



“Getting Ready to Succeed” Washington State Integrated Data System (WA-IDS) Early Childhood Data Review Report

Prepared for:

**Education Research
and Data Center**
Washington State Office
of Financial Management
PO Box 43113
Olympia, WA 98504-3113

Submitted by:

Abt Associates Inc.
4550 Montgomery
Avenue, Suite 800 North
Bethesda MD 20814



“Getting Ready to Succeed” –
Washington State Integrated Data System
Early Childhood Data Review Report

Table of Contents

Introduction	2
The Opportunity at Hand	3
The limitations of previous studies	3
The promise and opportunity in IDS	4
Washington State as an exemplar for the nation	4
Work completed under this grant to date	5
Gaps remain in available data	6
Moving forward: three proposed studies	8
Analytic Plan #1: ‘Patterns of Access’	9
Analytical Approach	9
Analytic Plan #2: ‘Patterns of Use’	14
Analytical Approach	15
Analytic Plan #3: ‘Patterns of Outcomes’	19
Analytical Approach	20
References	25

Introduction

This document proposes a set of studies examining the early care and education (ECE) experiences—and the concurrent health and well-being of—Hispanic families and children in the State of Washington. These proposed studies are presented to Washington’s Education Research & Data Center (ERDC) by Abt Associates in fulfillment of Task 2.4.3.6 under Contract No. K1899, under which Abt was contracted by ERDC to describe a set of proposed studies that could be executed with the current data, and/or specify what additional studies may be performed with additional data.

The document proceeds in four parts. The first part (‘The Opportunity at Hand’) presents a summary of the social and policy context for the proposed set of studies, a statement of the significant opportunity to advance public knowledge on this important and timely subject that we believe to be currently at hand, and a review of work completed to date under this project.

The second, third, and fourth parts present, in sequence, draft analytic plans for the three separate proposed studies. The first proposed study (‘Patterns of Access’) examines potential disparities in the degree to which low-income Hispanic families currently have access to adequate ECE experiences. The second (‘Patterns of Use’) concerns the degree to which these families ‘take up’ or utilize the publicly funded resources that are available to them within their communities. The third (‘Patterns of Outcomes’) concerns the degree to which various health services and ECE experiences help to address early risk factors and facilitate children’s optimal health and school readiness outcomes later in life.

We believe that any of the proposed studies would, if effectively executed, meaningfully advance public understanding of the complex web of relationships that exist among race/ethnicity, community characteristics, public benefit use, ECE utilization, and children’s subsequent health and well-being. Information from such studies would provide policymakers with new levers with which to advance the health and well-being of key target populations.

Suggested Citation: López, M. L., Watt, R., Litwok, D. & Grindal, T. (2017). “Getting Ready to Succeed”; Washington State Integrated Data System ((WA-IDS) Early Childhood Data Review Report. Education Research and Data Center, Washington State Office of Financial Management, Olympia, WA.

The Opportunity at Hand

Children from Hispanic families lag behind their white peers on numerous measures of educational success.ⁱ From performance on standardized achievement tests early in elementary school all the way up to high school graduation rates, disparities between Hispanics and whites have stubbornly persisted in Washington State and across the US from generation to generation, despite a range of policy interventions intended to close or lessen the gap.

Evidence suggests that these gaps emerge early, with Hispanic children less likely than white children to take early advantage of critical public services, such as quality preschool programs, that have been shown to support positive development early in life and even through into adulthood.ⁱⁱ These gaps continue once children begin school, with Hispanic students more likely to be retained in grade, receive special education services because of learning or behavioral disorders, and be subject to harsh disciplinary procedures such as suspension and expulsion.ⁱⁱⁱ

The limitations of previous studies

Policymakers' and practitioners' efforts to understand the root causes of these disparities have heretofore been limited in two important ways. First, prior research has typically examined children and families within a single service delivery context. That is, researchers have usually studied the experiences of families within (for example) ECE settings, the child welfare system, the range of individual public assistance systems, the education system, and/or the juvenile justice system separately. This approach, driven partially by a lack of access to quality integrated data systems, has had the unfortunate effect of creating a fragmented picture of Hispanic family life which ends up telling only part of the story.

A second limitation of prior work is that it typically uses the terms 'Hispanic' or 'Latino'¹ to encompass a rather diverse collection of families who vary substantially in terms of important demographic characteristics (e.g., country of origin, immigration status, educational level, language proficiency, etc.). Each of these demographic characteristics, if studied individually, is often differentially associated with families' needs and related outcomes. As a result, the oversimplification of racial/ethnic categories in studies may result in inaccurate findings, and/or mask important variations necessary to address the policy relevant research questions of interest. In contrast, a more comprehensive and integrated approach to examining data on the heterogeneous Hispanic population is likely to yield a more accurate understanding of the Hispanic population and ensures that public resources are maximally responsive in addressing the needs and outcomes of Hispanic children and their families.

¹ Although we use the terms Hispanic and Latino interchangeably, because federal data use *Hispanic*, we use that term when referring to estimates produced in federal data sets.

The promise and opportunity in IDS

Many state and federal agencies already collect a wealth of information regarding families' social, cultural, and economic characteristics, along with information on their participation in health and social service programs. Until recently, however, this information often was collected and analyzed separately by each specific program or service delivery agency. Moreover, practical barriers, including limitations in data sharing agreements across state agencies, have made it difficult or impossible for researchers to track the progress of individuals simultaneously, over time, and across different service delivery programs, and thus to take advantage of the range of demographic and program information collected by individual agencies.

The recently-developed Washington State Integrated Data System (WA IDS) allows users to link individual-level K-12, preschool, and higher education databases cross-sectionally and longitudinally with previously discrete data sets from health, wage, benefit, and subsidy agencies. This WA IDS affords users an exciting new resource to address long-standing challenges, as well as a crucial opportunity to break new ground in the study of social conditions, educational progress, and critical outcomes in early childhood among Hispanic families in the State of Washington. The studies proposed herein are designed to make best use of this valuable new resource to address important, policy-relevant questions regarding the health and well-being of children and families in the State of Washington.

