CHILD AND FAMILY SERVICES REVIEW
TECHNICAL BULLETIN #8–AMENDED

Re-Issued on May 13, 2015

This Technical Bulletin is designed to accompany the Final Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews (CFSRs) published in the *Federal Register* on October 10, 2014. This information is specific to the third round of the CFSRs. It provides technical detail on how we calculate whether a state has met the national standards. It also provides technical information on establishing program improvement goals relative to the statewide data indicators for states not meeting national standards.

This guidance is in accordance with 45 CFR 1355.35(a)(1), which requires that a state's program improvement plan (PIP) describe methods that will be used to evaluate progress. If the Children's Bureau (CB) and the state cannot reach consensus on the content of a PIP or the degree of program improvement to be achieved, CB retains the final authority to assign the contents of the plan and/or the degree of improvement required for successful completion of the plan [45 CFR 1355.35(a)(2)]. We provided guidance for calculating state performance on the national standards for the first two rounds of the CFSRs.[1] We also provided technical information on how we would evaluate states' attainment of goals regarding statewide data indicators in PIPs from those first two rounds.[2] This technical bulletin provides information for states relevant to CFSRs in federal fiscal years 2015 through 2018. We are applying new logic in categorizing state performance as well as guiding attainment of goals regarding the data indicators.

Section I provides descriptions of the various methodologies we used in generating the national standards on the data indicators and developing initial assessments of state performance relative to the national standards. Section II contains technical information concerning our approach to determining and evaluating the degree of improvement for PIPs specific to the statewide data indicators.

Section III contains information about item-specific measurement for PIP items other than the statewide data indicators. Section IV contains guidance for determining the amount of improvement necessary toward meeting the national standards through an approved PIP for the second round of CFSRs using revised national data indicators developed for Round 3. Finally, section IV provides technical information on anticipated content of the CFSR data profiles.

[1] See ACYF- CB-IM-00-11, *National Standards for the Child and Family Service Reviews* (December 28, 2000); ACYF-CB-IM-01-07, *Updated National Standards for the Child and Family Service Reviews and Guidance on Program Improvement Plans* (August 16, 2001); and ACYF-IM-07-05, *Measuring Program Improvement Plan (PIP) Improvement for the Child and Family Services Reviews (CFSRs) National Standards* (June 29, 2007).
[2] See CFSR Technical Bulletins #1 and #3.

Section I. CB's Methodologies for Calculating and Categorizing States' Risk-Standardized Performance Relative to the National Standards

This section describes how we will calculate and categorize state performance.

A. Calculating State Performance

State performance on each statewide data indicator will be assessed using a multi-level (i.e., hierarchical) model appropriate for that indicator. A multi-level logistic regression model will be used for indicators in which the outcome for a child either occurred or did not occur. A multi-level Poisson regression model will be used for indicators in which the outcome is a count per unit of time. We chose multi-level modeling because it is a widely accepted statistical method that enables fair evaluation of relative performance among states with different case mixes. The multi-level models that we employ when assessing each state's performance takes into account: (1) the variation across states in the age distribution of children served for all indicators, and the state's entry rate for select indicators; (2) the variation across states in the number of children they serve; and, (3) the variation in child outcomes between states.

The result of this modeling is a performance value that is a more accurate and fair representation of each state's performance than can be obtained with simply using the state's observed performance. We refer to this performance value as the state's risk-standardized performance (RSP). It is akin to risk-standardized morbidity and mortality rates often used in epidemiology and health care. The RSP is the ratio of the number of predicted outcomes in the state over the number of expected outcomes (both obtained from the model), multiplied by the national observed performance. This ratio is similar to the observed over expected ratio used in other types of statistical analyses.

B. Risk Adjustment

This section describes how we risk adjust, and the key steps in the modeling approach, including the calculation of a state's predicted outcomes, expected outcomes, and RSP for a given indicator.

Child Age

We will risk adjust on child's age for each indicator (depending on the indicator, it is the child's age at entry, exit, or on the first day). Adjusting on age allows us to control statistically for the fact that children of different ages have different likelihoods of experiencing the outcome, regardless of the quality of care a state provides.

We use "dummy" variables for each age when calculating the risk adjustment for age. Use of dummy variables is a common strategy in regression models to measure the impact of a characteristic on an outcome. A dummy variable has a value of 1 or 0 to indicate the presence or absence of the characteristic. For example, a child who entered care at age 2 will have a "1" for the "age 2" variable and a "0" for all others. For all but the first day permanency indicators, 19 age dummy variables are used to represent the ages from birth to 3 months, 4 to 11 months, and each year from age 1 through 17.

The permanency measure for children in foster care 12 to 24 months uses 17 age dummy variables (ages 1 through 17), and the permanency indicator for children in foster care 24 or more months uses 16 age dummy variables (ages 2 through 17). The method requires

specifying a base or reference age group, which was set at the median age for all children included in the calculation.

State Foster Care Entry Rate

We will also risk adjust on the state's foster care entry rate (per 1,000 children in the general child population) for two indicators: permanency in 12 months for children entering foster care and re-entry to foster care in 12 months. Adjusting on entry rate allows us to control for the fact that states with lower entry rates tend to have children at greater risk for poor outcomes, presumably because these states carry a higher concentration of children with more risk factors.

We calculate the entry rate as the number of children entering foster care during the 12-month period divided by the number of children in the state's child population, multiplied by 1,000. We obtain the child population data from the population division of the U.S. Census Bureau. These Census data reflect population estimates as of July 1 of each year, whereas the 12-month periods CB uses to define children entering care are either October to September, or April to March. Therefore, we chose to use the Census year closest to the 12-month period the child entered foster care as the denominator. For example, if the indicator follows children who entered care between April 1, 2011, and March 31, 2012 (an "11B/12A" file in AFCARS file conventions), we use child population estimates from the July 2011 Census estimate. If the 12-month period spanned October 1, 2012, through September 30, 2013, we would use population estimates as of July 1, 2013.

