Interview with Dr. David Olds regarding his reanalyzed findings
January 23, 2006

NFP: You recently released updated, reanalyzed findings on some of the outcomes produced by the Nurse-Family Partnership. What prompted your reanalysis?

David Olds: Our research team has uncovered a better, more accurate method of analyzing what statisticians refer to as "low-frequency-count" outcomes. We have reanalyzed all of the "low-frequency-count" outcomes that we previously published on the NFP and found that some of the program-control differences reported earlier are not statistically significant. In the reanalysis, the estimates of the nurse-visited-control-group differences were pretty much the same, but the significance of those differences was altered. In addition, some of the program effects that had been confined to the higher-risk portions of the sample in the original analysis were found to be significant for the whole sample in the re-analysis. We have a responsibility to the scientific and public policy communities to make sure that they have the most scientifically accurate results we can produce.

NFP: Why do you feel it is important to provide this updated analysis to a broad audience?

David Olds: It has always been our commitment to provide the most accurate findings and this provides an opportunity to continue that tradition. Our team has communicated with the editors of journals that published findings on "low-frequency-count" outcomes and offered to submit the reanalysis. The journal editors responded that the original statistical methods were appropriate for the time we reported the findings and that we would not need to submit an updated version. They noted that in all areas of science, as new statistical methods become available, findings reported earlier are likely to change. To ask authors to update findings in publications as new methods become available would put undue burden on authors and journals. We feel a responsibility, however, to make sure that the public and scientific communities have the benefit of our best work.

NFP: What exactly are "low-frequency-count" outcomes?

David Olds: These are events that do not occur to most people, but that when they do occur may happen more than once. Low-frequency-count outcomes in our studies include events such as the number of arrests, convictions, and days of incarceration that mothers reported for themselves in the 15-year follow-up of the Elmira trial, and the number of child abuse and neglect substantiated reports found in official records through child age 15 in Elmira (all of which remain statistically significant under the re-analysis).

NFP: We have always thought that the NFP had unusually strong evidence. Does this change that belief?

David Olds: No. These new analyses provide a level of scientific scrutiny that is highly unusual for this type of work and this reanalysis increases our confidence in those findings that are significant in the new analysis. Some of the details for some

small set of outcomes are altered, but the basic message is strong. Some of our colleagues have called this reanalysis standard-setting.

NFP: Can you fill us in on the history of these analyses?

David Olds: Historically biostatisticians have had difficulty testing the effects of interventions when the outcomes have this "low-frequency-count" aspect. At the time we published the results on these types of outcomes, the method we employed for analyzing these types of outcomes (generalized linear models, assuming Poisson error) was considered state-of-the-art. Our chief biostatistician incorporated these statistical methods into our program of research in the mid-1980's, about 10 years before these methods became widely available through standard statistical software. Using these state-of-the art methods presented certain challenges, but we knew about these challenges and made adjustments in the analysis to address these challenges that were considered acceptable at the time. The methods passed peer review. Since then, newer statistical methods have been developed to analyze these kinds of outcomes and the availability of these new methods led us to reconsider the earlier methods.

NFP: So what exactly did you do?

David Olds: We reanalyzed each of the variables for which we had employed the earlier statistical method (generalized linear models assuming Poisson error) and compared the results obtained with a newly developed method for handling these types of data (zero-inflated Poisson), as well as with five other methods of analyzing low-frequency count outcomes, including a method known as the negative binomial. For each of the methods, we conducted extensive analyses (simulation analyses or randomization tests) for each of the seven methods for each of the "low-frequency-count" outcomes to determine the likelihood that the statistical test produced an accurate result. These extensive analyses revealed that some of the tests produced inaccurate estimates of statistical significance. In addition, in the original analyses we had truncated all outlying values to a common value to reduce the likelihood that the results would be unduly affected by these outlying values (one of the challenges of employing Poisson methods); in the new analyses we conducted additional simulation analyses (anywhere from 1000 to 5000 separate analyses) for each possible point of outlier truncation for each variable to help determine the point that would give us the most accurate result.

NFP: And what did you find?

David Olds: First, we found that the new method (zero-inflated Poisson) that we thought was going to give us more accurate results did not provide any more accurate tests of statistical significance than did the method we had employed originally. We discovered that an older method (generalized linear models, assuming negative binomial error) was better. Some researchers had considered the negative binomial and Poisson error assumptions to be equally appropriate, if corrections for "over-dispersion" of the data in the Poisson case were invoked, as we had done. Our analyses now show that assuming the negative binomial provides a better "fit" of our data than did Poisson error assumptions. Moreover, we found that even the superior method (generalized linear models, assuming negative binomial error) required additional truncation of some outlying values to produce accurate results.