Washington State as an exemplar for the nation

As it happens, Washington provides a rich and dynamic test case for advancing public understanding of the early educational achievement, health and well-being of the diverse and rapidly growing U.S. Hispanic population. The state has a large and diverse Hispanic community, with substantial variability in geographic location (urban/rural), nativity, and age. In 2016, Washington had the 15th-largest Hispanic population in the country, at 12%.^{iv} The percentage of Hispanic residents varies by location, ranging from 20% in some areas to upwards of 80% in others. In 2016, 20% of all students served by Washington public schools identified as Hispanic. And the Hispanic population is not limited to the coastal cities: In Yakima County, a rural county in eastern Washington, 48% of the local overall population was of Hispanic origin in 2016.

The rate of growth, relative youth, and demographic diversity of the Hispanic population highlights the relevance of Washington State to stakeholders interested in better understanding the needs of Hispanic children and families nationwide. As in other parts of the U.S., the academic achievement of Washington's Hispanic children relative to white students has remained unacceptably low. According to several recent ERDC reports, Hispanic students score significantly below white and Asian students on every academic measure tested in Washington State. This achievement gap begins in the early grades and persists over time.^v Just 32.3% of Hispanic students, for example, met standards in all six domains of school readiness on the WA Kids assessment, compared to 53.3% of their white peers.^{vi} Meanwhile, at the other end of the

educational spectrum, on-time graduation rates among Hispanic students in 2016 were 72.3%, compared to 81.5% for their white peers, and just 70.1% for low-income Hispanic students.^{vii}

The studies proposed in parts two, three, and four of this document seek to utilize integrated data systems to answer important questions about racial/ethnic disparities in early childhood education access, with a particular focus on the growing Hispanic population in Washington State. These proposed studies highlight the potential of the WA IDS system to allow researchers and practitioners to effectively answer practical, policy-relevant questions previously beyond our grasp.

Work completed under this grant to date

In 2016, Washington’s ERDC contracted with Abt Associates to conduct a review of the data presently available for use in studies of the type proposed here, and gaps therein, to summarize the results of that review, and to propose steps needed to fill any gaps identified in the process. In the service of that goal, ERDC provided our team with ten data codebooks relating to information data sources that are potentially linkable using the ERDC WA IDS, and likely available for use in the proposed studies. These codebooks are summarized in Table 1.

Table 1 – The Abt team reviewed nine separate codebooks

1	Washington All Payer Health Care Claims Database (APCD)
2	Court Contact and Recidivism Database (CCRD) Codebook
3	Comprehensive Education Data and Research System (CEDARS) Data Manual
4	Department of Early Learning (DEL) Early Learning Management System (ELMS) Data Dictionary
5	Department of Early Learning Early Support for Infants and Toddlers (ESIT) Codebook
6	Department of Early Learning (DEL) Subsidy Data Codebook
7	Washington Center for Health Statistics (CHS) Birth Data Dictionary
8	Washington Kindergarten Inventory of Developing Skills (WaKIDS) Fields List
9	Washington State Board for Community and Technical Colleges (SBCTC) Data Dictionary
10	Working Connections Child Care (WCCC) Licensed Centers Codebook

Our team reviewed these codebooks for information on 30 separate key data items (e.g., ‘Child Hispanic Status,’ ‘Child Language,’ or ‘Center-Based Preschool Participation’) necessary to complete future analyses of potential interest. In this way, we were able to determine (for example) that ‘EthnicityCode,’ a variable listed in the **CEDARS Data Manual**, was likely to be a reasonably good proxy for our item ‘Child Hispanic Status.’ Similarly, ‘RaceType’ (a variable listed in the **ELMS Data Dictionary**) likely represents a good proxy for ‘Child Racial Category.’

Although we linked each data source variable to only one of our 30 key data items for the proposed studies, some items were linked to more than one data source variable. For example, although we thought that **CEDARS’** ‘EthnicityCode’ was likely to be a good proxy for ‘Child Hispanic Status,’ the same was true of ‘EthnicityType,’ found in the **ELMS Data Dictionary**, and of ‘Hispanic,’ found in

the **DEL Subsidy data**. In such cases, we proposed that the final analytic data file contain all of the data source variables, and that the eventual research team use the multiple records as checks in later analyses, thus improving the convergence and fidelity of the data that is finally used.

ERDC reviewed the Abt team’s preliminary assessment of data gaps and provided additional information regarding child kindergarten entry and 3rd grade test scores, center-based preschool participation, and child care subsidy participation. A list of confirmed source items was provided to ERDC in January of 2017.² In addition to data source variable and item names (which were listed in the first two columns of the table), this list also contains a record of the variable’s type, a brief description of the variable, the variable’s units, if known, whether the variable is time invariant or not, the variable’s observation point, and which codebook we found the variable in. Although some data was later sent to Abt (specified in footnote 2, below), this table represented Abt’s best understanding of data was readily available to ERDC going forward.

Gaps remain in available data

Based upon our review of the data, we believe that ERDC already has sufficient data in hand to complete a great deal of productive and valuable research (we describe, for example, one study we believe to be wholly possible at present in detail in Analysis Plan #1). We also describe two remaining key data gaps that, if filled, would significantly advance the team’s capacity to conduct further path breaking research.

Low-income families. The first data gap concerns a source of information on the population of children born in Washington to low-income families. Although it would be ideal to use individual-level state **Medicaid participation data** to identify children from low-income families at the earliest point in time, ERDC have indicated that it is unclear whether it will be possible to access these data for the proposed studies. If state **Medicaid participation data** is not available for use in these studies, we propose two alternative approaches:

Option 1: WIC participation data—currently available as part of the **Birth Data Dictionary**—might serve as a proxy for household income and residence in the state of Washington in the birth to age three period prior to the possibility of participation in publicly funded ECE at either age three or four. Although Washington’s WIC program eligibility is restricted to those families whose household income falls at or below 185% of the federal poverty line (compared to Washington Apple Health [Medicaid]’s 138% threshold) these records would provide an indicator low income status, as well as residence in Washington in the years prior to preschool and/or kindergarten entry. A larger concern is that the utilization rate for WIC (approximately 70%^{viii}) among families with eligible infants is lower than the take up rate for Medicaid (approximately 93%^{ix}). Thus, using WIC data to limit the sample to low-income families would exclude a larger number of eligible children from the analyses, than would be the case if Medicaid data was used.