Steps for Calculating Risk Standardized Performance

The process for arriving at a state's RSP involves the following steps:

Step 1. Calculate each child's predicted probability for experiencing the outcome. The regression model determines the predicted probability each child in the state will experience the outcome given his or her age and the state he or she is in. Probabilities range from 0.0 (0%, or outcome will never occur) to 1.0 (100%, or outcome will always occur). These probabilities reflect a child's given level of "risk" of experiencing the outcome. For example, the regression model might calculate that the expected probability of a 17-year-old in State X exiting to permanency by 12 months is .34, or 34%. For a 1-year-old in State X, the expected probability might be .54, or 54%.[3]

Step 2. Calculate the number of children in the state predicted to experience the outcome. We sum the predicted probabilities for all children in the state to get the number of children we predict will, on average, experience the outcome. This predicted number is the number of outcomes (e.g., exits to permanency by 12 months, number of placement moves) we would predict the state to have, on average, based on the state's performance with its actual, observed case mix. Compared to the actual number of children who had the outcome, the predicted

[3] A child's predicted probability is based on two values obtained from the model: (1) the risk associated with the child's age (i.e., the age-specific beta coefficient, or β) plus (2) the state's intercept, which reflects the underlying risk of experiencing the outcome in that state after accounting for child's risk. The result of this sum is transformed to a probability to ease interpretation. The state's intercept is calculated based on the state's actual observed performance relative to states with similar children—considering how many children it served, the age of these children, and how many of these children experienced the outcome. For the permanency indicators, in which a higher number is more desirable, the intercept will be positive for a higher-than-average state, negative for a lower-than-average state, and close to zero for an average state. For the remaining indicators, the opposite is true. If there were no differences among states, then after adjusting for risk, the intercepts would be identical across all states.

number is a better estimate of how the state is likely to perform, on average, assuming no change in case mix, policy, and practice.

Step 3. Calculate each child's expected probability for experiencing the outcome. The same regression model then determines the expected probability each child in the state will experience the outcome, given his or her age, if he or she were in the "average" state. It is similar to the predicted probability obtained in Step 1 but ignores the specific state the child is in. It does this by using data on all children in the nation and ignoring the state or he or she is actually in. In other words, each age reflected in the nation's case mix has a given "risk" or probability of achieving the outcome.[4] For example, the regression model might calculate that the expected probability of a 17-year-old exiting to permanency by 12 months—if he or she were in the "average" state—is .23, or 23%. For a 1-year-old, the expected probability might be .41, or 41%.

Step 4. Apply the expected probabilities. The expected probabilities are applied to the children in each state: infants in the state are assigned the probability of the outcome associated with being an infant in an average state, 17-year-olds in the state are assigned the probability for 17-year-olds in an average state, and so on.

Step 5. Calculate the number of children in the state expected to experience the outcome. We sum the expected probabilities for all children in the state to get the number of outcomes we would expect if the state's children were treated by the "average" state. This expected number is the number of outcomes expected if the "average" state had the state's same case mix.

Step 6. We take the ratio of the number of "predicted" outcomes over the number of "expected" outcomes. For a measure like permanency, where higher numbers are more desirable, a state with more predicted than expected exits can be said to have a higher-than expected permanency rate; the state's ratio will be greater than 1 (e.g., 500 predicted / 400 expected = 1.25). A state with fewer predicted than expected permanent exits can be said to have a lower-than-expected permanency rate; the state's ratio will be less than 1 (e.g., 400 predicted / 500 expected = .80). A state with the same number of predicted as expected exits will have a ratio of 1 (e.g., 500 / 500 = 1.00), which suggests they perform no differently than the "average" state.

Step 7. Convert the ratio. To convert the ratio into a more meaningful value, we multiply it by the national observed performance. This puts the ratio in the same units (a percentage or rate per days in care) as the national observed performance. The result is the state's risk-standardized performance. As a point of reference, a state with "average" performance (a ratio = 1) will have an RSP equal to the observed national performance (i.e., 1 x the national observed performance). All other RSPs will be above or below the national observed performance. This multiplying step is a form of indirect standardization, so called because the national case mix is the "standard" population against which all states' RSPs can be compared.

[4] A child's expected probability is based on two values obtained from the model: (1) the risk associated with the child's age (i.e., the age-specific beta coefficient, or β) plus (2) the average intercept of all the state's, which can be interpreted as the unique level of care the "average" state provides for its children.

In sum, the RSP is the ratio of the number of "predicted" outcomes over the number of "expected" outcomes, multiplied by the national observed performance. RSPs can be compared to the national observed performance and are relative to RSPs of other states. However, a state's RSP should not be compared to its observed performance. For example, if a state's RSP for recurrence of maltreatment is higher than its observed performance, this does not mean the state's performance declined after risk adjustment. The converse is also true.[5]

Because the national observed performance is essentially the "weighted mean" of the state means, the national performance will be higher than most states if the larger states have higher rates. This tends to occur with the recurrence of maltreatment and maltreatment in foster care indicators. Therefore, when we scale a state's ratio by multiplying it with the national observed performance, it will scale the ratio quite strongly, producing a RSP value that is much higher than the state's observed performance, in which no scaling factor was applied. That the RSP is a meaningful number is an artifact of multiplying the ratio by the national observed performance. It has no statistical properties in and of itself.

C. Categorizing State Performance

Because the states' RSP and the national observed performance are based on the same national case mix, a state's RSP can be compared directly to the national observed performance to determine if the state performed above or below the "average" state.

To determine whether a state's RSP is statistically higher or lower than the national observed performance, CB calculates approximate 95% interval estimates around each state's RSP.[6] Whether it is desirable for a state to be higher or lower than the national performance depends on the indicator. For the permanency measures, a higher value is more desirable; for the remaining measures, a lower value is desirable.

CB will compare each state's interval estimate to the national observed performance[7] and assign each state to one of three groups:

- "No different than national performance" if the 95% interval estimate surrounding the state's RSP includes the national observed performance.
- "Higher than national performance" if the entire 95% interval estimate surrounding the state's RSP is higher than the national observed performance.

[5] To determine the impact of risk adjustment, one strategy is to look at how a state's ranking changed before and after risk adjustment.

[6] The RSP is a complex function of parameter estimates, and calculating exact interval estimates requires a computationally intensive bootstrapping process. Therefore, CB calculates approximate confidence intervals by using each child's beta coefficient, each state's intercept, the standard error of the intercept, and the traditional 1.96 multiplier. These confidence intervals produce results identical to those we obtained when using 95% confidence intervals around each state's intercept, which is a less communicable metric but a recognized approach to identifying groups that are statistically above or below a standard rate, like the national observed performance.

[7] Comparing the upper and lower interval estimates to the national observed performance is done using rounded versions of these data points. For indicators expressed as a percentage, the interval estimates and national observed performance are rounded to one decimal place. For indicators expressed as a rate (maltreatment in foster care and placement stability), the interval estimates are rounded to two decimals places. We chose to round these values because we do not believe it is appropriate to attempt to measure state performance to the degree of precision implied by three or more decimal places. The rate indicators are rounded to two decimal places due to the small rates associated with them.

