

4-30-2004

The Impact of Fidelity on Program Quality in the Healthy Families America Program

Stacey R. Kessler

University of South Florida

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

Scholar Commons Citation

Kessler, Stacey R., "The Impact of Fidelity on Program Quality in the Healthy Families America Program" (2004). *Graduate Theses and Dissertations*.

<https://scholarcommons.usf.edu/etd/1110>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

The Impact of Fidelity on Program Quality in the Healthy Families America Program

by

Stacey R. Kessler

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Major Professor: Carnot Nelson, Ph.D.
Michael Brannick, Ph.D.
Judy Bryant, Ph.D.

Date of Approval:
April 30, 2004

Keywords: variability, program evaluation, homevisitors, child abuse, neglect

© Copyright 2004, Stacey Kessler

This project would not have been possible without the help of numerous individuals. First, I would like to thank my committee members Dr. Carnot Nelson, Dr. Michael Brannick, and Dr. Judith Bryant. They provided a tremendous amount of support, knowledge and encouragement along the way. Their suggestions truly improved the quality of this project.

The collaboration of the Healthy Families America staff in Chicago was invaluable. Kathryn Harding, John Holton, Wendy Mitchell, Holly Seymour, and Cyd Wessel all spent an enormous amount of time and effort in order to see this project through to completion. Their dedication to the continuous improvement of the Healthy Families America Programs is admirable.

Next, I would like to thank my colleagues and professors in the I/O program at USF. These people assisted me in countless ways. Their cooperation, helpfulness and teachings proved invaluable to my personal growth.

My advisor, Carnot Nelson, was the driver of this project. He worked tirelessly with me to see this project through. Throughout my time in this program, Carnot has been a true mentor to me; one that I will not forget for years to come.

Last, I would like to thank my family who has always supported me in anything that I have done. My parents have set tremendous examples for me through their hard work in both their careers and volunteer activities. My brother has always been a close friend who has demonstrated an ability to succeed academically while maintaining other areas of his life.

Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Chapter 1: Introduction	1
Background and Large-scale Program Implementation	3
Blakely et al's (1986) Study	8
Fidelity	11
Validity and Fidelity	13
Examples of Programs Using Fidelity	14
Home Visitation Programs	18
Variability Between Program Sites	18
Important Variables in Home Visitation Programs	19
Olds' Model	20
Healthy Families America	23
Current Study	26
Variable Selection	26
Hypotheses	26
Chapter 2: Method	28
Programs	28
Measures	29
Healthy Families America Credentialing Tool (1999)	29
Healthy Families America Credentialing Tool (2003)	32
Annual Site Profile Update	34
Procedure	35
Chapter 3: Results	37
Healthy Families America Credentialing Program Self-Assessment Tools	37
Transformation of Scores	38
Test of Hypotheses	53
Hypothesis Testing	54
Exploratory Analyses	58
Chapter 4: Discussion	60
Limitations	63
Future Research	65

References	66
Appendices	70
Appendix A: The Critical Elements of Healthy Families America	71

List of Tables

Table 1	A Comparison of the 1999 and 2003 Credentialing Tools	33
Table 2	Means, Standard Deviations (raw form)	55
Table 3	Means, Standard Deviations, and Intercorrelations	56

List of Figures

Figure 1.	Histogram of Credentialing Element 1	39
Figure 2.	Histogram of Credentialing Element 2	40
Figure 3.	Histogram of Credentialing Element 3	41
Figure 4.	Histogram of Credentialing Element 4	42
Figure 5.	Histogram of Credentialing Element 5	43
Figure 6.	Histogram of Credentialing Element 6	44
Figure 7.	Histogram of Credentialing Element 8	45
Figure 8.	Histogram of Credentialing Element 9	46
Figure 9.	Histogram of Credentialing Element 10	47
Figure 10.	Histogram of Credentialing Element 11	48
Figure 11.	Histogram of Credentialing Element GA	49
Figure 12.	Histogram on Total Credentialing Elements	50
Figure 13.	Histogram of the DV: Percentage of Children with PCP's	51
Figure 14.	Histogram of the DV: Percentage of Children Immunized	52
Figure 15.	Scatterplot of the Two Dependent Variables	61

The Impact of Fidelity on Program Quality in the Healthy Families America Programs

Stacey Kessler

ABSTRACT

The current study examined the relationship between program fidelity, or adherence to the program model, and program outcomes using the Healthy Family America Program Model. Specifically, 103 program sites were evaluated based on their adherence to the program model. The outcome indices included the percentage of children with updated immunizations and the percentage of children with primary care physicians. First order correlations, multiple regression, and canonical correlation were used to analyze the data. The results of the study are mixed. Specifically, an overall index of fidelity is positively related the percentage of children with updated immunizations, but not to the percentage of children with primary care physicians. Additionally, only one of the 11 facets of fidelity was related to the percentage of children with updated immunizations. Both the implications for these findings and future avenues for research are discussed.

Chapter One

Introduction

Diffusion of innovation refers to the way in which new programs spread from one location to another. The current study has its origins in this process. Specifically, Blakely et al. (1987) discuss two discrete schools of social program innovation. These include the “profidelity” school and the “proadaptation” school. The former school, “profidelity,” advocates for implementing programs as planned, or in other words, exhibiting fidelity to the original model (Calsyn, Tornatzky, & Dittmar, 1977). Advocates of this school argue that changing or diluting the program will lead to a decrease in its effectiveness (Blakely et al., 1987). The opposing school, “proadaptation,” argues that some degree of modification in a program is beneficial when implementing a program model at a local site (Glaser & Backer, 1977; Larsen & Agarwala-Rogers, 1977). However, most argue that this reinvention or adaptation cannot change the core of the program model because the integrity of the program must be maintained in order to yield effectiveness (Blakely et al., 1987). Thus, these two opposing viewpoints, offering varying methods of implementing a program model, comprise the adaptation/adoption debate. Blakely et al. (1987) empirically tested these schools of thought and discovered that high fidelity adopters have more effective implementations of the model (i.e., better program outcomes) than do programs that do not adhere to the basic model. Regarding program reinvention, this study also found that high fidelity programs that incorporate additions to the program model, as opposed to changing or deleting the model, are more successful

than programs that do not have additions to the model. Thus, Blakely et al. (1987) offered a compromise between the two schools of thought.

Blakely et al.'s (1987) study offers a unique way to examine program models. The focus of the current study is to test part of Blakely et al.'s findings using the Healthy Families America (HFA) Model. Healthy Families America is a multi-site home visitation program that targets families at-risk for abusing and/or neglecting their children. The program consists of paraprofessional caseworkers visiting these families in order to address the issue of child abuse and neglect, to educate the parents, and to support them throughout the early years of raising their child. HFA has over 500 of these sites throughout the country, serving a variety of populations within a variety of cultures. Therefore, there is inherently great variability between all of these program sites. Thus, HFA is not a stagnant "one size fits all" program model, but rather a model that local programs adopt. Although HFA founders acknowledge that local programs adopt the model, they sought to maintain the integrity of the model by implementing a credentialing process to "certify" local programs. The goal of the current study is to examine the effect of local programs' fidelity to the HFA model on the outcomes of these sites.

*Background on Large-scale Program Implementation
and Replication in Early Childhood Programs*

Before further reviewing Blakely et al. (1987) and the concept of fidelity, it is necessary to examine program implementation and replication because the large-scale implementation of any program model is a formidable task requiring careful planning. The fundamental claim in program replication is that a program implemented under ideal conditions will be successful. However, when a program is implemented at a new site, researchers do not have the freedom to implement that program under ideal conditions. Therefore, it is necessary to ascertain whether that program model can succeed at a new site under less than optimal conditions. The key issue then is to determine to what extent the new program, implemented at the new site, follows the original program model—this is fidelity (On a side note, program or site refers to the local program implemented and program model refers to the theoretical program ideology or concept.).

At the onset, it is important to acknowledge the variety of successful methods through which to expand and replicate large-scale program models (Yoshikawa, Rosman, & Hseuh, 2000). The first, “staged replication,” consists of expanding a single-site program following extensive testing of that program. Essentially, researchers evaluate a single site and if they deem it successful, they implement the program model at multiple sites. Then, if the evaluators pronounce each additional implementation favorably, practitioners implement the program model on a grand scale—this is large-scale implementation. A specific example of this process is Olds’ Nurse Home Visitation Program (Olds, Henderson, Tatelbaum, & Chamberlin, 1986). Olds and his colleagues first tested the program model in Elmira, New York. Following successful evaluations,

they implemented the model in Memphis, Tennessee (Kitzman et al., 2000). Another type, “franchise replication,” occurs when a single, central organization governs local program sites throughout the country. The governing agency sets the performance standards of the model and then monitors local programs implementation of the model. One example of this is the Healthy Families America program model. The third approach is the “multi-site demonstration,” and this involves an initial evaluation of multiple sites before further replication. This process is similar to the staged replication, except it lacks the first step, and implementers evaluate multiple programs at the same time. An example of this approach is the Infant Health and Development Program, which provides care for children born at a low birth rate (Yoshikawa, Rosman, & Hseuh, 2000). A fourth type of replication is the “mandated replication” and involves investing in government-mandated programs (Yoshikawa, Rosman, & Hseuh, 2000). The Head Start program is an example of this type of replication. The final type, “government-supported private sector expansion,” involves the government supporting private organizations to implement needed programs. A noteworthy example is the allocation of government funds to private sector childcare programs (Yoshikawa, Rosman, & Hseuh, 2000).

Regardless of the type of replication used, there are a number of problems that researchers and practitioners encounter when trying to implement new programs (Yoshikawa et al., 2000). First, when experts deem the implemented program successful, they often draw the conclusion that the program can automatically be implemented effectively in other locations. However, this is not an accurate conclusion. One example illustrating this problem is the initial implementation and replication of the Head Start program throughout the country (Yoshikawa et al., 2000). This program was originally

replicated at an expedited pace, based solely upon a few randomized trials. The program was replicated “top down” using federal money. Additionally, because the replications were done in a hasty fashion, it is not clear whether the new programs were truly based upon the Head Start model. This problem is similar to the issue with the initial implementation and replication of the HFA program model. At the onset, program implementers were replicating HFA programs; however, it remained unclear whether these programs truly constituted HFA programs. The HFA program designers recognized this issue and inserted a credentialing process, involving the basic elements of the HFA model, in order to “certify” local programs.

Another problem is that the results of vigorous testing are unclear as to whether the implemented programs produce the desired results (Yoshikawa et al., 2000). Although some program evaluations demonstrate improvement for both parents and children, others fail to find significant outcomes for participants. A solution to this problem involves examining the effect of specific program implementations on outcomes for parents and children. Unfortunately, there seems to be a lack of this type of research in the field. On a similar note, little attention is paid to the cultural relevance of a program when implemented in a new population. It must be recognized that different cultures will have different needs from one another and that implementers of a program in a new area must be sensitive to these differences. Last, in an attempt to address the former issue, program implementers have been tempted to dilute or alter the original program model. For the purposes of evaluating large-scale implementations, this alteration is problematic because it becomes unclear whether the original program model is successful or whether a modified version of the program model is successful within a

specific population.

Yoshikawa et al. (2000) also refer to “underlying paradoxes” that exist in the program implementation and replication process. First, programs are often replicated based upon whether the program model succeeds under ideal conditions as opposed to whether it can succeed under real-world working conditions. This is a clear flaw in the methodology because demonstrating that a program model can work under ideal conditions is not helpful in determining whether the program model will work in an actual community under average conditions. Therefore, researchers and practitioners alike must exercise caution and ensure that their research question addresses whether a program model can succeed under actual conditions as opposed to succeeding under ideal conditions.

