Chapin Hall at the University of Chicago
Policy research that benefits children, families, and their communities

The Child and Family Services Review Composite Scores: Accountability off the Track

John R. Schuerman
Barbara Needell

2009

child welfare

The Child and Family Services Review Composite Scores: Accountability off the Track

John R. Schuerman
Barbara Needell

Recommended Citation

Schuerman, J. & Needell, B. (2009). *The Child and Family Services Review Composite Scores: Accountability off the Track.* Chicago: Chapin Hall at the University of Chicago

ISSN: 1097-3125

© 2009 Chapin Hall
at the University of Chicago

Chapin Hall
at the University of Chicago
1313 East 60th Street
Chicago, IL 60637

773-753-5900 (phone)
773-753-5940 (fax)

www.chapinhall.org

Acknowledgments

We are most grateful to Christine Wei-Mien Lou, Carol Peng, and Emily Putnam-Hornstein, graduate student researchers at the Center for Social Services Research at University of California at Berkeley, for their assistance in conducting the analyses presented here and for their notes on drafts of this paper.

We are also grateful to readers Terry V. Shaw, Amy D'Andrade, Fred Wulczyn, Mark Courtney, and Matt Stagner for many helpful comments.

About the Authors

John R. Schuerman is emeritus professor at the School of Social Service Administration at the University of Chicago and faculty associate emeritus at Chapin Hall at the University of Chicago. He has been involved in a number of evaluations of child welfare programs and other social programs.

Barbara Needell is a Research Specialist at the Center for Social Services Research at the University of California at Berkeley. As Principal Investigator of the Child Welfare Performance Indicators Project, she has worked extensively with statewide and county-specific administrative data.

Disclosures

John Schuerman has served as a consultant to the California and Florida child welfare departments in their appeals of sanctions under CFSR1. Some material above is adapted from memoranda he wrote in those actions. Barbara Needell works in part under an interagency agreement with the California Department of Social Services, and in this role provided support to California in its CFSR1 appeal.

Introduction

In 1994, Congress enacted revisions to the Social Security Act; they included provisions directing the Department of Health and Human Services to review and evaluate state foster care and adoption services supported by federal funding under Titles IV-B and IV-E of the Act.[1] The review process, carried out by the Children's Bureau of the Administration on Children and Families (ACF), was to determine whether each state was in "substantial conformity" with Titles IV-B and IV-E, HHS implementing regulations, and the approved state plan for child welfare services. Thus was born the Child and Family Services Review (CFSR) process. ACF subsequently adopted regulations governing the reviews.[2] ACF has also promulgated a number of other directives and policy guidelines regarding the review process.[3]

The goal of improving the performance of state child welfare systems is clearly desirable. Too often these programs are failing our society's most vulnerable children and it is important that the programs be held accountable. It is also desirable that these efforts at accountability be data driven, insofar as possible, and the attempt by the federal government to use data is laudable. However, the history of the CFSR process has been rocky. Over time both the data and the process have undergone several revisions.

In this paper, we take a look at the history of the CFSR process and examine the current federal outcome measures. We review the methods used to derive the national standards—six numerical targets that are key components of the review. After an introductory overview of the CFSR process, we discuss the variation among states and how that impacts the quality and use of the data reported to ACF for determining national standards. We then turn to the analytic approach that has been applied to this data, discuss its shortcomings, and make some suggestions for improvement.

[1] Codified now in 42 USC 1320a–2a.

[2] Found in 45 C.F.R. 1355.31–39.

[3] See Appendix for a listing of CFSR documentation relevant to this article.

CFSR Overview

There are two phases to the CFSR process: the Statewide Assessment and the Onsite Review.[4] Six months before the onsite review, ACF sends each state a State Data Profile containing descriptive statistics and outcome measure calculations based on data states have submitted to the National Child Abuse and Neglect Data System (NCANDS) and the Adoption and Foster Care Analysis and Reporting System (AFCARS)[5] in the most recent 3 years. NCANDS has data about children who have been the subject of child abuse or neglect investigations, and AFCARS has data on children in foster care. The Data Profile contains information on how well the state is doing relative to the national standards and other data related to safety and permanency. States use the Data Profile as a foundation on which to develop a Statewide Assessment—a report on child and family outcomes and systemic factors. The Statewide Assessment is submitted to ACF 60 days prior to the Onsite Review. The Onsite Review takes place in three sites, including the largest metropolitan area in the state. It involves a case review of a limited number of cases (50 in the first round of CFSRs and 65 in the second round), as well as interviews with various child welfare stakeholders.[6]

States are held responsible for meeting what ACF refers to as "substantial conformity" on seven outcomes under the three broad domains of safety, permanency, and well-being, and seven systemic factors.

Safety Outcomes

S1. Children are, first and foremost, protected from abuse and neglect.

S2. Children are safely maintained in their homes whenever possible and appropriate.

[4] See National Resource Center for Organizational Improvement, T/TA Related to the Child and Family Services Review (CFSR) Process, CFSR Process Overview Presentation, http://muskie.usm.maine.edu/helpkids/cfsrta.htm.

[5] See http://www.acf.hhs.gov/programs/cb/systems/ for links to more information about these data systems.

[6] We do not consider here flaws in the Onsite Review process, including the inadequate numbers of cases.

Permanency Outcomes

P1. Children have permanency and stability in their living arrangements.

P2. The continuity of family relationships and connections is preserved for children.

Child and Family Well-Being Outcomes

WB1. Families have enhanced capacity to provide for their children's needs.

WB2. Children receive appropriate services to meet their educational needs.

WB3. Children receive adequate services to meet their physical and mental health needs.

Systemic Factors

Statewide information system

Case review system

Quality assurance program

Staff and provider training

Service array

Agency responsiveness to the community

Foster and adoptive parent licensing, recruitment, and retention.

During the CFSR process, 45 items are reviewed using information from the State Data Profile, the Statewide Assessment, Case Record Reviews, and Stakeholder Interviews.[7] States that are not in substantial conformity on one or more of the seven CFSR outcomes or the seven systemic factors must develop a Program Improvement Plan (PIP). To date, no state has passed the CFSR at the initial stage; every state has had to develop a PIP.[8] The PIP is negotiated between ACF and the state, but ultimately the goals and action steps must meet ACF approval. Financial penalties may be incurred if the PIP is not successfully completed.

As part of the requirements for substantial conformity on two of the outcomes (S1 and P1), ACF developed six national standards. The five remaining outcomes (one safety, one permanency, and all three well-being), along with the seven systemic factors, have no national standards associated with them.

---

7 http://www.acf.hhs.gov/programs/cb/cwmonitoring/tools_guide/procedures/appendixb.htm

8 According to ACF, "All 50 States, the District of Columbia, and Puerto Rico completed their first review by 2004. No State was found to be in substantial conformity in all of the seven outcome areas or seven systemic factors. Since that time, States have been implementing their PIPs to correct those outcome areas not found in substantial conformity. The second round of reviews began in the spring of 2007." (http://www.acf.hhs.gov/programs/cb/cwmonitoring/recruit/cfsrfactsheet.htm).

Although the second round of the CFSR process (CFSR2), like the first round (CFSR1), contains six national standards, this paper focuses mostly on the permanency outcome (P1. Children have permanency and stability in their living arrangements) and the four national standards associated with it as they are defined in CFSR2. In CFSR2, the data used to set standards were taken from 2003–2004 data, and up to four linked 6-month AFCARS files were used, depending on the measure.

In part because of severe criticism of the initial CFSR process (CFSR1), HHS revised it for the second round of reviews (CFSR2). CFSR2 attempted to overcome some widely recognized limitations of CFSR1 but as we show here, it has come up short. Many of the CFSR1 flaws were repeated in CFSR2, and additional problems were introduced.