- "Lower than national performance" if the entire 95% interval estimate surrounding the state's RSP is lower than the national observed performance.

The methodology described above is similar to that used by the Centers for Medicare & Medicaid Services to measure hospital performance as part of its Hospital Inpatient Quality Reporting program.[8] The methodology is also consistent with the use of such models in education and health care to distinguish statistically high- and low-performing schools and hospitals.[9]

[8] Yale New Haven Health Services Corporation, & Center for Outcomes Research & Evaluation. (2013). 2013 Measures Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Readmission Measures for Acute Myocardial Infarction, Heart Failure, and Pneumonia (Version 6.0). Retrieved from: http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic/Page/QnetTier4&cid=1219069855841; The COPSS-CMS White Paper Committee (2012). Statistical Issues in Assessing Hospital Performance. Retrieved from: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf

[9] See Goldstein & Spiegelhalter (2007). League Tables and Their Limitations. 159(3), 385–443; Normand & Shahian (2007). Statistical and Clinical Aspects of Hospital Outcomes Profiling. Statistical Science, 22(2), 206–226.

Section II: Technical Information About Evaluating States' Attainment of Required Improvement Regarding National Standard Data Indicators in the PIP

This section pertains to how we will evaluate state progress toward meeting the PIP improvement goals when a state does not meet the national standards for the statewide data indicators.

A. Data Periods for Evaluation

To evaluate a state's performance toward PIP goals, we will use the state's AFCARS and NCANDS data submissions from a period no earlier than the non-overlapping 12-month period following the completion of CFSR Round 2. We will also ensure that there is a 12-month non-overlapping period following CFSR Round 2 PIP implementation for our final determinations of attainment of required improvement. This is consistent with prior guidance provided for earlier CFSR rounds in CFSR Technical Bulletin #3A regarding non-overlapping periods.

If a state is monitoring a statewide data indicator for the PIP, CB will evaluate the state's performance every 6 months with available rolling AFCARS and/or NCANDS data against the national standard. If, at any point during the PIP implementation or non-overlapping period, a state meets the national standard but has not yet met its established PIP target, we will consider the state to have met its PIP goal.

States may submit corrected or more complete AFCARS and/or NCANDS data in accordance with the procedures for those file submissions for us to use in evaluating the state's CFSR progress. However, states must resolve data quality concerns prior to the end of the 12-month non-overlapping period following PIP completion.

B. Methodology for Setting Program Improvement Goals

CB will require states that do not meet the national standard for a statewide data indicator to include improvement on that indicator in their PIPs. If CB is unable to determine a state's performance on an indicator due to data quality issues, we will also require the state to include that indicator in its PIP along with key strategies to correct the quality of the data. See the data quality thresholds listed in Attachment D to the Federal Register notice for additional information on how we make these determinations.

For two of the statewide data indicators, permanency in 12 months for children entering foster care and re-entry to foster care in 12 months, CB will determine performance for program improvement purposes on one indicator in concert with the other as a companion measure.

If a state does not meet national standards in a risk-adjusted statistical model, the state will need to demonstrate improvement during the PIP. We rely on unadjusted, observed performance over time to monitor state performance during the PIP. The key components for setting improvement goals and monitoring a state's progress over the course of a PIP involve calculating baselines, setting improvement goals, and when companion measures are included in an improvement plan, also establishing thresholds. CB will set improvement goals and thresholds in part relative to each state's past performance.

The method for establishing PIP baselines and improvement goals is described in the following steps:

Step 1. Calculate 7 estimates of past performance. We use a state's data from the past three years (we call these Year 1, 2, & 3) to obtain seven values: the state's actual performance in Year 1, Year 2, and Year 3, and the averages of Years 1 through 3, 1 and 3, 2 and 3, and 1 and 2.

Using all seven values instead of the original three gives us a more reliable estimate of the state's overall past performance and amount of fluctuation that is typical for that state.

As an example, consider that in the most recent three entry cohorts we could track, State Y discharged 29.4%, 28.9%, and 27.4% of its children to permanency within 12 months. From this we create 7 estimates as displayed in table 1.

Table 1. Example of Calculating 7 Estimates of Past Performance

Year 1	Year 2	Year 3	Years 1–3 (Average)	Years 1 & 2 (Average)	Years 1 & 3 (Average)	Years 2 & 3 (Average)
29.4	28.9	27.4	28.6	29.1	28.4	28.1

Step 2. Estimate the variability in past performance. Some states, particularly larger states, show very slight fluctuation in performance from year to year. Other states have shown much larger changes in performance over time, often due to smaller and more variable population sizes.

Therefore, we want to take into account the state's *variability* in past performance. A well-known measure of variability is the standard deviation (SD). In our case, the SD will tell us how much our estimates of past performance vary from the average (mean) of these estimates. A low SD indicates very little variability—the data points tend to be very close to the mean; a high SD indicates a lot of variability—the data points tend to be spread out over a larger range.

Step 2.1. Calculate a sample mean and standard deviation, based on the 7 values of past performance. The mean reflects an estimate of each state's past performance based on the past 3 years, and the standard deviation (SD) reflects the overall variation in this past performance. The SD tells us how far away the 7 values tend to be from the mean. To increase the reliability and accuracy of the estimates we use a technique called bootstrapping. Bootstrapping involves taking the original seven values and sampling with replacement to generate a large number of resamples of 7 values. We generate 1,000 resamples of 7. The process does not create any new numbers: each resample is simply a random sample selected from those same seven values that we calculated earlier. Because we resampled with replacement, there is no guarantee that any of our 1,000 resamples will contain all of the seven values, and some of the seven values will be repeated.

Step 2.2. Bootstrap the mean and standard deviation 1,000 times. We repeat the bootstrap process 1,000 times to create 1,000 sample means and 1,000 sample standard deviations. Repeating the process 1,000 times is a common technique of bootstrapping, and 1,000 is a reasonable number of bootstraps. Although it does not increase the amount of information in the original dataset (we are still working with only seven values), it does allow us to assign a more accurate measure of the underlying variability in each state's past performance.

Step 3. Calculate a grand mean and standard deviation. We then average the 1,000 sample means to get one number (a grand mean) that represents the state's past performance on the measure. We also average the 1,000 SDs to get one number (average standard deviation) that represents the state's typical variability in past performance for the measure.

