

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

This submission by Andrew Barclay is in response to the April 23, 2014 Department of Health and Human Services (HHS) request for public comment regarding Statewide Data Indicators And National Standards For Child And Family Services Reviews under 45 CFR Part 1355.34(b), FR-2014-04-23 v. 79, no. 78, Proposed Rules, pp. 22604-22615.

I am a biostatistician, engineer, cofounder of the national Fostering Court Improvement project, and cofounder of the Barton Child Law and Policy Center at Emory University School of Law. I maintain child welfare outcomes reporting sites based on NCANDS and AFCARS data for 18 states as a volunteer. I also consult for states, foundations, and federal grantees. I would like to comment on the technical and policy merits of the proposed rules, and in particular the proposed indicators, national standards, risk adjustment, and Program Improvement Plan (PIP) targets.

Statewide Indicators Proposed by DHHS in FR-2014-04-23 (my naming scheme):

1. CFSR3 S1, Maltreatment in foster care: Of all children in foster care during a 12-month period, what is the rate of victimization per day of foster care?
2. CFSR3 S2, Re-Report of Maltreatment: Of all children who received a screened-in report of maltreatment during a 12-month period, what percent were reported again within 12 months from the date of initial report?
3. CFSR3 P1, Permanency in 12 Months for Children Entering Foster Care: Of all children who enter foster care in a 12-month period, what percent discharged to permanency within 12 months of entering foster care?
4. CFSR3 P2, Permanency in 12 Months for Children in Foster Care for 2 Years or More: Of all children in foster care the first day of the year who had been in foster care (in that episode) for 2 years or longer, what percent discharged to permanency within the next 12 months?
5. CFSR3 P3, Re-Entry to Foster Care: Of all children who entered foster care in a 12-month period who were discharged within 12 months of that entry to reunification, living with a relative, or guardianship, what percent re-entered foster care within 12 months of their discharge?
6. CFSR3 P4, Placement Stability: Of all children who enter foster care in a 12-month period, what is the rate of placement moves per day of foster care?

Summarized Comments:

1. I support the shift from composites to simple measures. As implemented, CFSR round 2 composites proved to have little or no utility in the field.
2. I strongly support parsimony, but I would shift the emphasis implied by the measures from permanency to safety. Many state agencies have transformed from primarily foster care agencies into true child protection agencies, and I attribute much of that transformation to 14 years of CFSR, and the 6-month re-victimization safety measure in particular. For many state agencies, the CFSR prompted the first measurement of their primary mission, protecting children. I support improved and increased measurement of the safety and permanency of the

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

90% of alleged victims not removed to foster care. I favor a 3/3 safety/permanency measure balance over the proposed 2/4 safety/permanency balance.

3. I strongly support the use of observed national rates as the basis for standards.
4. I support the use of each states' historical performance as the basis for PIP targets. However, the proposed rule's sketch of interval estimation with a bootstrap is concerning. If the simpler, mainstream methods of interval estimation for proportions based on a Bernoulli trial assumption were ruled out, I would expect some discussion of the reasons for that in the proposed rule. There is none.
5. I find the information in the proposed rule inadequate for comment on the merits of the proposed risk-adjustment. The comparison to hospital risk-adjustment, perhaps appropriate to provider contracting, may be entirely inappropriate to subjects who are selected without consent through a court-ordered treatment process. The degree of control a state agency exerts over its removals to foster care is fundamentally different from a hospital's control over patients seeking treatment. State agencies judge risk, choose their "patients", and often defend those decisions before a judge prior to admission -- a very different process from hospital admission. The proposed rule would add a complex third layer of adjustment to the "adjustments" agencies and courts make in their patient selection. I have submitted a FOIA request for the expert panel report in the hope that it might shed light on the logic and provide some specifics of the proposed methods. As requested in the proposed rule, I include some variable recommendations below. These were chosen by a fairly objective (but hastily implemented) variable/feature selection procedure using cross-validated ensemble machine learners operating on FFY2012 CFSR2 outcome measures interacting with 2012 Census Current Population Survey SES data. The topics of safety and foster care risk adjustment merit far more research investment and comment before proposing rules.
6. I strongly support the proposed measure of maltreatment in foster care, S1. This will be a tremendous benefit to the field. I encourage the CB to provide specific technical guidance to clarify the definition of incident date to indicate the most recent known day of maltreatment. The quality of links between NCANDS and AFCARS records based solely on AFCARS record number can be estimated using other identifiers time-bracketed by removal and incident dates. This measure is very sensitive to the quality of those links, so checks are critical.
7. I strongly disagree with the proposed use of screened-in reports as a subsequent safety event in S2. I can find no basis in logic, data or the literature for this major policy shift. The rule strongly suggests that the CB believes public reporting is a better indicator of harm than the judgment of agency investigators. The proposed rule contradicts itself by employing one standard of evidence for maltreatment in foster care and another for overall maltreatment. I encourage the CB to continue to consistently use state agency determinations to assess harm to children, as it has for 20 years, and improve upon that by using them for all agency responses, especially AR and non-removal.
8. I support the measures P1, P2, and P4 as proposed. I strongly support measuring all types of permanency with the same yardstick.

