design. We randomly selected our 46 qualitative cases from the larger New Hope population, ensuring that our smaller sample was representative of the larger group. The design was also enhanced because several graduate research assistants received intensive training in both methods and took on the highly productive task of reconciling quantitative and qualitative data.

Our mixed methods enabled us to understand program impacts to a degree that would not have been possible had we been relying on one type of data alone. For example, we used the qualitative data to understand some of the "black-box" program impacts emerging from the quantitative data, one of the most important of which was the program's favorable impacts on the teacher-reported academic performance and positive behavior of boys. We also used the qualitative data to identify important subgroups for which New Hope program impacts were particularly strong. And even with its small (n=43) sample, the qualitative interviews provided estimates of impacts of factors that are difficult to measure in a conventional survey (e.g., family problems that interfere with work).

Insights from our qualitative work also facilitated future New Hope-based research by improving the design of our five-year follow-up survey, and by suggesting topics worthy of exploration with the quantitative data.

At the same time, we discovered limits to this synergy. When considered individually, qualitative cases, no matter how richly explored, do not identify program impacts. Qualitative data can be misused if Procrustean efforts are employed to force participants' stories to fit

particular quantitative results. Complex personal experiences should not be manipulated for illustrative purposes.

We begin with a description of the New Hope program and the key components of its evaluation. We then present the lessons we learned in each phase of the project, from the design of the evaluation, to the analysis of program impacts, to benefits for future work. We conclude with some thoughts on the implications of our findings.

II. The New Hope Program

Enrolling its first volunteers in August of 1994, the New Hope program offered its participants a comprehensive package of benefits. In exchange for a proven work effort of 30 hours a week over a given month, participants were eligible for four benefits. First, a wage supplement ensured that the net income of families increased as they earned more on the job. The supplement did not phase out until families were at 200% of the poverty line and ensured that nearly all working families had incomes above the official poverty line (currently $13,290 for a family of three). Second, the program offered subsidized health insurance through a Health Maintenance Organization. The third benefit was a childcare subsidy that enabled families with at least one dependent child under age 13 to choose any state-licensed or county-certified child care provider, including providers of both preschool programs for young children and extended-day program for school-age children.

Fourth, if a private sector job could not be found, then New Hope participants were entitled to a community service job for two renewable periods of up to six months each. Although paying only the minimum wage, these community service jobs ensured that families could meet their 30+ weekly hour work requirement and thus qualify for the rest of the New Hope benefits. All told, the benefits package cost about $4,000 per year, which is at the high end of what states are spending to implement the reforms required by 1996 federal welfare-reform legislation. A final, less tangible, benefit was that New Hope had competent and caring "project representatives" who offered intensive case management as well as emotional and instrumental support.

Although its designers conceived of New Hope as a permanent package of benefits to which low-income working families should be entitled, budget realities limited its duration to three years. Budget constraints also limited its geographic scope: eligibility for the New Hope Project was confined to two low-income ZIP-Code-defined neighborhoods in Milwaukee. The "north side" neighborhood was predominantly African American; the "south side" neighborhood was primarily Hispanic.

Eligibility for New Hope benefits was restricted to individuals who: i) lived in one of the two targeted ZIP-Code neighborhoods; ii) were over 18; iii) had family income at or below 150% of the poverty line; and iv) were willing to work 30 or more hours a week. Individuals living in the two targeted neighborhoods were informed about the New Hope program in a number of ways, both formal (e.g., through the recommendation of other social service agencies) and informal (e.g., signs in Laundromats). Those interested in the program attended an orientation session at which benefits were explained, but so too was the fact that a lottery would be run and half of them would be assigned to a "control" group that did not receive any of the New Hope benefits. Between August 1994 and December 1995 the New Hope project enrolled and assigned 1,357 participants.

By all accounts New Hope was well implemented (Brock et al., 1997). The randomization process was carried out successfully; the demographics of New Hope participants matched those of their larger neighborhoods; and the program's wage subsidies, community service jobs, HMO and childcare subsidizes were well implemented and readily available.[1] Independent critics have also noted the innovative nature of the program, the widespread local community and political networks supporting New Hope, and the importance to participants of the "case representatives" (Mead, 2000).

III. The New Hope Evaluation

The evaluation of New Hope's impacts on work, family life and child well-being was based on data gathered from both quantitative and qualitative sources. The quantitative data came from survey as well as administrative sources, while the qualitative data were gathered as part of a longitudinal ethnography.

New Hope quantitative data. The evaluation of New Hope's impacts on work, family life and child well-being was based on data gathered from six quantitative sources. First, just prior to random assignment, all volunteer families completed a baseline survey form that provided information on an array of socio-demographic characteristics. Second, an extensive survey administered two years after random assignment asked members of the control and program groups about their employment experiences and work-related outcomes during the time of the intervention. All 1,357 participants filled out the baseline information form; some 1,086 (80%) responded to the two-year survey.

Third, a subset of 812 experimental and control parents also answered additional questions about their family practices and children's well-being for the Child and Family Survey (CFS). Every family who had at least one child between the ages of 1 and 10 at baseline qualified; up to two children were chosen from each family. Of the 812 families, 578 were used in the final analysis.[2] Baseline demographic characteristics of the full and the CFS sample are presented in Table 1 and show that most families had low incomes, single parents, and income from welfare sources.