Step 4. Calculate the improvement goal. We use the state's overall SD to identify a value we think is far enough from its overall performance to be meaningful, but not so far as to be unattainable.

A common approach to identifying extreme changes is with statistical significance testing. In a normal distribution, 95% of the data is within two SDs from the mean. A common rule of thumb in statistics is that values more than two SDs from the mean—the other 5% of the data—are considered extreme; they are unlikely to occur by chance (given what we know about the data) and therefore represent something statistically significant.[10] In our case, values beyond two SD might be good candidates for improvement goals.

The problem with this approach to statistical significance testing is that it assumes we have a normal distribution, which we cannot assume. The overall grand mean of these means can be used to estimate the population mean, and the average of the 1,000 SDs can be used to estimate the population SD. We can say with confidence that at least a certain percent of values will be within k standard deviations of the mean. For example, using Chebyshev's Inequality, at least 94% of the values in any distribution are within four SDs of the mean, and at least 75% of the values in any distribution are within two SDs of the mean.

As an example, State Y's grand mean performance was 28.6% with a standard deviation of .6101%.

Table 2. Example of Calculating the Improvement Goal

Year 1	Year 2	Year 3	Grand Mean (GM)	Standard Deviation (SD)	GM + 2 SDs
29.4	28.9	27.4	28.6%	.6101%	29.8% (i.e., 28.6% + 2*.6101%)

If we were to use these values as improvement goals, the baseline would be the grand mean. There are, however, a couple of concerns with using the grand mean as the baseline. First, the grand mean is an average of the last three years of performance, rather than reflecting a state's

[10] Popular levels of significance are 5% (0.05), 1% (0.01), 0.5% (0.005), and 0.1% (0.001).

most recent performance. Secondly, if there have been substantial increases in performance in the most recent year, this approach could result in a goal that is actually worse than performance in the most recent year. Conversely, states that have had substantial declines in performance in the past year could be assigned targets that are simply out of reach.

The solution to this issue is to calculate an improvement factor using the grand mean, but apply it to the estimate for the most recent year (Year 3 in the table, which serves as the baseline). Using the example above, with a grand mean of 28.6%, a grand mean plus 2 SDs of 29.8%, and the most recent year's value of 27.4%, we would obtain an improvement factor this way:

29.8%/28.6% = 1.042

Then we apply that improvement factor to the Year 3 value, to get our goal: 27.4% x 1.042 = 28.6%.

Step 5. If the indicator is being used as a companion measure in the PIP, calculate the threshold. Thresholds are calculated as the inverse of the improvement goals, to identify the point at which the state is performing substantially worse on a particular outcome. So instead of adding 2 SDs as we did earlier, this time we subtract 2 SDs. Keeping with the previous example, with a grand mean of 28.6%, a grand mean *minus* 2 SDs (27.4%), and the grand mean value of 28.6%, we would have a threshold factor of 27.4%/28.6% = 0.958, which, when applied to the Year 3 baseline, give us a threshold of: 27.4% x 0.958= 26.2%.

Table 3. Example of Threshold Calculation

Year 1	Year 2	Year 3 (Baseline)	Grand Mean	Standard Deviation	GM + 2SDs	Improvement Factor	Threshold Factor	Improve-ment Goal	Threshold
29.4	28.9	27.4	28.6%	.6101%	29.8	1.042	0.958	28.6%	26.2%

Step 6. Setting Caps and Floors. Sometimes the approach described above is all that will need to be done to calculate a goal or threshold for a state's PIP. However, this approach can occasionally yield performance goals that are too aggressive, or conversely, too modest. To handle this issue, we set a cap and a floor on the improvement factors that are generated, using the distribution of all states' improvement factors as a guide.

Step 6.1. Calculate improvement factors for *all* states (even if they are not including the statewide data indicator in a PIP), and rank and re-order the states according to the size of the improvement factor, from largest to smallest.

Step 6.2. Determine the 50th percentile, which is the middle of the distribution. This will be the cap on the improvement factor required in a PIP. If a higher percentage is the preferred outcome (as in the example here), then replace all improvement factors greater than the one at the 50th percentile with the one at the 50th percentile. This will be referred to as an "adjusted improvement factor."

Step 6.3. Determine the 20th percentile if the desired outcome is a higher number, and determine the 80th percentile if the desired outcome is a lower number (such as Re-entry rates). This will be the floor on improvement factors. The purpose of establishing a floor at the upper or lower quintile is to set a minimum limit on how much improvement will be required.

Step 6.4. For those states with an improvement factor beyond the floor, replace it with the value at the 20th or 80th percentile to obtain the "adjusted improvement factor."

Step 6.5. All states with improvement factors that fall between the cap and floor will use their original improvement factor as their adjusted improvement factor, as generated from their own data.

Step 6.6. Round the baseline to one decimal place if it is an indicator expressed as a percent, and round it to two decimal places if the indicator is a rate. Round the adjusted improvement factor to three decimal places. We chose to round the state's performance to only one decimal point for most of the indicators because we do not believe it is appropriate to attempt to measure state performance to the degree of precision implied by two decimal places. The exception to this rounding procedure is maltreatment in foster care and placement stability, which are both rounded to two decimal points due to the small ratios reported by states. Further, we do not require a level of precision beyond three decimal places for improvement factors.

Step 6.7. Multiply the rounded adjusted improvement factor with the rounded baseline to obtain the final improvement goal.

If the indicator serves as a companion measure in a PIP:

Step 6.8. Calculate a threshold to indicate the point at which the state's performance cannot decline. When calculating a threshold, instead of adding 2 times the SD to the grand mean, you would subtract it (or vice versa, depending on the measure). This allows you to obtain a "threshold factor."

Apply the same caps and floors to the threshold factor as are applied to improvement factors, so that they are simply the inverse of the final improvement goal. Round the adjusted threshold factor.

Multiply the rounded baseline and the rounded threshold factor. The result is the threshold, which is used for companion measures that are monitored only to see if performance on that indicator declines during a PIP. The percent change between the baseline and the improvement goal is the same as the percent change between the baseline and the threshold.

See table 4 for an example of how these calculations apply for all states. Note that the shaded areas indicate where the caps or floors on the adjusted improvement factors have been applied.

