CHAPIN HALL
CENTER FOR CHILDREN
AT THE UNIVERSITY OF CHICAGO

Monitoring Child Welfare Programs: Performance Improvement in a CQI Context

Fred Wulczyn

2007

Chapin Hall Working Paper

Chapin Hall Center for Children at the University of Chicago

1313 East 60th Street

Chicago, IL 60637

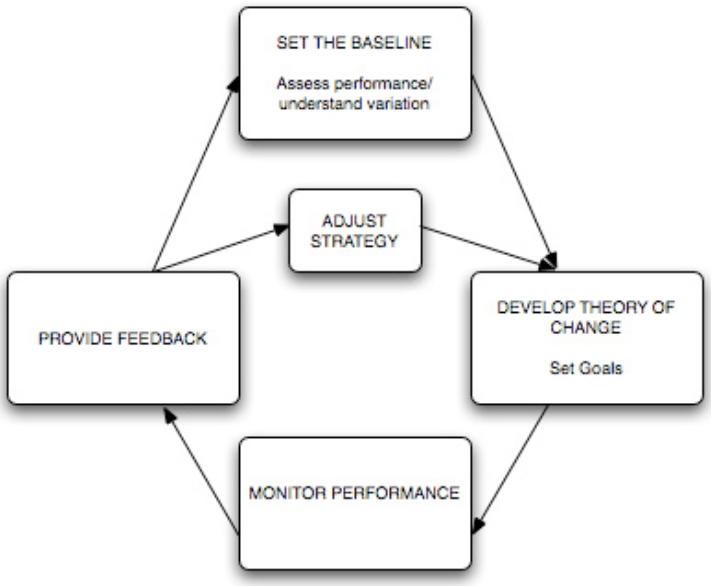
773-753-5900 (phone) 773-753-5940 (fax)
www.chapinhall.org

CS-137

INTRODUCTION

In practice, the continuous quality improvement process in child welfare systems is no different from other areas of system design. The process begins with a set of core outcomes, which typically define the central mission of the organization. The core, or mission-critical, outcomes in child welfare involve child safety, permanency, and well-being. The second element of the continuous quality improvement process involves a statement of performance or a baseline. The baseline refers to systematically gathered data that describes in current and historical terms how well the organization achieves the core outcomes. Baseline data can come from a variety of sources, including administrative data and case records. The baseline is related directly to the third element of the CQI process: setting goals and deciding on a theory of change. To the extent that an organization understands how well it accomplishes its goals, the CQI process implies continuous work to improve performance. Goals usually relate to a *gap* between current performance (the baseline) and future performance (where the agency would like to be). The theory of change describes the steps the organization plans to take in order to close the gap. The steps may be organizational (fiscal, policy, etc.) or practice-based (i.e., effective service models). The theory of change often emerges out of stakeholder interaction. The last step in the CQI process involves monitoring and feedback. In the parlance of systems change, the theory of change represents hypotheses that describe the relationship between inputs (changes in the organization of services) and outcomes. The hypotheses behind the CQI plan reflect the simple belief that changing the inputs will produce the intended impact on the outcomes. Monitoring provides a way to discern whether the intended changes are taking place; whereas feedback refers to the distribution of information back through the system to key actors as part of a systematic effort to keep the stakeholders informed of progress. In short, monitoring and feedback provide the information needed to understand whether system changes are "working." When all the pieces are working together the process will look something like this:

Figure 1
The Cycle of Improvement



Using permanency outcomes for children placed in foster care (e.g., reunification or adoption), we illustrate some of the issues encountered when attempting to use baselines to establish goals and monitor progress. Our approach is based on an approach developed at Chapin Hall Center for Children, in conjunction with a number of states, including New York, Tennessee, and New Jersey. The approach used here stresses the use of administrative data, but this should not be construed to mean that other types of data cannot be used to reinforce the basic approach. On the contrary, administrative data properly assembled and used in conjunction with other forms of data substantially improve the insights obtained through a systematic review of a local child welfare system. In fact, building a robust connection between administrative data and other sources of information is critical to developing a plausible theory of change.