[1] MDRC oversaw all aspect of the evaluation efforts, and coordinated the efforts of the other researchers involved. Hans Bos, Aletha Huston, Robert Granger, Greg Duncan and Vonnie McLoyd co-directed the quantitative study of the child and family portion of the evaluation. Tom Weisner and Lucinda Bernheimer co-directed the qualitative study.

[2] Some 67 Hmong participants in the program were not included in the family study because of cultural and language differences. Disregarding this portion of the sample, the final response rate was 78%.

Table 1

Descriptive Characteristics at Baseline For the Full Sample, The Child and Family Study (CFS) and the New Hope Ethnographic Study (NHES)

	Full Sample	*CFS*	*NHES*
% Black	51.4	55.0	51.2
% Hispanic	26.5	29.3	34.9
Parent's average age	31.8	29.4	29.6
% Female	71.6	89.8	95.3
% Married parents	21.8	20.8	20.1
% of Parents who have a GED or high school diploma	57.3	59.4	65.1
% of Parents working thirty hours or more	37.5	36.5	39.5
% Earning less than $5000 in previous year	72.2	76.2	81.4
% Families receiving aid	62.9	80.7	83.7
% Three or more children in family	31.5	46.0	44.2
Sample size	*1357*	*745*[a]	*43*

a: 812 participants qualified for the Child and Family Study; 67 Hmong families participants were not included in the CFS study due to cultural and language differences.

Fourth, when parents signed written consent forms, evaluators sent self-enumerated questionnaires to teachers of school-age children (5-12 years old). Teachers were asked to rate the children on a variety of academic behaviors and skills. The teachers were blind to the purpose of the study and informed only that the child was involved in a study about children and their families. Parents granted permission for teachers to be contacted for 557 out of the 627 age-eligible children. Of the 557 possible responses, 418 were returned for a response rate of 75%.

Fifth, a database maintained by New Hope as part of its management information system provided data on the use of benefits by all program participants. Sixth, again using signed consent forms (in this case obtained just prior to random assignment), evaluators obtained administrative data on receipt of benefits from the state TANF welfare and food stamp programs, earnings data reported by employers to the Social Security system and, in aggregated form, state Earned Income Tax Credit payments.

New Hope qualitative data. The qualitative data come from the New Hope Ethnographic Study (NHES). The NHES is a longitudinal study of 46 families, most of who were randomly sampled from the 812 program and control families with children aged 1-10 at baseline. While we would have preferred to begin the NHES when families first enrolled in New Hope, we were unable to secure research funding prior to the program's third year of operation. Thus, the NHES began in April of 1998 and will continue until the spring of 2001. Owing to incomplete information, three families were dropped from the study, which left a total of 43 NHES families: 22 experimental and 21 control group members. Baseline demographic characteristics for the NHES are also presented in Table 1 and are similar to those of the larger set of 812 families in the Child and Family Study. Fieldworkers, in their monthly visits to families, listened to parents tell their stories, conducted participant observation in homes, took families out for lunch and dinner, went with them to church and also visited children's schools.

To guide the topics of these visits, researchers generated a set of topics that explore a family's daily activities and routine as well as beliefs and values (Weisner, 1997). These topics were then organized into a template, which fieldworkers used to structure their notes. In all, there were twenty main headings in the template, as well as additional sub-headings (see Appendix Table A for a list of main template headings). Examples of main headings are "Beliefs about and use of child care"; "Relationships with partner/spouse"; and "Barriers to employment." An example of a sub-heading would be "Alcohol/substance abuse", listed under the "Barriers to employment" main heading.[3] If a template area was not covered on a particular visit, fieldworkers raised these issues on subsequent visits and systematically probed for information, ensuring that fieldwork data were as complete as possible and that there were no "false negatives" in the field data.

[3] For further details on the organization of the ethnography, see Weisner et al. (1999).

IV. Design Lessons

Our mixed method approach was enhanced by several decisions we made about the design of the evaluation. In this section, we discuss the advantages of two key decisions - randomly sampling our qualitative cases and using the same individuals to analyze both types of data.

The wisdom of randomly sampling qualitative cases. Sampling statisticians long ago established the power of sampling at random a very small subset of the population as a basis for obtaining population estimates of interest (e.g., means, correlations, regression relationships). For many purposes, samples as small as 1,000 to 1,500 cases provide acceptably precise estimates of the attitudes and behavior of national populations.

Random selection is a key element of this power. A useful analogy is to an attempt to infer the taste of a pot of soup from one sip. Stirring the soup prior to its sampling ensures that the sip will be representative.

Qualitative researchers rarely adopt this kind of sampling perspective in their case selection, perhaps believing that lessons from large-sample studies do not apply to the much smaller samples typically employed in ethnographic studies, or that attempts to recruit cases randomly will fail more often than not. In some cases, their non-random sampling plans are purposeful and based on a theory-driven belief that selecting cases from a design matrix based on combinations of certain attributes (e.g., economic status, race AND neighborhood type) is key for gathering cases that vary along important dimensions (Bernard, 1995; Johnson, 1990; Pelto and Pelto, 1978).