Table 4. Example of PIP Adjusted Improvement Goals and Thresholds Applying Caps and Floors for Permanency in 12 Months, for Children Entering Care

State	State will PIP?*	Year 1 - 2009B10A	Year 2 - 2010B11A	Year 3 - 2011B12A (Baseline)	Grand Mean	Avgs	Grand Mean + 2 *Avgs	Grand Mean - 2 *Avgs	Improvement Factor	Adjusted Improvement Factor	Adjusted Improvement Factor (Rnd)	Threshold Factor	Adjusted Threshold Factor	Adjusted Threshold Factor (Rnd)	Improvement Goal	Threshold
State 1	Yes	39.0%	34.4%	26.9%	33.4%	4.0%	41.3%	25.5%	1.2369	1.0632	1.063	0.7631	0.9368	0.937	28.6%	25.2%
State 2	Yes	26.6%	34.0%	29.4%	30.0%	2.4%	34.8%	25.2%	1.1606	1.0629	1.063	0.8394	0.9368	0.937	31.3%	27.5%
State 3	No	46.7%	36.0%	38.9%	40.5%	3.6%	47.6%	33.4%	1.1762	1.0632	1.063	0.8238	0.9368	0.937	41.4%	36.4%
State 4	No	48.8%	43.9%	45.6%	46.1%	1.6%	49.3%	42.9%	1.0698	1.0632	1.063	0.9302	0.9368	0.937	48.5%	42.7%
State 5	No	47.2%	50.6%	44.9%	47.6%	1.9%	51.3%	43.8%	1.0786	1.0632	1.063	0.9214	0.9368	0.937	47.7%	42.1%
State 6	No	53.4%	49.6%	45.2%	49.4%	2.6%	54.7%	44.1%	1.1070	1.0632	1.063	0.8930	0.9368	0.937	48.0%	42.3%
State 7	No	49.9%	52.3%	45.9%	49.4%	2.1%	53.6%	45.2%	1.0856	1.0632	1.063	0.9144	0.9368	0.937	48.8%	43.0%
State 8	No	49.0%	46.7%	51.6%	49.1%	1.6%	52.2%	45.9%	1.0640	1.0632	1.063	0.9360	0.9368	0.937	54.9%	48.3%
State 9	No	36.6%	36.4%	40.0%	37.7%	1.3%	40.2%	35.1%	1.0684	1.0632	1.063	0.9316	0.9368	0.937	42.5%	37.5%
State 10	Yes	33.0%	32.1%	25.3%	30.1%	2.7%	35.6%	24.7%	1.1801	1.0632	1.063	0.8199	0.9368	0.937	26.9%	23.7%
State 11	Yes	40.0%	41.2%	36.7%	39.3%	1.5%	42.3%	36.3%	1.0764	1.0632	1.063	0.9236	0.9368	0.937	39.0%	34.4%
State 12	Yes	36.6%	41.6%	36.8%	38.3%	1.8%	41.9%	34.7%	1.0940	1.0632	1.063	0.9060	0.9368	0.937	39.1%	34.5%
State 13	Yes	40.6%	37.4%	37.2%	38.4%	1.2%	40.9%	36.0%	1.0635	1.0632	1.063	0.9365	0.9368	0.937	39.5%	34.8%
State 14	Yes	39.4%	38.4%	32.5%	36.8%	2.4%	41.6%	32.0%	1.1306	1.0632	1.063	0.8694	0.9368	0.937	34.5%	30.4%
State 15	Yes	37.3%	37.8%	33.6%	36.2%	1.5%	39.2%	33.2%	1.0828	1.0632	1.063	0.9172	0.9368	0.937	35.7%	31.5%

State	State will PIP?*	Year 1 - 2009B10A	Year 2 - 2010B11A	Year 3 - 2011B12A (Baseline)	Grand Mean	Avgs	Grand Mean + 2 *Avgs	Grand Mean - 2 *Avgs	Improvement Factor	Adjusted Improvement Factor	Adjusted Improvement Factor (Rnd)	Threshold Factor	Adjusted Threshold Factor	Adjusted Threshold Factor (Rnd)	Improvement Goal	Threshold
State 16	Yes	37.5%	34.7%	30.8%	34.3%	2.2%	38.6%	30.0%	1.1257	1.0632	1.063	0.8743	0.9368	0.937	32.7%	28.9%
State 17	No	51.0%	46.6%	45.6%	47.7%	1.8%	51.4%	44.0%	1.0775	1.0632	1.063	0.9225	0.9368	0.937	48.5%	42.7%
State 18	No	51.7%	46.8%	46.0%	48.2%	2.0%	52.2%	44.2%	1.0831	1.0632	1.063	0.9169	0.9368	0.937	48.9%	43.1%
State 19	Yes	33.1%	36.5%	38.3%	36.0%	1.7%	39.4%	32.6%	1.0948	1.0632	1.063	0.9052	0.9368	0.937	40.7%	35.9%
State 20	No	51.3%	51.1%	46.7%	49.7%	1.7%	53.0%	46.3%	1.0676	1.0632	1.063	0.9324	0.9368	0.937	49.6%	43.7%
State 21	Yes	33.9%	32.8%	37.2%	34.6%	1.5%	37.6%	31.6%	1.0861	1.0632	1.063	0.9139	0.9368	0.937	39.5%	34.8%
State 22	Yes	40.5%	36.5%	32.3%	36.5%	2.6%	41.7%	31.2%	1.1444	1.0632	1.063	0.8556	0.9368	0.937	34.3%	30.3%
State 23	No	53.6%	58.9%	55.1%	55.9%	1.8%	59.4%	52.3%	1.0629	1.0629	1.063	0.9371	0.9371	0.937	58.6%	51.6%
State 24	No	43.6%	40.2%	40.6%	41.5%	1.2%	43.9%	39.0%	1.0592	1.0592	1.059	0.9408	0.9408	0.941	43.0%	38.2%
State 25	Yes	33.0%	31.1%	30.5%	31.5%	0.9%	33.3%	29.8%	1.0548	1.0548	1.055	0.9452	0.9452	0.945	32.2%	28.8%
State 26	Yes	31.4%	29.2%	29.6%	30.1%	0.7%	31.6%	28.6%	1.0498	1.0498	1.050	0.9502	0.9502	0.950	31.1%	28.1%
State 27	No	44.6%	45.0%	47.9%	45.8%	1.2%	48.1%	43.5%	1.0503	1.0503	1.050	0.9497	0.9497	0.950	50.3%	45.5%
State 28	No	45.6%	42.3%	43.9%	43.9%	1.1%	46.1%	41.8%	1.0488	1.0488	1.049	0.9512	0.9512	0.951	46.1%	41.8%
State 29	Yes	39.8%	37.9%	37.3%	38.3%	0.8%	40.0%	36.6%	1.0437	1.0437	1.044	0.9563	0.9563	0.956	38.9%	35.7%
State 30	No	46.3%	43.3%	44.4%	44.7%	1.0%	46.6%	42.7%	1.0437	1.0437	1.044	0.9563	0.9563	0.956	46.4%	42.5%
State 31	No	59.5%	59.9%	63.1%	60.8%	1.3%	63.4%	58.3%	1.0415	1.0415	1.042	0.9585	0.9585	0.958	65.8%	60.5%