Exploring the Data

State Variation

As in the initial round of CFSRs, the national standards for the CFSR2 were developed from state data, which vary in both character and quality. Because there is great variation in the operation of state systems, it is problematic to combine state data to form national standards. Child welfare statutes vary across states in definitions of such basic things as what constitutes child maltreatment.[9] Beyond differences in laws, regulations and practices differ between states (and even within some states). These differences arise from history, differences in culture, political climate, population characteristics, incidence of community and family problems, and (most likely) some random, inexplicable factors. Furthermore, states differ in which populations are included in their child welfare caseloads, some including and some not including mentally ill and developmentally disabled children, or children in juvenile corrections systems. In addition to different populations, policies, and practices, the meanings of various data elements in the datafiles submitted to ACF and used in the CFSR process differ among states. Some of this variation is due to differences in the laws governing the operation of state child welfare agencies. But another major source of variation stems from differences in practice and in the interpretation of terms. In states with county-administered systems, there may be variation among the counties in these matters.[10] There are also more technical differences: some states include trial home visits, runaways, respite care, preadoptive placements, etc. in their count of placements and status changes, while others do not. Sometimes counting a placement depends on whether it is paid for by the

[9] The provisions of state child welfare laws in a number of areas are summarized at: http://www.childwelfare.gov/systemwide/laws_policies/state/. For definitions of child maltreatment, see: http://www.childwelfare.gov/systemwide/laws_policies/statutes/defineall.pdf

[10] The Child Welfare League of America has studied these variations and their impact on the development of national standards. See CWLA, National Working Group to Improve Child Welfare Data Highlights, "Placement Stability Measure and Diverse Out-of-Home Care Populations," April 2002 and CWLA, National Data Analysis System, Issue Brief, "Can States be Compared Based on Child Welfare Data?" May 2006.

state, other times not. Still further, changes in law or regulation within a state may have significant effect on how placements or transitions are counted, making comparisons of data over time problematic.[11]

Despite these differences, ACF treated all states the same in establishing standards in CFSR2, just as it did in CFSR1.[12] The failure to take into account the differences among states means that the measures developed by ACF capture much more than just the performance of the states. If data from states are to be used to set national standards and determine state adherence to those standards, efforts should be made to remove such factors in order to produce cleaner measures of each state's performance.[13]

Data Quality

States also vary in the quality of data collected and reported to ACF. In 2003, the General Accounting Office issued a report entitled Most States are Developing Statewide Information Systems but the Reliability of Child Welfare Data Could be Improved.[14] On the "Highlights" page accompanying this report, the GAO cites a number of factors that affect data quality, including "inaccurate and incomplete data entry by caseworkers, insufficient caseworker training, differences between state and federal data definitions, and lack of clear, documented guidance from HHS." The report goes on to say that these problems with quality "may lead to inaccurate measures of state performance on federal outcomes." In its response to the GAO report, ACF did not dispute this assertion.

At about the same time the GAO report was published, the Inspector General (IG) of the Department of Health and Human Services examined the AFCARS system.[15] This examination found many of the same shortcomings identified in the GAO report. The IG found that published reports (based on the data used for the CSFR process) contained incomplete data (pp. 5–6). Again the IG cited factors affecting data quality (pp. 6–7). Significantly, the report stated that "[s]tates reported that key AFCARS data elements

---

[11] For a dramatic example of this in Illinois, see Patricia Martin Bishop, Lawrence Grazian, Jess McDonald, Mark Testa, and Sophia Gatowski, "The Need for Uniformity in National Statistics and Improvements in Outcome Indicators for Child and Family Services Reviews: Lessons Learned from Child Welfare Reform in Illinois," *Whittier Journal of Child and Family Advocacy*, v. 1 (2002), no. 1, pp. 1–36, see pp. 8–9.

[12] For a discussion of the importance of adjusting for population and caseload characteristics, see Ramesh Raghavan, "Risk Adjustment Practices," paper presented to Planning Meeting on the Metrics of Performance Assessment in Child Welfare Systems, National Research Council—Institute of Medicine Board on Children, Youth, and Families, Washington, March 13, 2009.

[13] The Children's Bureau website, http://www.acf.hhs.gov/programs/cb/pubs/cwo05/state_data/ contains the Child Welfare Outcomes 2002–2005: Report to Congress. This includes comments by some states regarding their submissions of data (presumably for FFY 2005). A number of states comment on the uniqueness of their definitions of data elements and changes in how they have handled data over time and warn of problems in comparisons both between states and over time.

[14] GAO-03-809, July 2003. This report concerns the quality of Statewide Automated Child Welfare Information Systems (SACWIS) which are the basis of NCANDS and AFCARS.

[15] Office of the Inspector General, DHHS, Adoption and Foster Care Analysis and Reporting System (AFCARS): Challenges and Limitations. March 2003.

are not clearly and consistently defined, resulting in inconsistent reporting" (p. 7). The issue of the handling of juvenile justice populations is of particular concern because they are included in some states' child welfare data and not in others. Placement moves in juvenile justice often occur for reasons quite different than in child welfare, such that the data are not comparable. Again, ACF did not question the IG's conclusions about the quality of data.

Improvements in data have been made in the last few years, but concerns about data quality remain.[16] Judgments about the adequacy of data should be governed by the consequences of decisions made on the basis of those data. The more severe the consequences, the better the data should be. Because the CFSR process can lead to quite severe monetary consequences for states, the data ought to be quite good. Whether or not fiscal penalties that impact vulnerable children and families are appropriate even with data that are quite good is an issue we do not address in this paper.

Conflicts among Measures

As in CFSR1, CFSR2 continues to have multiple measures that are sometimes at cross purposes.[17] In particular, the principles of avoiding placement and effecting reunification as quickly as possible conflict with the measure of placement stability. That is, a state that prevents placement whenever possible and effects early reunifications will be left with a placement population that is more difficult, with greater problems, and therefore more likely to encounter multiple placements. Furthermore, a state adhering to the principle of "least-restrictive alternative" will tend to place children in "lower-level" placement situations whenever possible. In some of these cases, a more restrictive placement will prove necessary (e.g., treatment foster homes or group homes), requiring a change in placement.

We simply do not have assessment technology in child welfare that allows for determining the optimal placement for a child's needs at the time the child is first taken into care. On the other hand, moves from more-restrictive (e.g., group homes) to less-restrictive settings (e.g., family foster care), sometimes made possible by additional services will also result in lower measures of placement stability, even though such moves may clearly be in the best interests of the child. In addition, states that bring many children (often unnecessarily) into care for very short time periods will show inflated performance on this measure, since those children will almost always have two or fewer placements. A state that is moving toward increased permanence for children who have already been in care for long periods of time will demonstrate decreased performance on some measures but improved performance on others. These conflicts are inherent in child welfare practice and policy and states vary in the way these considerations are weighed.

[16] In Child Welfare Outcomes 2002–2005: Report to Congress, cited above, states also comment on improvements in data quality as well as ongoing problems with it.

[17] ACF made some attempt to deal with this problem in CFSR2 through the use of composites, but as we will show, this effort fell short.

Equal Weighting of States

In CFSR2, ACF treats all states the same in establishing national standards, as it did in CFSR1, with no weighting by population or child welfare caseload size. Vermont and California are weighted equally, despite the fact that California serves nearly 60 times as many children in foster care as does Vermont. Thus, some states have inordinate effects on the national standards.

The Lack of Longitudinal Data

A major problem in CFSR1 is the use of what amounted to cross-sectional data. It is generally accepted by child welfare researchers that longitudinal data are far superior for the purpose of measuring system performance and that the use of cross-sectional data may cause serious distortions. In a 2004 article, Courtney, Needell, and Wulczyn demonstrated how point in time data can significantly distort a true picture of child welfare systems.[18] In CFSR2, some of the measures approach the ideal of longitudinal data (e.g., measure C1.3, the measures are listed in the Appendix), but others do not (e.g., C2.1).

[18] Mark E. Courtney, Barbara Needell, and Fred Wulczyn, "Unintended Consequences of the Push for Accountability: The Case of National Child Welfare Performance Standards," *Children and Youth Services Review*, v. 26 (2004), pp. 1141–1154. This article critiques the national standards of CSFR1 by showing that point-in-time data for a number of the national standards measures give significantly different results from those obtained through longitudinal analyses, such that states are likely to rank differently under the two approaches.

The CFSR2 Composites

Overview of Principal Components Analysis

In this section we review and comment on the use of Principal Components Analysis (PCA) for 4 of the national standards in CFSR2. CFSR2 includes 15 measures of permanency, arranged under four areas called "permanency composites": 1) timeliness and permanency of reunification; 2) timeliness of adoption; 3) permanency for children in long-term care; and 4) placement stability.[19] These measures are based on data submitted to AFCARS, data rife with all the problems discussed earlier.

Perhaps because the large number of measures (15) was thought to be unwieldy, the decision was made to combine them into four composites. Combining measures can be done in a number of ways. Usually, it involves the development of "linear combinations," or weighted averages. The simplest such combinations are simple sums or averages (which, from a statistical standpoint, are essentially the same thing) of either the raw measures or of standardized versions of them. Alternatively, they could be weighted in some way, based on theoretical or other considerations. ACF decided to employ a statistical method to determine weights: principal components analysis (PCA). PCA is sometimes thought of as a "data reduction technique" used to simplify a large number of similar measures.[20]

A PCA analysis begins by determining the linear combination (weighted average) that has the greatest variation among all possible linear combinations of the measures.[21] This is the first principal component (PC). Next, of all linear combinations that are uncorrelated with the first PC, the one with the highest variation is selected. This is the second PC. This process is repeated, each successive PC is uncorrelated with the preceding PCs, and the PCs are in descending order of variation. Generally the number of PCs is the same as the number of original variables. Of course, if the number of principal components derived is

[19] There were also two safety measures that we do not deal with here. Apparently more measures were considered. We do not know the process by which these 15 were settled upon. Probably the availability of adequate data entered into the decision. The 15 measures are listed in the Appendix.