SETTING BASELINES AND GOALS

Generally speaking, the performance goal is based on a baseline that captures what would have happened (e.g., measured as an outcome, such as the proportion of children admitted who leave

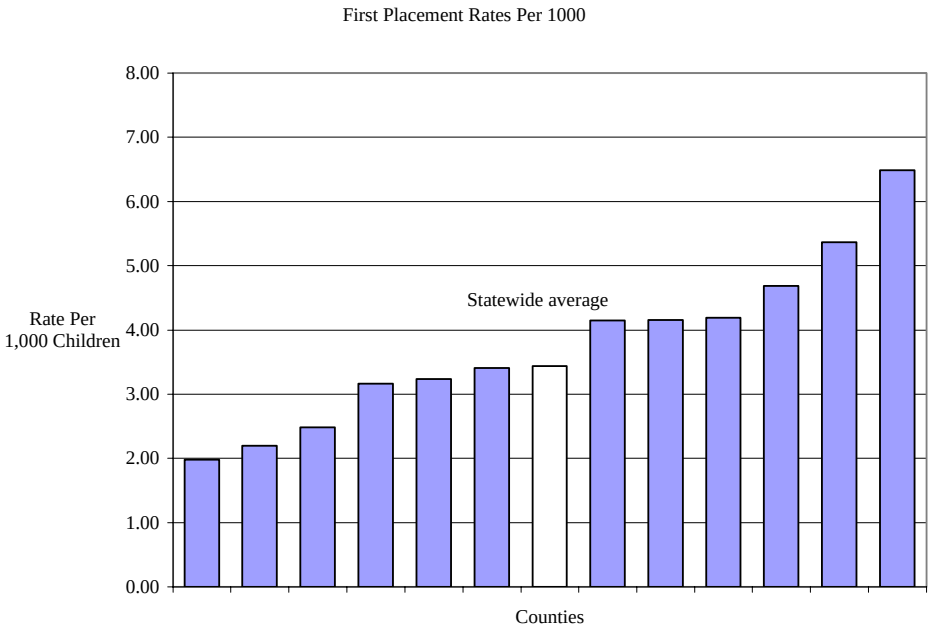
placement for permanency) if the proposed system changes were not implemented. In essence, the baseline serves as a counterfactual, in the absence of an experiment or some other comparison group that allows for judging whether the organization (e.g., a child welfare agency) is making progress toward improving outcomes.

There are a number of approaches—some better than others—to setting the baseline and establishing goals. For example, baselines and goals should be adjusted to reflect patterns of variation in the local service delivery system. This means two things. With respect to child welfare outcomes, baseline expectations differ depending on the various populations of children served within the system. Babies, for instance, are much more likely to be adopted than older children, whereas older children are much more likely to run away. Baselines should reflect such basic differences--in part because such adjustments influence how one thinks about doing a better job. Baseline expectations (and goals) also differ for different parts of a state or county. In many states, performance in urban areas differs from performance in rural areas. To the extent these differences are a function of different approaches to helping children and families, adjusting baseline for geographic variation may, again, improve the thinking that goes into the theory of change.

Speaking broadly, one important byproduct of adjusting the baseline to reflect how children and families use child welfare services is the knowledge gained about underlying patterns of variation. In a state child welfare system, whether the programs are state- or county-administered, county (or regional) performance will be distributed around the mean for the state. For example, in Figure 2 below, the county placement rates are distributed above and below the statewide average. An appropriate baseline for the counties above the mean may be different from the baseline used for counties where the placement rates are already below the state average. Figure 2 also points out a basic feature of system improvement – the goal of improving systems is to reduce the wide variation in experience that exists in some child welfare systems. Assuming

that children in the counties with below average placement rates are being adequately protected, reducing the placement rates in the counties with above average placement rates while holding placement rates level in the remaining counties would reduce the variation and lower the mean/statewide placement rate. In short, similarly situated children in different parts of the state would have roughly the same experience, all else being equal.

Figure 2



Adjusting the baselines and corresponding goals is best done empirically. That is, to the extent that administrative data are available, those data should be analyzed to ascertain the extent to which there are meaningful differences in the pattern of care. Because the number of distinct groups or subpopulations is potentially large, we recommend parsimony and clinical relevance as two guiding principles when deciding how many discrete groups ought to be created. Parsimony is important because each group has to be monitored separately and too many groups increases complexity. Clinical relevance is meaningful because the composition of the groups should convey something about the children/families that helps practitioners organize an appropriate response. For example, babies and adolescents are two groups that often stand out as distinct

populations for which separate baselines make sense because they have different experiences and will likely need different services.