State	State will PIP?*	Year 1 - 2009B10A	Year 2 - 2010B11A	Year 3 - 2011B12A (Baseline)	Grand Mean	Avgs	Grand Mean + 2 *Avgs	Grand Mean - 2 *Avgs	Improvement Factor	Adjusted Improvement Factor	Adjusted Improvement Factor (Rnd)	Threshold Factor	Adjusted Threshold Factor	Adjusted Threshold Factor (Rnd)	Improvement Goal	Threshold
State 32	No	44.3%	41.6%	42.6%	42.8%	0.9%	44.6%	41.1%	1.0407	1.0407	1.041	0.9593	0.9593	0.959	44.3%	40.9%
State 33	No	48.9%	47.4%	45.9%	47.4%	0.9%	49.3%	45.5%	1.0398	1.0398	1.040	0.9602	0.9602	0.960	47.7%	44.1%
State 34	No	51.9%	51.5%	54.3%	52.6%	1.0%	54.5%	50.6%	1.0371	1.0371	1.037	0.9629	0.9629	0.963	56.3%	52.3%
State 35	No	49.1%	49.0%	46.8%	48.3%	0.8%	50.0%	46.7%	1.0341	1.0341	1.034	0.9659	0.9659	0.966	48.4%	45.2%
State 36	No	66.3%	66.1%	63.8%	65.4%	0.9%	67.2%	63.6%	1.0271	1.0313	1.031	0.9729	0.9687	0.969	65.8%	61.8%
State 37	Yes	15.5%	15.9%	15.9%	15.7%	0.2%	16.1%	15.4%	1.0204	1.0313	1.031	0.9796	0.9687	0.969	16.4%	15.4%
State 38	Yes	40.1%	39.3%	40.8%	40.0%	0.5%	41.0%	39.1%	1.0236	1.0313	1.031	0.9764	0.9687	0.969	42.1%	39.5%
State 39	No	49.5%	49.5%	49.7%	49.6%	0.1%	49.7%	49.4%	1.0031	1.0313	1.031	0.9969	0.9687	0.969	51.2%	48.1%
State 40	Yes	32.1%	32.3%	32.6%	32.3%	0.2%	32.7%	32.0%	1.0105	1.0313	1.031	0.9895	0.9687	0.969	33.6%	31.6%
State 41	No	61.3%	60.3%	58.9%	60.2%	0.8%	61.7%	58.7%	1.0250	1.0313	1.031	0.9750	0.9687	0.969	60.7%	57.1%
State 42	No	49.5%	49.7%	47.9%	49.0%	0.7%	50.4%	47.7%	1.0269	1.0313	1.031	0.9731	0.9687	0.969	49.4%	46.4%
State 43	Yes	29.3%	29.4%	28.7%	29.2%	0.2%	29.6%	28.7%	1.0159	1.0313	1.031	0.9841	0.9687	0.969	29.6%	27.8%
State 44	No	46.5%	47.2%	45.4%	46.3%	0.6%	47.5%	45.2%	1.0248	1.0313	1.031	0.9752	0.9687	0.969	46.8%	44.0%

*Status as represented in the table 4 is based on a preliminary assessment against the national standards. The determination of whether a state needs to include a statewide data indicator in PIP may change when we run the applicable data at the time of the state's review.

Section III. Guidelines for Determining and Approving PIP Item Measurement Methods and Degrees of Improvement

This section provides guidance on which items a state must make quantifiable improvement in during a PIP and therefore must be measured when the state is not in substantial conformity for an outcome. We also provide here our preapproved methodological approaches that a state can use to measure such improvement. States may request that we consider alternative measurement methods prior to PIP approval as we recognize that states have different capacities for measuring program improvement. Any state-specific proposals will be considered individually and we will advise states whether they are approved.

A. General Guidelines

When a Safety Outcome is not in substantial conformity, the state must include quantifiable measurement in the PIP for all items that are areas needing improvement. This is consistent with guidance we have provided in prior rounds to comport with regulatory requirements for states to prioritize areas of nonconformity affecting child safety.

We will require states to include quantifiable measurement for Well-Being Outcome 1 items when that outcome is not in substantial conformity. We will negotiate with the state the specific Well-Being Outcome 1 items that the state will measure based on the state's CFSR findings and the state's proposed strategies to address the outcome. We will not require states to specify particular measures for Permanency Outcome 1 items given the variety of statewide data indicators associated with that outcome and that may already be required for inclusion in the PIP.

To the extent that a state does not believe that its CFSR case review performance is indicative of its statewide performance, the state may request that CB consider state-provided aggregate data that demonstrates this. For example, if a state has statewide data for a comparable time period that is of good quality, and which indicates that the state's child protective services investigations are timely at a rate of 95%, CB will not require the state to include the related Item 1 in its PIP.

The state must develop a measurement plan for outcome items that are identified as areas needing improvement that do not require quantifiable measurement in the PIP as outlined above. This means that at a minimum, the state will include a key activity in its PIP to develop a method of measurement by the conclusion of the PIP for ongoing measurement of identified areas needing improvement within Permanency Outcome 2, Well-Being Outcome 2, and Well-Being Outcome 3. The resultant measurement plan and implementation steps must be included in the state's subsequent title IV-B Annual Progress and Services Report and/or Child and Family Services Plan (APSR/CFSP). CB will monitor future APSRs and CFSPs for the state's reported implementation progress, adjustments to the measurement plan, and/or results.

States must provide indicators of improvement on systemic factors, which may include data measures as appropriate. States are encouraged to address the most challenging items identified within a systemic factor versus all areas designated as areas needing improvement within a systemic factor.

B. Development of Baselines and Review Periods for Onsite Review Instrument Item Measures

CB will negotiate baselines from the states' most reliable data source for onsite review instrument item measures other than statewide data indicators. We want to build on state efforts to institutionalize capacity consistent with CB's Information Memorandum on Continuous Quality Improvement (CQI) as described in ACYF-CB-IM-07, CFSR Technical Bulletin #7, and the requirements of the CFSP.