[20] Much of the remainder of this article contains considerable detail about the use of PCA and the setting and application of the four permanency national standards. Some of these details appear to be missing from publicly available information from ACF.

[21] Because it is possible to obtain linear combinations of indefinitely large variation simply by choosing very large coefficients, it is necessary to impose constraints on the coefficients.

the same as the number of initial variables, there is not any simplification or data reduction. Data reduction occurs by ignoring some of the PCs, specifically those with lower variation.

The data that go into a principal components analysis are either the correlation matrix of the original variables or the covariance matrix (correlations are covariances of measures that have been put into standard score form, so the correlation matrix is a covariance matrix of standard score forms of the original measures). Using the correlation matrix for PCA means that all the variables are treated as if they have the same variance, namely 1.

ACF used principal components analysis on correlation matrices to develop summary measures within each of the permanency areas. The goal of PCA is to capture as much of the variation in the original measures as possible in a smaller number of synthetic variables, variables that are linear combinations of the original variables. Dropping the PCs with low variation means that not all of the variation in the original variables is captured, but offsetting this loss is the simplification obtained by having a smaller number of variables. The adequacy of a principal components analysis is indicated by the amount of original variance captured. The meaning of an individual principal component is contained in the weights, and this meaning is sometimes not straightforward or intuitive. In general, in a PCA on a correlation matrix, those variables with the highest correlations with other variables are those with the highest weights.

It is not clear why ACF used a statistical manipulation, PCA, to obtain combinations of variables rather than a simpler approach, such as a simple average (perhaps after standardizing the original variables). It is not evident that PCA produces a superior result for the purposes for which the combinations were intended.[22]

A Closer Look at the PCA Analysis

By means of a Freedom of Information Act request, we have obtained the county-level data that ACF used to develop the national standards for each of the permanency areas.[23] The dataset provided included data for each county for each of the 15 measures (these data are percentages or median lengths of time for

[22] Another analytic technique, factor analysis (FA), has some similarities to principal components analysis. FA also computes new synthetic variables as linear combinations of observed variables. However, FA is somewhat more structured than PCA. Generally, in FA it is assumed that the computed synthetic variables reflect underlying forces that in some sense “cause” the variations in the observed variables. We do not believe that FA would be appropriate in the present situation.

[23] We have posted the dataset we received at cssr.berkeley.edu/ucb_childwelfare/CFSR2data

FY 2004), together with the number of children served in foster care in each county in FY 2004.[24] These data allowed us to explore the statistical processes that ACF used to develop the national standards and to provide additional information about the process that is not included in the material published by ACF. We analyzed the data using the FACTOR routine in SPSS, the software used by ACF.[25]

Components Retained

The cutoff point for the PCs that are "kept" is arbitrary, but a commonly used criterion when correlation matrices are used is to keep all PCs with variation greater than the variation of the original measures (called the "eigenvalue one" criterion).[26] From our examination of the data, it is evident that this is not the criterion that ACF used. For composites 1, 2, and 3, ACF chose to keep more components than indicated by the eigenvalue one criterion. It appears that this was done because some variables that ACF wanted to include in the national standards would not have been adequately represented in the final results if the eigenvalue one criterion were used.

Rotations

A procedure sometimes used in principal components analysis is rotation of the principal components. Rotations are used in order to make the PCA more interpretable. ACF did not discuss a rotation step in any published ACF material that we have found. However, in trying to reproduce the component score coefficients published by ACF, we discovered, by trial and error, that a rotation was used. The introduction of rotations introduces certain complications in the analysis. There are a number of methods for performing rotations. The method used here was VARIMAX. Although VARIMAX is the most common rotation method, the selection of rotation method is subject to debate. The use of another method would have produced different county scores for the components, so the result is somewhat arbitrary. The rotation step is quite significant for the eventual results and in our view it is unfortunate that it does not appear to have been disclosed.

---

[24] ACF says that it combined ("rolled up") counties with fewer than 50 children served in foster care in the county in FY 2004. However, there are a few "counties" with fewer than 50 children in the dataset, and none with less than 40. In Federal Register Announcement, Amendments to June 7, 2006 announcement, Attachment B, ACF says that the total number of counties in the analysis was 2,984. However, in the next sentence is says that the number of "counties" in the analysis (after roll up) was 2,119. This is incorrect; there were 2,141. According to http://www.usgs.gov/faq/faq.asp?id=785&category_id=31, there are 3,141 counties in the U.S. including D.C. plus 78 counties (municipios) in Puerto Rico (there are 65 "counties" in these data for Puerto Rico) making a total of 3,219 counties. It is evident that data are missing for a number of counties.

[25] ACF first assigned each county to matched samples, which it calls set A and set B, and ran principal components analyses separately on these sets. This was a good idea, allowing for testing of the robustness of the solutions. After determining that the two subsets produced similar results, the samples were combined and the analyses run on all counties; this analysis was used in subsequent steps. We have not attempted to replicate the A and B analyses (we did not have the A and B assignments of counties, which was done randomly) as it was not necessary for our attempt to reproduce ACF's ultimate results.

[26] Eigenvalues give the variance of the principal components. If the correlation matrix is used, all variables in the analysis have a variance of 1, so if this criterion is used, we keep all principal components with variance greater than 1.

Combining the Principal Components

The principal components could have been used to establish standards for the states. However, in the three areas with more than one PC, ACF proceeded to average the principal components, so that there would be only one measure in each area. This resulted in another linear combination of the original measures, in which the coefficients were simply the averages of the coefficients across the components.[27] This step abrogated the idea behind using PCA in the first place, which is to construct uncorrelated variables capturing the variation in the original variables. It also destroyed the differences in variation among the principal components, ignoring the fact that the first principal component captured the most variation in the original variables. This new linear combination (the average of the principal components) has a variance, but, by definition, it does not maximally capture the variation in the original variables (its variation is smaller than that of the first PC).[28]

County-Level Data

An additional problem with the principal components analysis in this situation lies in the use of county-level data as the input information. Although it is desirable to have a large number of observations, and the use of county-level data is no doubt an improvement over the state-level data used in CFSR1, this had the effect of weighting data from states with large numbers of counties. It is probably desirable to weight states by population or by child welfare population, but the number of counties does not result in a meaningful weighting because it is not closely correlated to the population of states.[29] The numbers of children served in each county were used as weights in determining state scores, but not in the step determining the principal components.

Implications: How Will States Use these Measures?

The linear combination constituting a composite measure does not have a concrete, intuitive meaning; it is an abstraction.[30] These combinations do not provide immediately evident guidance to policymakers as to

[27] In the computational spreadsheets provided by ACF to the states to compute composites by county and for the state overall, the procedure ACF uses is to compute, for each county, the component scores using the principal component score coefficients, and then average these component scores. The step of computing the component scores is not necessary. It would be easier and more straightforward for the states to compute the composite scores using the average principal component score coefficients, which we provide in the Appendix (it would also be slightly more accurate, given the rounding error that occurs in computer routines). It is possible that a state might want to look at the PC scores for counties and the state as a whole, but it would probably be better to examine the data for the original measures.

[28] SPSS computes principal components score coefficients that, when multiplied by standard score forms of the original measures, result in standard score forms of the component values. Hence, these components have variances of 1. The variances of the averages of these components (the composites), across counties are less than 1. They are, for C1: 0.4997; for C2: 0.3331; and for C3: 0.4997. Because there was only one component for C4, averaging was not involved.

[29] Southern and Midwestern states tend to have more counties than others. In this analysis, Texas has the most counties, 136, Georgia is next with 127. The District of Columbia was also included in the analysis as a single county. Puerto Rico has 65 counties in the analysis.

[30] The composite scores were further scaled to range between 50 and 150. More on this follows.

what should be done to improve the well-being of children. However, the composite linear combinations may encourage gaming of the system or other inappropriate actions by the states. In an attempt to "pass" their PIPs, states can analyze the linear combinations, weighing two factors: those measures that have the most impact on a composite and those measures that it is easiest to affect with policy and program changes. Choosing measures to focus on in this manner may not lead a state to address changes most important for children and families.[31]

PCA Validation

ACF has asserted that it tested and validated its Composite Scores in a manner that we cannot substantiate. In a Federal Register Announcement, ACF says it did a "consolidated variable PCA in order to cross-validate the solutions that emerged from the separate PCAs."[32] That is, a PCA with all 15 original measures was computed. ACF goes on to say, "The results from the consolidated variable analysis were identical to those that emerged from the separate PCAs; thus the overall four-composite solution was identical across different data analyses."