² Please note that this list did not include ESIT data, access to which has not yet been confirmed by ERDC at the time of completion, or APCD data, which was sent to Abt in early September 2017.

Option 2: Secondly, we might be able to use a combination of individual-level household residence and community-level demographic data sources to statistically estimate the probability that a family is low-income. Such an estimate might draw upon the household ZIP code (which we have from four different codebooks) and neighborhood demographic data (drawn from the **American Community Survey (ACS)**) to predict the likelihood that a child's household falls below a targeted income threshold. Although this method would be less robust than using actual, individual-level data, there is some evidence that it can be workable, if appropriately qualified and handled with caution.^x

Head Start. Across all the sources of data above, many of the children participating in different types of publicly funded ECE options will be captured and available for use in analytic samples. However, the one key source that is currently missing is data on children served by Head Start programs. Data from the National Institute of Early Education Research (NIEER) indicate that approximately 5% of three-year-old children and 7% of four-year-old children (out of approximately 180,000 three and four year olds in total) in Washington were enrolled in Head Start (HS) programs in 2016.^{xi} Separately, the 2015 Head Start Program Information Report indicates that approximately 14,000 three- to five-year-old children were served by Regional Head Start, Migrant and Seasonal Head Start (MSHS) and the American Indian and Native Alaskan (AIAN) Head Start programs in that year. Unfortunately, while Head Start programs already collect information on the total aggregated number of enrolled children per grantee, they do not yet share individual child-level data with ERDC. Thus, not having Head Start data would exclude a modest number of the total number of publicly funded ECE slots, thereby reducing the generalizability of the findings on ECE utilization by Hispanic families.

Over the first quarter of 2017, we have worked closely with ERDC to help identify and overcome previous barriers regarding Head Start grantees' willingness to share their individual-level data with ERDC. Washington HS grantees have previously been concerned that sharing their student-level data might put them out of compliance with the Federal Office of Head Start. This concern seems to have been linked to some unclear legislative language contained within the 2007 Head Start Act. However, more recent guidance, from the recently released Head Start Performance Standards, clarifies this issue and actually encourages HS grantees to work with appropriate state agencies to set up data sharing agreements that would allow them to share their data.

Given that the individual-level data would need to come from each separate grantee, and at the grantees' own initiative, we will continue to support ERDC's efforts to develop collaborative relationships with the various Head Start grantees in order to get access to as much of this data as possible. To this end, we have helped to facilitate relationships among the WA Department of Early Learning (DEL), the WA Head Start Association, and the ACF Regional Office staff in the Region 10 office in Seattle, as well as ACF Central Office staff, the National Migrant and Seasonal Head Start Association (NMSHSA), and the National Head Start Association.

Moving forward: three proposed studies

The next three parts of this document present, in turn, three separate proposed studies:

- (1) “Patterns of Access” examines variations in the availability of ECE programs and/or potential disparities in ECE supply across different geographic areas or communities, for Hispanics versus non-Hispanics;
- (2) “Patterns of Use” examines variations in ECE and public assistance utilization patterns from birth to kindergarten entry for Hispanics versus non-Hispanics;
- (3) “Patterns of Outcomes” examines variations in developmental trajectories, risk profiles, and health service utilization from birth to kindergarten entry for Hispanics versus non-Hispanics.

We believe that the proposed “Patterns of Access” study could be effectively executed with the data ERDC currently has access to, whereas the “Patterns of Use” would be possible with the addition of Medicaid and Head Start data, and the “Patterns of Outcomes” would be possible with expanded access to health data, in addition to all other data previously discussed. For each of the three research studies in what follows we have categorized research questions into three groups: (1) a description of the Hispanic population; (2) a comparison of the Hispanic and non-Hispanic population; and (3) other conditional relationships between an outcome and covariates estimated with a regression model.

Analytic Plan #1: ‘Patterns of Access’

This section presents a draft analytic plan for the proposed “Patterns of Access” study, which would examine variations in the availability of ECE programs for preschool age children and/or potential disparities in ECE supply across different geographic areas or communities within the state of Washington, for Hispanics versus non-Hispanics. This proposed study builds upon work conducted with respect to disparities in the availability of high-quality child care options by the Center for American Progress in 2016 across eight states, and related work on “child care deserts” conducted by Child Care Aware in the same year.^{xii} Although understanding patterns of access to high-quality childcare is far from a panacea, it is an important first step in identifying and limiting potential disparities in ECE access along socio-economic or racial/ethnic lines.

The “Patterns of Access” study is designed to answer the following policy-relevant research questions:

1. What percentage of incoming Hispanic kindergartners in the state of Washington live within a ZIP code characterized as an “ECE desert” (geographic areas which do not contain an adequate supply of high-quality ECE options)?
2. Are Hispanic kindergartners in Washington more likely to live within a ZIP code characterized as an ECE desert than non-Hispanic kindergartners?
3. Are there within-group differences for Hispanics in the likelihood of living in a ZIP code characterized as an ECE desert, based on child and family characteristics (e.g., parental nativity status, home language, etc.)?

In contrast to our second and third proposed analytic plans (‘Patterns of Use’ and ‘Patterns of Outcomes’) we believe that it is possible to effectively execute a rigorous and policy-relevant study of the research questions above using only administrative data already available to ERDC. Below we outline the analytic plan for the proposed study.

Analytical Approach

The proposed approach would start by identifying a cohort of children entering kindergarten in the public schools and then link to data from several other sources of individual-level data, such as the **Birth Data Dictionary** and **Medicaid** data. The DEL Licensed care provider data and/or **Child Care Resource and Referral (CCR&R) data** will then be used to identify the universe of licensed ECE programs and providers, along with data from the **American Community Survey (ACS)** for geographic mapping of the location of ECE programs and providers in the state of Washington. Together, these various sources of data would help to paint a picture of potential differences in access to ECE among low-income Hispanics in the State of Washington.

