

## Texas Child Care Market Rate Survey Study Methodology

### Background

45 Code of Federal Regulations 98.45(c) requires that at least every three years as part of the Child Care Development Fund state plan, lead agencies develop and conduct a statistically valid and reliable survey of the market rates for child care services. A market rate survey (MRS) will be considered “statistically valid and reliable” if the study methodology meets the following benchmarks:

*Provides complete and current data* – The survey uses data sources that fully capture the universe of providers in the priced child care market. In addition, the survey should reflect up-to-date information that is based on price data collected within a specific time period.

*Includes the priced child care market* – The providers charge parents a price directly that is established through an arm’s length transaction. In an arm’s length transaction, the parent and the provider do not have a prior relationship that is likely to affect the price charged.

*Represents geographic variation* – The survey includes providers from all geographic parts of the State, Territory, or Tribal service area. It should also collect and analyze data in a manner that links prices to local geographic areas.

*Uses rigorous data collection procedures* – The survey uses good data collection procedures that includes a response from a high percentage of providers (generally, 65 percent or higher is desirable and below 50 percent is suspect); implements a strong sample design; conducts analyses of potential response bias to ensure that the full universe of providers in the child care market is adequately represented in the data and findings; and survey based on the languages used by child care providers.

*Analyzes data in a manner that captures market differences* – The survey examines the price per child care slot so that all child care facilities are not equally weighted because some serve more children than others. The also collects and analyzes price data separately for each age group and category of care to reflect market differences.

## Texas Workforce Commission Child Care Market Rate Survey Methodology

TWC contracts with the University of Texas at Austin School of Social Work and the Ray Marshall Center to conduct the Texas MRS using a methodology that complies with the required federal benchmarks. The TWC methodology creates statistically valid and reliable rate distributions of part-day and full-day child care market rates by:

- Geographic area (i.e. for the entire state of Texas and for each of the 28 local workforce development areas (LWDAs);
- Each of the three provider facility types (i.e. licensed child care centers (LCCC), licensed child care homes (LCCH) and registered child care homes (RCCH)); and

- Age groups (i.e. infants, toddlers, preschool-age and school-age children).

The survey also has rigorous data collection and rate estimation processes, adequate response rates; is available in Spanish and Vietnamese; and weights rates by the number of slots the rate represents. As a result, development of a new MRS methodology is not necessary. The full explanation of the proposed methodology follows.

#### Data Sources and Sample Population

The MRS sample will come from a list of all licensed and registered child care facilities provided by the Texas Department of Family and Protective Services (TDFPS) and the TWC administrative database of providers with agreements.

The MRS study sample will be drawn from the list of all licensed and registered child care facilities provided by the TDFPS a month before the surveys will begin. The surveys are typically conducted between September and March of the study cycle which captures complete and current information for the rates estimates.

The list of facilities from which the sample is drawn is typically between 15,000 and 18,000 and includes the three types of child care providers (LCCC, LCCH, RCCH) in each of the 28 LWDAs.

#### TWC Administrative Data

Although the DFPS list of facilities includes all licensed facilities in the state, TWC also has a database of administrative data that contains agreed upon rates charged by providers. Thus, starting with the 2018 MRS methodology, the TWC administrative data of known rates will be used to increase the precision of the rate estimates by boosting the sampling of non-subsidized providers and the total number of independent rates obtained.

The TWC administrative data will help improve the precision of rate estimates, it cannot be used alone to provide statistically valid and reliable rate estimates because:

- it does not fully capture the universe of providers in the priced child care market. A match of past surveys indicates it only represent one third of the providers that complete the MRS; and
- it does not contain the number of slots for each facility by rate and age type. So, the price per child care slot is not available and correct facility weights cannot be applied.

#### Determining the Priced Child Care Market

Since the population of interest for the MRS study is the priced child care market, the sample will not include relative child care providers. In addition to relatives, providers will be excluded from the DFPS list of licensed providers if they:

- do not have an established price or charge the public directly for services;
- are not available to all families in the market;

- are not available at adequate times or hours; and
- discount rates based on prior relationships with families.

Thus, the following providers will be excluded from the list of available provider prior to the sample being drawn or at the time of the survey interview:

- Head Start facilities - do not charge families directly;
- Drop-in care facilities - charge very high hourly rates and not typically used by employed parents for regular care;
- part-day care with no after-school care, summer camps, care provided to specific populations only (i.e. children with special needs, children of teen moms, children of staff at a company, etc.)
- free/family-discounted child care services
- kindergartens that do not offer regular afterschool care; and
- summer camps, and nursery schools.