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

9. I disagree with the logic and design of P3, reentry. State efforts to decrease re-entry very often apply to children with long lengths of stay, or are fully independent of length of stay. In some cases cohorts with long lengths of stay are at highest risk of reentry. A national standard should not limit the risk pool as proposed without more basis. I see no valid reason to limit measurement of the quality of permanency to only removals during the period. I would replace this with a simple cohort entering permanency independent of length of stay. The measure and results would be consistent with and easily interpreted in the context of the round 2 reentry measure, P1.4.
10. I strongly support the addition of companion measures to PIP plans. For example, a national reentry standard based on the full risk cohort entering permanency could be paired to PIP companion measures targeting specific sub-cohorts appropriate to each state.
11. I suggest greater specificity in the cohort terminology used by the CB to provide clarity and prevent misunderstandings. With a possible exception (birth?), all entry cohorts are also exit cohorts, so the term is not specific without context. I would call the P1 and P4 cohorts *removal cohorts* to improve specificity (and emphasize the parent/judicial/due-process perspective that entries to foster care are removals from family). I would call the P2 cohort a *point-in-time* (or cross-sectional or in-care) cohort. I would call the P3 cohort a *removal-discharge* cohort or a *removal-permanency entry* cohort (the complex name reflecting the cohort complexity). I would call the S1 cohort a *served in foster care* cohort, with a person-time denominator, and I would call the S2 cohort a *screened-in* cohort.
12. I suggest greater specificity and consistency in the indicator definitions. I think the intent of S2 is to measure time from report to re-report as (subsequent reported incident date - index report date). The proposed rule alternately uses disposition date (p 12) and report (Attachment A) in the S2 measure definition, but does not mention incident dates. Please clarify.

Detailed comments and suggestions for the proposed measure S2 and the risk-adjustment strategy follow. I am sorry I don't have time to comment on all the good decisions and work that has been done by CB and its partners in getting to this point.

CFSR3 S2, Re-Report of Maltreatment:

This rule would be bad public policy. I strongly disagree with the use of screened-in reports as an indicator of harm in safety events following an index report. Perhaps I am stating the obvious, but I am not aware of any maltreatment reporting statute that asks reporters to investigate, substantiate, swear to, or otherwise affirm harm to a child. I hope that will never be the aim of reporting statutes, or our reporting systems. In general, I would like reporting to be triggered by the lowest level of evidence, e.g. mere suspicion, to alleviate what I believe, based on 4 rounds of the NIS and some research, to be large-scale under-reporting of maltreatment.

In contrast, I think that the ideal "measure of the agency's attempts to prevent maltreatment" would be (using the language of diagnostic testing) 100% sensitive and 100% specific in detecting harm to a child. In other words, the opposite end of the evidence spectrum from public reporting. I think that perfect

sensitivity and specificity should be an aim of state agency investigative processes, making the determinations from those processes the highest level of evidence of harm available to us. I think the CB should prefer that highest level of evidence in judging the success of state agency prevention efforts. Quoting Drake 2003, quoting Drake 1999: Substantiation is "a statement by the worker that 'I have enough evidence to believe that child maltreatment has occurred.'" If the CB asks states to determine harm, and then uses something entirely different for a national safety standard, I think that is the textbook definition of arbitrary and capricious rule making. For now, the state agency dispositions remain the highest and best evidence we have to detect harm. Imperfect as they are, agency determinations are far more reliable indicators of harm than reports from the general public that are explicitly not intended to determine harm.

