The Children's Healthy Living (CHL)
Center of Excellence

CHL DATA CLEANING AND PROCESSING PROTOCOL

Vol. 1 Individual-Level Data
for the CHL Community Randomized Trial and FAS Prevalence Study

Developed by the CHL Data Work Group

for use in the CHL Pacific Region

For further information, please contact:
CHL Data Center
Children's Healthy Living Center of Excellence
Ag Sciences 302, 1955 East West Road, Honolulu, HI 96822
Email: chldata@hawaii.edu
Website: www.chl-pacific.org

This project is supported by the Agriculture and Food Research Initiative Grant no. 2011-68001-30335
from the USDA National Institute of Food and Agricultural Science Enhancement Coordinated Agricultural Program.

United States Department of Agriculture
National Institute of Food and Agriculture

CHL Data Cleaning and Checking
Vol. 1 Individual-Level Data

for the CHL Community Randomized Trial and FAS Prevalence Study

Table of Contents

Data Cleaning and Checking

Data were collected at the five jurisdictions for the community randomized trial and six jurisdictions (four states of FSM, RMI, and Palau) for the FAS prevalence study. In most cases, these jurisdictions entered the data into CHL Data Entry System and into PacTrac3. Data are stored on a password-protected secure server and on secure computers. Ultimately, the jurisdictions sent the data to the CHL Coordinating Center in Hawaii through secure transfer methods.

The CHL Data Manager was responsible for preliminary data validation cleaning and analysis for the CHL Measurement study.

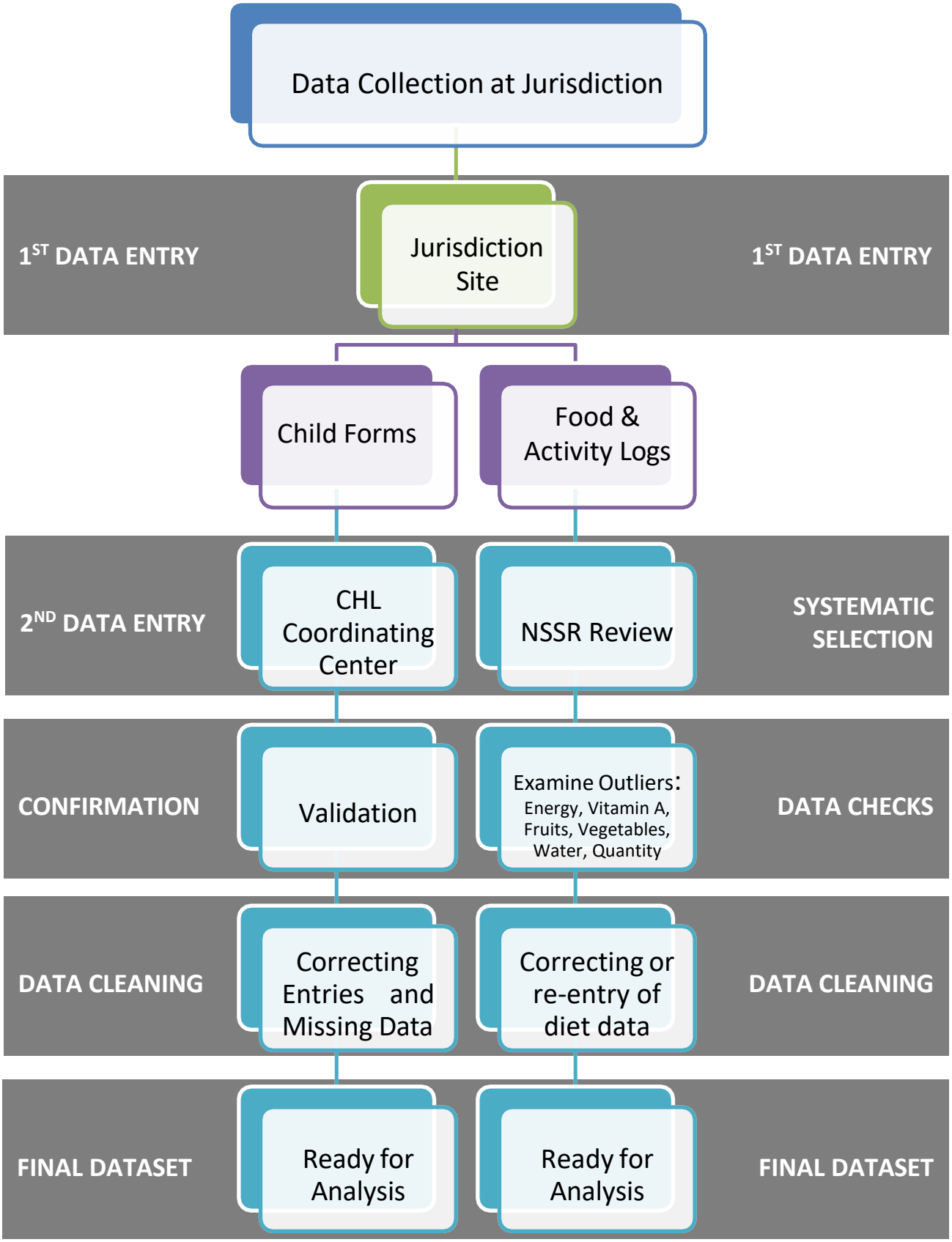
At the validation stage, any discrepancies found between the first data entry and second data entry datasets were corrected at the data set of the second data entry. Scanned images of individual forms were examined to validate the corrections.