#### Sample Size and Selection Methodology

After the facilities identified prior to the interview are removed from the statewide universe of facilities, a small statewide representative sample of 1,000 facilities will be randomly selected in order to:

- 1) provide greater statistical power for estimating statewide rates and trends and
- 2) increase the precision of rate estimates for the largest local boards that provide the bulk of care in the state.

In addition to the statewide sample, a second random sample will be drawn from the remaining facilities to estimate rates at the LWDA level. To concentrate the statistical power of the sample where it is needed most, and thus make efficient use of study resources, the second sample involves over-sampling of LWDAs with fewer facilities.

For the LWDA sample, a target number of each type of facility to be selected from each LWDA based on projections of the number needed to yield an average of at least 40 independent rate observations in full-day rate categories which in as increase over prior years. This increase will initially boost the sample size and improve the precision of rate estimates. The new projections will be greater than those based on detailed response rates to the last three annual surveys which indicated that, on average, 75 LCCCs, 10 LCCHs, and 59 RCCHs per LWDA was sufficient to reach the prior goal of 35 independent rates per category.

These target numbers will be increased by between 7 percent and 12 percent for LWDA and facility type combinations that showed low response rates in recent surveys. The target number of facilities will then be randomly selected from each LWDA, or if fewer providers are available in any given combination of LWDA and facility type, the entire population will be selected (a complete census). The statewide sample and LWDA-based samples will then be combined. Implementation of this sampling scheme will result in all facilities of any type being selected from about a third of the LWDAs and a complete census of LCCHs or RCCHS in about half of the LWDAs.

### *Identify TWC Providers in the Sample*

Before the sample is finalized, a match will be performed between the list of sampled facilities and a database indicating which facilities are serving subsidized children, whose rates are already known because they have been agreed upon. Since rates are already known for these subsidized facilities, they will be randomly split into two groups, with one half of those serving subsidized children being retained in the sample and the other half set aside to be excluded from further study. The net effect of the increased target and removal of a fraction of subsidized facilities, will be a ten to twelve percent reduction in the total sample according to projections, but the survey sample itself will be more heavily inclusive of non-subsidized facilities. Another way of describing this is that we will under-sample subsidized facilities, whose rate data are largely redundant, and over-sample non-subsidized whose rates constitute the bulk of the paid market and are not otherwise available without surveying. Despite the reduced overall sample size for the survey, both subsidized and nonsubsidized rates will be estimated with substantially greater precision.

Before local rates are estimated, survey data from the statewide and local samples will be combined with administrative data representing the rates of all subsidized facilities who were not surveyed. The weights on the administrative records are fixed so that in the aggregate subsidized care carries the same weight as its share of care provided statewide. This adjustment is important, since without it the additional data would cause the resulting rates to more closely resemble subsidized than non-subsidized rates.

## **Survey Data Collection Standards and Process**

### **Data Collection Process**

After the sample is determined, surveys will be conducted by telephone interview between September 2017 and March 2018. The telephone interviews will last approximately 10 minutes and will be used for all three types of providers. Most child care facilities are extremely busy, and to both maintain a high response rate and to be responsive to the pressures on child care staff, a telephone interview is most convenient and least demanding method to obtain results. Before the telephone interview, however, a letter—in both English and Spanish or English and Vietnamese—will be sent to each provider explaining the survey goals and objectives, confidentiality policies, and the voluntary nature of participation. Most the interviews will be conducted in English, but the interview guides also will be translated into Spanish and Vietnamese. The survey questions may be sent to providers in advance of the interview if the provider requests them.

The survey will collect information on:

- percentage of children enrolled that are receiving TWC/Board-subsidized child care;
- enrollment numbers and the daily rates charged for full-day (six hours or more) and part-day (fewer than six hours) care for infants (0–17 months); toddlers (18–35 months); preschool age children (36–71 months); and school-age children (72 months and above).
- information about hours of operation;

- availability of care for children with disabilities;
- additional charges (registration/activity fees) and transportation fees;
- administrative category (e.g., sectarian/nonsectarian, community-based, for-profit/nonprofit);
- national accreditation or Texas Rising Star (TRS) provider status;
- data on payment practices among those who participate in the subsidy program; and
- barriers to participation for those who do not participate in the program.