Another paradox concerns the need for program fidelity versus the need for flexibility to adapt to the local culture (Yoshikawa et al., 2000). Yoshikawa et al. acknowledge this issue and offer a solution in which implementers of the model must systematically decide how to facilitate such a balance. This solution is problematic because it is necessary to examine program reinvention. Specifically, when attempting to adapt a program model to a local situation, researchers need to decide carefully whether certain elements can be modified. Yoshikawa et al. partially address this issue by stating that fidelity is a paramount concept and that certain elements of a program, such as training and education of the staff, are connected to child outcomes. They also indicate that these elements should be considered when implementing a program model and that future research in this area is imperative because it can elucidate the ways in which these elements relate to program outcomes.

The “representativeness versus feasibility paradox” (Yoshikawa et al., 2000, p. 19) is another important consideration in program implementation and replication. This paradox refers to the issue that evaluators of multisite programs often focus on a few sites and not necessarily a representative sample. Evaluators often commit this error because they have cooperation from only some of the sites and therefore do not have access to review the number of sites necessary to comprise a representative sample. This is problematic because, in the aftermath, researchers and practitioners often make generalizations about a program model based upon this small number of sites that are not necessarily representative of the program model.

Based upon their research, Yoshikawa et al. (2000) advocate linking the assessment of fidelity to the outcomes of program sites. They also acknowledge the paradox of trying to implement a program with fidelity to the model, while allowing for flexibility to meet cultural needs. They conclude that further research on additions and adaptations is important and should be made a priority.

Blakely et al. 's (1987) Study

Blakely et al. (1987) empirically studied the positions of the “profidelity” school and the “proadaptation” school by examining 70 program sites disseminated from seven original models (10 sites within each model). In order to do so, they examined and measured fidelity, reinvention and effectiveness. At the time, the quantification of fidelity was a cutting edge concept because fidelity was most often used in theory-based studies.

Blakely et al. (1987) forwarded this concept by recognizing that social programs could be conceptualized and defined by their core items. In order to identify the core items of the program models, two of the authors visited the original site of the program model. They used a five-step process that involved conducting detailed interviews of program personnel and making careful observations of the programs. They taped all of the interviews with program staff in order to content analyze the discussion. The goal of this process was to identify a comprehensive list of core program components that were both observable activities and unique from one another. From this process, the researchers identified the core components of each of the seven program models. They then devised a separate three point Likert scale (0 = lowest, 2 = highest fidelity) for each of the core items of each of the seven models.

Two pairs of researchers then conducted all of the site visits and assessed a site’s fidelity to its respective model by rating the site using the scale developed for its respective program model. The raters then summed the score of each site and this number became the site’s index of fidelity. Interrater reliability was tested on 20% of the sites and was found to be .81.

For the purposes of examining reinvention, Blakely et al. (1987) used a similar process, but one that they describe as rudimentary. The two pairs of researchers conducted site visits and made audio recordings of their observations regarding reinvention at the sites. The raters then determined whether the activity in question was either a reinvention or a lack of fidelity. If the given activity was deemed a reinvention, the raters then made two additional classifications. First, these site visitors content analyzed the tapes and described the resulting reinvention as either an addition to the program or as a modification of the program. An addition to the program model meant that “the instance would not fall within the activities, materials, facilities, etc. defined by any of the existing fidelity components” (Blakely et al., 1987, p. 262). On the other hand, a modification of the model occurred when “the local site had implemented the activity, material, or facility required by the innovation, but had done so in a novel way” (Blakely et al., 1987, p. 262). In order to assess the magnification of the reinvention, the authors designed a three point scale which rated the reinvention as minor, moderate or substantial. Therefore, the resulting score for a program site’s reinvention (classified as either reinvention or an addition to the model) was the sum of this three point scale (applied to each instance of program modification).

In order to assess program effectiveness, the researchers used program evaluations of the local sites. Within these evaluations, the researchers were able to identify outcome variables unique to each program model and used these as indices of effectiveness. Since the actual outcome variables employed varied among the program models and even within program models, the authors of the study designed a quality ranking system. They

assigned each local site a score ranging from 1 to 10. These rankings were based upon staff judgments of program sites' records.

The researchers found a .38 ($p < .01$) correlation between fidelity and program effectiveness. Thus, an increase in program fidelity was related to an increase in program effectiveness. In order to evaluate the effect of reinvention, the researchers used a series of partial correlations and determined that additions to the model, as opposed to a modification of it, increased effectiveness.

An important limitation of Blakely et al. (1987) is that it examines education and criminal justice programs. Specific programs included, but were not limited to, tutorial reading programs, jury management programs, and juvenile delinquency programs. A second noteworthy limitation is that the methods (e.g., staged replication, franchise replication, etc.) through which implementers replicated these programs were unclear. Variability among program models along this dimension could be an important confounding variable. Therefore, although these results can serve as a starting point, we cannot make a generalization because criminal justice programs and education programs may not be analogous to home visitation programs.

Fidelity

Blakely et al.'s (1987) study clearly illustrates that the concept of fidelity provides one way to address the adaptation/adoption issue. It is interesting to note that the origins of the concept of fidelity are not in program evaluation. Rather, fidelity emerged in the 1960s in order to address the issue of outcomes in psychotherapy studies (Bond, Evans, Salyers, Williams, & Kim, 2000). In short, researchers had great difficulty interpreting the success of therapies because these therapies were not carefully defined and because various models overlapped. Therefore, researchers began to use the concept of fidelity in order to define treatments, which in turn allowed for more accurate outcome studies. Specifically, when treatments, or treatment models, are carefully defined, researchers are better able to make precise comparisons among various treatments. Although in existence for the last 40 years, fidelity is a concept that until recently was often ignored in the literature. For example, Moncher and Prinz (1991) found that only 55% of the 359 psychotherapy treatment outcome studies examined the concept of fidelity.

Fidelity emerged as an important concept in education, during a careful examination of the educational reforms in the late 1950's and 1960's. Specifically, after the Soviet Union launched Sputnik, the United States entered a period of reactionism in education (Paisley, 1973). The leaders of the United States believed that the launch of a Soviet satellite before the launch of an American satellite was as a direct result of failing schools in the United States. Therefore, the leaders of the country began to allocate money for the purposes of reforming the education system. Researchers, already engaged in educational research and development, began to produce recommendations for change

is the country's educational system. However, the researchers recognized that they had a problem disseminating the information to the teachers and principals in charge of educating the nation's youth. In response to this problem, the U.S. Office of Education developed the ERIC system, a document distribution service (Paisley, 1973). However, teachers and principals were not receptive to the service for a number of reasons and therefore did not use it to implement recommended changes. Upon recognizing the failure, the department employed a different strategy in which they published reports on topics of interest to teachers and principals. They delivered these reports directly to the state departments of education, which in turn disseminated them to the schools.

Although the U.S. Office of Education succeeded in transmitting the research to the practitioners, there still remained the problem of implementing the new ideas (Berman, 1981). Berman (1981) reported that the United States government believed that if researchers disseminated valid and reliable products to the schools, then the schools could implement these products. However, after the attempted implementations were carefully examined, it became clear that this was impossible. For example, many of the programs that schools implemented differed from the intended program model and, in turn, the implemented program could not produce the intended results. Therefore, researchers recognized that it was important to examine the way individuals implemented the program model and not just whether the program was implemented. More specifically, researchers began to understand that they could "consider the innovation to be successful if the installed innovation replicates the originally conceived innovation with a high degree of fidelity" (Berman, 1981, p. 264). Thus, the concept of fidelity became an important one in educational programs.

Following the recognition of the importance of fidelity in educational programs, researchers identified fidelity as an important factor in program evaluations, specifically multi-site programs (Paulson, Post, Herinckx, & Risser, 2002). Often local programs differ from each other along a number of dimensions. These include the local culture, the funding of the program, and the collaboration of local organizations. Fidelity provides a way to measure a local program's adherence to the model in order to compare the program across sites (Paulson et al., 2002).

Validity and Fidelity. Program fidelity is a paramount concept that serves a variety of research purposes related to validity. Specifically, fidelity relates to internal validity, external validity, predictive validity, and construct validity (Moncher & Prinz, 1991). The following section examines the effect fidelity has on validity and in turn on studies addressing outcomes.

Fidelity is necessary for internal validity because it ensures that programs adhere to the program model (Bond et al., 2000). A key issue for ensuring model adherence is for program designers and implementers to define the program model clearly at the onset. This allows for fair comparisons across program sites (Moncher & Prinz, 1991). Fidelity also affects internal validity because it helps to ensure that extra contaminants are not a part of the program. This process of identifying contaminants is important because it is difficult to determine whether the contaminant contributed to the success of the program or whether the success was a result of the original program.

Researchers and practitioners also affect external validity when they attempt to replicate a program. Problems in replication occur when the treatment is not carefully

defined and when there are inconsistencies in the implementation of programs (Moncher & Prinz, 1991).

The utilization of fidelity also assists researchers when attempting to resolve issues related to construct validity. Threats to construct validity exist when the underlying concepts and elements of the program are not adequately understood (Moncher & Prinz, 1991). Fidelity scales, developed for specific programs, allow researchers and practitioners to identify critical components of a program. Critical components are key elements of the program and often include training of staff, ratio of staff to number of cases, and the location of services (Bond et al., 2000). On a similar note, the identification of critical components also pertains to predictive validity when these elements are related to outcomes.

Therefore, fidelity is a necessary concept, and measures of fidelity have become a tool for program evaluation. Key issues to note include the following. First, designers of programs often fail to describe their programs specifically, thereby preventing researchers from drawing accurate and meaningful comparisons among programs (Paulson et al., 2002). Second, when evaluating reasons for a program's lack of success, it is often impossible to determine whether the failure was because of the implementation or because of the design of the original program.

Examples of Program Models using Fidelity Scales. Researchers have examined program fidelity in a variety of fields including education, psychotherapy, and rehabilitation. Furthermore, they have recognized that the concept of fidelity is useful. The following section will highlight a number of studies in which researchers have used

fidelity.

In one study, Jackson, Paratore, Chard, and Garnick (1999) examined program fidelity in a literacy program, the Early Intervention Project, aimed to assist elementary-aged students who struggled with literacy issues. Since the program required a significant amount of training and support for the teachers, the school district arranged a number of formal and informal training sessions. Following these meetings, the researchers decided to collect data in order to examine both the fidelity of the implementation of the program model as well as the outcome data pertaining to the students' reading level. In order to examine fidelity, the researchers used a number of data sources. First, they had a member of the research team observe each teacher while he/she delivered the program lessons. During these observations, the researcher completed a careful log regarding the implementation of the lesson. Researchers visited each teacher a minimum of three times within the period in which they conducted this study. Second, the teachers kept weekly logs regarding their lesson plans and the delivery of these lessons. Third, in order to measure the literacy level of the students following the delivery of this curriculum, the researchers used a variety of literacy tests administered to the students as well as the teachers' evaluation of the students. The results of this study indicate that the teachers implemented the literacy program with a high degree of fidelity and that a majority of the students had success with this new program.

One particular noteworthy issue is the way in which the researchers measured the teachers' implementation of the program model. For example, although the researcher kept a log during his/her evaluation of the teacher during the teacher's lessons, subjectivity is still a problem. Specifically, the researcher could be prone to bias during

these sessions. Another problem is that the researcher did not observe all teachers during every lesson they delivered. Therefore, it is possible that the teacher delivered the lessons with a high degree of fidelity only when the researcher was present. This is even more probable if the teachers knew when to expect each visit from the researcher. Another way the measure of fidelity of the implementation is flawed is in the use of the teachers' logs as indices. The teachers could be prone to bias and might not accurately describe the way in which they delivered the lesson. More objective measures, such as an index of the critical elements of the literacy program, would be beneficial in examining fidelity.

A specific example of a successful use of a fidelity measure is the case of the Assertive Community Treatment (ACT) program, regarded as a fairly effective treatment for persons suffering with mental disorders (Teague, Bond, & Drake, 1998). Originally, the program model was replicated and implemented in a variety of settings with different clientele. Due to the strong possibilities of program reinvention, it was nearly impossible for researchers to evaluate and compare these programs accurately. Therefore, Teague et al. attempted to develop a measure to evaluate the local programs' fidelity to the ACT model. In doing so, Teague et al. specifically defined the program model and identified 13 of its key elements. The resulting scale, the Dartmouth ACTS Scale (DACTS), was developed, tested and found to be psychometrically respectable.