1 – Identify the Analytic Sample of children

Data from the **Comprehensive Education Data and Research System (CEDARS)** will be used to identify a cohort of students enrolled in kindergarten in Washington State during a particular academic year (specific year to be determined).

2 – Identify race/ethnicity of children

With an initial analytic sample of children in hand, children who were categorized as Hispanic in the State of Washington **Birth Data Dictionary** will be distinguished from those who were not.³ The following variables might be used for this task:

Table 1 – Multiple Hispanic variables are available in the Birth Data Dictionary

hispcdes	A categorical variable which indicates the child's Hispanic status as designated by mother at birth.
hispcal	A binary variable which indicates the child's calculated Hispanic status.
racecdes	A categorical variable which indicates the child's race as designated by mother at birth.

For children whose records do not appear on the **Birth Data Dictionary** (typically, children who were not born in Washington, or more rarely, children who were not born in a hospital), race and ethnicity characteristics will be drawn from **CEDARS** records. It is unclear whether these children should be included in final analysis, as their presence in Washington is time variant. However, some children in this category may have moved to Washington before aging into pre-kindergarten eligibility, and would likely be retained.

3—Describe parents’ nativity status

The following variables (and associated data sources) could be used for this subtask:

Table 2 – Nativity variables from the Birth Data Dictionary

Label	Name	Source
Mother’s Birthplace	bctrymom	Birth Data Dictionary
	birplmom	Birth Data Dictionary
Father’s Birthplace	bctrydad	Birth Data Dictionary
	birpldad	Birth Data Dictionary

4—Impute subgroup characteristics

³ For further discussion of our initial work on this and other data sources, please see part 1 of this document (‘The Opportunity at Hand’), section 5 (‘Work completed under this grant to date’).

In order to answer the fourth research question, attributes—notably, a best estimate of the children’s families’ income status—will be associated with the children’s home geographies using geocoded data on Washington kindergarteners’ home addresses from the **CEDARS** data manual. The level of geographic resolution would likely be at the nine-digit ZIP level. Thereafter, **American Community Survey (ACS)** data will be used to:

- 1) Characterize neighborhoods. Neighborhood attributes relevant to later sections of the proposed study—including the average racial/ethnic composition, and rates of publicly-funded social service use—will be associated to individual children and their families on the basis of the average values of attributes of the ZIP codes in which they and their families reside.
- 2) Impute attributes of families in the study sample.⁴ There is some evidence that it is possible to use census data to estimate poverty level and inequality for small geographic areas, and then impute those attributes to families living in those areas—a technique which has already been applied in a prior study of Hispanic ECE utilization in Miami.^{xiii} In the absence of a good proxy for individual-level family income status, like **Medicaid** data, these techniques will be applied to generate estimates of family income level for the kindergartners still in the analytic sample, and then limit that sample further to families below a certain income level.

5—Define ECE Deserts within Geographic Areas

CCR&R data and **DEL licensed care data** will be used to identify the universe of licensed ECE providers in Washington state, and build upon the model developed by Malik et al.^{xiv} to characterize each ZIP code in Washington state as an ECE desert (or not). Malik et al. characterized a ZIP code as a “desert” if there were at least 30 children under the age of 5 within the ZIP code and no child care center (in our case, ECE provider), or if there are at least 30 children and more than one ECE provider, but the ratio of children to cumulative capacity is greater than 3:1. This study initially replicates, but then also expands upon that model in two ways. First, by introducing key measures of child care quality in addition to capacity ratios:

- Whether or not a provider participates in the WA Early Achievers (EA) Quality Rating and Improvement (QRIS) program; and
- Whether or not a provider operates on non-traditional hours, which are important to low-income families where parents may be more likely to have non-traditional work hours and/or irregular work schedules.

⁴ As previously discussed in part 1 of this document (‘The Opportunity at Hand’), section 6 (‘Gaps remain in available data’).

This study will also incorporate geographic mapping techniques developed by the Advancement Project (and related to similar techniques used by the Illinois Early Childhood Asset Map Project) to accurately capture the degree to which ECE supply in Washington state meets the needs of the population in different geographic areas.^{xv} At the end of this process, each ZIP code in Washington state will have a ratio of the number of licensed child care slots (supply) relative to the total estimated number of eligible young children (relative demand). This information will be used to characterize each zip code as either an ECE desert—in which sufficient ECE options are not available—or not.

6—Analyze and Report

With these data in hand, we can address the policy-relevant research questions listed above as follows:

Descriptive analyses of the Hispanic population:

- 1) *What percentage of incoming Hispanic kindergartners in the state of Washington live within a ZIP code characterized as an “ECE desert”?*

The analyses would identify a cohort of incoming Hispanic kindergartners enrolled in the public schools and tabulate the proportion living within a ZIP code characterized as an ECE desert.

Analyses comparing the Hispanic vs. non-Hispanic population:

- 2) *Are Hispanic kindergartners in Washington more likely to live within a ZIP code characterized as an ECE desert than non-Hispanic kindergartners?*

This research question can be answered in two ways. One approach is to directly compare the unadjusted proportion of incoming Hispanic kindergartners who are living within a ZIP code characterized as an ECE desert, to the unadjusted proportion of incoming non-Hispanic kindergartners.

Alternatively, we can model the probability of living within a ZIP code characterized as an ECE desert as a function of observable characteristics. This modeling strategy allows us to test whether Hispanic kindergartners are more likely than non-Hispanic kindergartners to live within a ZIP code characterized as an ECE desert, *conditional upon the other covariates included in the regression model*. A linear regression model can use an indicator for living in a ZIP code characterized as an ECE desert as the outcome variable.

The outcome is modeled as a function of descriptive covariates, which may include variables such as age, gender, density of Hispanics living within the community, Spanish spoken in the home, a measure of socioeconomic status, and an indicator variable that equals 1 if the child is Hispanic and equals 0 if the child is not Hispanic. Tests of statistical significance for the coefficient on the Hispanic indicator variable indicate whether the

conditional difference in the probability of living in an ECE desert between Hispanic and non-Hispanic children is statistically significant.