Additional data cleaning was conducted on the validated data set when new variables were calculated or a variable was particularly examined, e.g., in the calculation of bmi, bmiz, bmipct and other related variables. Variables which need particular attentions include the following new variables:

1. Child's measures on ht (ht1, ht2, ht3, etc), or wt (wt1, wt2, wt3, etc.) or waist (waist1, waist2, waist3, etc.): if obvious errors were found on those measures, the errors were corrected before the calculation of bmi and related variables.

Figure 1 displays the process of data preparation prior to analysis -- the CHL data preparation flow chart. NSSR is the Nutrition Support Shared Resource at the University of Hawaii Cancer Center.

Figure 3. CHL Center Data Preparation Flow Chart



Measures: Description and Variables

Child and Household Demographics (Form 23-02)

The child and household demographics questionnaire, answered by the parent / caregiver, contained questions about the child, the caregiver, and the household. This demographic questionnaire, Form 23-02, --*"Information about your child and household"* was used at all CHL measurement sessions in all communities – intervention, comparison (delayed optimized), and temporal. These questions measured the dimensions and dynamics of our population and allow us to compare the variation among the jurisdictions.

Two versions exist of Form 23-02 – one is used throughout all of CHL except for FAS. This descriptive section and the following variable tables correspond to Form 23-02.

At the end of this section is a section specifically for FAS. It includes all the Form 23-02 variables used in FAS, which also covers the changes made to adapt the form for FAS. The actual variables and their associated response codes are included.

Many of these questions were adapted from The Center for Alaska Native Health Research Demographic and Medical Screening Questionnaire. *

Information collected included **household** composition and each member's relationship to the child as well as household income.

Questions about the **caregiver** included their relationship to the child, their marital status, educational attainment, employment status, food assistance, and religion. Addressing food security and availability was also included in the demographic questionnaire, to help understand the support services used by participants in our geographically varied jurisdictions. The food security questions were adapted from NHANES (cdc.gov/nchs/data/nhanes/nhanes_11_12/fsq_family.pdf). In some of the CHL community meetings, religion was identified as an important factor for group gatherings and decision making in the community. Therefore, we included inquiries about religion.

Questions about the **child** included his or her birthdate, sex, racial and ethnic background, languages spoken, place of birth, early child life information -- their birth weight and early feeding, and medical conditions. The demographic form is also used to collect numbers of hours of sleep / day and medical information. Sources include the Behavioral Risk Factor Surveillance System 2011 survey questions and the 2011 Middle School Youth Risk Behavior Survey.