ACF does not give details about this "consolidated" analysis, in particular the number of components extracted and whether a rotation step was employed. However, we attempted to reproduce this analysis. We forced the extraction of eight components, the total number of components ACF found in the four composites. We did the analysis using listwise deletion of missing data (as was done in the case of the separate analyses), resulting in an analysis involving 1,412 counties.[33] About 80 percent of the variance in the original measures was captured in this analysis, but four of the principal components had eigenvalues of less than 1 (56 percent of the variance was captured by the four principal components with eigenvalues of greater than 1). The unrotated solution was problematic to interpret, so we focused on the rotated solution. Interpretation usually focuses on the (rotated) component matrix, which, in the case that the correlation matrix is used in the analysis, is the matrix of correlations between the original variables and the components.[34]

---

31 We provide the coefficients for the composite measures in the Appendix.

32 Amendments to June 7, 2006 announcement, http://www.acf.hhs.gov/programs/cb/cwmonitoring/legislation/fed_reg.htm Attachment B, Step 11.

33 We also did the analysis using pairwise deletion (not really a legitimate procedure in these circumstances) and found quite similar results.

34 In this case, the component score coefficient matrix (the coefficients used to produce component scores) yields similar results. The component score coefficient matrix is the product of the inverse of the correlation matrix of the original measures and the component matrix ($\mathbf{R}^{-1}\mathbf{A}$, where $\mathbf{A}$ is the component matrix).

Our efforts to confirm ACF's claims regarding this analysis were only partially successful.[35] It is possible that ACF did this analysis in another way that more clearly validates its assertions. But we have been unable to confirm the claim that "results from the consolidated variable analysis were identical to those that emerged from the separate PCAs." [36]

Setting the National Standards

As indicated above, ACF computed county scores for each of the four composites.[37] These were weighted by the number of children served in foster care in the county.[38] These weighted county scores were then summed over each state and divided by the total number of children in foster care in the counties involved in the computation of the composite. This is the state unscaled score for each composite. These scores were then scaled so that they ranged from 50 to 150.[39]

---

[35] We show the rotated component matrix in the Appendix. The components C1A and C4 are relatively clean, the measures "load" highly on components 1 and 2 and on no other components. (In a rotated solution, the order of the components is relatively arbitrary. They are not necessarily ordered in terms of the amount of variance of the original variables, unlike the unrotated matrix.) However, C1B, which in the four-composite analysis consists primarily of C1.4, is not so clean. Its highest loading is on component 5, but C3.3 is more highly loaded on this component, the primary element in C3B. As to composite C2, the first component is relatively clean, measures C2.1 and C2.2 are highly loaded on component 4. However, C2B, made up primarily of C2.3 and C2.4, is murky, C2.3 is highly loaded on component 3, along with C3.1, while C2.4 is highly loaded on component 7. The last component of C2, C2C, made up primarily of C2.5, is again quite clean, since C2.5 is loaded highly on component 8, with no other measures contributing significantly. In regard to composite C3, things are even murkier. C3A is made up primarily of C3.1 and C3.2. They are split, C3.1 on component 3 and C3.2 on component 6, with C3.1 sharing component 3 with C2.3. C3B, made up primarily of C3.3, shares component 5 with C1.4.

[36] We also tried an analysis using the eigenvalue one criterion, which resulted in four components. On this analysis, the measures in composites 1 and 4 were reasonably "clean," being loaded most highly on components 1 and 3 respectively. C3.1 and C3.2 were loaded on component 2, along with C2.3 and C2.4, while C2.1, C2.2, and C2.5 were loaded on component 4. C3.3 was not loaded on any component.

[37] The coefficients were multiplied by the standardized values (z-scores) of the original measures. Some of the original measures were reversed, that is, the signs of their z-scores were reversed. This was done for those measures in which a low value represented better performance. For years after FY 2004, the z-score calculations embedded in ACF's spreadsheets do not actually produce z-scores for those years, since they use the 2004 means and standard deviations to compute the scores and later years will have different means and standard deviations. In some cases, data missing for counties in FY 2004 are likely to be available in later years.

[38] The weighting was by the number of children served in foster care in FY 2004. This is a very rough weighting. It would have been better to weight the county scores on each measure in the composite by the number of children involved in that measure. An interesting example is composite 4, which combines three measures, with different groups of children, having different numbers of children in each. Nonetheless, ACF treated all three groups as if they had the same number of children.

[39] The formula for this scaling was 50 + 100* (state value – lowest of all state values)/(highest of all state values – lowest of all state values). There is no statistical reason to do this transformation; it was done purely for esthetic, political, and public relations reasons. But it does no harm in considering a single year's distribution of state values. However, ACF has indicated that it will use the same values in computing composites in future years. This means that for future years, the composites will not range between 50 and 150, since the highest and lowest state values will change.

Missing Data

The data used for setting the national standards in CFSR2 come from 51 jurisdictions, all of the states except Alaska (from which apparently no data were available) plus the District of Columbia and Puerto Rico. But the data from the states were uneven across the composites. Data from all states were available for composites 3 and 4, but only for 47 states on composites 1 and 2 (the same states for both composites). Missing were New Hampshire, Oregon, Wisconsin, and Puerto Rico. However, our analysis of the data indicates that for composites 2 and 3, a number of states reported data for only a few counties (see Table 9 in the Appendix), accounting for small numbers of children in care in those states. This is yet another indication that for CFSR2 data quality remained problematic.

An Arbitrary Bar

In the effort to improve the functioning of child welfare systems, the data on these composites could have been used in various ways. The government could have decided to focus on those states that were performing quite badly, say, below the 25th percentile. Alternatively, a triage approach could have been used, focusing efforts on those in the middle, giving high-performing states a pass while essentially giving up on low-performing states as unredeemable. Obviously, such a course of action is politically problematic, but triage is often the best use of resources. Still another approach would be to focus on those states with declining scores, requiring them to at least stabilize their performance. Instead, it was decided to establish national standards at the 75th percentile.[40] This is an arbitrary number, apparently chosen to "set the bar high," and we know of no other justification for it.

But, the 75th percentile for the states was not actually used. Rather, what would be the 75th percentile of the normal curve was used (i.e., the 75th percentile of a normal curve with this mean and standard deviation). The reasons for using a normal distribution for this determination are not clear.[41] The normal curve is a theoretical statistical distribution. Very few quantities in nature are normally distributed. There is no reason to believe that state outcome data are or should be normally distributed.[42] It would have been perfectly legitimate to simply determine the 75th percentile of the distribution as it stood.

In CFSR1, ACF went through a somewhat laborious routine in which certain states were dropped from the determination of the national standards because their data did not fit the normal distribution and so were assumed to be flawed. In CFSR2, ACF set the standard for inclusion of states in the calculation at a point at which all states (having data on the particular composite) were included.

---

[40] In Lake Wobegon all children are above average. Apparently, here, all states are expected to be significantly above average.

[41] Distributions are sometimes transformed to be more normal because some inferential statistical techniques assume a normal distribution in the population from which the sample was taken. It is possible that was the rationale here, since confidence intervals were later constructed (confidence interval construction is an inferential technique). There are, however, significant problems with this procedure, as indicated below.

[42] Technically, most of these data, those concerning percentages, cannot be normally distributed. The normal distribution extends from minus infinity to plus infinity, while percentage data such as these must be between 0 and 100 percent.

Error in the Error Adjustment

As in CFSR1, ACF recognized that the resulting 75th percentile was subject to error and attempted to adjust the value accordingly. ACF used the standard error of the mean for this adjustment, computing a "sampling error" by multiplying the standard error of the mean by the z-score for the upper limit of a confidence interval and subtracting this sampling error from the 75th percentile. However, the standard error of the mean is not the standard error of the 75th percentile, the latter being somewhat larger than the former.[43] Thus, using the standard error of the mean underestimated the adjustment to the 75th percentile by over one-third.[44] In CFSR1, the upper limit of a 95 percent confidence interval was used for this adjustment (z = 1.96). In CFSR2, the upper limit of an 80 percent confidence interval (z = 1.282) was used, a stricter requirement, resulting in states being given less of a "break" in CFSR2.[45] It is not clear why this change was made, other than to "raise the bar" further. Nor is any justification given for the selection of an 80 percent confidence interval (why not 90 percent or 70 percent?). ACF chose one limit in CFSR1 and another in CFSR2, neither of which were justified, both of which are arbitrary, and there is no adequate explanation for the change.