We will measure improvement based on the information that a state can provide and replicate through its PIP implementation period. Sampling approaches must include all children served in foster care and receiving in-home services on the same basis as provided in CFSR Technical Bulletin #7. The state may be able to use CFSR onsite review findings as baselines for a program improvement plan if the state conducted its own reviews and such reviews are aligned with its ongoing statewide CQI monitoring plan. CB will make this determination on a case-by-case basis. Similar to Round 2, states in Round 3 that have traditional onsite reviews will not be able to use the CFSR onsite review findings as a baseline because of the inherent differences between it and the state's ongoing CQI process.

For CB to make a determination of an appropriate baseline for PIP measurement purposes, the state must identify the sources of its baseline data and the data's alignment with the state's ongoing CQI approach. This includes the review instruments the state will use, the sample frame of cases included in the baselines, and the number of applicable cases by item.

C. Preapproved PIP Item Measurement Approaches

CB encourages states to use its state-generated data from its CQI system or management information systems for PIP monitoring and measurement. States should provide to CB documentation of their proposed PIP measurement plans, with a description of their specific case review criteria or aggregate data methods employed. Such documentation must include the state's baseline sample source and sample frame, sample size, the review period and locations, instruments, reports and a description of goal measurement. After we have approved the state's PIP measurement plan, the state must notify CB if it intends to change its review instruments, reports, or sampling methods or approach as we must confirm that it remains approvable.

We are providing below the specific measurement methodology we have preapproved for states to establish and measure improvement toward achieving PIP item-specific measurement goals. The first method, the retrospective data method, addresses situations where a state's baseline data is available prior to CB approving the state's PIP and the state outlines a process for determining the baseline and target goals from existing data. The second method, the prospective data method, addresses situations where the state's baseline data will be collected during the PIP implementation period and the state has a process for developing a minimum case sample prior to setting the goal of improvement. A third method provides information on the methodology for using a state case management data or other aggregate data to measure a universe larger than a sample review approach.

For the retrospective and prospective methods, we recommend an 80% confidence level. CB believes this will allow states more flexibility in demonstrating improvement with somewhat smaller and less labor-intensive case samples than a 90% or 95% confidence level. In these methods we recommend that the state samples be equal to or greater than the number of applicable cases for the item from the state's CFSR onsite review. At a minimum, states should include their largest metropolitan area and a representative cross-section of counties or jurisdictions in their sample, including Tribal or other significant populations.

The baseline and measurement samples should include case types similar to the distribution and ratio used for the CFSR onsite review. Once a baseline sample size is established, the ongoing monitoring measurement sample size and ratio must be comparable to the baseline. The number of applicable cases used for a baseline would be the minimum required for ongoing monitoring measurement for CB to determine that goals are met. We will apply a 2% tolerance when comparing the applicable cases to the baseline applicable cases and a 5% tolerance on the distribution of case types and metropolitan area proportion between the baseline and subsequent reviews.

Method 1—State retrospective data with minimum improvement determined by sampling error

This method utilizes the available state review percentage findings and tests whether the state's quarterly performance exceeds the original baseline proportion plus the sampling error. States use 12 months of practice findings beginning no earlier than the first quarter of the AFCARS submission used for the CFSR onsite review sample to establish a baseline. The minimum sample for a given item should be equal to or greater than the applicable cases for the item from the state's CFSR onsite review findings. The actual percentage satisfying the given item is computed from the state's 12 months of practice review sample, and that sample size would be used to compute the actual sampling error using an 80% confidence level. The actual applicable case sample should be greater than or equal to the minimum number of applicable cases reviewed during the CFSR onsite review. The state's baseline would be the computed percentage.

We would establish the goal for improvement by adding the sampling error to the baseline percentage. Larger samples would result in lower improvement goals because of the smaller sampling error. Once a baseline is established, the sample size must remain comparable through the monitoring and measurement period. The state would use percentages computed from 12 months of practice data/findings to determine whether the state satisfied its improvement goal. In situations in which a state has consistently measured a particular item from the baseline and demonstrated the minimal improvement outlined above prior to PIP approval, the item will not require further measurement goals during PIP implementation. States must still address the agreed-upon activities in their PIPs for each item that contributed to a determination of nonconformity for each outcome or systemic factor.

Table 5 provides an example of how this method may be applied to a state's retrospective data. If the state's 12 month practice sample results do not reach the minimum applicable cases, the state could increase its sample size in the next quarter to achieve the threshold. CB can assist states to compute the sampling error and improvement goal using the process outlined in the examples below.

Table 5. Example Applying Retrospective Data Method

CFSR OSRI Item #	Applicable Cases From Onsite Review (Minimum Sample Size)	Actual Number of Applicable Item Cases Over Baseline Year	State Baseline Year Proportion	Baseline Year Actual Sampling Error	12-Month Goal % (Baseline + Sampling Error)
12	65	74	0.60	0.0728	67.3%
1	28	26	0.72	0.0909	Sample not sufficient

Method 2—State prospective data with baseline and goal established during PIP implementation

This method establishes a baseline from a minimum sample using the state's 12 months of practice findings beginning after PIP implementation. The minimum sample size as determined by the applicable cases for the item from the state's CFSR onsite review would be required for the 12-month baseline. As the baseline would be established during the period of PIP implementation, the improvement target would be reduced by up to half of the sampling error to allow for baseline overlap with improvement strategies and reduced time to measure improvement. We will calculate the sampling error reduction rounding to the number of overlapping months using .041167 per month or .125 per quarter for a maximum of 12 months. States would then use rolling quarters or months of findings encompassing 12 months to determine whether they have met their improvement goals.

Table 6 provides an example of how method 2 may be applied to a state's prospective data.

Table 6. Example Applying Prospective Data Method

CFSR OSRI Item #	State Baseline Year Proportion (BYP)	Number of Item Applicable Cases During Baseline	Baseline Year Sampling Error (BYSE)	Number of Months of PIP Implementation Overlap With Baseline	12-Month Goal % (BYP + 0.X/8= BYSE)
12	0.60	74	0.0729	12	0.60 + 0.5 x BYASE
1	0.72	50	0.0813	5	0.72 + 0.2083 x BYASE = 78.1%

Method 3—Use of state data collected from the Statewide Automated Child Welfare Information System (SACWIS) or Other Case Management Data

Consistent with our approach in the second round of CFSRs, we recommend that a minimal amount of improvement for item measures derived from a state's SACWIS or case management information system reporting be based on the sampling error, at a 95% confidence interval. This interval is recommended because statewide universe data are used and a lower confidence level would yield very minimal improvement goals. The minimum improvement amount will be computed by adding the sampling error to the 12 months of data or a weighted proportion for a

12-month period using quarterly data reports. Reports proposed by the state under this method should include design syntax and/or extraction methodology that must be approved prior to inclusion of the measurement in the PIP.