Press coverage, statute, policy, and practice all induce large short-term jumps in reporting. We hope, with some basis, that agency determinations of harm are insulated and independent of short-term fluctuations that will erroneously indicate improvements and declines in safety through the proposed measure. Asking workers to prevent harm to children is far more appropriate than asking them to prevent subsequent reports. Under the proposed rule, a system that encourages prevention and limited government intervention by safely preserving families in non-removal responses with enhanced monitoring and reporting will be indistinguishable from a dysfunctional system that removes unnecessarily and discourages subsequent reports.

The disposition level fields in NCANDS could be improved to better reflect the level of evidence, and to separate agency findings from the agency's response. I find them to be confused with respect to findings and responses (these should be separated), but adequate for the current purpose, including safety on alternative response tracks. I have found those data to be very useful in measuring the safety of various alternative responses to maltreatment in many jurisdictions for many years. I have commented previously on their weaknesses and suggested improvements from the perspective of a data consumer.

The change to a 12-month follow-up window comes at too great a cost. I think that I appreciate the many advantages of following an index case for a longer period, but I often search for ways to decrease that follow-up to minimize cost. Longer follow-up usually comes with costs in resources, relevance/sensitivity to intervention effects, interpretation, and, with annual NCANDS submissions, analysis complexity. When measuring recidivism at the level of states, the costs and compromises of 12 months of follow-up are unnecessary and unjustified.

For most measures of recidivism, the 12-month follow-up rates are predictable with 6 or even 3 months of follow-up because they follow a parametric function well. In other words, the 12-month rates contain no additional information over 6-month rates, because the underlying processes are the same. The Gompertz cumulative distribution is one such function. It is "often applied to describe the distribution of adult lifespans by demographers and actuaries."¹ That has some intuitive appeal, though other forms of decaying exponentials and failure-time functions may fit more closely.

¹ http://en.wikipedia.org/wiki/Gompertz_distribution, visited 5/20/2014.

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

To illustrate how it can be that less than 12 months of follow-up data can predict the 12-month recidivism rate, I linked 24 months of NDACAN NCANDS records on child ID. Figure 1 shows the Kaplan-Meier survival estimates from index report to subsequent report (the left chart, the proposed measure) and subsequent victimization (the right chart, victim=1+ substantiated/indicated, AR-victim or death). The black solid line represents 24 months of data (FFY2010+FFY2011 linked, 7.2 million records), while the red solid line represents FFY2010 data only. The 15-day time resolution of NDACAN data produces the stair-step artifact. The blue and green dashed lines are parametric functions fit to 24 months and 12 months of data respectively.²