In the case of the New Hope evaluation, we debated at length the wisdom of random versus purposive sampling for the qualitative study. Since we had a complete list of program and control families, it was a simple matter to generate a simple random sample of cases. But some members of the research team argued that since an n=43 sample was too small to detect program impacts and the experiences of New Hope experimental families were so much more interesting than experiences of control families, the NHES sample should consist only of families in the experimental group. We also considered the wisdom of including "exemplar" cases – participants whom New Hope project representatives could readily identify as embodying exactly the kinds of experiences that program designers either hoped for or feared.

In the end, we opted for a stratified[4] random sample of cases plus a handful of exemplar cases: 51 randomly sampled cases plus three exemplar families. Response rates were not a problem: our intensive recruitment efforts led to success in recruiting 86% of the sampled families that had not moved out of the greater-Milwaukee area.

Experience has repeatedly confirmed the wisdom of our random sampling decision. The ethnographic sample has proved invaluable for understanding experiences (e.g., how families view and use the Earned Income Tax Credit – Romich and Weisner, forthcoming) that were not at all anticipated when the ethnographic study began. It can also correct for false assumptions by offering an invaluable form of insurance in representation of cases. Our *a priori* theoretical expectations about "interesting" and "uninteresting" situations proved depressingly inaccurate in

[4] Stratification is a method used by sampling statisticians that involves ordering the population by group prior to sampling. To ensure that we had an even balance of program and control cases, we randomly sampled within these two groups. We also stratified by race/ethnicity and residential location prior to random sampling.

the light of what subsequent analysis of both quantitative and qualitative data revealed to be truly interesting situations for understanding New Hope program impacts. All in all, we came away from our experiences believing that most ethnographic studies would profit from serious consideration to a random case selection.

Using the same individuals to gather and analyze both kinds of data. Although hard to prove, our experience suggests that it is vital to integrate the two methods of data collection. For three graduate students who both gathered and analyzed qualitative data and analyzed the survey data, the integration was complete. We cannot imagine achieving the same degree of integration between our two groups of methodologists, each of which specializes in one form of data collection and analysis. Individuals trained and actively engaged in both methods must constantly confront the productive tensions resulting from the two. The qualitative dimension provides a deeper level of meaning to the quantitative variables and analysis, while the larger quantitative sample provides a needed population perspective on the relatively small and potentially idiosyncratic nature of families in the qualitative study. We document below several instances where the synergy between the two methods deepened our understanding of family process and child development in New Hope families.

Another advantage these students offered was that they were simultaneously qualitative and quantitative researchers. In meetings consisting of only quantitative analysts, for example, they could offer insights into the progress and findings of the qualitative study. This was done at no additional cost to the project (as the students would have attended anyway), but greatly enhanced the knowledge base of the meetings. Other members of the project thus benefited from their methodological multi-tasking, as they were more informed of the findings from the other branch of the project than they might otherwise have been.

V. Understanding Program Effects

We gained a much richer understanding of New Hope's impacts than we anticipated by relying on the combination of qualitative and quantitative data. Researchers used the insights of one methodology to inform the findings of the other.

Qualitative data can understand program effects. One of the most important — and initially puzzling — impacts of the New Hope experiment was on the teacher-reported achievement and behavior of pre-adolescent children. In the experimental group, boys, but not girls, were rated by their teachers 0.3 to 0.5 standard deviations better behaved and higher achieving than their control-group counterparts. These are quite large impacts, and important to understand. Based on the survey data alone, however, the analysts were unable to understand why the New Hope program, focused as it was on the parents' work and income, should have such differential impacts for boys than for girls.

The qualitative data revealed instances where parents referenced the gender of their children and suggested that mothers believed that gangs and other neighborhood pressures were much more threatening to their elementary-school boys than girls. As a response to these pressures, mothers in the experimental group channeled more of the program's resources (e.g., child care subsidies for extended-day programs) to their boys. A 35-year old African-American mother of four, quoted in the field notes, observed:

> Not all places have gangs, but [my neighborhood] is infested with gangs and drugs and violence. My son, I worry about him. He may be veering in the wrong

> direction...it's different for girls. For boys, it's dangerous. [Gangs are] full of older men who want these young ones to do their dirty work. And they'll buy them things and give them money.