The Child Information section uses validated questions for race/ethnicity. In developing the ethnicity questions, we were informed by multiple sources including the CDC Race and Ethnicity

code set version 1.0 (Centers for Disease Control and Prevention, 2000), the Demographic Information form from the 2010 Census Questionnaire (U.S. Census Bureau, 2009), and Dr. Novotny's previous study, the Healthy Living in the Pacific Islands (2001-9) questionnaire. Sources used in this section include the 2010 Census Questionnaire (U.S. Census Bureau, 2009), the Demographic Information form from the Federal Trade Commission (2012), and The National Health Plan Collaborative Toolkit (2008). The race questions will allow proper reporting to the USDA, based on OMB requirements (http://www.whitehouse.gov/omb/fedreg_race-ethnicity).

Created New Variables from Form 23-02

During the data cleaning process, many new variables were created under various reasons. Whenever a correction was made to an original variable, then a new variable was created in order to differentiate it from the original variable. Some new variables were created to address the needs of various research questions. Those new variables should always be used in place of the original variables. If the methods or protocols used in creating those new variables are simple enough, they are then described in the following table (Table 1) under the column "Notes". For some variables, e.g., new race or ethnicity variable, the methods/protocols used were complicated. The following section in particular, describes the protocols used in create the new race or ethnicity variables.

Proposed methods for creating a new race_ethnicity variable

From the original data collection form, there are 45 variables related to race and ethnicity. From those 45 variables, we have created a new variable (*race_new_omb)*, which is based on US Office of Management and Budget (OMB) Definition and has 6 categories: White, Black, AIAN, Asian, NHPI, and More than one race. Whether a child is of Hispanic origin or not is not considered when creating this variable.

Another new variable, "*ethnicity_new*" is created in order to capture or reflect a study participant's race/ethnic composition. This variable is currently in the format of "RACE1-SUBCATEGORIES_RACE2-SUBCATEGORIES…." in alphabetic order, e.g.,

"ASIAN-CHINESE_ASIAN-FILIPINO_ASIAN-INDIAN_ASIAN-JAPANESE_ASIAN-THAI_NHPI-CHAMORRO". The reason to show the larger race category before the smaller ethnic group is to meet the needs of further researchers who may not be familiar with the populations in the Pacific. The variable "*ethnicity_new*" currently describes all available ethnicity/race group a participating respondent selected. To create this variable, one of the first steps is to check each of the open space variables and make corrections as needed. For example, for one of the study subjects, AIAN was selected; however under aian_descr, it is written "BGD, Bangl". There are no other race or

ethnic categories selected for this subject. Therefore, we assumed that this child actually was Asian, instead of AIAN, and the following correction was made:

```
(a). IF aian_descr= 'BGD, Bangla' THEN DO;
        AIAN='0'; AIAN_OTHER='0'; AIAN_DESCR='';
ASIAN_other='1';ASIAN_DESCR='BANGLADESHI';
  END;
```

As a result, a total of 45 new variables were created to differentiate them from the original 45 race/ethnicity related variables (see Table 1).

The variable "*ethnicity_new*" currently has 457 categories. To further reduce the categories, we created another race/ethnic variable. The variable was named "*race_new_pacific"* and was created using the following protocols:

I. Single race group: some of the subgroups are combined. This includes the following single race groups:

a. Among Asian only (3 groups: Filipino, Asian-East and Asian-other):

i. FILIPINO: This category included only the Filipino ethnic group.

ii. EAST-ASIAN: including Chinese, Japanese, Korean alone or mixed within those 3 ethnic groups;

iii. ASIAN-OTHER: All other single Asian ethnic group or mix within those groups

b. Among NHPI only (9 groups based on CHL participating jurisdiction main ethnic groups):

i. CHAMORRO, CAROLINIAN, CHUUKESE, KOSRAEAN, MARSHALLESE, NATIVE HAWAIIAN, PALAUAN, POHNPEIAN, SAMOAN, YAPESE, and NHPI-OTHER (all other single subgroups of NHPI, e.g., KIRIBATI, TOKELAUN, TONGAN, TAHITIAN, etc.)