Although they did not use the DACTS, McHugo, Drake, Teague, and Xie (1999), examined the relationship of fidelity in ACT programs to outcomes. Specifically, they assigned 87 patients to either high fidelity (61) or low fidelity (26) treatment. The participants all had drug use disorders mostly involving cocaine. The experimenters identified a number of ways to measure patient outcomes, including interviews, length of

substance abuse, objective and subjective indices of quality of life, current psychiatric symptoms, as well as both self-reports and objective measures of substance abuse. In order to assess model fidelity, the researchers identified nine critical components of the ACT Model and four additional components pertinent to the treatment of this disorder. Subject matter experts made a number of ratings on these components. The researchers used factor analysis to identify two higher order factors. They found a significant variation between programs on one of the factors and used component scores for this factor to identify four “high-fidelity” programs and three “low-fidelity” programs. The researchers found that fidelity had an effect on program outcomes. Specifically, clients in the high fidelity programs experienced less substance abuse, greater retention rates, and reduced hospital admissions. Therefore, programs exhibiting high fidelity to the ACT Model experienced increases in the quality of patient outcomes.

Home Visitation Program Models

Home visitation programs (or program models) are a special genre of social action programs in which fidelity is also an important concept. This type of a program model employs individuals to visit the home to meet with the parent or parent(s) who are at risk of abusing and/or neglecting their children. It is important to recognize that “home visiting is not a single, uniform intervention, but rather a strategy for service delivery” (Gomby, 1999, p. 3). Therefore, various program models implement home visiting services in many ways. Differences among models, and sometimes even within models, include the characteristics of the home visitors, processes for identifying “at risk” families, and the time of the initiation of services.

Variability among Program Sites. Variability among program sites is an inevitable occurrence. Gomby (1999) addressed the issue of variability within home visitation programs. Obviously, variability exists in most, if not all programs, but home visiting programs have a certain set of unique characteristics and dilemmas. Specifically, Gomby (1999) explains that one reason for this variability is the local clientele. Programs throughout the country serve different communities that reflect different cultures. Furthermore, within the HFA model, each program can choose the portion of the population they wish to serve. Specifically, some HFA programs (or sites) initiate services prenatally, some postnatally, and some both. Additionally, each local site has different entrance policies regarding such issues as drug problems or involvement with child protective services. Therefore, Gomby (1999) states that if local programs want to achieve strong outcomes, they need to implement the program model with fidelity.

Important Variables in Home Visitation Programs. Although variability among program sites is inevitable, Gomby (1999) and Yoshikawa et al. (2002) maintain that fidelity to certain key program elements is necessary to facilitate program success. Additionally, researchers (Gomby, 1999; Yoshikawa et al., 2002) state that it is imperative to connect these program elements to tangible outcomes.

Gomby (1999) specifically asserts that the training of home visitors, the supervision of the home visitors, and the frequency of the visits are imperative to the success of the program. Additionally, she notes that in most cases, the home visitors only make half of the scheduled visits. The failure to complete the number of required home visits is a major shortcoming of local programs and a failure to maintain fidelity.

Gomby (1999) also states that it is necessary to recognize the inherent difficulty in assessing outcomes for child abuse and neglect programs. At the onset, it would seem logical to use an index of child abuse and neglect because the goal of the program is to prevent such occurrences. However, the utilization of this variable has been problematic. Specifically, it is difficult to use the general population as a control group for a couple of reasons. First, government agencies often do not have accurate rates of child abuse and neglect in the overall population (Gomby, 1999). This occurs because child abuse and neglect is underreported within the overall population. To compound this issue, because the home visitors are required to report child abuse and neglect and are delivering intense services to these families, the HFA families are under closer scrutiny than the population at large and might have more reports of child abuse and neglect. Therefore, using child abuse and neglect rates as outcome indices for home visiting programs could make the program appear less effective than it actually is (Gomby, 1999).

As a result, Gomby (1999) states that it is necessary to select other variables to serve as outcome indices. Common outcome indices that researchers use include whether children have a primary care physician and children's immunization rate (Gomby, 1999). These outcome variables can be especially informative given that Gomby (1999) reports that, of the six major home visiting models she examined, evaluations of these programs yielded no significant benefits for immunization rates or the number of well-child visits. In other words, although a goal of home visiting models is to provide each child with well-child visits as well as immunizations, many programs fall short of this goal.

Olds' Model. One home visitation program model, designed by Olds, uses nurses to visit mothers at risk for abusing and/or neglecting their children (Olds, Henderson, Tatelbaum, et al., 1986). They used specific criteria to identify those in need of this type of services. They included pregnant women who had no previous live births and fit into all of the following categories: "(1) young age (less than 19 years), (2) single-parent status, or (3) low socioeconomic status" (Olds, Henderson, Tatelbaum, et al., p. 18). The program model included nurses visiting the home to provide parental education, skills to develop and strengthen the pregnant woman's informal social networks, and education for the woman on how to access community services. Upon the delivery of the child, the nurses conducted weekly visits and once the infant aged to approximately 16 to 24 months, the nurses conducted visits with less frequency. Each visit lasted approximately one and quarter hours.

Olds' program model is particularly important because it is one of the first to run systematic evaluations of home visiting programs. Olds, Henderson, Tatelbaum et al. (1986) used a randomized design and employed four treatment conditions. The first

treatment condition was the control group that received no home visitation services. The infants in this group received a sensory and development screening at one and two years of age. The second group was provided free transportation to prenatal and well baby physicals. The infants also received a sensory and developmental screening at one and two years of age. The third group received the same services as the second treatment group, but in addition, they received the home visitation service. This service included a nurse home visit every other week with an average of nine visits during the mothers' pregnancy. The fourth group received the same services as the third treatment group, but continued to receive nurse visits until the baby was two years of age.

Using this randomized design, Olds and his colleagues conducted studies, over the past two decades, which demonstrated support for the program. In order to do so, they contrasted groups one and two with groups three and four. In their initial study, Olds, Henderson, Tatelbaum, et al. (1986) discovered that the pregnant women who were visited by the nurse (treatment groups three and four) were more aware of community services, used these services with greater frequency, had better informal social support and exhibited better health habits than women in the control group (treatment groups one and two). A follow-up study which examined the mothers with their children showed that the women who had the nurse home visitor (treatment groups three and four) had fewer instances of child abuse and neglect in the first two years, punished their children less frequently, provided the children with more appropriate toys, and had fewer accidents and poisonings than did the control group (treatment groups one and two) (Olds, Henderson, Chamberlin, Tatelbaum, 1986). Later studies (Olds, Henderson, Kitzman, & Cole, 1995; Olds et al., 1997) indicated that the results of the previous study extended to

four years of age of the child and that the program, for the mother, decreased the number of subsequent pregnancies, decreased her usage of welfare, decreased engaging in child abuse and neglect, and even decreased criminal behaviors for the low-income mothers.

Researchers have also discovered positive, long-term effects of this program. Specifically, the adolescent offspring of the women who received the nurse visits (treatment groups three and four) had fewer instances of running away, fewer arrests and convictions, fewer sexual partners, and consumed less alcohol than their peers in treatment groups one and two (Olds et al., 1998).

It is important to recognize that these studies were all conducted in upstate New York and used a predominantly Caucasian population. In an effort to generalize the results, Kitzman et al. (2000) used the same basic design with primarily African-American participants residing in Memphis, Tennessee. The results of this study indicated some support for Olds' program model, although smaller in magnitude than with previous populations. Specifically, the women who received home visits from the nurses (treatment group four) had fewer subsequent pregnancies, longer gaps between births of subsequent babies, and fewer months using food stamps than women from treatment group two.

Although there is support for Olds' program model, there are a number of important limitations. First, the program model used only nurses as home visitors. By employing nurses, they severely limited the applicant pool of home visitors. Furthermore, it is also possible that this population of women would resist nurses because the nurses might not be able to identify with this at risk group (Leventhal, 2001). Second, whereas the results are strong for the population of Caucasian women living in upstate New York,

they are less substantial for the population of African-American women living in Tennessee. The inability of the program model to generate comparable outcomes for other portions of the population raises serious doubts as to the feasibility of replicating the original program implemented in Elmira, New York. It is important to note that this is a serious limitation of this program model as child abuse and neglect is a nationwide problem. Therefore, it is necessary to design and implement programs that generalize to other portions of the population. Another limitation of Olds' program model is the universal entrance criteria. These criteria only cover a small portion of the population at risk for child abuse and neglect.

Healthy Families America. Healthy Families America (HFA), a home visitation program model also seeking to prevent child abuse and neglect, addresses the issue of variability in its local programs. The HFA model is based upon the premise that child abuse is the result of a variety of factors and these factors affect parents' ability to properly care for and raise their children (Daro & Harding, 1999). The designers of the program model recognized this limitation and therefore, instead of creating a stringent model to be implemented uniformly throughout the country, they structured HFA as general model for local programs to adopt. Furthermore, this general model consists of a core set of elements that the local sites must follow. However, the HFA program model allows sites flexibility as to the way through which to accomplish this end. For example, one of the core elements requires HFA sites to implement a curriculum to use for home visits, although the model does not specify the exact curriculum they need to choose. Therefore, the HFA model is not identical in all sites. Rather, local sites adopt the model, a series of general principles, in order to serve their unique populations.

The core principles upon which HFA is based are referred to as the 12 critical elements (see Appendix A). These 12 elements are grouped into three different key areas: participant identification and engagement, program content and structure, and program staffing and supervision (Daro & Harding, 1999). These core elements are the result of extensive research in the child abuse and neglect literature and incorporate important theories from areas such as human development and parent-child interaction (Daro & Harding, 1999).

The concept of the 12 critical elements differs from Olds' model and is important because it demonstrates HFA's recognition that local programs vary from one another. Specifically, each site has a different setting and a unique population. For example, some programs are set in rural areas, while others operate in urban centers; some have predominantly Hispanic populations, while others have Caucasian or African American populations. Additionally, each program is permitted to use different entrance criteria to meet the needs of the local population. For example, some programs accept only teenage mothers, some work with first time mothers only, and yet others accept a wide variety of clientele. Third, each program is allowed some leniency regarding the initiation of service; the process begins either prenatally or at the birth of the child, depending upon the local program. Fourth, Olds' model solely relies upon nurses to serve as home visitors, whereas HFA employs home visitors with a variety of backgrounds. The latter distinction is important because it can affect the recruitment and selection of home visitors as well as the quality of the services delivered to the families.

Although HFA recognized the need for flexibility among local sites, they also recognized the need to have an intact program model based upon the 12 critical elements.

Therefore, they designed a credentialing process, based upon these core components, in order to certify and evaluate local programs or sites. This certification is an intense process of peer rating that examines all aspects of the program. Thus, the credentialing process serves as a technique for evaluating the accuracy of the adoption of the program model to local sites. The process begins when a local site conducts a self-study evaluation based upon the 12 critical elements. Following this study, two trained raters examine the self study, conduct an on-site visit, interview personnel at the local program, review the program's files and then rate the program along the 12 core elements. Within each of the 12 categories, there are a number of criteria designed to evaluate the broad dimension accurately. These criteria are referred to as second-order and third-order standards (third-order are the most specific criteria). Raters evaluate each criterion (over 100 divided across the dimensions) and assign it a numerical rating from 1-3. The rating is based upon strict guidelines within the credentialing process. In the event of disagreement between raters, these subject matter experts discuss the assignment of a rating until they reach consensus. Following the raters' evaluation, the site receives these initial ratings and then responds to each order receiving a 1. Then, a panel meets to discuss the awarding of the credentialing to the site. Specifically, based upon the reviewers' ratings of 1 and the local site's responses, the panel decides whether to award the credential, defer the site, or not to credential the site. Thus, the HFA program model allows for variability, and the credentialing process provides a way to examine whether local programs accurately adopt the model.