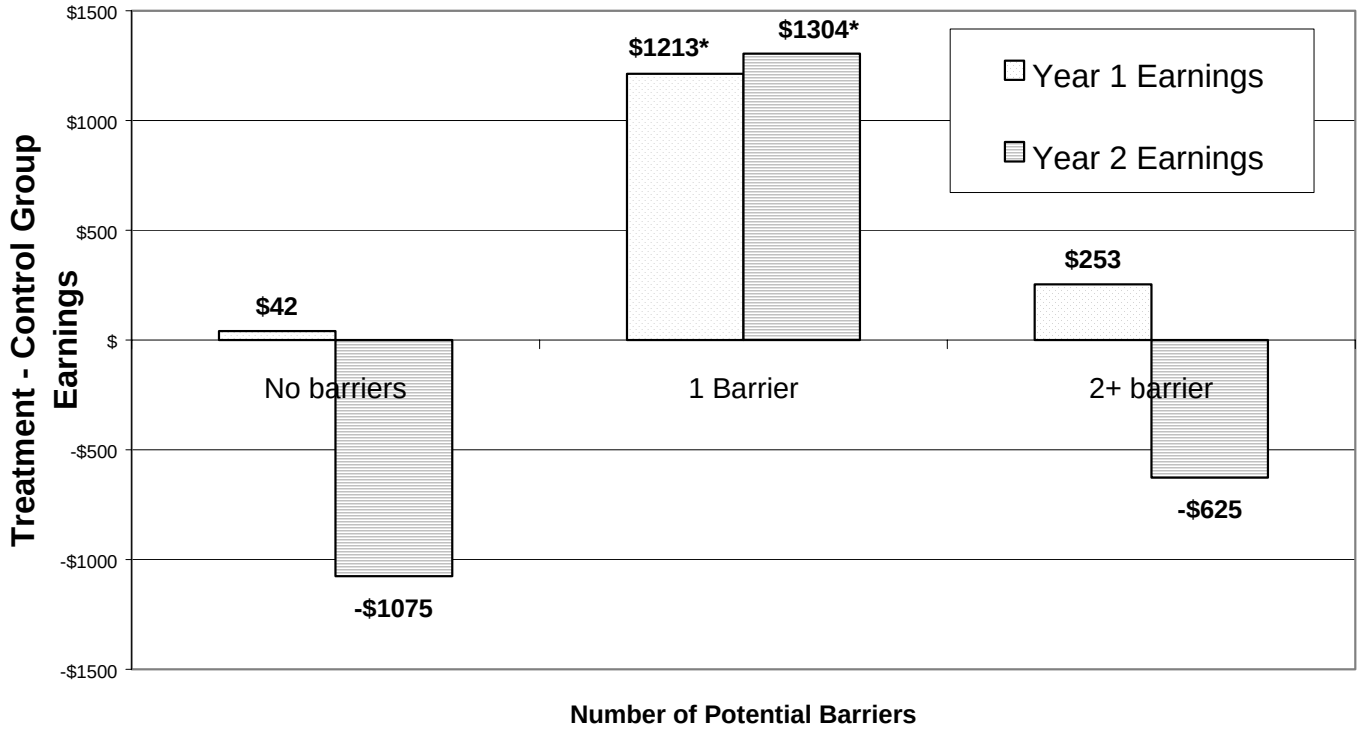
These kinds of sentiments appeared consistently in the qualitative data. Further quantitative analyses of both New Hope and national-sample survey data support the interpretation that parents living in bad neighborhoods do indeed devote differential time and other resources to their boys relative to their girls (Romich, 2000). We would not have had this insight, however, had we not had the ethnographic data.

Qualitative data can isolate program effects. Qualitative interviews suggested important heterogeneity among the experimental families. Some, perhaps one-fifth, appeared to have so many problems (e.g., drug dependence, children with severe behavior problems, abusive relationships) that New Hope's package of economic benefits was unlikely to make much of a difference. A second group was at the other end of the spectrum: they had no such apparent problems and were able to sustain employment on their own. In this case, control families in this group might be expected to do so well in Milwaukee's job-rich environment that it would be difficult for comparably unconstrained experimental families to do better.

A third group, however, with only one of the problems of the sort that New Hope might be able to address (e.g., Maria's difficulties in arranging for childcare, a minor criminal record that experience in a community-service job could overcome) appeared poised to profit from the New Hope package of benefits. Extensive quantitative work on subgroups defined according to the number of potential employment-related problems they faced at the beginning of the program confirmed the wisdom of these qualitatively derived insights. Using data gathered from the baseline interviews, Magnuson (1999) constructed an index of potential employment barriers based on past history of employment, completed schooling, arrests and the presence of either many or very young children.[5] She then estimated treatment-control differences for subgroups defined according to whether the family faced zero, one, or two or more barriers to employment. The results are presented in Figure 1.

[5] Although the qualitative interviews pointed to an additional set of maternal mental health and child behavior problems, the baseline interview did not ask the kinds of questions that would have provided this kind of information.

Figure 1: Impact of New Hope on Earnings, by Number of Potential Employment Barriers



This figure confirms what we had suspected based on our qualitative evidence. Experimental members who had either no barriers or multiple barriers did not earn a significantly different amount than their control counterparts. Program impacts on the earnings of families with only one barrier, however, were large and statistically significant in both years.

Promising quantitative uses of N=43 qualitative data. From a quantitative perspective, a sample of 43 cases seems pitifully inadequate to support inferences about the larger population. But given the qualitative interviews' unique ability to generate measures of important constructs, it is important to ask whether any quantitative inferences can be supported by a sample as small as 43 cases.

In statistical terms, a 43-case sample produces large but still informative confidence intervals. Suppose a circumstance (such as drug abuse) is suspected to be nearly universal in a population such as New Hope's and that only extensive qualitative interviews, rather than surveys, will provide valid data on its frequency. Suppose further that reports of drug abuse appeared in only 15% of cases in the qualitative sample. A 43-case sample produces a 95% confidence interval around the .15 proportion that spans .00 to .35. Thus, one cannot reject the hypothesis that there is virtually no drug abuse in the larger population, but one can be confident that the incidence is not as high as was initially suspected.