Interviewers record information from the interviews on a printed copy of the survey instrument and enter the results in the CATI system after the interview is complete because providers often answer questions in a non-linear manner. The paper copy of the interview contains only the sample identification number and no other provider identifying information. After the paper copies of surveys are entered into the system, they will be stored in a locked file cabinet, in a locked office. The project managers and principal investigators have keys to both the file cabinet and office.

#### Data Collection Standards

Protocols for data collection have been developed to ensure accuracy of results throughout the project. During data collection, interviewers will input their completed surveys into the CATI system. Data will be stored in the system until it is extrapolated for analysis. During interviews, the project manager and project coordinator will listen to interviews at random and provide feedback to interviewers to ensure that interviews are conducted according to the interview guides. Any interviewer questions will be answered by the project coordinator who makes decisions regarding discrepancies or rare issues that sometimes arise.

All data will be stored on password and virus protected computers. The investigators and project manager will have access to identified data stored in the CATI system. Interviewers will only have access to interviews they are currently completing and therefore, do not have access to identifiable data after they have entered the information for their interview into the system. Therefore, only the project manager and project coordinator will have the ability to edit data that has already been entered. Hard copies of the surveys completed by the interviewers will be stored in a locked file cabinet where only project manager and principal investigators will have keys.

To check for data entry, the project coordinator will spot check 1% of the completed interviews each week. The project coordinator will randomly pull call sheets from files and will verify that the call sheet has been entered correctly. If the call sheet is not entered correctly, that interviewer will be given permission in the CATI system to go back and re-enter all their call sheets correctly.

After the data are collected, all rates that appear extremely high or low (known as outliers) will be identified using statistical procedures, and researchers will individually check each rate report and correct any errors they find. Subsequent to these corrections, a small portion of the most extreme outliers remaining in the sample will be corrected to remove their excessive influence on statistical measures of the rates distributions. In addition, similar procedures will be

implemented to detect instances in which the number of slots reported for a given rate was too extreme, whether too high or too low, and to correct these extremes to more reasonable values. In general, as one might expect in such a large and diverse state, extreme rates, whether at the low or high end of the scale, are typically the result of unusual services provided by an individual facility and/or variations in subsidies and financing, rather than research error.

#### Response Rates

Over the years of conducting the survey, the research team has developed ways of improving response rates by providing quality communication with child care providers. In 2016, 77% of center providers in the sample completed the survey and 52% of home providers completed the survey. These response rates are particularly high given providers are not compensated for participating. This is due to availability of detailed information about the survey online, interviewer experience and training; and follow up with providers who have not participated in the survey with additional letters explaining the survey and requesting participation, particularly in LWDAs with low numbers of facilities.

#### Response Bias

Prior to computing final rate estimates, analyses of potential response bias will be conducted to ensure the full range of facilities is represented in the findings. Characteristics of facilities that are known from their DFPS listings, including items like whether they offer transportation, weekend care, etc., will be utilized in a response model. Should significant response bias be found, corrections can be made by adjusting weights to account for the bias.

#### Analytic Methodology and Levels of Analysis

Project researchers will summarize the survey data and conduct analysis at several distinct levels. The most basic analyses will be done at the level of individual providers. These results document the proportion of providers responding to the survey that offer various services (e.g., non-traditional hours, transportation), or that fall into different administrative categories (e.g., sectarian, community-based, and for-profit). The exception to this rule is the measure of the “percent of enrollment that is subsidized,” which is analyzed at the level of the individual child. A similar item concerns the percentage of *facilities* that serve any subsidized children. When this number is higher than the “percent of enrollment” number, it indicates that the subsidized children in an area are widely dispersed among facilities, and when it is lower, it indicates that the subsidized children are relatively concentrated among a smaller number of facilities in the area.

The remaining analyses were conducted at either the rate observation level or at the level of the child care slot, as described below. Each center or home-based facility can contribute more than one rate observation to the analysis, and each rate observation can represent more than one slot. However, the nature of this relationship depends on the type of facility.

Daily market rates for licensed centers were gathered for all categories of care offered, regardless of whether any children were currently being served in such categories. The categories consisted

of all possible combinations of age groups (gathered for actual age categories in which each center offered rates, but aggregated to the four standard categories for reporting) by full-time status (part day or full day). Thus, one center could contribute as many as eight independent rate observations, each representing any number of children (including zero). These rate observations were then weighted by the number of child care slots they represent when calculating market rate percentiles.