Analyses of within group differences among Hispanics:

- 3) *Are there within-group differences for Hispanics in the likelihood of living in a ZIP code characterized as an ECE desert, based on child and family characteristics (e.g., parental nativity status, home language, etc.)?*

To understand whether there are within-group differences in the likelihood of living in a ZIP code characterized as an ECE desert, we can add additional covariates to the regression model described in the outline for Research Question 2 above. Specifically, we can add an interaction between the variable that indicates whether or not the child is Hispanic and the indicator variable that describes the particular child or family characteristic we are investigating.

When this additional variable is added to the regression model, the coefficient of the interaction term reflects a comparison of the average likelihood for Hispanic kindergartners who have a particular characteristic to Hispanic kindergartners who do not have that characteristic. Tests of statistical significance of this coefficient indicate whether the difference between these two groups is statistically significant.

Analytic Plan #2: ‘Patterns of Use’

This section presents an analytic plan for the proposed “Patterns of Use” study which examines patterns of publicly-funded social services participation from birth to age three, as well as subsequent patterns of ECE utilization at age three and/or four, for Hispanic children in the State of Washington. More specifically, the proposed study is intended to examine whether the timing and/or pattern of early use of publicly funded social services for families with children between birth and age three, predicts subsequent ECE utilization at age three and/or four.

This policy relevant study may help to improve the knowledge base related to the historical patterns of underutilization of critical social services and ECE by Hispanics and other low-income families and likely is of interest to early education policymakers and practitioners in Washington State and beyond. Although we believe that it will be possible to effectively execute a modified version this plan given access only to the data currently available to ERDC, the addition of state **Medicaid** and **Head Start** data would greatly improve the quality and comprehensiveness of the proposed study.

The “Patterns of Use” study is designed to answer the following policy-relevant research questions:

Descriptive questions regarding Hispanic service use:

1. Among a cohort of low-income, Hispanic kindergartners in the Washington public schools, what were their families’ utilization patterns of publicly-funded social services (e.g., TANF, SNAP, WIC, etc.) during the child’s first three years of life?
2. Within the subgroup of Hispanics, are there within group differences in publicly-funded social service usage patterns over the first three years of life, based on child, family, and community characteristics (e.g., parental nativity status, home language, percent Hispanic density within the community, etc.)?
3. Are there differences in utilization patterns of publicly-funded social services for Hispanics over the first three years of life by type of subsequent publicly-funded ECE use (e.g., ECEAP, Head Start, WCCC, or family child care/center-based preschool using child care subsidies) in the two years prior to kindergarten entry (at ages 3 & 4)?
4. What percentage of low-income Hispanic kindergartners in Washington were enrolled in one or more publicly-funded ECE programs (e.g., ECEAP, Head Start, WCCC, or family child care/center-based preschool using child care subsidies) in the two years prior to kindergarten entry?

Comparisons between Hispanic and non-Hispanic service use:

5. Are there differences in publicly-funded social service usage patterns over the first three years of life between low-income Hispanic versus non-Hispanic children?

Predicted service use patterns:

6. Among low-income, Hispanic kindergartners in Washington, does prior publicly-funded social service use from birth to age three predict the use of publicly-funded ECE in the two years prior to kindergarten?
7. Are differences in ECE usage patterns over the first three years of life associated with the availability and accessibility of publicly-funded ECE services (e.g., child care supply, service density, service accessibility) and/or publicly-funded social services within different neighborhoods or communities?

Analytical Approach

The overall proposed approach would start by identifying a cohort of children enrolled in kindergarten in the public schools, and then link to a range of demographics, public assistance (TANF, SNAP & WIC), ECE, and other data on the characteristics and location of the communities where families live.

1 – Identify the Analytic Sample of Children

CEDARS data will be used to identify a cohort of students enrolled in kindergarten in Washington State during a particular academic year (specific year to be determined).

2 – Identify best approach for defining a low-income subsample

If available, state **Medicaid participation data** will be used to determine individual-level family income for children in the sample. If **Medicaid data** is not available, the alternative approach to calculate the best estimate of the children’s families’ income status is to use children’s home addresses in combination with aggregated community-level geographic and demographic variables available from the ACS data, in much the same way proposed in Analytic Plan #1.

3 – Identify race/ethnicity of children

With an initial analytic sample of families in hand, **Birth Data Dictionary** data will be used to distinguish those children who are Hispanic vs. non-Hispanic. The following variables might be used for this task:

Table 2 – Multiple Hispanic variables are available in the Birth Data Dictionary

hispcdes	A categorical variable which indicates the child's Hispanic status as designated by mother at birth.
hispcal	A binary variable which indicates the child's calculated Hispanic status.
racecdes	A categorical variable which indicates the child's race as designated by mother at birth.

For children whose records do not appear on the **Birth Data Dictionary** (typically, children who were not born in Washington or, more rarely, children who were not born in a hospital), **CEDARS data** will be used to determine race and ethnicity. It is unclear whether these children will be included in final analysis, as their presence in Washington is time variant. However, some children in this category may have moved to Washington before aging into pre-kindergarten eligibility, and will likely be retained.

4 – Describe how the population of Hispanic children in our sample differs in terms of demographics, health, and neighborhood attributes from non-Hispanic children

The following variables (and associated data sources) may be used for this subtask:

Table 3 – A range of variables are already available to our researchers

Label	Name	Source
Child Birth Country	BirthCountry	CEDARS
Child Disability Status	ChildIEP	DEL ELMS
	DISABILITY	SBCTC
	DisabilityType	DEL ELMS
	QualificationCode	CEDARS
Child Health at Birth	wt_grams	Birth Data Dictionary
	apgar10	Birth Data Dictionary
Child Language	LanguageType	DEL ELMS
	PrimaryLanguageCode	CEDARS
	WaK_Bilingual	WaKids
Father’s Birthplace	bctrydad	Birth Data Dictionary
	birpldad	Birth Data Dictionary
Father’s Education Level	educ_dad	Birth Data Dictionary
Mother’s Birthplace	bctrymom	Birth Data Dictionary
	birplmom	Birth Data Dictionary
Mother’s Education Level	educ_mom	Birth Data Dictionary
Parental Language Status	FamilyMemberLanguage	DEL ELMS
Parental Marital Status	married	Birth Data Dictionary
Primary Home Language	LanguageSpokenAtHome	CEDARS

We will also use existing variables to identify patterns of enrollment of children in our sample in a range of publicly-funded social service programs, including AFDC, WIC, and possibly SNAP and TANF, pending data availability.