Inappropriate Use of Inferential Statistics

This use of inferential statistics is most appropriate in situations where samples are randomly drawn from a large population. Here, we are not sampling from a population; the 51 "states" are the population. It is sometimes argued in situations such as this that the data points are samples in time, and therefore the data can be taken as a sample. But random variations across time within states are not captured by variations between states at a particular point in time. Sometimes inferential statistics are used in the situation in which the data, while not a sample, are thought to be subject to random influences. But this is largely not the case here. Although some random measurement error is involved, most of the variation among the states is due to non-random factors, many of which are known (see above). Although the computed 75th percentile was undoubtedly subject to error, it is not the kind of error (that is, random) that can be accounted for by inferential statistics.

---

[43] See Maurice Kendall and Alan Stuart, *The Advanced Theory of Statistics*, New York: MacMillan, 1977, pp. 251–252 (or later editions of Kendall). Kendall gives the asymptotic (for large samples) variance of any quantile from a sample from the normal distribution as $p(1 - p)/(nf^2)$, where p is the percentile (here, .75), n the sample size, and f the ordinate (the height of the normal curve corresponding to the percentile). For the 75th percentile the ordinate is about .3175s, where s is the standard error of the mean. Doing the arithmetic shows the standard error of the 75th percentile to be about 1.36 times the standard error of the mean.

[44] Means of large samples are roughly normally distributed, close enough to use the distribution. The distribution of means of moderately large samples (as in this case, 50 or so data points) is usually closer to Student's t distribution. The government used a normal distribution while it would have been better to use a t distribution in the calculation of the confidence interval. But it would have made only a slight difference.

[45] The adjustment to the lower bound of the 80-percent confidence level meant that the adjusted national standards were set at about the 69th percentile. Using the lower bound of the 95-percent confidence interval would result in setting the adjusted national standards at about the 66th percentile. These determinations require the assumption of a normal distribution of the original values for the states.

Expected Improvement of States

States falling short of the national standard are expected to improve during the CFSR period.[46] However, they are not necessarily expected to achieve the national standard; rather, the expectation is that they improve by at least a certain amount. As in other matters, specifying this required progress could have been done in a number of ways. For example, goals might have been developed through federal–state discussions, taking into account the circumstances of each state, their starting points (baselines), their resources, and their limitations. Instead, ACF imposed a minimum percentage of improvement for each state.

In CFSR1, every state falling short of the (adjusted) national standard was expected to achieve the same amount of improvement, regardless of how close or far the state was from the national standard and without regard to circumstances in the state that might affect how much improvement might be expected. States far below the national standard would have an easier time to achieve the expected improvement (they have more room to improve) than states closer to the standard.

In CFSR1, the expected amount of improvement was based on the standard error of the mean of all states. All states below the national standard were expected to improve by the "sampling error," the standard error of the mean times the z value for the upper end of a 95-percent confidence interval (1.96), the same sampling error used in the adjustment of the 75th percentile to set the national standards. Apparently, this was considered to be a "significant" improvement, in the sense of statistical significance, but that is a quite improper use of the idea of statistical significance. It is based on the standard error of the mean, which is not a measure of error in measuring the performance of an individual state. Attempts might have been made to assure that states made "statistically significant" improvements, in the sense that the change would be unlikely to be due to random influences over time, but (as noted earlier) variations among the states do not capture those random influences.[47] In any event, it is not evident that assuring a statistically significant improvement is the best way to set goals. Then there is the problem of setting the significance level. Variation among the states might have been used to determine the significance of change in the states as a whole (their means on various indicators), but not the significance of change in an individual state.

In setting improvement goals for CFSR2, ACF recognized the problem of requiring all states to achieve the same amount of progress. Hence, CFSR2 requires a percentage improvement by states, based on their baseline performance in a certain year. A somewhat more complex calculation was used to determine the

---

[46] We show the national standards for each composite in the Appendix.

[47] One would have to think further about the source and nature of those random influences in devising such a solution. It is possible that variations in a measure over a relatively short period of time within a state could be used in such a calculation. It is possible that proportional hazards models or multilevel analytic techniques could be used to separate "performance" from the variations among states that arise from other sources, see Fred Wulczyn, Lijun Chen, and Britany Orlebeke, "Evaluating Contract Agency Performance in Achieving Reunification," Chapin Hall Center for Children, undated. More research is needed on this possibility.

amount of required improvement. The process began with the sampling error (as in CFSR1), based on a one-sided 80-percent confidence interval ($z = 1.282$), as in the calculation of the national standards. But then, the average performance of the five states just below the national standard was determined. The ratio of the sampling error to this average was then computed and used as the percentage improvement expected of states.[48]

Little justification is given for the process in the relevant ACF Information Memorandum other than "The average of the five States was used to avoid undue influence from States that performed at the extremes on a particular indicator."[49] As we have suggested above, the use of the sampling error to determine expected improvement is highly flawed. It is retained in CFSR2, with the wrinkle that now five states are used to determine the fate of all other states falling short of the national standard. The selection of five states is quite arbitrary, why not 2, or 10, or 20? Most importantly, why should such a procedure be used to determine the expected improvement of states?

The Problem of Proportions of Proportions

This approach encounters another problem. The four permanency national standards are based largely on percents of time some event happens (call this p). They could just as well be based on the percent of time that event does not happen (100% – p). Whichever way one looks at it should not matter in further manipulations of the data. But this is not the case when we take proportions of proportions. For example, one of the national standards for safety in CFSR2 is the proportion of children who were victims of substantiated maltreatment in a particular period who were not repeat victims in the following 6 months. It should not matter if we were to phrase this in terms of the proportion who were victims. Suppose that a state had a baseline rate of 80 percent not subsequent victims (below the national standard of 94.6 percent) and was required to improve by 10 percent of its baseline. This would mean that the state would be required to improve by 8 percent. But if we look at this in terms of the percent of children who were subsequent victims (20 percent) they would have had to improve by 2 percent.[50]

---

[48] This process is explained in ACYF-CB-IM-07-05. Our explanation of this process is simpler than that in the IM, which is unnecessarily complex.

[49] Ibid.

[50] The figure for required improvement is used for illustrative purposes; it is not the requirement derived by ACF. Although this argument is made in regard to a safety measure (where there is only one measure, not a composite), it also applies to the composites, since they are made up of individual measures that include percentages.

Conclusion

As we stated at the outset of this article, the goal of improving the performance of state child welfare systems is clearly desirable. However, the current national standards in the CFSR process do not optimally support that goal. The composite scores that comprise the permanency national standards in CFSR2

- ignore important state variation in the demographics of the children and families served;
- fail to account for systemic state differences in caseload inclusion criteria;
- draw on data derived from a database that was not designed to measure longitudinal performance, and is still not of the quality to justify imposing fiscal penalties;
- disregard the inherent practice and policy conflicts between measures;
- count/weight states equally despite enormous differences in the size of child populations;
- employ a complicated statistical method, principal components analysis, when it is not evident that such a method is in any way required or superior to simpler and more transparent approaches to measurement;
- make many arbitrary and statistically inappropriate decisions in the use of the PCA procedure;
- arbitrarily set the national standard at the 75th percentile, and then rely on ill-conceived rules that set the standard at a different level; and
- use a flawed method to develop a minimum improvement requirement.

It is not at all clear that reasonable outcome indicators can be developed from currently available AFCARS data, given the limitations in those data identified above. One of those limitations, the lack of fully longitudinal data, could be fixed. A Notice of Proposed Rulemaking (NPRM) was released in 2008 proposing that AFCARS be converted to a longitudinal database, along with other reforms.[51] At the time this article was written, the process appears to be on hold. Most, if not all, states have the data needed for

[51] Federal Register Vol. 73, No. 8, Friday, January 11, 2008, Proposed Rules.

this kind of database, and a review of the public comments to the NPRM suggested that this change is welcome.[52] But other problems will remain, most notably differences among states (and within states) that make their data not comparable.

A functional, useful CFSR process would encourage improvement with a clear and coherent use of available data, with an eye to ways to improve both the quality and quantity of data related to children and families who come to the attention of the child welfare system.

[52] http://www.regulations.gov/search/Regs/home.html#docketDetail?R=ACF-2007-0125

Appendix

List of Original Measures for the Composites[53]

Composite 1:

C1.1: Of all children discharged from foster care to reunification in the year who had been in foster care for 8 days or longer, what percent were reunified in less than 12 months from the date of the latest removal from home?