D. High Performance Plateau Adjustment

In situations where the state's PIP item goals are above 90%, we will apply consideration of a plateau effect in determining whether a state has met its goal. If the state has an improvement goal above 90% and is able to sustain performance above the baseline for three quarters, we will consider the goal met even if the state does not meet its actual goal.

Section IV. Guidelines for Measuring Round Two Improvement Using Round 3 National Standards

This section addresses how we will consider a state's performance on the national standards in Round 3 in determining whether the state has met its CFSR Round 2 PIP.

At the conclusion of a state's Round 2 PIP, if a state has not attained its negotiated amount of improvement with regard to a Round 2 statewide data indicator, we will consider the state's performance on the related Round 3 statewide data indicator in determining whether the state met the goal. We have determined that 4 of the CFSR Round 2 national standards have a direct relationship with the CFSR Round 3 statewide data indicator, as displayed in table 7.

Table 7. Related CFSR Round 2 and Round 3 Statewide Data Indicators

CFSR Round 2 Statewide Data Indicator	Related CFSR Round 3 Statewide Data Indicator
Absence of Maltreatment Recurrence	Recurrence of Maltreatment
Absence of Child Abuse or Neglect in Foster Care	Maltreatment in Foster Care
Composite 1 -Timeliness and Permanency of Reunification	Permanency in 12 Months for Children Entering Foster Care Re-Entry to Foster Care in 12 Months ** Note that both indicators would need to be met to replace composite 1*
Permanency Composite 4—Placement Stability	Placement Stability

CB will calculate the state's national standard performance for the related statewide data indicators for the state's non-overlapping period following the implementation of the state's CFSR Round 2 PIP. If a state meets the national standard for the related CFSR Round 3 statewide data indicator(s), we will consider the state to have satisfied the improvement requirement and successfully completed that portion of the PIP for Round 2. If the state does not meet or exceed the replacement national standard indicator, the applicable penalties for the outcome associated with the data indicator will be withheld as specified in 45 CFR 1355.36.

Section V. Anticipated Data Profile Content

We will provide states with an initial data profile for the purposes of developing the statewide assessment. We will also provide states with updated data profiles after each 6-month AFCARS file is processed. The format and content of data profiles may change over time, based on efforts to improve the profile as a vehicle for communication.

A. Performance on CFSR 3 Statewide Data Indicators

For each statewide data indicator, for the three 12-month periods prior to the PIP and every 12-month period thereafter:

State Data

- Number of cases in the denominator and numerator
- Observed performance (numerator/denominator)
- Risk-standardized performance with 95% interval estimate (from the multi-level model)
- Number of children excluded from the measure due to minor data quality problems

Comparative Data

- National standard (i.e., national observed performance: national numerator / national denominator)
- An indication as to whether the state met the national standard, was no different than the national standard, or did not meet the national standard

B. Data Quality Related to CFSR 3 Statewide Data Indicators

For each data quality check, for the three 12-month periods prior to the PIP and every 12-month period thereafter:

State Data

- Number of questionable cases
- Percentage of questionable cases

Comparative Data

- The data quality limit established by the CB
- The statewide data indicator(s) that, for accurate measurement, require the DQ standard
- An indication as to whether the state's percentage of questionable cases did not meet the data quality standard, in which case performance for the affected indicator(s) was not calculated

C. Performance Improvement Plan Data

For each statewide data indicator included in a performance improvement plan:

- Time frames associated with the PIP (e.g., start date, end date)
- The state's baseline
- The state's improvement goal

- The threshold (i.e., if a companion measure)
- Observed performance relative to the baseline and goal shown for rolling 12-month reporting periods (when available) through the PIP measurement period

D. Context Data

For each context item below, for the same 12-month periods used to calculate performance on the statewide data indicators, show (for both the state and the nation):

Safety—Reports

- Reports that CPS screens in and out (N & %)
- Children who are subjects of an investigation or alternative response (N & rate per 1,000 children)
- Children who are subjects of an investigation or alternative response (N & %)
- Time between report and initiation of investigation or alternative response, Child File (Mean and SD)
- Performance on the originally proposed Re-report of Maltreatment indicator

Safety—Dispositions

- Children who are confirmed by CPS as victims of maltreatment (N & rate per 1,000 children)
- Children by CPS disposition (substantiated/indicated, unsubstantiated, other (N & %)
- CFSR 2 indicator—Absence of maltreatment recurrence (N & %)
- CFSR 2 indicator—Absence of child abuse and neglect in foster care (N & %)
- Child victim cases opened for post-investigation services (N & %)
- Child victims entering foster care based on a CAN report (N & %)
- Child fatalities resulting from maltreatment (N & %)

Foster Care—Caseload

- Children who enter care during the 12-month period (N & rate per 1,000)
- Children in care on the first day of the 12-month period (N)
- Entries during the 12-month period (N)
- Exits during the 12-month period (N)
- Children in care on the last day of the 12-month period (N)

Foster Care—Other

- All CFSR 2 individual measures, for three years (so that trends may be seen)
- Children in care by number of removal episodes (N & %)
- Children in care by placement type (N & %)
- Children in care by permanency goal (N & %)
- Children in care by number of placement settings in most recent episode (N & %)
- Children by exit type
- Children in care by length of stay
- Number of children waiting to be adopted
- Time from TPR to adoption, for those adopted

E. Observed Performance on the CFSR 3 Statewide Data Indicators for Specific Subgroups and Populations

Maltreatment in foster care by:

- Child age
- Perpetrator type
- County

Recurrence of maltreatment by:

- Child age
- Perpetrator type
- County

Permanency in 12 months by:

- Length of stay
- Child age
- Permanency types (reunification, live with relative(s), guardianship, adoption)
- Current placement setting
- County

Re-entry to foster care in 12 months by:

- Child age
- Permanency types (reunification, live with relative(s), guardianship)
- County

Placement stability by:

- Child age
- County