To take an example from New Hope, one of our interests was in understanding the extent to which children had temperamental, behavioral, or health problems and the links between these problems and their mothers' ability to obtain and keep a job. The survey revealed problems with a relatively small fraction of families. In contrast, work in progress exploiting the fact that the qualitative interviews were able to probe extensively on this topic is finding that 60% of families reported at least one child with one or more troubles (Weisner and Bernheimer, 2000).

Thus, the "stylized facts" that emerge from simple descriptions based on n=43 samples can be useful for conceptualizing important processes that may be at work, placing anecdotes and stereotypes into perspective, and thinking about policy levers.

A more formal statistical use of a matched qualitative/quantitative design to estimate important parameters in the larger population is based on the conception that a randomly-sampled qualitative study is a kind of planned missing-data design (Little and Rubin, 1987). Survey measures are available for both the qualitative and larger survey sample, but qualitative measures are available only for the n=43 qualitative sample. If the qualitative sample is a random subset of the larger survey sample, then it is possible to impute measures uniquely found in qualitative sample to the larger survey sample.

Suppose that we want to estimate for the larger New Hope sample the mean of Y, a measure (e.g., drug abuse) found in the qualitative sample (number of cases = m) but not in the larger survey sample (number of cases = n). Suppose further that both data sets contain a set of predictors of Y (X_1,...X_k). Little and Rubin (1987) show that the mean of Y can be estimated as the average of the observed and predicted values of Y in the survey sample. The variance of this regression estimator is:

$$V_{reg} = \frac{\sigma^2}{m}(1 - \frac{n-m}{n}\rho^2) \qquad (1)$$

where σ^2 is the variance of Y, σ^2/m is the variance of the NHES sample mean and ρ is the multiple correlation coefficient between Y and X_1,...X_k (Little and Rubin 1987). If the X's are completely predictive of Y (i.e., ρ=1), then the method is virtually equivalent to measuring Y on

the full survey sample. On the other hand, if $\rho=0$, then there is no information gained from the survey-sample cases that is not included in the qualitative subsample.

Suppose $\rho^2 = .20$, m=43 and n=800. The effective sample size for the estimation of the mean of Y increases from 43 to [43 + (.2 x 755)] = 194, which reduces the standard error by a factor of about 2.1. Analogous methods can be applied to estimation of subgroup and conditional means. Thus, under some circumstances it is possible to use the rich measures only available in the n=43 qualitative sample to estimate their prevalence in the larger population.

VI. The Benefits to Future Work

Even after we had completed the evaluation of two-year New Hope data for the program report (Bos et al., 1999), we used the synthesis between the qualitative and quantitative data in additional ways. The first was our ability to improve the five-year follow-up survey based on knowledge gleaned in the ethnography, and the second was the generation, from the qualitative data, of additional quantitative research.

Use qualitative data to generate survey measures. As our ethnographic work began, we discovered that certain aspects of family functioning were measured poorly in the two-year follow-up survey. This is because the NHES did not begin until just after the two-year survey data collection was complete. In contrast, the five-year survey will profit greatly from lessons learned from the ongoing qualitative interviewing. In particular, by listening to how New Hope families understand their daily routines, we constructed quantitative measures that will offer a more complete account of family well-being. Some items that are included in the five-year follow-up that were not part of the two-year survey include: the role of male partners; beliefs about the welfare system; budget questions; and the role of family support.

For example, our ethnographic work revealed that the presence of troubled children and inflexible jobs account for some of the variance in labor force participation. Neither dimension was listed as a possible job impediment on the two-year survey. On the five-year survey, however, we expanded the section that measures job attitudes by including two new items. Participants are now asked to rate the influence of "having a child with a serious health, emotional, or behavioral problem" and "being allowed to deal with family problems or emergencies that may come up while you're at work or school." Neither of these items would have been included had it not been for the ethnographic research.

Explore quantitatively ideas generated from the qualitative data. The beauty of ethnographic work is that the data provide detail on topics that cannot be fully explored by survey data. If, however, there is no larger data set on which to test an hypothesis, then findings from an analysis of ethnographic data are limited in their generalizability and replicability. However, if both types of data are present, then one can use the qualitative data to explore program dimensions, which can then be analyzed in the larger survey data.

An example is the take-up of New Hope benefits by program participants. The intent of New Hope was to centralize assistance for participants, so that they did not have to deal with several different agencies to receive the services that they needed (Brock et al., 1997). Soon after the qualitative study began, however, fieldworkers noticed great heterogeneity in the New Hope families that affected their use of the benefits. Very few used the program as was intended (as a continuously-used bundle of benefits); most used individual benefits selective and intermittently.

Both New Hope evaluators and designers were puzzled by this disconnect between program intent and program use.

A systematic qualitative analysis of benefit usage by experimental members of the NHES revealed that participants evaluated the usefulness of New Hope according to different standards (Gibson and Weisner, 2000). Some saw its usefulness in cost-benefit terms (the advantages of receiving the supplements versus the demands of working full-time), while others measured its usefulness by how well the program coincided with personal ecological concerns (not using a community service job because it was considered too demeaning). As these standards varied, so too did service use. Gibson (2000) analyzed the larger survey data and confirmed the great heterogeneity in program use among experimental members. Not only was this heterogeneity of use related to socio-demographic characteristics at baseline, as fieldworkers had suspected, but it also shaped the effect that the program had on individual families.[6] These quantitative analyses, however, were undertaken only after the rich qualitative data revealed their likely potential value.