5 – Describe, cross-tabulate, and analyze the use of publicly-funded ECE programs in Washington State with the children in our analytic sample.

With these data in hand, we can address the policy-relevant research questions above as follows:

Descriptive questions regarding Hispanic service use:

1. *Among a cohort of low-income, Hispanic kindergartners in the Washington public schools, what were their families' utilization patterns of publicly-funded social services (e.g., TANF, SNAP, WIC, etc.) during the child's first three years of life?*
2. *Within the subgroup of Hispanics, are there within group differences in publicly-funded social service usage patterns over the first three years of life, based on child, family, and community characteristics (e.g., parental nativity status, home language, percent Hispanic density within the community, etc.)?*
3. *Are there differences in utilization patterns of publicly-funded social services for Hispanics over the first three years of life by type of subsequent publicly-funded ECE use (e.g., ECEAP, Head Start, WCCC, or family child care/center-based preschool using child care subsidies) in the two years prior to kindergarten entry (at ages 3 & 4)?*
4. *What percentage of low-income Hispanic kindergartners in Washington were enrolled in one or more publicly-funded ECE programs (e.g., ECEAP, Head Start, WCCC, or family child care/center-based preschool using child care subsidies) in the two years prior to kindergarten entry?*

The first four descriptive questions can be answered using tabulations of the data. In some cases these tabulations will be stratified by other characteristics, such as social services used or type of ECE program.

Comparisons between Hispanic and non-Hispanic service use:

5. *Are there differences in publicly-funded social service usage patterns over the first three years of life between low-income Hispanic versus non-Hispanic children?*

As in Analytic Plan #1, the comparison of Hispanic to non-Hispanic children can be made in using either unadjusted data (i.e., a comparison of means) or adjusted data (i.e., a regression model). A regression modeling strategy would allow testing of whether Hispanic kindergartners differ from non-Hispanic kindergartners in their use of publicly funded social services *conditional on the other covariates included in the regression model*. A linear regression model can use an indicator for using a publicly funded social service as the outcome variable.

The outcome is modeled as a function of descriptive covariates, such as age, sex, density of Hispanics living within the community, Spanish spoken in the home, a measure of socioeconomic status, and an indicator variable that equals 1 if the child is Hispanic and equals 0 if the child is not Hispanic. Tests of statistical significance for the coefficient on the Hispanic indicator variable indicate whether the conditional difference in the probability of use between Hispanic and non-Hispanic students is statistically significant.

Predicted service use patterns:

6. *Among low-income, Hispanic kindergartners in Washington, does prior publicly-funded social service use from birth to age three predict the use of publicly-funded ECE in the two years prior to kindergarten?*
7. *Are differences in ECE usage patterns over the first three years of life associated with the availability and accessibility of publicly-funded ECE services (e.g., child care supply, service density, service accessibility) and/or publicly-funded social services within different neighborhoods or communities?*

Research questions six and seven can be addressed using regression models with interaction terms. For example, to understand whether prior publicly-funded social service use from birth to age three predicts ECE use in the two years prior to kindergarten among low-income Hispanic children we can regress an indicator for ECE use on the same series of covariates described above, but add an interaction between an indicator for prior use of a publicly-funded social service and an indicator for whether the child is Hispanic and low-income.

When this additional variable is added to the regression model, the coefficient of the interaction term reflects a comparison of the average likelihood for low-income Hispanic kindergartners who used publicly-funded social services to low-income Hispanic kindergartners who did not use publicly-funded social services. Tests of statistical significance of this coefficient indicate whether the difference between these two groups is statistically significant. A similar analytic model can be used to estimate the relationship between prior ECE usage patterns and current use of ECE programs or publicly funded social services.

Analytic Plan #3: ‘Patterns of Outcomes’

This section presents an analytic plan for the proposed “Patterns of Outcomes” study, which would examine the complex set of relationships among several key areas of interest—early risk factors at birth, receipt of early intervention and related health services during the first five years of life, and subsequent kindergarten readiness and health and well-being outcomes at kindergarten entry—that, taken together, fully leverage the unique power of the IDS to provide a complete picture of children’s experiences in the years between birth and kindergarten. Whereas the previous two studies described in this document will clearly advance understanding of more established fields of inquiry (ECE access and use), this study’s results are more exploratory, but provide an innovative linking of early health and risk status variables with subsequent examination of the potential role of early intervention services in helping to reduce the need for later services and improve longer term health, well-being and academic outcomes into the early elementary school years.

The proposed “Patterns of Outcomes” study is designed to answer the following policy-relevant research questions:

Descriptive questions regarding patterns of risk factors at birth and health service use for Hispanic children:

1. Among a cohort of low-income Hispanic kindergartners in the State of Washington, what are the patterns of risk factors at birth?
2. Within the Hispanic subgroup, are there within-group differences in the patterns of use of health services and ECE programs over the first four years of life among children based on child, family, and community characteristics?

Comparisons between Hispanic and non-Hispanics on risk factors at birth and health service use:

3. Among low-income kindergartners in Washington, are there differences in the risk factors at birth between Hispanic and non-Hispanic children?
4. Among low-income kindergartners in Washington, are there differences for Hispanics versus non-Hispanics in the patterns of use of health services, early identification services, home visiting programs, and ECE programs, from birth to kindergarten entry?

Predicted relationship between health services and later health and well-being outcomes:

5. Among low-income kindergartners in Washington, does utilization of health services, early intervention services, home visiting programs, and/or ECE programs predict positive health outcomes, lower levels of behavioral problems, and a reduced need for special education services at kindergarten entry? Do these relationships vary for Hispanics and non-Hispanics?