Among the attributes of children or their placement experience that are candidates for use in adjusting the baseline, age, placement type, and reason for placement are useful. Race/ethnicity is another option; however, stratifying baselines by race or ethnicity suggests that race-based differences in the experience of children in the child welfare system are tolerable. For instance, African American babies spend more time in foster care than white babies. Establishing a separate baseline for each group creates the expectation that the differences will be smaller in time but persistent. There are three other issues to keep in mind when setting baselines and establishing goals. First, as mentioned earlier, the interactions between outcomes have to be taken into account. Increasing permanency (e.g., reducing length of stay) may influence the reentry rate. In part, the interaction between outcomes is an issue of monitoring. As counties systematically improve in one area, other outcome domains have to be monitored to determine whether there are other, adverse consequences. Second, each county (or administrative unit) has its own performance trajectory. That is, two counties with different performance today may have performed similarly 5 years ago, and counties that are different today may have started 5 years ago from a similar position. The point is this. Setting goals for a county with steadily improving performance may be different from setting goals for counties with steadily eroding performance. The challenge is sorting counties based on a trended baseline. Finally, in some smaller counties, certain outcomes are relatively rare. In such instances, a baseline and goal may be accompanied by a corridor that expresses a range of performance above and below the baseline (or goal) that represents acceptable performance. For example, in some populations in smaller counties, reentry to foster care is relatively rare. Reentry rates in smaller counties can be affected adversely by sibling groups returning to care such that the reentry rate exceeds the goal. A corridor (sometimes called a risk-corridor) provides for subjective judgments regarding

performance relative to a goal. Perhaps more importantly, the use of corridors creates flexibility in a field where precise prediction at the individual level is often difficult. After all, the baseline performance and goals serve as guidelines for decision makers, not as a substitute for decision-making.

USING HISTORY TO ESTABLISH A BASELINE AND SET THE GOAL

To illustrate some of the lessons learned helping states establish a baseline and set a goal, we have constructed the following example. The task is to determine what the permanency baseline should be for the 2003 admission cohort in a fictitious region in a fictitious state. The basic data, displayed in Table 1, show what happened in the 3 preceding years for children admitted as babies. The data on permanency shows where the children were at the end of 2 years. For example, 428 babies were admitted to care in 2000. [1] Two years later (12/31/2001), 166 children (39%) were still in care. In other words, "2 years later" represents a calendar marked from the beginning of the time period through the end and does not mark time from the child's date of entry.

These data are risk adjusted in that they refer only to children admitted as babies. A similar chart for a different group of children would show different exit patterns. Adolescents for example would expect to have many more exits to other exit types (i.e., running away) and probably fewer adoptions. The data also represent a single region as opposed to an entire state. Again, other parts of the same state may show different results for children admitted as babies. The goal is to make "apples-to-apples" comparisons, an objective facilitated by controlling for attributes of the

[1] In this example and the one that follows it, we do not want to create the impression that setting goals and baselines falls within a 2-year window of time. On the contrary, performance of the child welfare system has to be observed over the full life of each cohort. We selected 2 years in part because the federal Child and Family Services Reviews use a two window, even though the approach here is quite different from the baselines developed for the state performance improvement plans.

population.[2] Finally, the data reflect recent trends. In this particular region, outcomes have changed slightly. In other regions, one might find a more distinct time trend, even over short periods of time. For example, if performance had improved dramatically in the past few years, expectations for future improvement may have to reflect that fact, especially if the county's performance is already substantially above the state's statistical average.[3] The opposite is also true.

[2] Although these data are risk adjusted, we do not mean to imply that within-group heterogeneity has been addressed entirely. Babies coming into foster care in one part of the state could be different from the babies entering in some other part of the same state. Judgment is an important part of the CQI process and local knowledge has to be applied to the interpretation of baselines and goals.

[3] The use of the term *average* in the context of system improvement is sometimes interpreted to mean mediocre. Here, we are talking about the statistical average, a feature of every system of performance measurement, unless of course performance is identical across the entire system. The statistical average is a mathematical construct, not an assessment of quality. Even in systems that provide high-quality services, performance will center near the statistical average. What often distinguishes high-quality service systems is the limited range of variation around the average, across different service units.