VII. Limits of Qualitative Data

Until now, we have described only the advantages of our mixed method approach. We also discovered some limits to this approach.

The limits of N=43 qualitative data for understanding experimental impacts. An important goal for the New Hope evaluation is to estimate program impacts on family and child outcomes of interest. These take the form of differences in means and proportions between the experimental and control group. Is an n=43 sample, evenly divided between experimental and control families, helpful for this purpose? For example, given that important measures such as the sustainability of family routines were gathered only for the qualitative subsample, it would be extremely useful if this were indeed the case. But the task of estimating subgroup differences with a mere 43 cases is much more difficult than an estimation of a mean or proportion for the entire population.

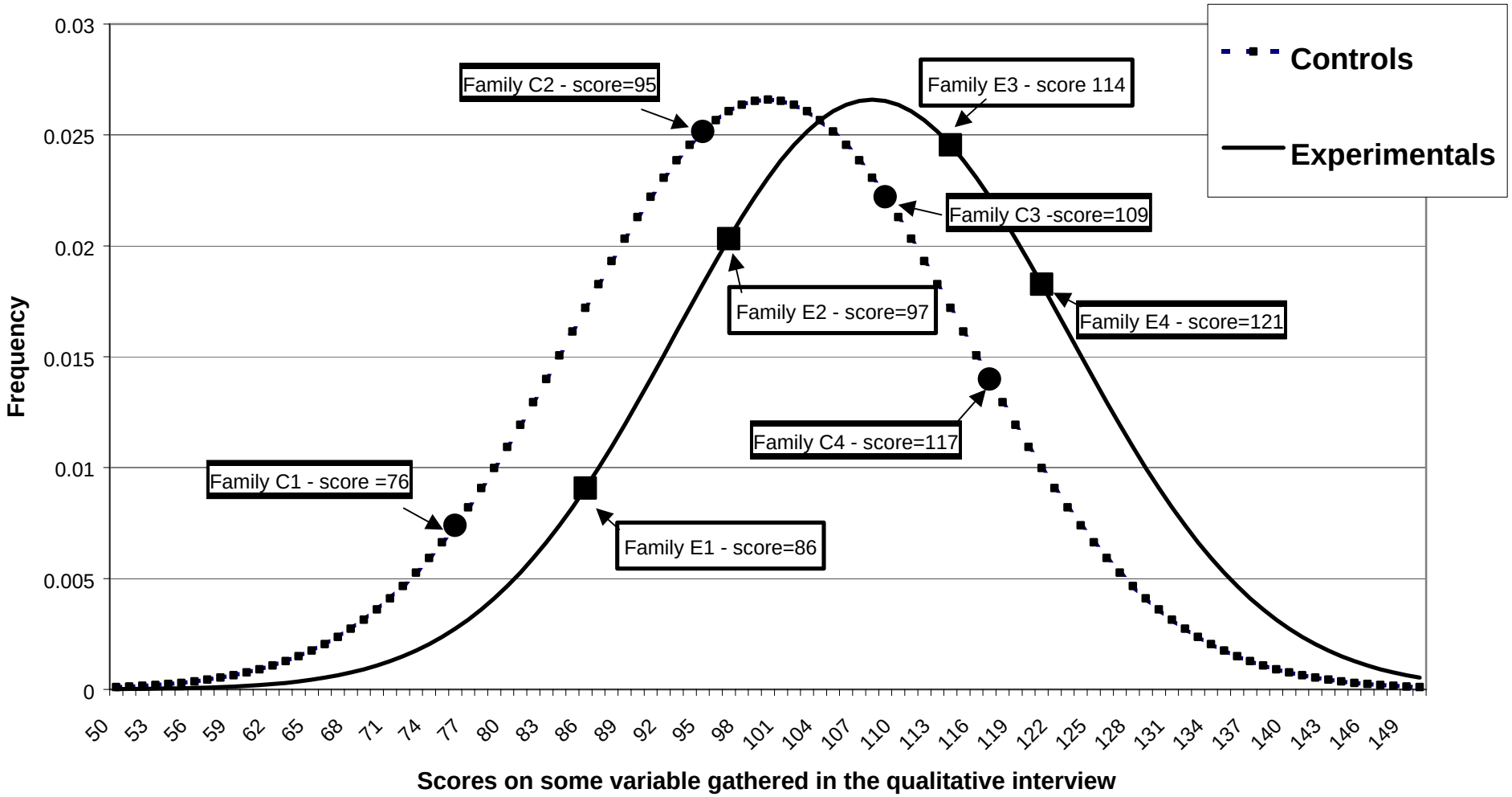
To illustrate the power of an n=43 sample to estimate program impacts, we coded several of the template themes from the NHES and converted the measures into a 0-8 scale.[7] We then standardized and regressed each on the experimental/control status of each case. The standard errors for the estimated coefficients were around .30, indicating that the N=43 NHES sample is capable of detecting a program impact of about .6 of a standard deviation at a 95% confidence level. This is a very large effect size, and larger than any of the impacts actually estimated from the quantitative data. Thus, the n=43 sample has a very limited ability to identify all but the largest program impacts for the quantitative measures derived from the qualitative interviews.