6. Among low-income kindergartners in Washington, do prior use of health services, early identification services, home visiting programs, and ECE programs from birth to kindergarten entry predict school readiness outcomes once the child arrives in kindergarten, as well as academic achievement outcomes at the end of third grade? Do these relationships vary for Hispanics and non-Hispanics?

In order to execute this plan effectively, it will be necessary to secure a more robust source of child health and early intervention service data throughout the first five years of life—including health services, early identification efforts, and home visiting data—than ERDC currently has on hand, as well as full access to **WaKids** and **CEDARS** data. If executed, however, we believe the study would help to capture the degree to which early intervention and access to health services in the first five years of life might help to reduce the prevalence of identified problems and costly special education services once the child arrives in kindergarten. Below we outline the stages and algorithm of tasks that we propose.

Analytical Approach

The overall proposed approach for the Patterns of Outcomes study would start by identifying a cohort of children entering kindergarten in the public schools via the **CEDARS** data set and then link that data to health and early intervention services, as well as ECE program use.

1 – Identify the Analytic Sample of Children

Using the **CEDARS** data, a cohort of children in kindergarten in Washington State during a particular academic year (specific year to be determined) will be identified.

2 – Identify race/ethnicity and family income status of children

With an initial analytic sample of children in hand, those children who were categorized as Hispanic in the State of Washington **Birth Data Dictionary** will be distinguished from those who were not. The following variables might be used for this task:

Table 1 – Multiple Hispanic variables are available in the Birth Data Dictionary

hispcdes	A categorical variable which indicates the child's Hispanic status as designated by mother at birth.
hispcal	A binary variable which indicates the child's calculated Hispanic status.
racecdes	A categorical variable which indicates the child's race as designated by mother at birth.

For children whose records do not appear on the **Birth Data Dictionary** (typically, children who were not born in Washington or, more rarely, children who were not born in a hospital), race and ethnicity attributes will be drawn from **CEDARS data**. It is unclear at whether these children will be

included in final analysis, as their presence in Washington is time variant. However, some children in this category may have moved to Washington before aging into pre-kindergarten eligibility, and will likely be retained.

As before, in the event state **Medicaid** participation data is available, that data source will be used to characterize family income. If **Medicaid** data is not available, attributes—notably, our best estimate of the children’s families’ income status—will be associated with the children’s home geographies using geocoded data on Washington kindergarteners’ home addresses, in much the same way proposed in Analytic Plan #1 and #2.

3 – Link Health Data (intervention and outcome)

The WA IDS does not currently have full access to data sets that include variables on receipt of health interventions for children age 0-5, beyond (potential) access to the **Early Support for Infants and Toddlers (ESIT)** and **Washington All Payer Health Care Claims Database (APCD)** data sources, and possibly some data on home visiting services and/or Electronic Health Records (EHR) capturing data on the receipt of health services, which may yet become available. However, the **CEDARS** data manual will allow users to develop a good sense of the degree to which children in the analytic sample are receiving special education services during the kindergarten years. **CEDARS data** also allows, in some limited cases, a sense of whether a child received special interventions or treatment before kindergarten entry.

Table 2 – The CEDARS data manual includes a number of key health data

Early Childhood Home Setting (aged 3-5)	Report those special education students (aged 3 through 5) who received the majority of their special education and related services in the principal residence of the child's family or caregivers, and who did not attend a regular early childhood program or a special education program provided in a separate class, separate school, or residential facility.
Early Childhood Home Residential Facility (aged 3-5)	Report those special education students (aged 3 through 5) who received education programs in publicly or privately operated residential schools or residential medical facilities on an inpatient basis (do not include children who also attended a regular early childhood program).
Early Childhood Separate School (aged 3-5)	Report those special education students (aged 3 through 5) who received education programs in public or private day schools designed specifically for children with disabilities (do not include children who also attended a regular early childhood program).

ESIT data, if available, would provide an excellent source of information about the differing conditions which children might be diagnosed with in the years between age 0 and 5 (including asthma, autism, diabetes, epilepsy, nutritional deficiency, and visual impairment), and would be a

robust source of variables for this proposed study. Similarly, access to additional data on **home visiting** services would significantly improve the scale and scope of the proposed study, as would full access to the **APCD** database briefly described above.

As with the school readiness outcome variables—whatever health data becomes available will be linked to the initial analytic sample in order to run descriptive reports and conduct the analyses proposed below in stage 5 of this analytic plan.

4 – Link school readiness (outcome) data

Using extant data on school readiness and growth data contained in the **CEDARS** data manual and the **WaKids** field list, each study child’s school readiness and growth will be assessed on a number of key dimensions, including academic competence, socio-emotional development, and special needs status. A number of key variables already exist for this purpose in both data sources, including 75 different data fields in the **WaKids** manual and a range of progress monitoring assessments in the **CEDARS** data manual. It may also be possible to link summative performance data on academic achievement gathered at the end of third grade to these data in order to gain a more complete picture of longer term outcomes a few years beyond kindergarten.

Taken together, these various measures can be used to paint a robust picture of each individual child’s school readiness and academic achievement growth upon kindergarten entry, and at the end of third grade. Using the robust power of the IDS, we will be able to link these data to our initial analytic sample, as well as (if desired) other child characteristics already discussed in prior analytic plans.

5 – Analysis

With these data in hand, we can address the policy-relevant research questions above as follows:

Descriptive questions regarding patterns of risk factors at birth and health service use:

- 1. Among a cohort of low-income Hispanic kindergartners in the State of Washington, what are the patterns of risk factors at birth?*
- 2. Within the Hispanic subgroup, are there within-group differences in the patterns of use of health services and ECE programs over the first four years of life among children based on child, family, and community characteristics?*

These first two descriptive questions can be answered using tabulations of the data and statistical tests for comparisons of means. In some cases these tabulations will be stratified by other characteristics, such as types of health services used or type of ECE program.