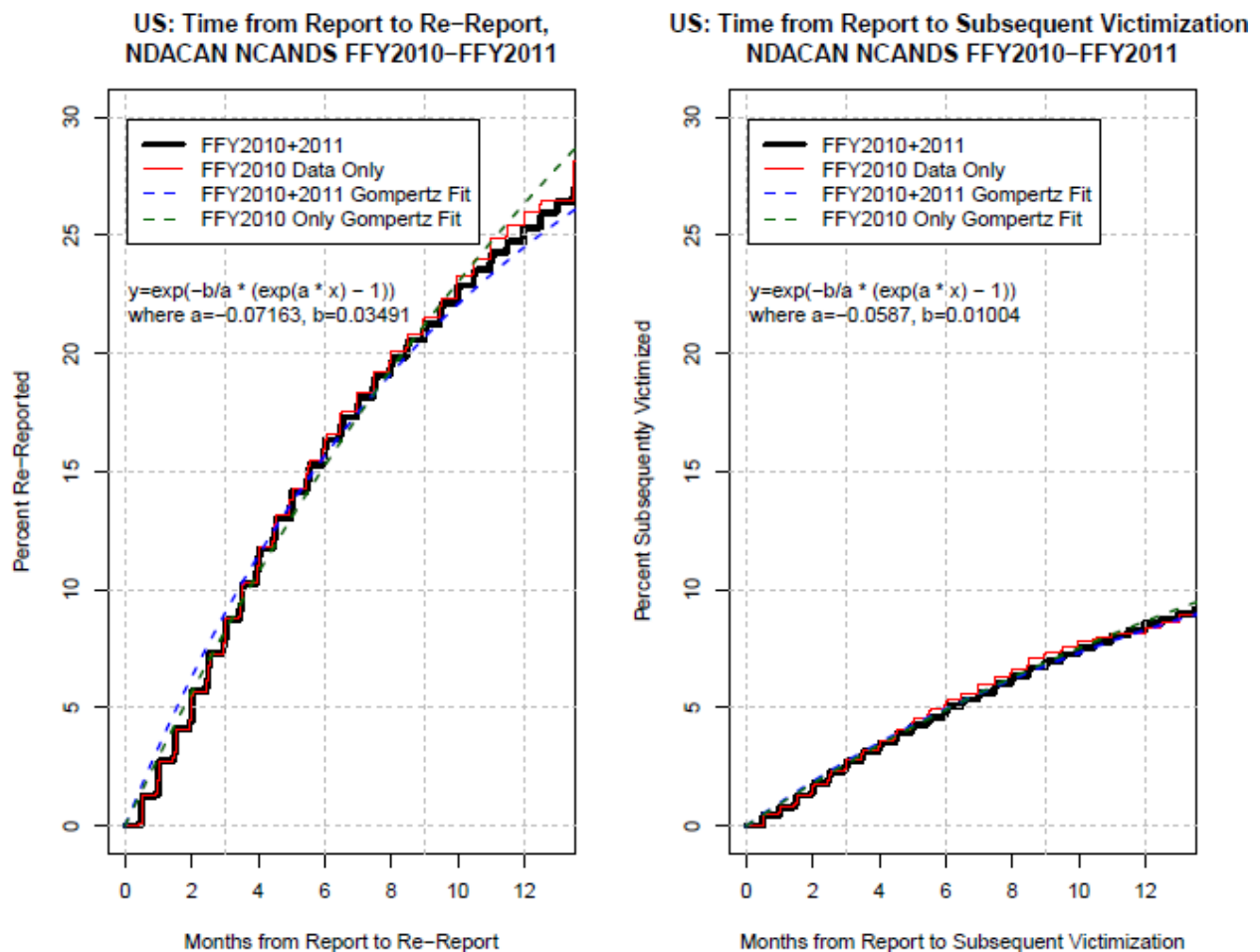


Figure 1: Illustrating goodness of functional fit; Time from report to re-report or victimization.

Not perfect, but the measured 12-month re-report rate from FFY2010+FFY2011 data is 24.5% versus 26.3% from the fitted FFY2010-only data. Some may deem the 1.8% difference important, but I think the appropriate question is: Does this additional 6 months of follow-up time net enough additional information to justify the costs? These recidivism curves are always smoothly time-varying, always the

² R's "flexsurv" package fits a Gompertz distribution to survival data, <http://cran.r-project.org/web/packages/flexsurv/>, visited 5/20/2014.

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

same shape, even in the smallest states. If 90%+ of the information is contained in half the data, I don't see any way the additional follow-up can justify the costs. There is no cause to wait an additional 6 months and incur additional costs to confirm that the 12-month rate follows from the 6-month rate.

From the legal, policy and practical perspective, I find many reasons to prefer short-term government interventions with prompt reporting feedback that target high-risk cases. To strike a balance with due-process rights, I think we should prefer a light-touch with brief government involvement. I think it is reasonable to expect our state-of-the-art, limited-engagement safety interventions to produce strong effects on high-risk cases in months 1-6. It appears that those effects stay with the family beyond closing of the case, but I think it is unreasonable at this stage in the development of our safety interventions to expect the effects to protect all reported children, the vast majority low-risk, in their families for more than 6 months beyond closing of the case.