[6] Community service job users, for example, tended to have poor employment histories at baseline, but were the only group of participants that showed increases in emotional well-being after two years. Other participants may have had stronger labor market connections at the time of random assignment, but did not show any emotional health benefits.

[7] For example, we coded the adequacy of a family's income to meet basic needs, where a "zero" would indicate that the family had no income, and an "eight" indicates that the family is able to meet basic needs every month. This is part of a much larger coding project in the NHES; see Weisner, Coots, and Bernheimer (1996) for further details.

A different, more person-centered, question is whether there is some way in which individual experimental cases in the qualitative sample can somehow represent the more general set of experimental families and be contrasted with individual control families in the qualitative sample who somehow represent the larger set of control families. We see no way in which this might be done. Imagine two normal distributions for an outcome of interest – say sustainability of family routines - one for the experimental and the other of the control group (Figure 2). Suppose further that the bundle of New Hope benefits has a fairly large positive impact on sustainability. In Figure 2, this is illustrated by the fact that the experimental group's distribution is shifted to the right of the control group's distribution, by 8 points on a normal distribution with a (control) mean of 100 and a common standard deviation of 15.

Figure 2: Four Experimental (E) and Control (C) Cases Drawn from Populations with an Experimental Impact of .5 Standard Deviations



But now consider some of the individual cases in the two groups. While the average sustainability of program group members will exceed that of control group members, there will be many instances of low-sustainability program families and high-sustainability control families. For example, experimental family E1's score of 86 is well below the control-group average of 100, while control-group family C4, with a score of 117, is well above the program group average of 108.

It is clear from Figure 2 that even a large program impact leaves a great deal of overlap between the treatment and control groups. The task of assessing program effects amounts to discovering the nature of these shifts, a task that no amount of intensive analysis of or reflection about individual cases can accomplish. Having hundreds of sample points along these two distributions is key; a richer description of a small number of sample points cannot help. Thus, there is simply no way of using the detailed data gathered from individuals cases in the n=43 ethnographic sample to assess overall program effects.

Proper and improper uses of qualitative data to illustrate quantitative findings. For reasons related to the timing of its funding, the qualitative evaluation of New Hope was not launched until the final year of New Hope's operation. As a result, the quantitative evaluation of the two-year effects was completed long before the qualitative evaluation. Nevertheless, when preparing the two-year report, the NHES provided participant narratives that exemplified and crystallized the statistical program effects of New Hope. Such narratives presented the context and dynamics of our families in a way that mere survey data could not.

These vignettes added an important dimension to the report by providing readers with vivid descriptions of families in the program. We were aware, however, that selective case description might not do justice to the subtleties of the cases and might also confirm the quantitative research community's suspicion that qualitative data is merely a collection of anecdotes. This is why when, in selecting individual families and constructing their vignettes, we opted for page-long, rather than one or two-sentence, descriptions of their circumstances. No family fits a "simple" story, and we believed that a too-brief description of any given family would fail to provide a realistic picture of family circumstances.

To illustrate our point, the following is part of a vignette that was used in the evaluation report:

> Janet is a single mother who lives on the south side of Milwaukee with her two young ones…She used the New Hope medical benefits supplement for only a year, and during that year she had her son's tonsils removed. The premiums were over $400 a month for her and her sons, so after the surgery was completed, she discontinued the care. She says she almost never takes her boys to the doctor. They are very healthy and her point of view is that paying so much money for a service she never uses is unreasonable. (Bos et al., 1999, page 74)

A simple telling of this story might indicate that the New Hope medical benefits were very useful, as they allowed parents to have medical procedures for their children that they otherwise could not afford. However, that would ignore the interesting detail that Janet stopped using the New Hope health insurance, even though she did not have other health care coverage. With this fact, Janet's story becomes more complex, but it is also likely to be more representative of the complicated response people had to the New Hope benefits (Gibson and Weisner, 2000).

VIII. Conclusion

We have delineated the advantages of using a mixture of qualitative and quantitative methodologies for evaluation projects. For the New Hope project, the use of both was critical, as the interplay between the two enhanced the evaluation effort throughout the process. Our evaluation benefited by randomly sampling our ethnographic cases and using the same graduate students to work with the quantitative and qualitative data. As we progressed in the assessment of the New Hope program, other advantages of using mixed methods emerged. We used the qualitative data to isolate and identify quantitative program effects and explored ways in which it might be possible to make quantitative inferences with a very small number of cases. Finally, once the evaluation of the two-year findings was complete, we used our qualitative work to better inform the construction of our five-year follow-up survey.

We have skirted some of the hard questions regarding the balance between the two methods. Given that evaluation efforts have limited resources, issues of the scope of the two types of data collection are quite important. For example, we had initially planned to use 60 families for the NHES, but budget constraints reduced that number to 46. What was the cost of forgoing those 14 families? Should those families have been kept, even if that had meant fewer resources to maximize our surveys' response rates or fewer minutes of interviewing time? We have not made these kinds of cost-benefit comparisons across methods and cannot say if we have correctly allocated resources between the quantitative and qualitative arms of the project.