C1.2: Of all children in foster care for 8 days or longer discharged to reunification during the year, what was the median length of stay (in months) from the date of latest removal from home until the date of discharge to reunification?

C1.3: Of all children entering foster care for the first time in a 6-month period, and who remained in foster care for 8 days or longer, what percent were discharged from foster care to reunification in less than 12 months from the date of latest removal from home?

C1.4: Of all children discharged from foster care to reunification during the year, what percent reentered foster care in less than 12 months from the date of discharge?

Composite 2:

C2.1: Of all children who were discharged from foster care to a finalized adoption during a year, what percent were discharged in less than 24 months from the date of the latest removal from home?

C2.2: Of all children who were discharged from foster care to a finalized adoption during the year, what was the median length of stay in foster care in months from the date of latest removal from home to the date of discharge to adoption? [Exit cohort]

---

53 http://www.acf.hhs.gov/programs/cb/cwmonitoring/legislation/fed_reg.htm

C2.3: Of all children in foster care for 17 continuous months or longer on the first day of the year, what percent were discharged to a finalized adoption by the last day of the year? C2.4: Of all children in foster care for 17 continuous months or longer and not legally free for adoption on the first day of the year, what percent became legally free during the first 6 months of the year? C2.5: Of all children in foster care who became legally free for adoption during the year, what percent were then discharged to a finalized adoption in less than 12 months?

Composite 3:

C3.1: Of all children in foster care for 24 months or longer on the first day of the year, what percent were discharged to a permanent home by the end of the year and prior to turning 18? C3.2: Of all children discharged from foster care during the year who were legally free for adoption, what percent were discharged to a permanent home prior to turning 18? C3.3: Of all children in foster care during the year who were either discharged to emancipation or turned 18 while still in care, what percent had been in foster care for 3 years or longer

Composite 4:

C4.1: Of all children who were in foster care for at least 8 days but less than 12 months during the year, what percent had two or fewer placement settings? C4.2: Of all children who were in foster care for at least 12 months but less than 24 months during the year, what percent had two or fewer placement settings? C4.3: Of all children who were in foster care for at least 24 months during the year, what percent had two or fewer placement settings?

Note on the following tables:

In the following, except for Table 1, which comes from ACF, the data are derived from our examination of the database we received from ACF. Some of these data have been published by ACF, others have not. Where ACF has published data, except where noted, the following numbers match those in ACF sources, although ACF has not provided as many decimal places. We provide more decimal places because of the use that might have to researchers wishing to pursue these matters further.

Table 1. Data on original measures, from ACF sources

	Range	Median	Mean	s.d.	*N* states
Composite 1					47
C1.1, %	44.3-92.5	69.9	72	18	51
C1.2, months	1.1-13.7	6.5	7.01	4.28	51
C1.3, %	17.7-68.9	39.4	43	22	47
C1.4, %	1.6-29.8	15.0	13	11	47
Composite 2					47
C2.1, %	6.4-74.9	26.8	32.5495	28.8808	51
C2.2, months	16.2-55.7	32.4	32.500473	13.203543	51
C2.3, %	2.4-26.2	20.2	22.7374	15.1451	51
C2.4, %	0.1-17.8	8.8	9.9281	14.1955	51
C2.5, %	20.0-100	45.8	48.3167	31.4212	47
Composite 3					51
C3.1, %	8.1-35.3	25.0	25.9397	17.1993	51
C3.2, %	84.9-100	96.8	91.4946	19.5984	51
C3.3, %	15.8-76.9	47.8	45.4943	29.175	51
Composite 4					51
C4.1, %	55.0-99.6	83.3	82.6134	11.7676	51
C4.2, %	27.0-99.8	59.9	59.2494	20.094	51
C4.3, %	13.7-98.9	33.9	35.1219	21.412	51

Sources: Range and median from Table A: Data Indicators for the Child and Family Services Review. Means and standard deviations from a "Composite Computational Spreadsheet" developed by ACF and distributed to the states with their Data Profiles and online.[54]

Notes: The above data appear to be for states as a whole, rather than counties, despite the statement quoted below. We do not have data that would allow us to reproduce state figures.

ACF notes: "The range and medians for each individual measure reflect the distribution of all counties that had data for that particular measure, even if that county was not included in the overall composite calculation. A State was excluded from the calculation of the composite national standard if it did not submit FIPS codes in its AFCARS submission or it did not provide unique identifiers that would permit tracking children across fiscal years for variables for which that was relevant."

[54] http://www.nrccwdt.org/resources/cfsr/data_tools.html

Table 1a. Our calculations of similar data, from county-level data

	Range	Median	Mean	s.d.	*N* counties
Composite 1					
C1.1, %	0-100	73.390558	71.586624	17.8240371	2139
C1.2, months	0.492813-36.665298	6.406571	7.009712	4.374811	2139
C1.3, %	0-100	42.105263	42.917458	21.892282	1980
C1.4, %	0-68.7500	11.572802	13.461120	11.399399	1980
Composite 2					
C2.1, %	0-100	25.000000	31.861496	29.869939	1799
C2.2, months	0.098563-143.868583	31.080082	33.397125	14.696473	1799
C2.3, %	0-100	16.666667	18.582034	16.261982	2138
C2.4, %	0-100	0	8.094364	13.668392	2130
C2.5, %	0-100	45.798023	46.9685	32.622284	1652
Composite 3					
C3.1, %	0-100	22.222222	24.139145	18.410156	2134
C3.2, %	0-100	100	91.860838	19.259343	1772
C3.3, %	0-100	45.454545	44.690208	30.214310	2027
Composite 4					
C4.1, %	13.0769-100	84.210526	82.612592	11.767651	2141
C4.2, %	0-100	60.000000	59.236130	20.103398	2141
C4.3, %	0-100	32.413399	35.121871	21.246181	2140

Table 2. Principal Component Coefficients

Composite 1, Varimax rotation, *N* = 1,975 counties

Original Measure	Component C1A	C1B	Average	Communality
C1.1	0.461854	0.084503	0.273179	0.864586
C1.2	0.450755	0.070006	0.260381	0.834910
C1.3	0.294982	-0.00542	0.144781	0.391766
C1.4	0.12893	1.025159	0.577045	0.998162

Composite 2, Varimax rotation, *N* = 1,512 counties

Original Measure	Component C2A	C2B	C2C	Average	Communality
C2.1	0.532637	-0.03192	-0.02635	0.158121	0.833224
C2.2	0.551356	0.106165	-0.0322	0.208441	0.826038
C2.3	-0.08716	0.525895	0.254709	0.231148	0.677423
C2.4	0.139705	0.669455	-0.25562	0.184514	0.779928
C2.5	-0.02971	-0.05936	0.929679	0.280204	0.923182

Composite 3, Varimax rotation, *N* = 1,681 counties

Original Measure	Component C3A	C3B	Average	Communality
C3.1	0.544779	0.136518	0.340649	0.545791
C3.2	0.745698	-0.21975	0.262973	0.726538
C3.3	-0.10786	0.979007	0.435572	0.955327

Composite 4, *N* = 2,140 counties

Original Measure	C4	Communality
C4.1	0.398193	0.653751
C4.2	0.417374	0.718249
C4.3	0.399650	0.658545

Notes: The coefficients and numbers of counties above match those in http://www.acf.hhs.gov/programs/cb/cwmonitoring/cfsr_composite.htm, although the numbers published there are to fewer decimal points. The document to which the above cited table is attached, Federal Register Announcement, Amendments to June 7, 2006 announcement, Attachment B, has quite different, incorrect numbers of counties.

The averages are coefficients that can be used to compute the composites from standard score forms of the original measures, rather than computing composites and then averaging them. The communalities are the percents of variance of the measures that are accounted for by the components. The relatively low communality for C1.3 indicates that statistically it may not belong with the other C1 measures.

Note that on C4 the weights for the three variables on the single principal component are very similar. Hence, a simple average of these variables would have produced a result very close to that of the PCA. However, the variation on C4.1 is much less than that of the other two variables (by about half). This variable contributes about as much as the other variables to the composite, when it could be argued that it should contribute less, given its relative variance.

Table 3. Eigenvalues and percents of variance

Principal Component	Number of variables	Eigenvalue	% of variance
C1A	4	2.204024	55.100604
C1B	4	0.885400	22.134989
Total for C1			77.235593
C2A	5	1.782919	35.658387
C2B	5	1.333158	26.663165
C2C	5	0.923718	18.474355
Total for C2			80.795906
C3A	3	1.362579	45.419296
C3B	3	0.865077	28.835895
Total for C3			74.255191
C4	3	2.030544	67.684816

The total percents of variance shown here differ slightly from ACF's figures in Federal Register Announcement, Amendments to June 7, 2006 announcement, Attachment B.