Under the proposed rule, the measure of protection from harm of a child reported at the beginning of FFY2015, 10/1/2014, will be available after 2 NCANDS cycles, in February 2017, 28 months later. With the leadership turnover in many states, a 2+ year reporting lag makes advocacy using that information extremely difficult. The shelf-life of our reporting information is short, and it's relevance to decision-makers and the field is strongly tied to its currency. If PIP interventions seek to produce long-term effects, then companion measures with longer follow-up will be justified.

Within the constraints of the 12-month NCANDS reporting cycle (I suggest this be changed to a quarterly cycle), I find that following index cases disposed during the first 9 months of the FFY followed by 3 months of follow-up (based on incident dates) strikes a good balance between variance and accuracy of the rate estimates in small states. However, for consistency, I recommend that CB retain the 6-month re-victimization rate (among high-risk cases) of previous CFSR rounds as S2 and add a similar measure as S3 limited to the cohort of victims (again, high-risk) not removed to foster care. This will give some support to states and encourage innovation in finding methods that safely avoid the trauma of removal from family.

I expect that the CB understands that measurement of recurrence using different index and subsequent determinations, e.g. screened-in reports followed by substantiated victimization, will exacerbate the variance across states. Under such a design, two states using the same level of evidence and services with the same prevention efficacy but differing in their reporting rates (due to infrastructure, policy, public awareness, etc.) could have vastly different measurements. I have not found any evidence that using screened-in reports rather than victims as index cases decreases variance between states.

If parsimony demands that CB drop a permanency measure to add a safety measure, then I would combine the 2 permanency timeliness measures (P1 & P2) into a single measure of the per-capita person-time that children spend without a permanent family: $\text{sum}(\text{child-days spent during the FFY in non-relative or congregate foster care}) / (\text{state child population})$. This rate is easily calculated from AFCARS and conveniently comes to about 1-2 days without family per child in the population.³

³ See, e.g., http://fosteringcourtimprovement.org/ga/DHRRRegion/incare_summary.html

Risk Adjustment:

Is risk-adjustment appropriate to the proposed application? Until now, I had not considered how different our methods of selecting children into our child welfare programs are from hospital admissions. I assumed the YNHSC/CORE methods could be re-purposed to our application, until I read the reasons that they do not adjust for population socioeconomic status (SES): "Risk adjusting for patient SES would suggest that hospitals with low SES patients should be held to different standards for patient outcomes than hospitals treating higher SES patient populations."⁴ If the CB includes SES-related variables, it may be proposing exactly the sort of adjustment that CMS forbids.

The proposed rule states "[t]he goal of risk adjustment is to minimize differences in outcomes that are due to factors over which states have little control, such as the age of children coming into foster care." The example posed presents a quandary. *DeShaney v Winnebago* 1989 (aka the "Poor Joshua!" decision) shapes much of our law regarding selection for services. In *DeShaney*, the SCOTUS found that state agencies are not required to protect children from non-state (e.g. parent) actors.⁵ Therefore, according to Chief Justice Rehnquist and Judge Posner, state child welfare systems have control over all characteristics of the children to whom they provide services, with the exception of children in state custody. In contrast, hospitals are required by EMTALA to provide emergency services.⁶

Of course, this is not the reality, and state agencies and courts are in the business of finding and protecting children, and hospitals treat patients without being ordered to do so. However, I think a state's right to select whom it will or will not serve and the differences between state selection processes (including courts as a component, as we always have) are important considerations in choosing appropriate variables and methods to improve the comparability of outcomes. I am sorry, but I cannot offer a method for adjusting for the range of selection processes, and I do not see a clear answer to this question yet.

Suggested risk-adjustment variables: Since measures P1, P2, and P3 closely mirror CFSR round 2 measures, I used the CB's state proportions published on <http://cwoutcomes.acf.hhs.gov/data/> to assess the associations of a few Census Current Population Survey (CCPS) rates to these 3 permanency outcomes. I used two different machine learning approaches, giving them equal votes as to strength of association: 1. a gradient boosting machine⁷, and 2. random forests⁸. I find that these algorithms provide more insight, with fewer assumptions, more efficiently than any survey, literature search or meta-

⁴ <http://goo.gl/1nJxUL> p 13.