Current Study

The goal of this study was to examine the fidelity of local programs to the Healthy Families Model. Blakely et al.'s (1987) findings suggest that the most successful programs are those that exhibit high fidelity to the program model as well as make additions to that model in order to meet the needs of the local population. Therefore, this study provided the first step in applying Blakely et al.'s findings to the HFA Model by examining the relationship between program fidelity and program outcomes.

Variable Selection. The independent variable, fidelity, was measured using 11 of the 12 critical elements of the HFA program model. Element number seven was not included as part of the index of fidelity because it contained information about children's medical outcomes. These include the percentage of children with updated immunization rates and the percentage of children with primary care physicians. Gomby (1999) supports the usage of these indices as outcome measures.

Hypotheses. Based upon Blakely and his colleagues' findings, it was proposed that higher fidelity adopters have better program outcomes than lower fidelity adopters. Specifically, it was hypothesized that adherence to the HFA model was positively related to the outcome indices of the percentage of children with updated immunizations and the percentage of children with primary care physicians. Specifically, the following hypotheses were proposed, based upon the selected variables. Please note that the two dependent variables, the percentage of children with updated immunizations and the percentage of children with primary care physicians, were reverse coded.

H1: Fidelity to the HFA model as indicated by a composite of 11 out of 12 of the HFA critical elements will be positively related to the percentage of children with primary care physicians.

H2: Fidelity to the HFA model as indicated by a composite of 11 out of 12 of the HFA critical elements will be positively related to the percentage of children updated on their immunizations.

H3: Separately, each facet of fidelity will be positively related to the percentage of children with primary care physicians.

H4: Separately, each facet of fidelity will be positively related to the percentage of children updated on their immunizations.

Chapter Two

Method

Programs

The programs in this sample consisted of 103 HFA programs that had either successfully undergone the credentialing process or had attempted to do so. This study did not consider sites that had not undergone the process simply because there is no metric by which to measure their fidelity to the HFA Model.

Another important consideration in this study was that some sites are considered part of a state system. This means that the state has a number of sites, and not all of them were visited during the credentialing process. Nonetheless, the credentialing classification was applied to all of the sites within this state system. For the purposes of this study, only sites that had undergone the site visits were used because there were no data on the other sites.

An additional point to note is that the original sample consisted of 129 sites that had undergone site visits at the time of data collection. However, only 103 of these sites provided useable data. Eliminated sites included: 17 sites for which credentialing files were missing or incomplete, three sites with no identifiable credentialing process, and six sites that were missing outcome data.

Measures

Healthy Families America Credentialing Tool (1999). The core principles upon which HFA is based were referred to as the 12 critical elements. As previously indicated, these critical elements provided the core basis of what each HFA program should include. Each of the elements was based on careful research in the field of child abuse and neglect.

The Healthy Families America program developed the 1999 Healthy Families America Credentialing Program Self-Assessment Tool based upon the 12 critical elements of HFA. The purpose of the tool was to measure the extent to which each HFA program adheres to the critical elements. These elements, also referred to as first-order elements, were general in nature. An example included, “the training of staff.” Each of the items consisted of second order and third order items. These items were more specific in nature, and over 100 of them were distributed across the 12 critical elements. In order to evaluate a program along each of these items, the credentialing assessors used an ordinal rating system of 1-3. A score of 3 indicates outstanding performance, a score of 2 indicating good performance, and a score of 1 indicating a need for improvement.

Although the scores are simple, the method of calculating these scores was an exhaustive process, involving approximately 10 subject matter experts over the course of approximately one year. The process began when a site conducts a self-study. This study consisted of the site reviewing all aspects of its program related to the 12 critical elements. In this review, the site provided written material explaining the ways in which the program included that element as well as supporting materials documenting this inclusion.

Following the completion of a self-study, two trained, peer reviewers examined the site's completed self-study, reviewed the site's files, and then conducted interviews with staff in order to assign preliminary ratings (1-3) to each order of the Healthy Families America Credentialing Program Self-Assessment Tool. For the second and third order elements, the raters based their determination upon specific indications within the credentialing process. Specifically, for the second and third order elements, the credentialing tool provided specific indicators for the raters to base their ratings. For example the third order item, 6-1.A., states that "the home visitor and family collaborate to identify family strengths and competencies." The score (1-3) were then based upon specific indices. A score of 3 indicated that "the home visitor and family routinely collaborated to identify family strengths and competencies. A score of 2 meant that the home visitor and family collaborated to identify family strengths and competencies. However, some instances found when collaboration did not occur." Finally, a score of 1 indicated that "the home visitor and family did not routinely collaborate. While the process of assigning third order scores was fairly objective, the award of the first order scores was based upon a subjective examination of the lower order elements as well as the description of the first order element.

After the peer reviewers assigned the initial ratings, the site received these ratings and responded to each order receiving a 1. Following this step, a panel of subject matter experts, consisting of two state representatives, two program managers, two trainers, two researchers, and one representative from the HFA Board of Directors, met and examined the peer reviewers' preliminary ratings and the site's response to these ratings. The panel

met to discuss the 1 ratings awarded to each site because a score of 1 is the only indication of poor performance. After reviewing all of the documentation, the panel decided whether to change each 1 rating or keep the rating that the initial reviewer gave it. Based upon these decisions, the panel decided whether to award credentialing, defer, or not to credential the site. In order for a site to be considered for the award of the credentialing, the first order ratings must all have been a 2 or 3 and 85% of the second and third order ratings must have been a 2 or 3. However, these standards provide only a minimum threshold, because although the site must achieve these minimum requirements in order to be considered for credentialing, the final decision rested with the credentialing panel. If a site is credentialed, then the certification was awarded for a period of four years. If the site is deferred, the site was told the length of time of the deferral (three, six or nine months) to correct their deficient areas as well as guidance for improvement. The site then had this period of time in order to make changes and to submit their improvements to the panel for review.

This study only used the ratings of the initial panel as indices of fidelity. The purpose of using the initial panel was to maintain consistency when measuring each site. This was formidable because if a site was deferred, the site would then undergo future reviews by the panel. These reviews, while important, were not used in this study.

The 1999 Credentialing tool was in use between 1998 and 2002. Fifty-nine sites in this study had been assessed using this tool.

Healthy Families America Credentialing Program Self-Assessment Tool (2003).

The 2003 credentialing tool is an updated version of the 1999 Credentialing tool. Its measurement purposes and the content are identical, except the wording and total number of items differ (Table 1). There are 44 sites in this sample that were assessed using this updated tool.

Table 1. A Comparison of the 1999 and 2003 Credentialing Tools

Variables	Sum of Possible 1's for 1999 Credentialing Tool	Sum of Possible 1's for 2003 Credentialing Tool
CE 1	8	8
CE 2	9	9
CE 3	9	9
CE4	9	10
CE 5	9	10
CE 6	19	21
CE 8	5	5
CE 9	10	13
CE 10	21	24
CE 11	8	12
CE GA	25	24
CE Tot ^a	132	145

^a Does not include Element Number 7

Annual Site Profile Update (1998, 1999, 2000, 2001, 2003). The “Annual Site Profile Update” was a survey sent to HFA sites each year for the purposes of ascertaining additional information about the characteristics of a site. Each year the survey was slightly different and the response rates have ranged from 75%-85%.

Each version of this survey collected two relevant variables for the present study: the budget of each site and the total number of families participating in each site within a given year. Out of the sites included in this survey, 86 reported these variables.

This study matched the budget and size of the site with the year that the site underwent the credentialing process. Please note that HFA did not administer a 2002 version because the 2001 survey was distributed late in the year. Therefore, for sites that underwent the Credentialing process in 2002 (a total of 23 sites), data were collected using the 2001 version.

Procedure

The credentialing data were used as an index of fidelity. Fidelity was measured by the number of third-order 1's within each critical element because scores of either 2 or 3 indicated that expectations were met, whereas, a score of 1 indicated a failure to meet expectation. Some second-order elements did not contain third-order elements. In such cases, the second-order standards were used. Thus, high fidelity adopters had fewer 1's, while low fidelity adopters had more 1's.

The outcome indices included the percentage of children with updated immunizations (Credentialing item 7-2.B) and the percentage of children with primary care physicians (Credentialing item 7-1.C/7-1.D.). The variable child immunizations was measured as the percentage of children with updated immunizations at the time of the evaluation process. It should be noted that these percentages did not include children who, for medical reasons, had been advised against immunization. To increase the accuracy of the measure, I used the actual percentages of immunized children and children with a primary care physician provided by the site. A legitimate concern was the integrity of sites' data. Therefore, all data were subject to verification by the HFA Credentialing review panel. In cases in which the panel's calculations were different from the site's reported rates, the panel's calculations were used. This occurred one time for the percentage of children with primary care physicians and four times for the percentage of children with updated immunizations. Additionally, there were times in which sites provided the outcome indexes, but they were unclear in the report. Specifically, the sites provided these dependent variables either semi-annually or quarterly. In these cases, the number was estimated based upon the data the site reported

and the rankings that the panel gave the site. For example, if a site was given a rating of 2 for their medical immunizations third order element, and a 2 corresponded to 80%-89% of children were updated on their immunizations, the average (85%) was reported. This occurred one time for the percentage of children with primary care physicians and six times for the percentage of children with updated immunizations.

Chapter Three

Results

Healthy Families America Credentialing Program Self-Assessment Tools

Since there are a small number of sites assessed by each tool, the study did not have enough power to analyze the results separately. Specifically, Tabachnick and Fidell (2001) state that to have a power of .80, the sample size for a correlation needs to be greater than $50 + 8m$ (m = the number of IV's) and to have a power of .80 for testing predictors, the sample size must be greater than $104 + m$. Therefore, in order to approach these sample sizes, the results of this study were analyzed by statistically combining the sites based upon the tool used.

Based upon the current sample, the overall alpha of both instruments combined together was .69. It should be noted that this reliability estimate included all twelve elements even though only eleven were used as indices of fidelity. The reliability estimate that included only the eleven items used in this study did not vary; it was .68. It should also be noted that these instruments were used in a different way than intended by the designers of the tool. Therefore, the reliability analysis should be viewed cautiously. Additionally, there are no previous estimates of reliability on either of these scales.

Transformation of scores

Since the scores on both the dependent and independent variables (the credential scores) were highly skewed, (see Figures 1-14), logarithmic transformations were conducted on these original scores in an attempt to normalize the data. However, since the scores were so strongly skewed, these transformations did not achieve that end. Based upon the failure of the logarithmic conversions and the multiple sources of the fidelity data (the data comes from two different versions of the Credentialing instrument), it was decided to standardize the scores. Specifically, the scores from the 1999 tool were converted to z scores based upon their respective means and standard deviations. Likewise, scores for the 2003 tool were also standardized based upon their respective means and standard deviations. Thus, the resulting transformations of scores from each tool have been statistically combined in order to examine their relationship with outcome variables.