Table 1
Baseline Permanency Outcomes for Infants
Admitted to Foster Care by Year: South Central Region

	All Admits	Number Reunified	Number Adopted	Number Other Exits	Number Still In Care
2000	428	215	36	11	166
2001	509	269	42	8	190
2002	551	282	44	5	220
		Percent Reunified	Percent Adopted	Percent Other Exits	Percent Still In Care
2000	100%	50%	8%	3%	39%
2001	100%	53%	8%	2%	37%
2002	100%	51%	8%	1%	40%
2003	**?**	**?**	**?**	**?**	**?**

The data in Table 1 show a system with relatively stable performance during this brief window of time, at least in this region and for this age group. Each year, about 40 percent of the original cohort is still in care at the end of the observation period. If one were to project how many children admitted in 2003 will still be in care at the end of 2004, 40 percent would be a reasonable baseline. Similarly, about 50 percent of the children will be reunified and another 8 percent will be adopted, provided that history repeats itself.

The baseline forms the basis of the *goal* or target for future performance. The goal expresses the anticipated level of performance given a set of service improvements. Stakeholders, looking at the data may decide that it is possible to completely eliminate other exit types and increase exits

to reunification.[4] Overall, the goal may be to reduce the percentage of children still in care to 35 percent, with an increase in reunification to 56 percent. By specifying population, geographic- and outcome-specific targets, the opportunity to develop a plausible theory of change is greatly enhanced because the effort is much more targeted. Put another way, using separate baselines avoids the one-size-fits-all approach that often characterizes reform initiatives.

MONITORING OUTCOMES OVER TIME

Once the goals for system improvement have been set, initiatives that embody the theory of change are rolled out. As noted, reform initiatives can operate at the systems level involving changes in policy, the locations of services, staffing patterns, or how funding is used. Change initiatives may also involve new case practice models (e.g., changing how visitation is conducted) and the use of evidence-based practices. From a monitoring perspective, the critical issue is knowing when the initiative starts so that monitoring reports reflect the timing of those service initiatives relative to when along a child's service trajectory change can be expected to happen.[5]

One important feature of child welfare services is the fact that outcomes at the individual level may take months if not years to observe, especially if a child is placed in foster care. In Table 1 above, 40 percent of the children were still in care at the end of the observation period. Real change in performance has to be measured against the time it takes to observe all the children in a

[4] Stakeholders with local knowledge ought to have some idea of what those other exit types include for this population and how exits of those types might be avoided. More importantly, setting the level of expected change, the goal, is a matter of both art and science.

[5] This is especially true for the children already in care when the initiative started. For example, when the Adoption and Safe Families Act was enacted, there were already 500,000 children in foster care. The impact of ASFA on various groups of children differs in part according to how long children had already been in care. Some children subject to ASFA's rules and regulations were placed earlier in the decade; other children were admitted in the year leading up to ASFA. ASFA's impact has to be understood with those differences in mind, especially since ASFA (or any other policy) cannot be expected to influence what has already happened. We recommend building a separate group around the children in care when the initiative was launched and tracking them separately from the children who entered care after the initiative started. Our examples, however, focus on children presumably admitted after initiation.

given cohort to leave foster care. At the same time, there is an expectation among advocates, policymakers, and families that service improvements will prove beneficial long before the last cohort member leaves care.

Our approach to this problem is to set a longer-range goal, with interim steps along the way that can be used to monitor progress. In this way, the monitoring occurs in something that approximates real time. Interim data can then be used to adjust strategies according to what the data suggest is happening. An example of this approach is provided in Table 2. As in the previous example, the data reflect a baseline for a specific population in a specific part of a state. In this case, the example follows the experience of 1,341 children admitted in a year marked at the beginning by the start of a new practice initiative.

The report consists of three basic pieces of information. The baseline represents a projection of where those 1,341 children are expected to be at the end of a 2-year window of time. In this particular report, we are showing only the baseline number of permanent exits; other columns could be added to the report to show the number of children by exit type (including other exits) and the number of children still in care. The first panel of data shows a "Summary of Two Year Performance Expectations." The *baseline* number of exits (created using historical data) describes how many children are expected to leave, given no changes in how well the system works. The first panel also provides a statement of the *goal* or how many exits are expected under the assumption that the program initiative has the impact it is designed to have. Both of these data points are expressed as a percent of the original population of children. The baseline and goal are also expressed in terms of placement days because reducing the number of children who are still in care at the end of a 2-year period does not always translate into fewer days used. For example, if all the children are discharged on the last day of the 2-year period, the target measured as a percent will have been met, but the time needed to reach the target will have been extended. For the children portrayed in this example, the baseline suggests that 751 of the

original 1,341 will have exited to one form of permanency or another (about 56%). If services improve the chances for permanency, a goal of 831 children seems reasonable (62%). If the increase in discharges to permanency happens in a timely fashion, placement days would drop from nearly 573,000 to slightly more than 544,000, over the full 2-year period.