⁵ http://en.wikipedia.org/wiki/DeShaney_v._Winnebago_County, visited 5/20/2014. Justice Rehnquist: "The affirmative duty to protect arises not from the State's knowledge of the individual's predicament or from its expressions of intent to help him, but from the limitation which it has imposed on his freedom to act on his own behalf... it is the State's affirmative act of restraining the individual's freedom to act on his own behalf - through incarceration, institutionalization, or other similar restraint of personal liberty - which is the "deprivation of liberty" triggering the protections of the Due Process Clause, not its failure to act to protect his liberty interests against harms inflicted by other means."

⁶ http://en.wikipedia.org/wiki/Emergency_Medical_Treatment_and_Active_Labor_Act, visited 5/20/2014.

⁷ <http://cran.r-project.org/web/packages/gbm/>, visited 5/20/2014.

⁸ <http://cran.r-project.org/web/packages/randomForest/>, visited 5/20/2014.

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

analysis. The parameter space is limited only by our imaginations, and the machine can perform a comprehensive search of the entire space to assess strengths of association, free of linearity or other assumptions, checked as it searches by extensive cross-validation. I ran a quick proof of concept for this comment, but I hope that the CB will combine many methods and many more variables, including linked NCANDS and AFCARS, American Community Survey, and other public data sources in an ensemble approach. I will post R code and data to recreate this analysis at <http://fosteringcourtimprovement.org/resources.php>. The fits of these models to the FFY2012 state proportions were poor (RMSE ~11%), so I recommend that many other variables be considered.

Tables 1 to 3 list the 5 strongest associations to the three FFY2012 CFSR round 2 state proportions:

Table 1: CFSR Round 2 Permanency Measure 1.3, 12-Month Reunification	
Variable	GBM+RF Importance Vote
CFSR2, C1.4: 12-Month Reentry	19
Adults, Bachelors or Higher	17
In Foster Care per 10K, 9/30/2012	14
Asian Alone	12
Child Poverty, Black/AA Alone	9

Table 2: CFSR Round 2 Permanency Measure 1.4, 12-Month Reentry	
Variable	GBM+RF Importance Vote
White Alone	20
Removals to Foster Care per 10K, FFY2012	18
Discharges from Foster Care per 10K, FFY2012	12
Child Poverty	11
CFSR2, C1.3: 12-Month Reunif	11

Table 3: CFSR Round 2 Permanency Measure 3.1, In-Care 24+ Months, Permanent within 12 Months	
Variable	GBM+RF Importance Vote
Adults, Bachelors or Higher	19
White Alone	12
Child Poverty, White Alone	12
Removals to Foster Care per 10K, FFY2012	10
Population 6-11	8

CFSR3 Indicators & National Standards Proposed Rule, 45 CFR 1355.34(b)

The 24 state-level variables considered by the 3 models were:

- | | | |
|-----------------------------------|-----------------------------------------------|--------------------------------------------------|
| 1. Population 0-5 | 10. Asian Alone | 19. Removals to Foster Care per 10K, FFY2012 |
| 2. Population 6-11 | 11. AI/Native Alaskan/PI Alone | 20. Discharges from Foster Care per 10K, FFY2012 |
| 3. Population 12-17 | 12. Mixed Race | 21. In Foster Care per 10K, 9/30/2012 |
| 4. Population 18-80+ | 13. Child Poverty | 22. CFSR2, C1.3: 12-Month Reunif |
| 5. Adults, No High School Diploma | 14. Child Poverty, White Alone | 23. CFSR2, C1.4: 12-Month Reentry |
| 6. Adults, Bachelors or Higher | 15. Child Poverty, Black/AA Alone | 24. CFSR2, C3.1: 24+ Months, 12-Month Perm |
| 7. Covered by Medicaid | 16. Child Poverty, Asian Alone | |
| 8. White Alone | 17. Child Poverty, AI/Native Alaskan/PI Alone | |
| 9. Black/AA Alone | 18. Child Poverty, Mixed Race | |

Thank you very much for proposing these improvements and for considering my comments.

Andy Barclay