Figure 1. Histogram of Credentialing Element 1

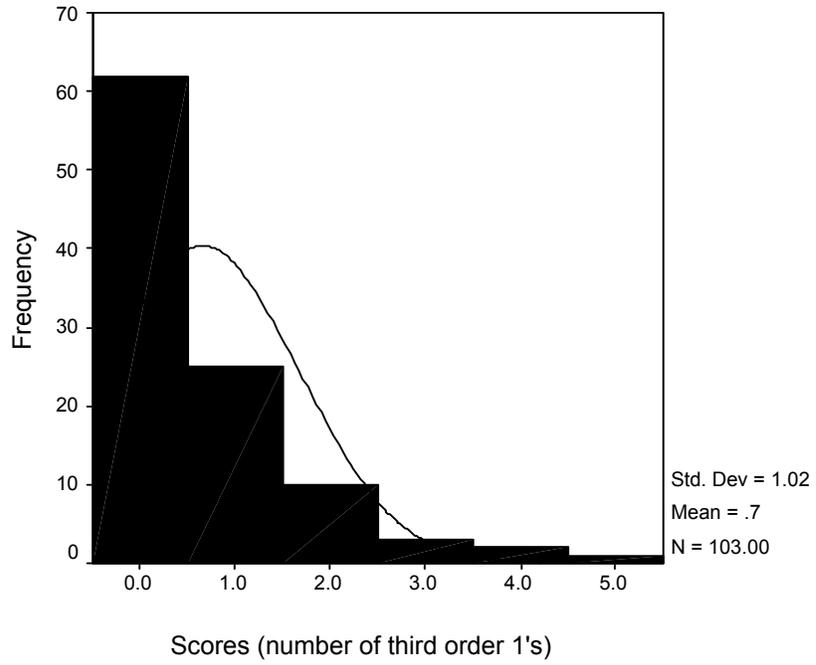


Figure 2. Histogram of Credentialing Element 2

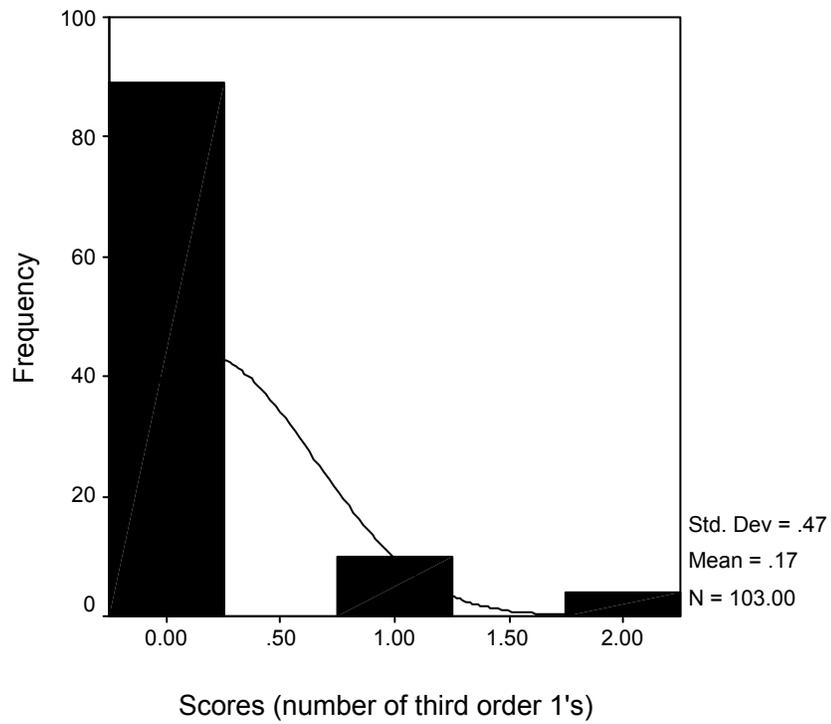


Figure 3. Histogram of Credentialing Element 3

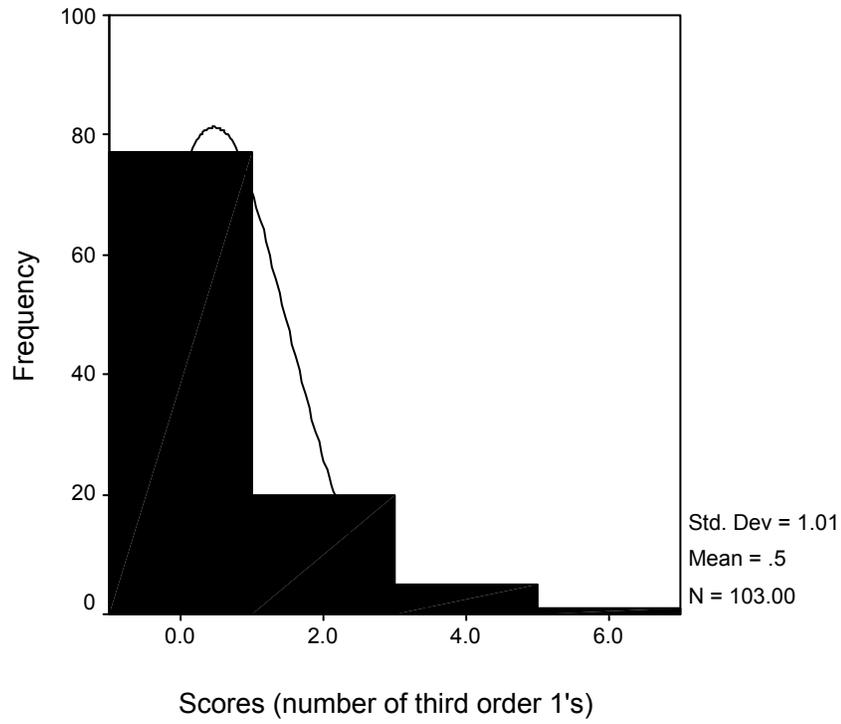


Figure 4. Histogram of Credentialing Element 4

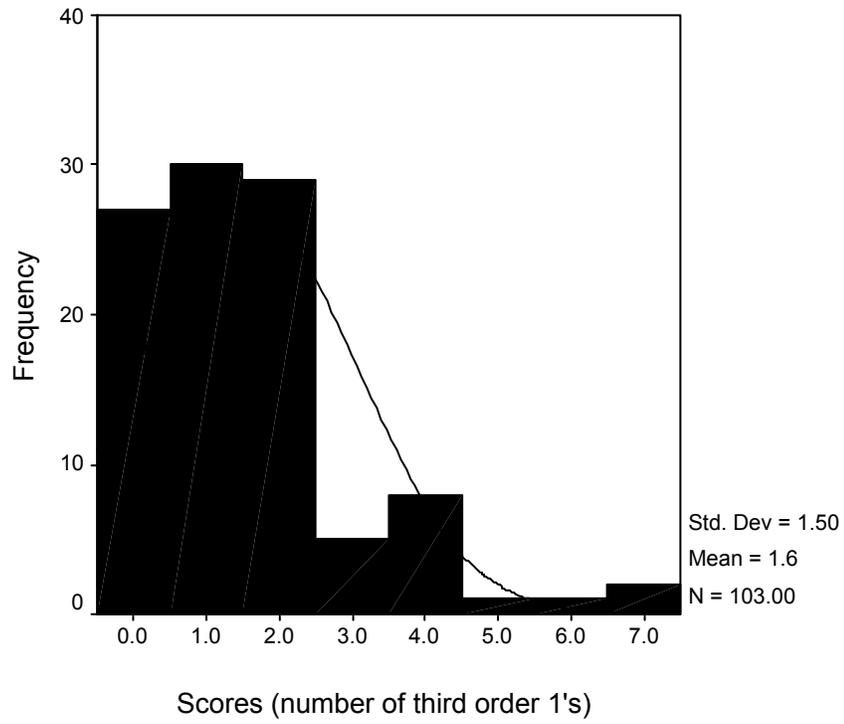


Figure 5. Histogram of Credentialing Element 5

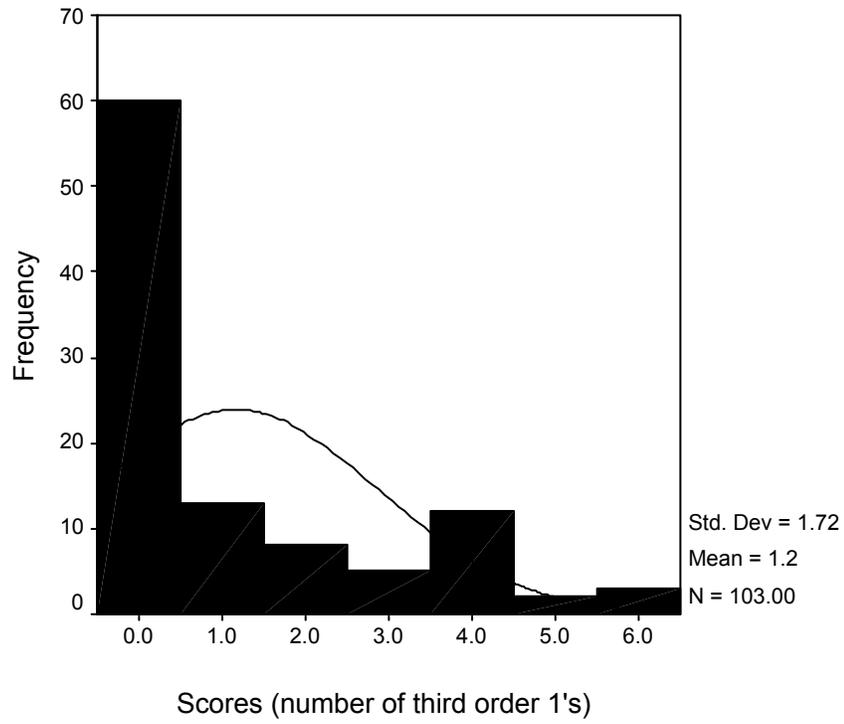


Figure 6. Histogram of Credentialing Element 6

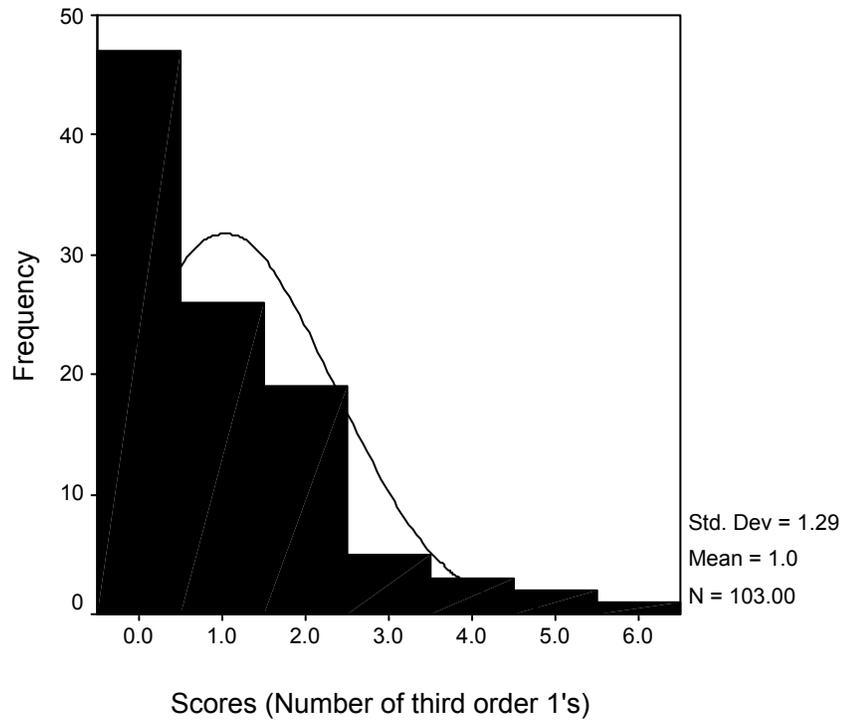


Figure 7. Histogram of Credentialing Element 8

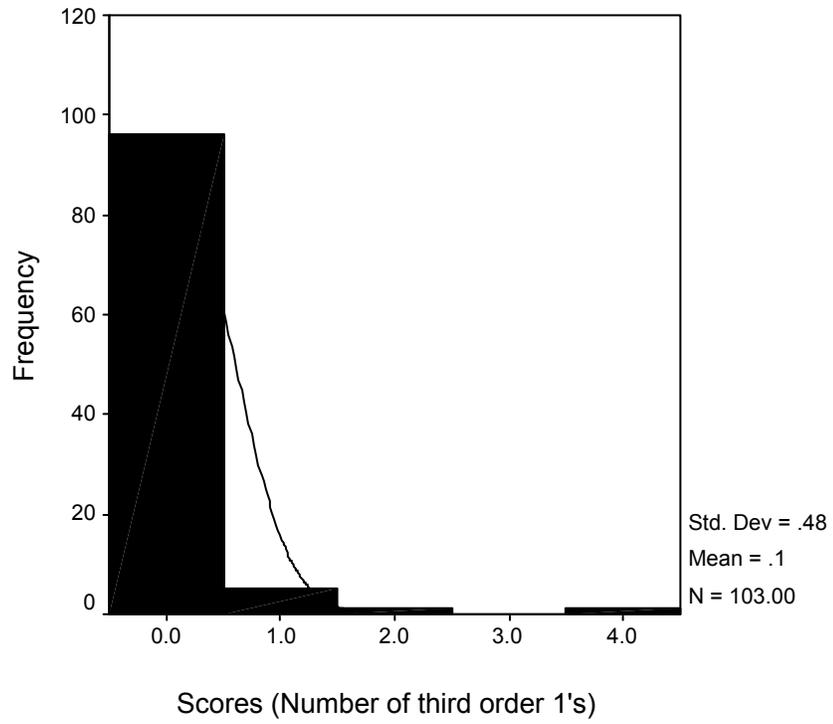


Figure 8. Histogram of Credentialing Element 9

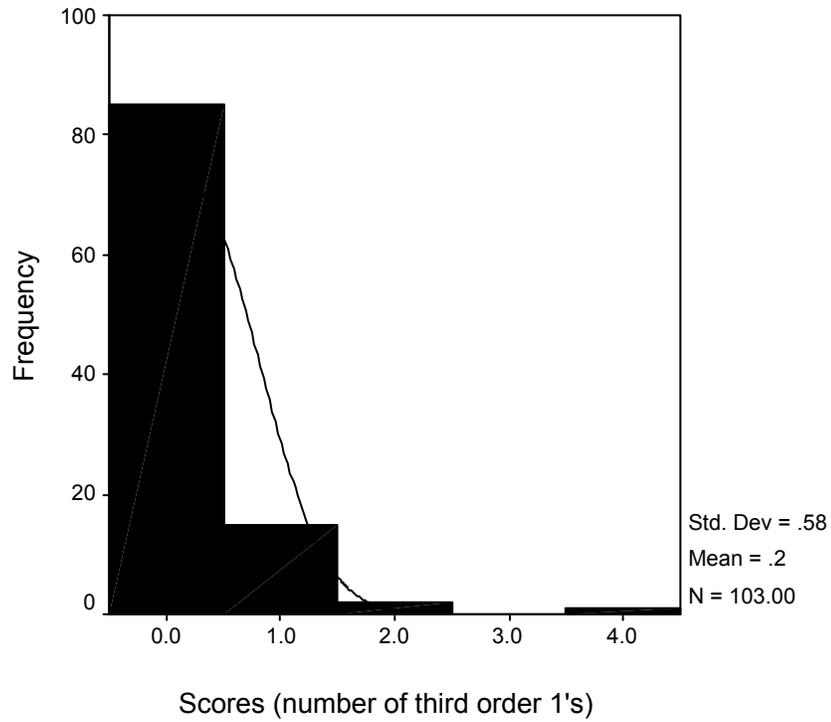


Figure 9. Histogram of Credentialing Element 10

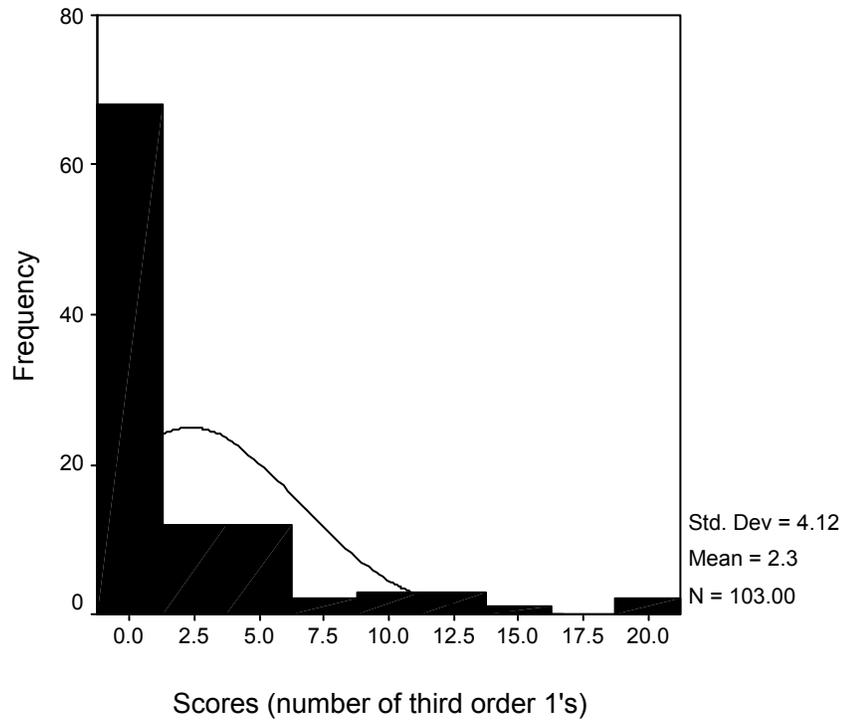


Figure 10. Histogram of Credentialing Element 11

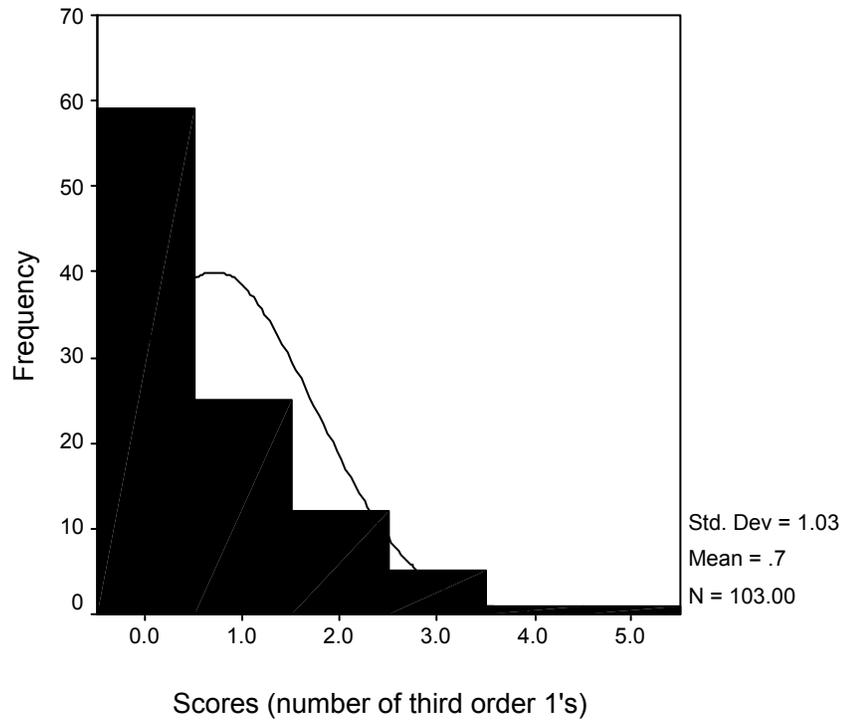


Figure 11. Histogram of Credentialing Element GA

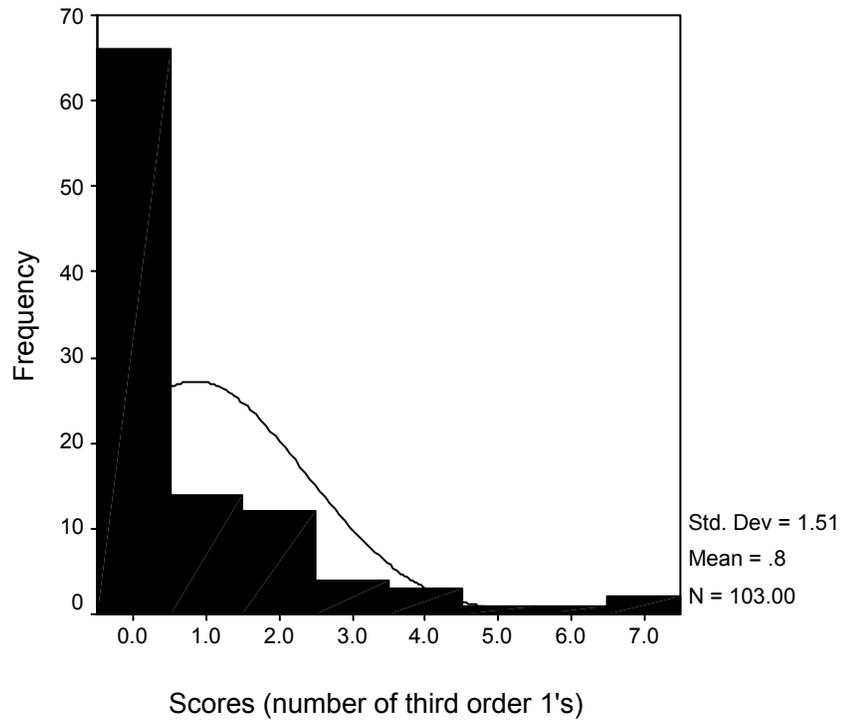


Figure 12. Histogram on Total Credentialing Elements

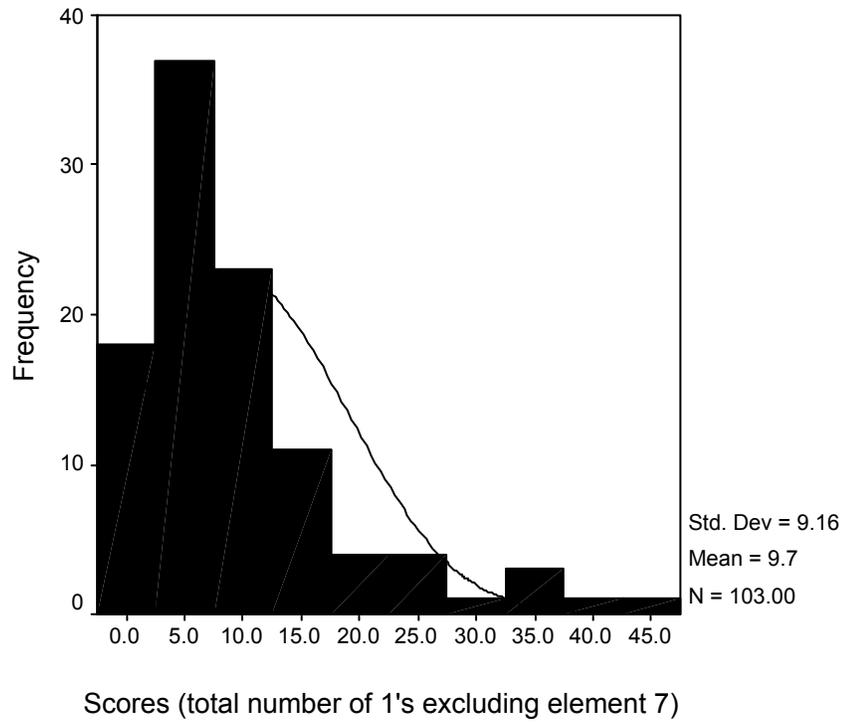


Figure 13. Histogram of the DV: Percentage of Children with PCP's

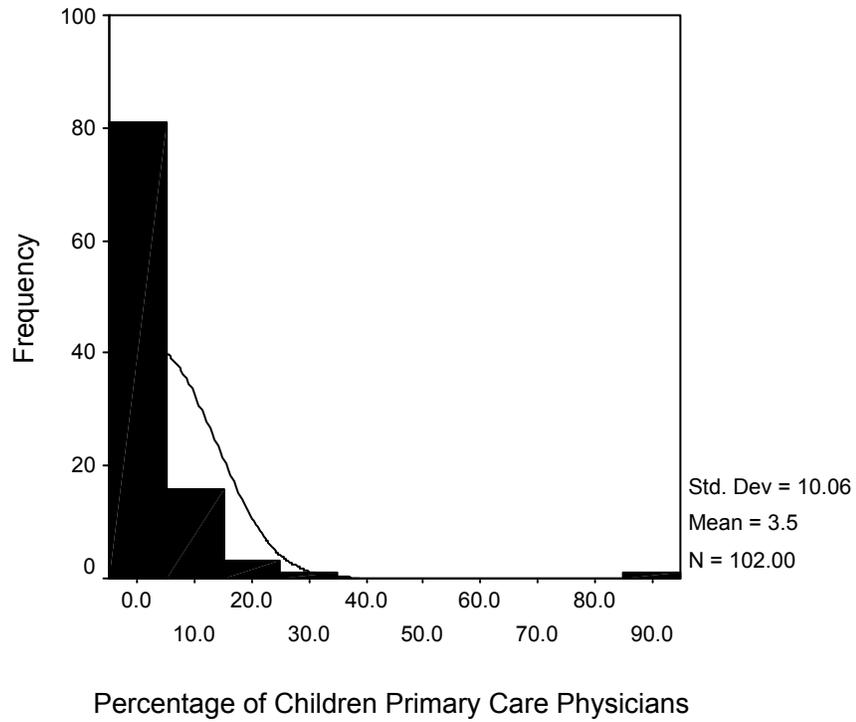
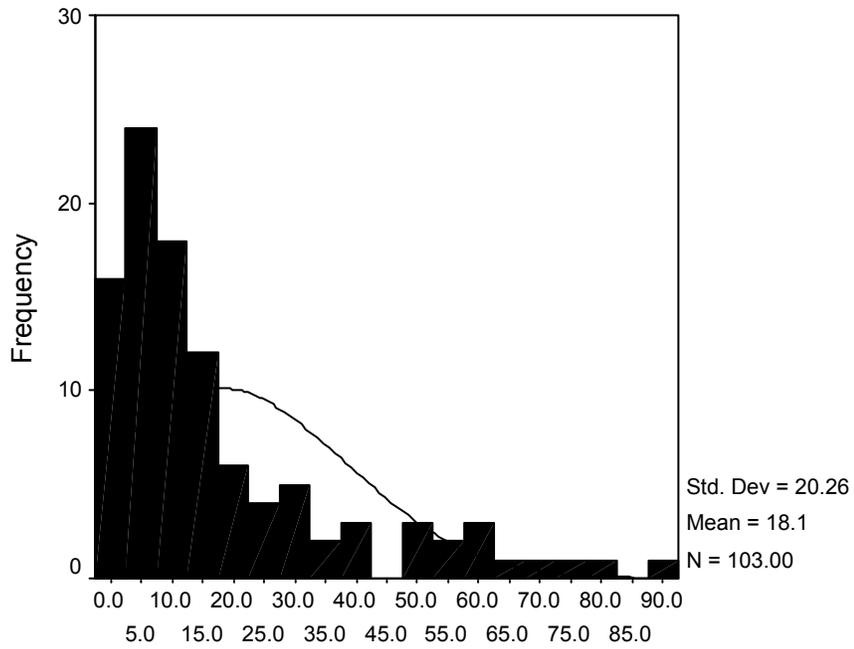


Figure 14. Histogram of the DV: Percentage of Children Immunized



Test of Hypotheses

Support for hypotheses one and two were tested using the Pearson product-moment correlation between a composite of all three independent variables and both of the dependent variables: the percentage of children with primary care physicians and the percentage of children with updated immunizations. Similarly, support for hypotheses three and four were tested using the Pearson product-moment correlation coefficient between each independent variable and each dependent variable. For ease of interpretation, the dependent variables were reverse coded. Therefore, the percentages actually indicated the number of children without a primary care physician and without updated immunizations. Accordingly, all positive correlations referred to a positive relationship (meaning that a high score on the dependent variables refer to lower outcome rates and a high score on the independent variables refer to a lower level of fidelity) and all negative correlations referred to a negative relationship.

Hypothesis Testing

The means and standard deviations of the credentialing elements and the outcome indices (reverse coded), separated by the credentialing tool used, can be viewed in Table 2. In order to statistically combine the scores on the two forms of the credentialing tool, the independent variables were converted to z scores, based upon the means and standard deviations of the respective credentialing tool. The means, standard deviations, and intercorrelations of all variables (combined across both credentialing tools) are reported in Table 3.

Table 2. Means, Standard Deviations (raw form)

Variables	1999 <i>M</i>	1999 <i>SD</i>	1999 N	2003 <i>M</i>	2003 <i>SD</i>	2003 N
CE 1 ^a	.680	1.151	59	.61	.813	44
CE 2 ^a	.170	.461	59	.18	.495	44
CE 3 ^a	.69	1.221	59	.16	.479	44
CE4 ^a	1.95	1.676	59	1.02	1.023	44
CE 5 ^a	1.39	1.781	59	.86	1.593	44