In contrast, market rates for home-based facilities, including registered homes and licensed homes, were gathered at the level of the individual child currently being served. Data on each child's age, detailed weekly schedule, and rates charged were gathered for purposes of calculating daily market rates. In this case, the individual children were treated as independent rate observations, and each facility could contribute as many as nine observations (or fewer, depending on the facility type and number of children enrolled). Because of this, there was no need to differentially weight the rate observations when calculating the local market rate percentiles for home-based facilities: each child (or rate) received a weight equal to one.

For the analysis of center rates, the number of child care slots for each rate category was determined in one of two ways. First, for categories of care in which children were currently being served, the number of children in each category served as a proxy for the number of slots. Second, rate categories in which no children were currently being served were also included in the analysis, since they were also theoretically part of the market. This was done by estimating the number of slots for each of these rate categories with a formula that multiplies the number of children served at each facility by the average proportion of children, across all licensed centers, served in each rate category. Thus, for example, a facility that served 100 children and had an existing part-day infant rate schedule, but did not currently serve any part-day infants, would have its number of slots for this category of care estimated at two (100 child capacity X 2 percent of children served in the part-day infant category across all licensed centers). If the same facility served no part-day preschoolers, its number of slots would be estimated at seven (100 child capacity X 7 percent served in this category across all licensed centers). This method allows fuller use of the rate information gathered, especially for rare forms of care in which rate observations were otherwise scarce.

#### Statewide Estimation

For statistics conducted at various levels of analysis, special procedures were necessary in order to compute statewide versions of the tables. Due to the fact that smaller areas were over-sampled for this survey (meaning that facilities in these areas were more likely to be selected for the survey), a correction had to be applied in order to estimate statewide rates without over-representing facilities in these smaller areas. Similarly, it was also necessary to apply a correction because of over-sampling of Licensed Homes, which are much less common than centers or Registered Homes. The correction involves computing weights that, when applied to the statewide analysis, equalize the probability that any given facility or rate might have been selected at random.

The calculation of weights for statewide estimation varies slightly depending on the level of analysis: facility level or rate level (see preceding discussion). For facility-level analyses, the

weight is the inverse of the probability of selecting the facility at random. Thus, for small areas in which all facilities were selected, the weight would equal one. For larger areas in which, for example, 20 percent of facilities were selected, the weight would equal five (1 divided by 0.20). When these weights are applied to any given statistical analysis, the facility with a weight of five would count five times as much as the facility with a weight of one, and the result would be representative of the state as a whole. Similar procedures are used for the calculation of statewide rates, except in this case it is more complicated because each rate must be weighted according to its share of the statewide child care slots.

#### Small-Area Estimation

One of the most difficult features of the Texas MRS is the need to produce stable child care rate estimates for LWDAs that vary widely in size, in terms of number of facilities, with the largest (Gulf Coast) being larger than many states, containing more than then times as many child care facilities as the four smallest Boards combined. In many smaller, less populous LWDAs and for some exceptionally rare rate categories (particularly part-day care in homes) the number of rate observations would have been too small to estimate stable rates using traditional statistical methods, even for the many areas in which the entire universe of facilities had been sampled. For instance, a traditional approach would be to randomly sample the entire state and simply compute an average as well as the percentiles of interest for each area. However, such a method would require a minimum of 30 observations, per area to calculate an accurate rate, and one would have to sample nearly all facilities in the state to achieve this for most areas. Small areas that have less than 30 observations per rate present a challenge to accurate rate calculations. For this reason, an estimation methodology first developed for the 2010 survey was used. Using this approach, parameters are estimated that fully characterize the distribution of rates in each category of care. These parameters are then applied, using an assumption of normality of the rate distributions, to estimate each rate percentile of interest. The primary advantage of this solution is that a medium to large sample size of rates is utilized to estimate each parameter, thus the estimates based on this methodology are more stable over time, and they fall into the expected patterns more frequently.

Prior to any estimation of parameters, the daily rates themselves are first transformed to make them assume a more normal distribution. This transformation is done by taking the positive square root of the daily rate. The effect of this is to minimize the effect of high-end outliers on estimates of the parameters of distributions<sup>1</sup>. This transformation is reversed later, following all estimation, by squaring the rates.