Comparisons between Hispanic and non-Hispanics on risk factors at birth and health service use:

3. *Among low-income kindergartners in Washington, are there differences in the risk factors at birth between Hispanic and non-Hispanic children?*
4. *Among low-income kindergartners in Washington, are there differences for Hispanics versus non-Hispanics in the patterns of use of health services, early identification services, home visiting programs, and ECE programs, from birth to kindergarten entry?*

As in Analytic Plan #1, the comparison of Hispanic to non-Hispanic children can be made in using either unadjusted data (i.e., a comparison of means) or adjusted data (i.e., a regression model). A regression modeling strategy allows us to test whether Hispanic kindergartners differ from non-Hispanic kindergartners in their risk factors at birth *conditional on the other covariates included in the regression model*. A linear regression model can use an indicator for a particular risk factor as the outcome variable.

The outcome is modeled as a function of descriptive covariates, such as age, sex, density of Hispanics living within the community, Spanish spoken in the home, a measure of socioeconomic status, and an indicator variable that equals 1 if the child is Hispanic and equals 0 if the child is not Hispanic. Tests of statistical significance for the coefficient on the Hispanic indicator variable indicate whether the conditional difference in the probability of the risk factor between Hispanic and non-Hispanic students is statistically significant.

Predicted relationship between health services and later health and well-being outcomes:

5. *Among low-income kindergartners in Washington, does utilization of health services, early intervention services, home visiting programs, and/or ECE programs predict positive health outcomes, lower levels of behavioral problems, and a reduced need for special education services at kindergarten entry? Do these relationships vary for Hispanics and non-Hispanics?*
6. *Among low-income kindergartners in Washington, do prior use of health services, early identification services, home visiting programs, and ECE programs from birth to kindergarten entry predict school readiness outcomes once the child arrives in kindergarten, as well as academic achievement outcomes at the end of third grade? Do these relationships vary for Hispanics and non-Hispanics?*

Research Questions five and six each have two parts; the first part can be addressed using a standard regression model, and the second part can be addressed using regression models with interaction terms. For example, to understand whether there is a relationship between use of health services and health outcomes, we can model the health outcomes as a function of the same covariates we have used throughout the analysis as well as a measure of health service use.

To further understand if this relationship varies for Hispanic and non-Hispanic students, we can add an interaction between an indicator for Hispanic and the measure of health service use to the model. The coefficient of this interaction term reflects a comparison of the measure of health service use for Hispanic and non-Hispanic students. Tests of statistical significance of this coefficient indicate whether the difference between these two groups is statistically significant. A similar analytic model can be used to estimate the relationship between health services and education outcomes.

References

- ⁱ See, for example, Hemphill, F.C., A. Vanneman, & T. Rahman (2011), “Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress—Statistical Analysis Report.” *National Center for Educational Statistics*.
- ⁱⁱ García, E.E., & Miller, L.S. (2008). Findings and recommendations of the national task force on early childhood education for Hispanics. *Child Development Perspectives* 2(2), 53-58.
- ⁱⁱⁱ Child Trends (2012). Children who repeated a grade. Policy brief; Hehir, T., Grindal, T., & Eidelman, H. (2012). Review of special education in the Commonwealth of Massachusetts. Policy brief; Wallace Jr, J. M., Goodkind, S., Wallace, C.M., & Bachman, J.G. (2008). Racial, ethnic, and gender differences in school discipline among US high school students: 1991-2005. *The Negro Educational Review* 59(1-2), 4.
- ^{iv} This and all subsequent demographic data drawn from US Census data and Pew Research Center (2016), “Hispanic Trends,” *Pew Research Center*.
- ^v Washington State League of Education Voters (2011). Washington’s Achievement Gap. Policy brief.
- ^{vi} Washington State Office of the Superintendent of Public Instruction (2016). “Kindergarten Preparedness Analytics.” Online resource.
- ^{vii} Washington State Office of the Superintendent of Public Instruction (2016). “Graduation and Drop-out Statistics: Annual Report to the Legislature.” Legislative reports.
- ^{viii} Ver Ploeg, M., & Beston, D.M. (eds.) (2003). Estimating Eligibility and Participation for the WIC Program: Final Report. National Academies Press.
- ^{ix} Henry J. Kaiser Family Foundation (2017). Medicaid/CHIP Child Participation Rates. Web tool.
- ^x Tarozzi, A. and A. Deaton (2008), “Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas.” Working Paper.
- ^{xi} National Institute for Early Education Research (2016). “State(s) of Head Start,” *NIERR*.
- ^{xii} Malik, R., K. Hamm, M. Adamu and T. Morrissey (2016). “Child Care Deserts: An Analysis of Child Care Centers by ZIP Code in 8 States.” *Center for American Progress*; Child Care Aware of America (2016), Child Care Deserts: Developing Solutions to Child Care Supply and Demand.” *Child Care Aware of America*.
- ^{xiii} Ansari, A. M. López, L. Manfra, S. Hartman, C. Bleiker, L. Dinehart and A. Winsler. “Differential Third-Grade Outcomes Associated with Attending Publicly Funded Preschool Programs for Low-Income Latino Children.” *Child Development* 2016; Tarozzi, A., & Deaton, A. (2008). Using Census and survey data to estimate poverty and inequality for small areas. Princeton University. Working paper.
- ^{xiv} Malik, R., K. Hamm, M. Adamu and T. Morrissey (2016). “Child Care Deserts: An Analysis of Child Care Centers by ZIP Code in 8 States.” *Center for American Progress*.
- ^{xv} Thomas, D., S. Fowler, B. Cesarone, and D. Rothenberg (2011). “The Impact of Publicly Funded Preschool in Illinois: An Analysis of Data from the Illinois Early Childhood Asset Map.” *The Illinois Early Childhood Asset Map Project*; Guzman, J. et al. (2015). “The Early Care and Education Landscape in Los Angeles County: Access, Workforce, and Quality. *Advancement Project*.



**BOLD
THINKERS
DRIVING
REAL-WORLD
IMPACT**

4550 Montgomery Avenue,
Suite 800 North
Bethesda MD 20814

ABOUT ABT ASSOCIATES

Abt Associates is a mission-driven, global leader in research, evaluation and implementing programs in the fields of health, social and environmental policy, and international development. Known for its rigorous approach to solving complex challenges, Abt Associates is regularly ranked as one of the top 20 global research firms and one of the top 40 international development innovators. The company has multiple offices in the U.S. and program offices in more than 40 countries.