CE 6 ^a	1.25	1.457	59	.75	.967	44
CE 8 ^a	.17	.62	59	.02	.151	44
CE 9 ^a	.31	.701	59	.11	.321	44
CE 10 ^a	2.31	3.94	59	2.41	4.39	44
CE 11 ^a	.900	1.094	59	.45	.875	44
CE GA ^a	1.10	1.626	59	.5	1.267	44
CE Tot ^b	11.644	10.103	59	7.159	7.045	44
PCPR ^c	2.644	5.861	59	4.951	13.920	44
ImmR ^d	18.066	21.642	59	18.080	18.483	44

^a Refers to the Independent Variables: The HFA Critical Elements

^b Sum of scores across all HFA Critical Elements excluding number 7

^c Refers to the dependent variable: % of children with pcp's (reverse coded).

^d Refers to the dependent variable: % of children with updated immunizations (reverse coded)

Table 3. Means, Standard Deviations, and Intercorrelations

Variables	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Z1 ^a	.000	.995	-															
2. Z2 ^a	.000	.995	-.044	-														
3. Z3 ^a	.002	.996	.417**	.001	-													
4. Z4 ^a	.000	.995	.223*	.321**	.288**	-												
5. Z5 ^a	.000	.995	.169	.015	.137	.208*	-											
6. Z6 ^a	.002	.995	.148	.019	.219*	.444**	.216*	-										
7. Z8 ^a	.007	.995	.092	.119	.057	.032	.046	.044	-									
8. Z9 ^a	.000	.995	-.017	-.089	.082	-.051	.037	.333**	-.024	-								
9. Z10 ^a	.010	1.050	.160	.045	.214*	.121	.331**	.212*	.255**	.093	-							
10. Z11 ^a	.001	.995	.337**	.214*	.301**	.320**	.287**	.293**	.185	.153	.479**	-						
11. ZGA ^a	.000	.995	.180	.066	.309**	.246*	.282**	.301**	.013	.179	.383**	.217*	-					
12. ZTot ^b	.000	.995	.393**	.157	.430*	.492**	.571**	.536**	.212*	.210*	.810**	.640**	.601**	-				
13. Budg	568632.81	745115.700	-.046	.107	-.073	-.123	.027	.172	.223*	.162	.105	-.157	.116	.099	-			
14. #Fam	105.490	96.833	.052	.134	-.062	-.086	.017	.166	.239*	.118	.183	-.139	.066	.131	.694**	-		
15. PCPR ^c	3.465	10.062	.028	-.031	-.012	-.003	-.053	.044	.015	.075	-.015	.044	-.029	-.005	.045	-.030	-	
16. ImmR ^d	18.072	20.257	.074	-.101	-.073	.077	.015	.174	.190	.202*	.181	.058	.146	.213*	.329**	.179	.136	-

Note. N's range from 86-103

^a Refers to the z score of each critical element based upon the respective credentialing form (1999 or 2003)

^b Refers to the total score without element 7. Also, the z score is based upon the respective credentialing form (1999 or 2003)

^c Refers to the dependent variable: Percentage of children with primary care physicians (reverse coded)

^d Refers to the dependent variable: Percentage of children with updated immunizations (reverse coded)

* p < .05

Hypothesis one, which predicted a positive relationship between a composite of 11 out of 12 of the HFA critical elements and the percentage of children with primary care physicians (reverse coded), was not supported. The second hypothesis, which predicted a positive relationship between a composite of 11 out of 12 of the HFA critical elements would be positively related to the percentage of children updated on their immunizations, was supported, $r = .213, p < .05$. Hypothesis three, which predicted a positive relationship between each facet of fidelity and the percentage of children with primary care physicians, is not supported. However, there was some support for hypothesis number four which stated that each facet of fidelity would be positively related to the percentage of children up-to-date on their immunizations. Specifically, there is a positive correlation of $r = .202, p < .05$ between element nine (the selection of service providers) and the percentage of children updated on their immunizations. It should be noted that while there was support for hypotheses two and four, only approximately 4% of the variance was accounted for. This indicates that the strength of the association between the variables is fairly small.