As we noted, 2 years is a long time to wait. To adjust for the need to have information sooner in the improvement cycle, we adapted interim data that reflect progress to date. These data are found in the second panel, labeled "Performance Summary as of 6/30/03." In brief, the data respond to the question, if system improvements produce a 10 percent increase in exits to permanency over 2 years, how much progress can be observed within the first 6 months? The length of the interim time period can be longer or shorter than 6 months. The objective is the same. Also, the 2-year end-point merely reflects the fact that a much longer period of time is needed to observe all of the children admitted during a given year leave foster care.

The monitoring report suggests that the baseline number of exits in the first 6 months would be 268 children, or about 36 percent of the expected exits. Because 6 months represents 25 percent of the 2-year window, the data suggests that more exits will occur in the first 6 months than in later periods. Given a goal of increasing exits by 10 percent over 2 years, the second panel of data also shows the corresponding figure for the first 6x months of the 2-year period (the goal). Last, the second panel introduces the *observed* number of exits. The observed exits are used as comparisons with the baseline and the goal to determine whether the observed rate of discharge is consistent with performance improvement that will, over a full 2-year period produce an overall increase in permanent exits. As before, the data are expressed in percentage terms and as care days. In this example, exits at 6 months (282) are behind the goal (295) but ahead of the baseline (268), suggesting that a careful review of the service strategy might point to ways the approach can be strengthened. Note as well that the impact on care days is relatively small, suggesting that those exits that did occur happened during the latter part of the first 6-month evaluation window.

Table 2
Monitoring Outcomes Over Time

Total Number of Children: **1341**	Permanent Exits
Summary of Two Year Performance Expectations	
Baseline # of Exits Over Two Years	751
GOAL: Total Expected # of Exits Over Two Years	831
Baseline Exits Over Two Years as a % of Total Children	56%
GOAL: Total Expected Exits Over Two Years as % of Total Children	62%
Baseline # of Days Used Over Two Years	572,892
GOAL: Total Expected # of Days Used Over Two Years	544,247
Performance Summary as of 6/30/03	
Baseline # of Exits as of 6/30/03	268
GOAL: Expected # of Exits as of 6/30/03	295
Observed # of Exits as of 6/30/03	282
Difference, # Observed-Expected	-13
Baseline Exits as of 6/30/03 as % of Total Children	20%
GOAL: Expected Exits as of 6/30/03 as % of Total Children	22%
Observed Exits as of 6/30/03 as % of Total Children	21%
Difference, Observed-Expected	-1%
Baseline # of Days as of 6/30/03	200,512
GOAL: Expected # of Days as of 6/30/03	190,486
Observed # of Days as of 6/30/03	199,011
Difference, Observed-Expected	8,525

These types of monitoring reports address a number of questions practitioners often have. First, how often should reports be produced? These data suggest that semi-annual reports may be adequate in that change tends to happen in small increments. Reports should be timed so that meaningful change is revealed in a timely fashion. Quarterly data are useful; every 6 months may be often enough. Practice wisdom should serve as a guide; the answer may differ depending on the outcome. It is important to remember that frequent reporting of data produces a burden of its own that may end up turning off the interest in using data to guide decisions. Granularity is another question that comes up frequently. State-level, region-level, and county-level data are the

obvious choices. In counties with large populations, within-county reports are useful as well. However, there is a point of diminishing return in that data reports for aggregations below the county level (e.g., supervisors and workers) often produce fewer and fewer children. As a result, the baseline becomes much less stable and harder to use as a barometer of the future. Finally, practitioners often ask, how long will it take change to happen? These data suggest that as change happens, a sensitive monitoring tool will show change as it happens.

SUMMARY

Monitoring performance in the absence of well-articulated goals is essentially useless. Improvement in complex systems happens through time. Understanding how much change to expect and just how hard one should push for change depends on knowledge about the past and the unique experiences of children in the child welfare system. Real, deep-seated system change is a marathon, not a sprint, so patience is a virtue.

CHAPIN HALL
CENTER FOR CHILDREN
AT THE UNIVERSITY OF CHICAGO

Chapin Hall Center for Children
at the University of Chicago
1313 East 60th Street
Chicago, Illinois 60637

www.chapinhall.org
phone: 773/753-5900
fax: 773/753-5940