We also know that we benefited from fortuitous circumstances, which may not apply in all situations. Both New Hope's directors and researchers in the evaluation team embraced the idea of a mixed-methods evaluation. Quantitative researchers, many of whom had had no prior experience with qualitative data, quickly grasped its potential importance. We also benefited from having excellent research assistants who were willing and able to be trained in both methodologies. This cross training proved invaluable, as most of the productive synergy between the two was either generated or first noticed by this group of graduate students. And, as noted above, New Hope was well implemented, making our task as evaluators that much easier.

It is unfortunate that project evaluations rarely involve both quantitative and qualitative methodologies. But this lack of interaction could stem from the perception that these methods are antithetical, rather than complimentary. Quantitative methodologists are skeptical that reliable conclusions can be gathered from the traditionally small sample sizes involved in qualitative work; and qualitative methodologists believe that survey data may oversimplify complex realities. However, one need not be employed at the expense of the other. A combination of the two, as we saw in our assessment of New Hope, can be used to greatly enrich evaluation efforts.

References

Bernard, H.R. 1995. Research Methods in Anthropology, Qualitative And Quantitative Approaches, 2nd Ed. Walnut Creek, CA: Alta Mira Press.

Bos, J., Huston, A., Granger, R., Duncan, G J., Brock, T., & McLoyd, V. 1999. New Hope for People with Low Incomes: Two-year Results of a Program to Reduce Poverty and Reform Welfare. New York: Manpower Demonstration Research Corporation.

Brock, T., Doolittle, F., Fellerath, V., & Wiseman, M. 1997. Creating New Hope: Implementation of a Program to Reduce Poverty and Reform Welfare. New York: Manpower Demonstration Research Corporation.

Gibson, C. M. "New Hope for New Hope: Service Utilization as Mediator of Family Well-Being." Poster presented at the Association for Public Policy Analysis and Management meeting, Seattle, WA. November 2000.

Gibson, C. M. & Weisner, T. 2000. " 'Rational' And Ecocultural Circumstances of Program Take-Up Among Low-Income Working Parents." Under review.

Johnson, J.C. 1990. Selecting Ethnographic Informants: Qualitative Research Methods Series No. 22. Newbury Park, CA: Sage Publications.

Little, R.J.A., & Rubin, D.B. 1987. Statistical Analysis with Missing Data. Wiley, New York

Magnuson, K. 1999. "Appendix K – The Barrier Indicator Index" in Bos, J., Huston, A., Granger, R., Duncan, G J., Brock, T., & McLoyd, V. 1999. New Hope for People with Low Incomes: Two-year Results of a Program to Reduce Poverty and Reform Welfare. New York: Manpower Demonstration Research Corporation, p. 346-55.

Mead, L. M. 2000. "The Twilight of Liberal Welfare Reform." The Public Interest 139 p. 22-34.

Pelto, P.J. & Pelto, G. 1978. Anthropological Inquiry: The Structure of Inquiry 2nd Ed. New York: Cambridge University Press.

Romich, J. L. 2000. "To Sons and Daughters: Gender, Neighborhood Quality and Resource Allocation in Families." Unpublished manuscript, Joint Center for Poverty Research, Northwestern University, Evanston, IL

Romich, J. L. and Weisner, T. (forthcoming). "How Families View and Use the EITC: Advance Payment Versus Lump Sum Delivery." The National Tax Journal. Special issue, Fall 2000.

Weisner, T. S. 1997. "The Ecocultural Project of Human Development: Why Ethnography and its Findings Matter." Ethos 25 (2): p. 177-190.

Weisner, T. S. & Bernheimer, L. 2000. "Troubled Children in Economically Poor Families: Evidence from the New Hope Ethnographic Sample." Paper presented at the annual Association for Public Policy Analysis and Management meetings, Seattle, WA, November 2000.

Weisner, T. S., Coots, J., & Bernheimer, L. 1996. "Ecocultural Family Interview Field Manual." Unpublished Manuscript, Center for Culture and Health, UCLA, Los Angeles, CA.

Weisner, T.; Bernheimer, L.; Gibson, C.; Howard, E.; Magnuson, K.; Romich, J.; & Lieber, E. 1999. “From the Living Rooms and Daily Routines of the Economically Poor: An Ethnographic Study of the New Hope Effects on Families and Children”. Paper presented at the Society for Research in Child Development biannual meeting, Albuquerque, New Mexico, April 1999

Appendix Table A

Main Template Headings

For the New Hope Ethnographic Study (NHES)

Influences on New Hope take-up	**Outcomes of New Hope participation**
Family background	Stability in participants' life
Work history/values of participant and relatives	Feelings of success and evidence of planfulness vs. procrastination
Education of participant and relatives	Participants' future orientation including investments in further training or education
Role of religion and spirituality	Meaning of work: job vs. career; resource vs. constraint
Paths to employment and pattern of work at entry, including role of underground economy	Equity
Number of and relationships with case reps, W-2 caseworkers, other social services	Social networks and community bridging, including involvement in school
Role of ethnicity	Children and childrearing
Beliefs about and use of child care	Political ideology
Relationships with partner/spouse	Job barriers
Life goals and ambitions including attitudes or values about work	Daily routine