Exploratory Analyses

In order to facilitate an understanding of the results, two distinct simultaneous regression analyses were conducted with the 11 predictors against each dependent variable. I also conducted a hierarchical regression equation, using the percentage of children with updated immunizations as the dependent variable, the budget of the site in the first block and the total fidelity score in the second block. Since the fidelity scores were in z scores, the budget was also transformed into z scores to be included in this equation. The reason that the budget was included only in this second series of regression equations was that the data were available for only 82 sites. Furthermore, since the size of the site and the budget of the site were highly correlated, only the budget was used. None of the regression equations were significant. Specifically, the simultaneous regression analyses with the 11 predictors against both the percentage of children with primary care physicians and the percentage of children with updated immunizations were $F(11, 90) = .134, n.s.$ and $F(11, 91) = 1.625, n.s.$, respectively. Additionally, the hierarchical regression equation, using the percentage of children with updated immunizations as the dependent variable, the budget of the site in the first block and the total fidelity score in the second block, was not significant $F(2, 82), = 2.899, n.s.$

Since the study contained multiple dependent variables and it was possible that these dependent variables would be highly correlated, canonical correlation was used to analyze the results further. The purpose of using this type of analysis was to examine all possible combinations of the independent variables and dependent variables. The results of the canonical correlations did not reveal any significant canonical variates, suggesting

that no meaningful combination of variables were evident.

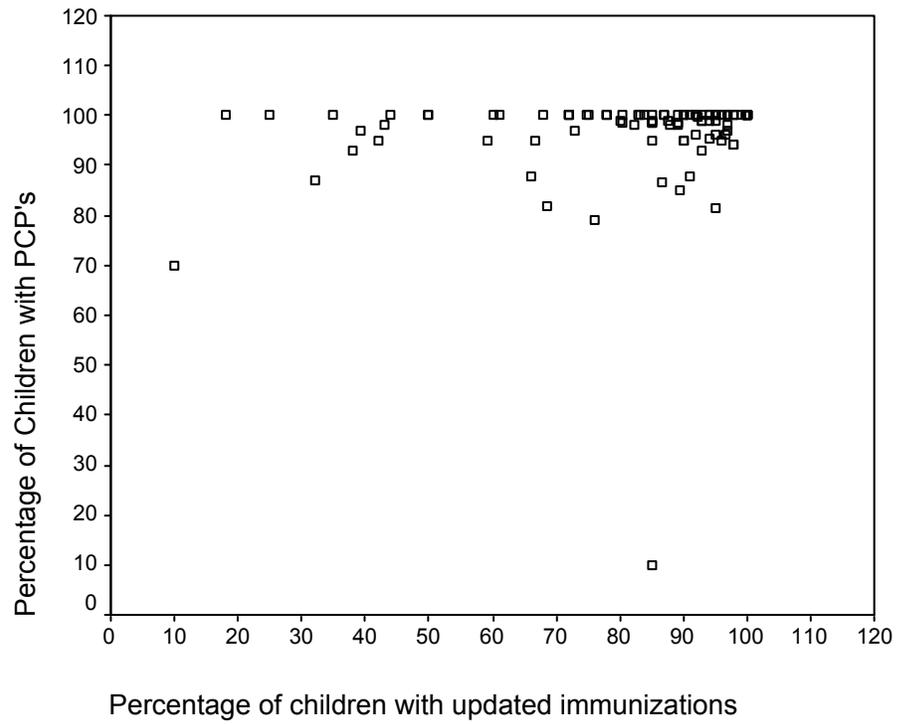
Chapter Four

Discussion

The results of this study indicated that there was a significant relationship between the total score on the credentialing process (without element seven) and the percentage of children with up-to-date immunizations. This means that the degree of adherence to the HFA program model was related to the percentage of children with updated immunizations. Additionally, the results indicated that there was a significant relationship between the score on element nine and the percentage of children with updated immunizations. This indicated that adherence to the HFA model regarding the selection of staff was related to the percentage of children with updated immunizations. However, these relationships each accounted for approximately 4% of the variance.