Table 4. Correlation matrices (listwise deletion of missing values)

Composite 1

	C1.1	C1.2	C1.3	C1.4
C1.1	1.000	-.832879	.397773	.230387
C1.2	-.832879	1.000	-.348799	-.238386
C1.3	.397773	-.348799	1.000	.169639
C1.4	.230387	-.238386	.169639	1.000

N = 1,975

Composite 2

	C2.1	C2.2	C2.3	C2.4	C2.5
C2.1	1.000	-.662412	-.209612	-.086982	.154560
C2.2	-.662412	1.000	.069	.005498	-.154539
C2.3	-.209612	.068687	1.000	.314609	.135964
C2.4	-.086982	.005498	.314609	1.000	.013997
C2.5	.154560	-.154539	.135964	.013997	1.000

N = 1,512

Composite 3

	C3.1	C3.2	C3.3
C3.1	1.000	.220593	-.180132
C3.2	.220593	1.000	-.140138
C3.3	-.180132	-.140138	1.000

N = 1,681

Composite 4

	C4.1	C4.2	C4.3
C4.1	1.000	.536488	.466507
C4.2	.536488	1.000	.541714
C4.3	.466507	.541714	1.000

N = 2,140

Table 5. States with small numbers of counties reporting data

State	Total number of counties	Composite 2 Counties present	% children present	Composite 3 Counties present	% children present
KY	84	1	1.1	15	39
LA	41	10	57	21	72
ME	10	2	40		
SC	39	9	47	21	77
SD	18	3	46	9	63
TN	80	39	64		
WY	16	9	71		
AR	45	24	74		
PR	65	0	0	6	28

Table 6. Permanency Composites

These figures are for state values. The four state composites have been rescaled to have ranges of 50 to 150.

Composite	Unscaled values		Scaled Values							
	Low	High	Median	Mean	Actual 75th Percentile	Normal 75th Percentile	National Standard	Sampling error	Average 5 states	Improvement (%)
C1	-1.706186	0.704621	113.1337	113.7001	121.906	126.0682	122.6034	3.4648525	121.2637	0.028573
C2	-0.6687398	0.611843	96.0386	95.3099	112.4896	110.6613	106.3607	4.3006221	103.9817	0.041359
C3	-1.203927	0.635777	115.8238	112.7142	124.2616	125.0529	121.7374	3.3154597	119.1728	0.027821

Sampling error is based on one side of an 80-percent confidence interval (z = 1.282) multiplied by the standard error of the mean.

“Average 5 states” is the average scaled values for the five states just below the national standard. The improvement factor is the ratio of the sampling error to the five state average. ACF shows these numbers plus 1 (e.g., for C1, 1.029) which is called the “improvement factor,” that is, the state’s baseline (the value in a year subsequent to FY 2004) is multiplied by this factor to determine the level that the state is expected to achieve during the CFSR period.

In Table A: Data Indicators for the Child and Family Services Review, ACF shows the figures shown above for means as medians. The data shown above for the National Standard, sampling error, average 5 states, and improvement percent match those in ACYF-CB-IM-07-05, although that memorandum shows fewer decimal places.

Table 7. Table of correlations of all measures

C1.1	C1.2	C1.3	C1.4	C2.1	C2.2	C2.3	C2.4	C2.5	C3.1	C3.2	C3.3	C4.1	C4.2	C4.3
1.000000	-0.838527	0.427401	0.272620	0.173585	-0.136859	0.030531	0.007658	0.134930	-0.061090	0.038708	-0.182142	0.063835	-0.021911	-0.026634
-0.838527	1.000000	-0.379268	-0.284252	-0.132649	0.128147	-0.024118	-0.024297	-0.133760	0.046044	-0.045967	0.170768	-0.079499	-0.002038	-0.008822
0.427401	-0.379268	1.000000	0.219849	0.092611	-0.048429	0.024755	-0.018000	0.076762	0.026564	0.033521	-0.147411	0.199316	0.038747	0.065309
0.272620	-0.284252	0.219849	1.000000	0.096609	-0.046513	-0.047278	-0.066236	0.042916	-0.046842	0.004451	-0.142806	0.233890	0.144212	0.122288
0.173585	-0.132649	0.092611	0.096609	1.000000	-0.664645	-0.212895	-0.095744	0.154152	-0.293053	-0.020629	0.000359	-0.028908	-0.073479	-0.165126
-0.136859	0.128147	-0.048429	-0.046513	-0.664645	1.000000	0.067230	-0.007155	-0.150764	0.208246	-0.006984	0.007106	0.047735	0.066854	0.123915
0.030531	-0.024118	0.024755	-0.047278	-0.212895	0.067230	1.000000	0.323056	0.150035	0.768250	0.218122	-0.136109	0.069283	0.060936	0.094800
0.007658	-0.024297	-0.018000	-0.066236	-0.095744	-0.007155	0.323056	1.000000	0.045925	0.270503	0.049293	-0.051208	-0.040678	0.013043	0.041804
0.134930	-0.133760	0.076762	0.042916	0.154152	-0.150764	0.150035	0.045925	1.000000	0.123533	0.143742	-0.035625	0.001865	-0.006594	0.017376
-0.061090	0.046044	0.026564	-0.046842	-0.293053	0.208246	0.768250	0.270503	0.123533	1.000000	0.173831	-0.173099	0.069832	0.071959	0.124820
0.038708	-0.045967	0.033521	0.004451	-0.020629	-0.006984	0.218122	0.049293	0.143742	0.173831	1.000000	-0.126294	0.076514	0.070256	0.103009
-0.182142	0.170768	-0.147411	-0.142806	0.000359	0.007106	-0.136109	-0.051208	-0.035625	-0.173099	-0.126294	1.000000	-0.056081	-0.026476	-0.055095
0.063835	-0.079499	0.199316	0.233890	-0.028908	0.047735	0.069283	-0.040678	0.001865	0.069832	0.076514	-0.056081	1.000000	0.569772	0.453434
-0.021911	-0.002038	0.038747	0.144212	-0.073479	0.066854	0.060936	0.013043	-0.006594	0.071959	0.070256	-0.026476	0.569772	1.000000	0.499554
-0.026634	-0.008822	0.065309	0.122288	-0.165126	0.123915	0.094800	0.041804	0.017376	0.124820	0.103009	-0.055095	0.453434	0.499554	1.000000

Table 8. PCA on all measures

Rotated Component Matrix

	Component							
	1	2	3	4	5	6	7	8
C1.1	.912	-.032	-.061	-.068	-.108	.033	.077	.090
C1.2	-.896	.000	.082	.032	.111	-.039	-.124	-.113
C1.3	.665	.127	.214	-.105	-.012	-.058	-.334	-.140
C1.4	.273	.260	-.129	-.005	-.564	-.310	-.090	.282
C2.1	.092	-.062	-.177	-.878	-.030	-.019	-.073	.085
C2.2	-.058	.052	.053	.906	.004	-.015	-.069	-.057
C2.3	.030	.047	.890	.058	-.026	.100	.175	.085
C2.4	.039	.020	.273	-.015	.015	-.017	.887	-.035
C2.5	.082	-.018	.153	-.138	.036	.111	-.018	.906
C3.1	-.046	.055	.896	.189	-.085	.048	.084	.070
C3.2	.040	.099	.113	.002	-.082	.919	-.017	.112
C3.3	-.096	.028	-.168	.031	.855	-.195	-.022	.134
C4.1	.111	.820	.087	-.038	-.043	-.031	-.154	-.043
C4.2	-.034	.846	.004	.011	-.007	.021	.046	-.008
C4.3	-.016	.767	.021	.149	-.032	.097	.097	.044