NFP: So how does this affect our understanding of the effects of the NFP?

David Olds: Some of the differences between the nurse-visited and control group previously reported as significant when tested with Poisson error assumptions are not statistically significant. For the 15-year follow-up of the Elmira trial, these outcomes include a previously reported 44% reduction in behavioral problems on the part of mothers due to their use of substances, a 56% reduction in the number of days that children reported drinking alcohol, and a 60% reduction in the number of times children reported running away from home (all effects concentrated in the group defined by the mothers being low-income and unmarried at registration). The differences between the control group and the nurse-visited group for these variables are no longer statistically significant in the reanalysis. We also reported an 81% reduction in number of youth convictions and a 63% reduction in the number of sexual partners for the 15-year old youth; these findings no longer meet the 5% cut-off for statistical significance but are now "trends" (i.e. there is between a 6% and 10% chance that the program effects observed are due to chance).

It's important to note that the majority of the most important outcomes from the 15-year follow-up of the Elmira sample remain statistically significant. These include program effects on the mothers' rates of arrests, convictions, and days incarcerated – with the arrest and conviction results based upon both self-report and official records from New York State. Analyses of outcomes that did not have the low-frequency count aspect are not affected by the reanalysis; positive program effects for these types of outcomes include reductions in the rates of subsequent pregnancy and births, greater intervals between the birth of first and second children, and reductions in use of welfare. In addition, nurse-visited 15-year old youth were significantly less likely to be abused or neglected (based upon official records), arrested, or adjudicated in Family Court as a person in need of supervision (PINS) for incorrigible behavior, such as persistent truancy or destroying parents' property. These significant effects, which held up under the new analyses, indicate that in Elmira, nurse-visited first-born youth and their mothers fared better than mothers and children randomly assigned not to receive the program. Moreover, in the earlier analyses the statistical significance of almost every difference between the nurse-visited group and the control group for these types of outcomes was concentrated in the higher-risk groups. In the new analysis, they are statistically significant for the entire sample. All previously reported significant effects in the Elmira trial analyzed with other methods remain significant.

NFP: How does this affect the overall picture of NFP program effects?

Some of the details of program effects on adolescents' antisocial behavior have been altered, with fewer statistically significant effects for those youth born to low-income, unmarried mothers. Those effects that remain statistically significant now hold for the entire sample. We have subjected these findings to a nearly unprecedented level of scientific scrutiny.

It's important to remember this: The NFP has been tested in three separate randomized trials, first with a sample of primarily white families living in a semi-rural setting (Elmira); then in a major urban area with a largely African American sample (Memphis); and most recently with a sample that includes a large portion of Hispanics

(Denver). The corrections produced by these new analyses affect only some of the findings from the 15-year follow-up of the Elmira sample.

The findings in which we have the greatest confidence are those found in at least two of the three trials. Given that the trials in Memphis and Denver have followed mothers and children only through younger ages so far, our confidence is greatest for findings earlier in the child's life, but these findings are of substantial importance. The replicated findings include program effects on women's prenatal health (especially cigarette smoking and hypertensive disorders of pregnancy), childhood injuries (the leading cause of death among children and youth in the United States over one year of age), the rates of subsequent pregnancy and intervals between first and second births (outcomes that can affect the mother's ability to care for her first child well and her ability to stay in school and find work because of fewer problems with child care), maternal employment (especially during the child's second year of life), and the children's readiness to enter school (reflected in outcomes such as child language development, attention, and behavioral regulation). We don't know yet whether the beneficial effects observed at child age 15 in Elmira will replicate in the subsequent trials of the program with minorities living in major urban areas, but we are optimistic.

We have a responsibility to the public and policy makers to make sure that the results of this program of research are as scientifically sound as possible and that we utilize the most updated methods of analysis available. It is replicated findings across contexts, populations, and time that increase our confidence in the impact of the NFP.

I am very pleased with the program's effects in the trials to date but promise to continue to conduct research to understand its long-term effects and to improve the NFP model. We continue to follow families from our original trials as the children get older to understand the long-term effects of the program and we are now conducting rigorous research in collaboration with our national program sites to make it even better.

NFP: It seems like your experience might be helpful to other researchers. Are you planning to share your experience with the research community in some way?

David Olds: Yes. We are writing a paper for publication that will share our experiences with other researchers who may benefit from our experience. In this way, we may help ensure that research findings reported in other intervention trials are of the highest caliber.