A surprising finding was that the two dependent variables were not highly correlated. This is most likely explained by the restriction in range of both outcome indices (Figure 15). Specifically, since both dependent variables had ceiling effects, it was difficult to detect a significant correlation. Another surprising finding was that the relationship of the site's budget and the percentage of children with updated immunizations. All of these findings need to be viewed cautiously because some of these significant correlations could have been spurious do to the large number of variables included in the correlation matrix.

Figure 15. Scatterplot of the Two Dependent Variables



None of the exploratory analyses were significant. Perhaps there is no actual effect or maybe there was a lack of power present within this study. Low power occurred because there were a large number of predictors and a small sample size as well as restriction of range in the dependent variables. The sample size was an issue in the hierarchical regression equations; these analyses had only a sample size of 82, whereas a sample size of 106 would have been necessary to detect a significant effect with a power of .80 (Tabachnick & Fidell, 2001).

The findings in this study loosely supported Blakely et al. (1987). Specifically, local sites' adherence to the program model was positively related to one of the outcome indices, the percentage of children with updated immunizations. On the other hand, the present findings differed from Gomby's (1999) findings. Specifically, she indicated that the supervision of staff, training of staff, and frequency of home visits were strongly related to program outcomes. However, the current study failed to support these findings.

Limitations

There are a number of limitations of this study. First, the sites reported their outcome variables as well as their budget and number of families enrolled in the program. The former were subject to review by independent raters. Specifically, if the raters' calculations of the outcome variables differed from those of the individual sites, the raters' calculations were used. This was potentially problematic because it was possible that not every rater recalculated the outcome indices or that they recalculated them in different manners. The latter variables, the budget of the site and the size of the site, were not subject to external review. Therefore, the integrity of individual site's data was questionable.

Another important limitation was the selection bias of sites included within this study. Specifically, only sites that were willing to pay for this review process and to expend the time to undergo it were included in this study. There are well over 400 HFA sites, and fewer than 150 of them have undergone this process. Therefore, sites included in this study were probably sites that believed they would be successful in the review process. Although it would have been beneficial to include all HFA sites, unfortunately, there is no metric by which to measure sites that have not undergone this review process.

A statistical limitation was the lack of power within this study. First, there was range restriction in the dependent variables. This was most likely due to the selection bias; sites undergoing this process might have had stronger outcomes than did sites that did not go through this process. This range restriction made the nature of the relationships between the independent variables and the dependent variables difficult to detect. Additionally, the small sample size and large number of predictors incorporated in

this study also reduced the amount of power. Specifically, Tabachnick and Fidell (2001) suggested a simple equation for estimating sample size: the sample size should be greater than or equal to $50 + 8m$ (m = the number of predictors) and an even higher ratio when the dependant variable is skewed. However, this study could not have included a higher sample size because it used archival data. Third, the data in this study were so highly skewed that logarithmic transformations could not normalize them. Therefore, there is a strong possibility that a Type II error was committed in this study.

A final important limitation of this study concerned the selection of the dependent variables. Specifically, previous research (Gomby, 1999) suggested that an index of child abuse and neglect was not an adequate outcome measure for a variety of reasons. Instead, Gomby (1999) proposed that the percentage of children with updated immunizations and with primary care physicians were better outcome indicators of home visitation programs. Although previous research supports the use of these outcome variables, a selection of different outcome variables might have been prudent. Specifically, one of the core goals of the HFA model was to strengthen families. To this end, a measure of family functioning or interaction might have been a more valid outcome measure. Unfortunately, not all HFA sites use such measures and even amongst the sites using this genre of measure, not all employ the same measure or even parallel measures.

Future Research

Future researchers should replicate this study with an increased sample size as more sites undergo the credentialing process. Additionally, it would be beneficial to explore whether modifications of the program model explain additional variance in the dependent variable. Specifically, it is possible that program additions to the model or subtractions from the model also predict certain program outcomes. For example, some HFA programs have added various nursing components to their local programs. This addition could explain the variance that is unaccounted for.

If the correlation between element nine, the selection of service providers, and the percentage of children with updated immunizations (reverse coded) was not spurious, a separate avenue for future research involves examining sites' selection of service providers. Specifically, the goal of element nine is to select quality service providers based upon their personal characteristics, their willingness to work with culturally diverse communities, and their skills to perform the job. Based upon this goal and its connection to one of the outcome indices, this avenue of research could have potential benefits for HFA sites.

References

- Berman, P. (1981). Educational change: An implementation paradigm. In R. Lehming and M. Kane (Eds.), *Improving schools: Using what we know* (pp. 253-286). Beverly Hills: Sage.
- Blakely, C.H., Mayer, J.P., Gottschalk, R.G., Schmitt, N., Davidson, W.S., Roitjman, D.B., et al. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253-268.
- Bond, G.R., Evans, L., Salyers, M.P., Williams, J., & Kim, H.W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research, 2*, 75-87.
- Caslyn, R.J., Tornatzky, L.G., & Dittmar, S. (1977). Incomplete adoption of an innovation: The case of goal attainment scaling. *Evaluation, 4*, 127-130.
- Daro, D.A. & Harding, K. A. (1999). Healthy Families America: Using research to enhance practice. *The Future of Children Home Visiting: Recent Program Evaluations, 9*, 152-175.
- Glaser, E.M. & Backer, T.E. (1977). Innovation redefined: Durability and local adaptation. *Evaluation, 4*, 131-135.
- Gomby, D.S. (1999). Understanding evaluations of home visitation programs. *The Future of Children: Recent Program Evaluations, 9*, 27-43.
- Jackson, J.B., Paratore, J.R., Chard, D.J., & Garnick, S. (1999). An early intervention supporting the literacy learning of children experiencing substantial difficulty. *Learning Disabilities Research & Practice, 14*, 254-267.

- Kitzman, H., Olds, D.L., Sidora, K., Henderson, C.R., Hanks, C., Cole, R., et al. (2000). Enduring effects of nurse home visitation on maternal life course: A 3-year follow-up of a randomized trial. *Journal of the American Medical Association*, 283, 1983-1989.
- Larsen, J.K. & Agarwala-Rogers, R. (1977). Re-invention of innovation ideas: Modified? Adopted? None of the above? *Evaluation*, 4, 136-140.
- Leventhal, J.M. (2001) The prevention of child abuse and neglect: Successfully out of the blocks. *Child Abuse and Neglect*, 25, 431-439.
- Moncher, F.J., & Prinz, R.J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, 11, 247-266.
- McHugo, G.J., Drake, R.E., Teague, G.B., Xie, H. (1999). Fidelity to assertive community treatment and client outcomes in New Hampshire dual disorders study. *Psychiatric Services*, 50, 818-824.
- Olds, D., Eckenrode, J., Henderson, C.R., Kitman, H., Powers, J., Cole, R., et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association*, 278, 637-643.
- Olds, D.L., Henderson, C.R., Chamberlin, R., & Tatelbaum, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics*, 95, 65-78.

- Olds, D., Henderson, C.R., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., et al. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized control trial. *Journal of the American Medical Association, 280*, 1238-1244.
- Olds, D., Henderson, C.R., Kitzman, H., & Cole, R. (1995). Effects of prenatal and infancy nurse home visitation on surveillance of child maltreatment. *Pediatrics, 95*, 365-372.
- Olds, D.L., Henderson, C.R., Tatelbaum, R., & Chamberlin, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Journal of the American Medical Association, 78*, 65-78.
- Paisley, W.J. (1973). *Post-Sputnik trends in educational dissemination systems*. Washington DC: U.S. Department of Health, Education, & Welfare National Institute of Education. (ERIC Document Reproduction Service No. ED088496).
- Paulson, R.I., Post, R.L., Herinchx, H.A., Risser, P. (2002). Beyond components: Using fidelity scales to measure and assure choice in program implementation and quality assurance. *Community Mental Health Journal, 38*, 119-128.
- Tabachnick, B.G., & Fidell, S.R. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- Teague, G.B., Bond, G.R., & Drake, R.E. (1998). Program fidelity in Assertive Community Treatment: Development and use of a measure. *American Journal of Orthopsychiatry, 68*, 216-232.

Yoshikawa, H., Rosman, E.A., & Hsueh, J. (2002). Resolving paradoxical criteria for the expansion and replication of early childhood care and education programs. *Early Childhood Research Quarterly, 17*, 3-27.

Appendices

Appendix A: The Critical Elements of Healthy Families America

1. Initiate services prenatally or at birth.
2. Use a standardized (i.e., in a consistent way for all families) assessment tool to systematically identify families who are most in need of services. This tool should assess the presence of various factors associated with increased risk for child maltreatment or other poor childhood outcomes (i.e., social isolation, substance abuse, parental history of abuse in childhood).
3. Offer services voluntarily and use positive, persistent outreach efforts to build family trust.
4. Offer services intensely (i.e., at least once a week) with well-defined criteria for increasing or decreasing intensity of service and over the long term (i.e., three to five years).
5. Services should be culturally competent such that staff understands, acknowledges, and respects cultural differences among participants; staff and materials used should reflect the cultural, linguistic, geographic, racial and ethnic diversity of the population served.
6. Services should focus on supporting the parent(s) as well as supporting parent-child interaction and child development.
7. At a minimum, all families should be linked to a medical provider to assure optimal health and development (e.g., timely immunizations, well-child care,

Appendix A (Continued)

etc.) Depending on the family's needs, they may also be linked to additional services such as financial, food, and housing assistance programs, school

readiness programs, child care, job training programs, family support centers, substance abuse treatment programs, and domestic violence shelters.

8. Services should be provided by staff with limited caseloads to assure that home visitors have an adequate amount of time to spend with each family to meet their unique and varying needs and to plan for future activities (i.e., for many communities, no more than fifteen (15) families per home visitor on the most intense service level. And, for some communities, the number may need to be significantly lower, e.g., less than ten (10).
9. Service providers should be selected because of their personal characteristics (i.e., non-judgmental, compassionate, ability to establish a trusting relationship, etc.), their willingness to work in or their experience working with culturally diverse communities, and their skills to do the job.
10.
 - a. Service providers should have a framework, based on education or experience, for handling the variety of experiences they may encounter when working with at-risk families. All service providers should receive basic training in areas such as cultural competency, substance abuse,

Appendix A (Continued)

reporting child abuse, domestic violence, drug-exposed infants, and services in their community.

b. Service providers should receive intensive training specific to their role to understand the essential components of family assessment and home visitation (i.e., identifying at-risk families, completing a standardized risk assessment, offering services and making referrals, promoting use of preventive health care, securing medical homes, emphasizing the importance of immunizations, utilizing creative outreach efforts, establishing and maintaining trust with families, building upon family strengths, developing an individual family support plan, observing parent-child interactions, determining the safety of the home, teaching parent-child interaction, managing crisis situations, etc.).

11. Service providers should receive ongoing, effective supervision so that they are able to develop realistic and effective plans to empower families to meet their objectives; to understand why a family may not be making progress and how to work with the family more effectively; and to express their concerns and frustrations so that they can see that they are making a difference and in order to avoid stress-related burnout.
12. The program is governed and administered in accordance with principles of effective management and of ethical practice.