Table 9. Table of State Composite Scores

						Unscaled scores				Scaled scores			
State	Children Served	C1 children	C2 children	C3 children	C4 children	C1	C2	C3	C4	C1 scaled	C2 scaled	C3 scaled	C4 scaled
AL	8889	8889	8118	8475	8889	0.20361	-0.5849846	-0.1186378	0.7561146	129.21812	56.540396	108.99259	109.83363
AZ	13970	13970	13758	13758	13970	-0.2398615	0.1257395	0.0376136	-0.3933742	110.82297	112.04043	117.48588	85.069605
AR	6218	6138	4590	4933	6218	0.7046209	0.0338268	0.0872726	-1.0216907	150	104.86302	120.18517	71.533463
CA	129678	129678	119841	117424	129678	-0.3579254	-0.4634097	-0.105534	-0.2428084	105.9257	66.034109	109.70487	88.313322
CO	14733	14733	14450	14260	14733	-0.1447092	0.1372446	0.2282909	-0.0592164	114.76988	112.93885	127.85044	92.268538
CT	8569	8569	7968	8569	8569	-0.2807069	-0.559553	-0.3722044	0.2445829	109.12871	58.526331	95.209579	98.81344
DE	1572	1572	1572	1572	1572	0.0414102	0.0309688	0.4697386	1.1993619	122.49009	104.63984	140.97471	119.38273
DC	3225	3225	3225	3225	3225	-0.949265	-0.2694991	-0.0358253	0.855364	81.396997	81.176472	113.49399	111.97181
FL	49834	49771	48873	49377	49834	-0.2055087	-0.0313413	0.2061221	0.2060371	112.24792	99.774077	126.64542	97.983028
GA	25088	25088	23840	23565	25088	0.2840438	-0.1320959	0.0322015	0.8006461	132.55451	91.906209	117.19169	110.79299
HI	5134	5134	5134	5134	5134	-0.1124453	0.1683991	0.1604069	0.3144479	116.10819	115.37169	124.1605	100.31858
ID	2897	2897	2703	2617	2897	0.2907345	-0.0640739	0.2688971	-0.0769606	132.83204	97.218012	130.05765	91.886265
IL	23411	23411	23229	22966	23411	-1.706186	-0.2834822	-0.3479869	-0.1419075	50	80.084546	96.525963	90.48708
IN	15033	15033	14499	14306	15033	0.6057554	0.2254803	0.2764811	0.188667	145.89907	119.82913	130.4699	97.608816
IA	10753	10753	10082	10052	10753	-0.3416726	0.3532392	0.1943405	-0.03429	106.59986	129.80575	126.00501	92.805539
KS	8146	8146	7469	7773	8146	-0.2920788	-0.3248448	0.050145	-0.9231826	108.65701	76.854562	118.16704	73.655673
KY	11737	11737	126	4537	11670	0.0132474	0.2053078	0.326749	0.010713	121.3219	118.25387	133.20229	93.775063
LA	7051	6964	4003	5075	7051	0.0020243	0.0313991	-0.5059666	-0.2479905	120.85637	104.67344	87.938728	88.201681
ME	3535	3535	1419	3385	3535	-0.678244	-0.0644667	-1.203927	-0.4250186	92.638916	97.187334	50	84.387874
MD	13187	13187	13056	12932	13187	-0.8308711	-0.6687398	-0.3981735	1.0842034	86.307962	50	93.797993	116.90181
MA	17923	17923	17923	17923	17923	-0.1523011	-0.3227923	0.0337975	-0.4276902	114.45497	77.014841	117.27845	84.330318
MI	29568	29568	29502	29317	29568	-0.484998	-0.1732974	0.1289138	0.0319274	100.65474	88.688813	122.44865	94.232095
MN	14371	14371	13403	13588	14371	-0.0028407	-0.1298528	-0.1713625	0.1590359	120.65457	92.08137	106.12666	96.970458
MS	4366	4366	3627	3799	4366	-0.243072	-0.3387539	0.2117521	0.1283689	110.6898	75.768406	126.95145	96.309784
MO	16896	16896	15436	15328	16896	-0.0191998	0.2882324	0.0178882	-0.5196575	119.976	124.7294	116.41367	82.349019
MT	2975	2975	2540	2621	2975	-0.6207817	0.0067039	-0.2222595	0.2161559	95.022447	102.74501	103.36007	98.201023
NE	9078	9029	8590	8927	9078	-0.1445121	-0.3692995	0.5855531	-0.3521016	114.77806	73.383118	147.26999	85.958762
NV	7521	7521	7334	7325	7521	0.0568693	-0.479836	-0.2656028	1.6415459	123.13134	64.751388	101.00408	128.90893
NH	1710	0	0	1660	1710			-0.2885034	0.2056589			99.759279	97.974881

						Unscaled scores				Scaled scores			
State	Children Served	C1 children	C2 children	C3 children	C4 children	C1	C2	C3	C4	C1 scaled	C2 scaled	C3 scaled	C4 scaled
NJ	19285	19285	19285	19285	19285	-0.1841547	-0.3415156	0.1641248	0.0091249	113.13369	75.552747	124.36259	93.740849
NM	3982	3982	3698	3457	3982	0.0053786	0.2246406	-0.2301139	-0.8016319	120.99551	119.76355	102.93313	76.274302
NY	46026	46026	45715	45924	46026	-0.8202077	-0.6009742	-0.0164996	0.8660917	86.75028	55.29178	114.54447	112.20292
NC	15033	15033	14826	14709	15033	0.4526951	0.2821835	-0.0083376	-2.0212243	139.55015	124.25705	114.98813	50
ND	2148	2148	1788	2007	2148	-0.2953708	-0.2207763	0.1809605	0.0205933	108.52046	84.981206	125.27772	93.987919
OH	30699	30699	29560	30417	30699	-0.3101026	-0.0892941	-0.0522702	0.0800617	107.90938	95.248579	112.60011	95.269076
OK	15624	15624	15328	14865	15624	-0.027459	0.009807	-0.013184	-0.7120713	119.63341	102.98732	114.7247	78.203751
OR	14218	0	0	14091	14218			-0.2197493	0.1756061			103.49652	97.327439
PA	33600	33600	32914	33470	33600	-0.8973493	-0.2737659	0.2775444	0.4281561	83.550456	80.843284	130.52769	102.76825
RI	3602	3602	3602	3602	3602	-0.4134826	0.4478444	0.1254388	0.1703188	103.62119	137.19341	122.25976	97.213531
SC	7692	7608	3597	5902	7692	0.440664	-0.1667593	-0.6173914	-0.5866604	139.05109	89.199366	81.882057	80.905541
SD	2680	2680	1242	1678	2680	-0.2076124	0.005614	-0.8244193	-0.6214979	112.16066	102.6599	70.628733	80.155019
TN	14265	14158	9165	11436	14265	-0.2221702	-0.1576558	0.0855871	-0.7872599	111.55681	89.910256	120.09355	76.583925
TX	35331	35265	34494	34260	35331	0.2004456	-0.0791774	-0.4620091	-1.2107795	129.08686	96.038583	90.328107	67.459825
UT	3827	3827	3535	3600	3827	0.5515241	0.6118434	0.0785025	-1.2277216	143.64956	150	119.70846	67.094833
VT	2163	2163	2163	2163	2163	-0.2377505	0.1180573	-0.2081755	-1.2615461	110.91054	111.44053	104.12563	66.366134
VA	9609	9609	8409	8994	9609	-0.0747273	-0.3729358	-0.1072243	0.1537764	117.67272	73.099165	109.61299	96.85715
WA	15413	15413	15296	15213	15413	-0.3534867	0.1495036	0.0417358	-0.1256904	106.10982	113.89615	117.70994	90.836455
WV	5101	5051	4316	4898	5101	-0.4301339	-0.3493579	0.0070356	-0.0474194	102.9305	74.940349	115.82377	92.522685
WI	12668	0	0	11966	12668			0.0048353	0.7241699			115.70416	109.14543
WY	2044	2044	1460	1787	2044	0.4147185	0.2965757	0.6357773	0.4528556	137.97488	125.38093	150	103.30037
PR	9459	0	0	2648	9459								

Some Sources on CFSR2

The Children's Bureau website, http://www.acf.hhs.gov/programs/cb/pubs/cwo05/state_data/ contains the Child Welfare Outcomes 2002–2005: Report to Congress. Data for each state on CFSR2 for 2004 and 2005.

Legislation, Policy, and Technical Bulletins Related to the CFSRs
http://www.acf.hhs.gov/programs/cb/cwmonitoring/legislation/index.htm

National Resource Center for Child Welfare Data and Technology
http://www.nrccwdt.org/resources/cfsr/cfsr.html

Database of state CFSR reports
http://childwelfare.net/cfsreview/hhs_docs/statereports/

University of California Child Welfare Dynamic Report System (page with CFSR2 raw data)
http://cssr.berkeley.edu/ucb_childwelfare/CFSR2data

About Chapin Hall

Established in 1985, Chapin Hall is an independent policy research center whose mission is to build knowledge that improves policies and programs for children and youth, families, and their communities.

Chapin Hall's areas of research include child maltreatment prevention, child welfare systems and foster care, youth justice, schools and their connections with social services and community organizations, early childhood initiatives, community change initiatives, workforce development, out-of-school time initiatives, economic supports for families, and child well-being indicators.

1313 East 60th Street
Chicago, IL 60637

T: 773.256.5100
F: 773.753.5940

www.chapinhall.org

Chapin Hall at the University of Chicago
Policy research that benefits children, families, and their communities