c. Among AIAN only: Due to the small sample size, all AIAN sub-ethnic groups were combined into one category as "AIAN". They may consist of a single AIAN ethnic group or a mix of more than one AIAN ethnic group.

d. Black only;

e. White only;

f. A new category is created and named in the following format "MIX_WITHIN_NHPI-PRIORITY ETHNIC GROUP". Due to sample size, this category was only created for NHPI ethnic groups as sample sizes for other race groups are small. The choice of the priority ethnic group was based first on the main ethnic group of interest of that

particular jurisdiction where the data is from; If for a jurisdiction where there were more than one priority ethnic groups, then priority was given to the ethnic group which had the largest frequency counts, e.g., Guam has two ethnic groups of interest: Chamorro and Chuukese. The following are some examples:

i. If a child is under category "NHPI-CHUUKESE _NHPI-OTHER" and the data is from Chuuk, we will then treat this as "MIX_WITHIN_NHPI-CHUUKESE";
ii. If a child is under "NHPI-CHAMORRO_NHPI-CHUUKESE_NHPI-PALAUAN" and data is from Hawaii, then we will use the frequency counts rule to set the order of priority. Highest priority will be given to those with the most frequency counts.

II. More than one race group: for those with two or more race groups, or currently under the category of "More than one race" of the variable *"race_new_omb".*

a. New categories will be given in the format of "MIX_PRIORITY ETHNIC GROUP". Priority was given using a combination of the following two methods: (1) interesting study ethnic groups depending on the source of data (Jurisdiction) and (2) frequency counts of each of these ethnic groups if there was more than one ethnic group. Interesting study ethnic groups for each of the participating jurisdiction are listed below:

i. Palau: Palauan;
ii. Yap: Yapese;
iii. Guam: Chamorro, Chuukese
iv. CNMI: Chamorro, Carolinian
v. Chuuk: Chuukese
vi. Pohnpei: Pohnpeian
vii. Kosrae: Kosraean
viii. RMI: Marshallese
ix. Am. Samoa: Samoan
x. Hawaii: Native Hawaiian
xi. Alaska: AIAN

b. In general, the order of priority is the following:

i. NHPI>AIAN>Filipino>Asian-east>Asian-other>Black>White for all jurisdictions except for Alaska;
ii. AIAN>NHPI> Filipino>Asian-east>Asian-other>Black>White for Alaska;

c. In some jurisdictions, like Guam and CNMI, there is more than one ethnic groups of study interest. In those cases, the priority was given based on frequency counts. The priority order for those three jurisdictions was:

i. For Guam: Chamorro>Chuukese>other PIs(order depends on frequency counts)>AIAN>Filipino>Asian-east>Asian-other>Black>White;
ii. For CNMI method one: This method follows the general CHL protocol where the top two priority ethnic groups are Chamorro and then Carolinian with the following order: Chamorro>Carolinian>Chuukese>Other PIs (order depends on frequency counts)>AIAN>Filipino>Asian East>Other PIs>Other Asians>Black>White;
iii. For CNMI method two: in this method, Carolinians are given the highest priority. If a child is of Carolinian mix with any other ethnic groups, e.g., Chamorro, then she/he is assigned as Carolinian, even though counts of Chamorro are higher than Carolinian. In this method, a new variable, "race_new_pacific_cnmi" was created. For CNMI specific reports, this variable can be used instead. Priority order under this method for CNMI is: Carolinian>Chamorro> Chuukese>Other PIs (order depends on frequency counts)>AIAN>Filipino>Asian East>Other PIs>Other Asians> Black>White;

III. The variable "*race_new_pacific*" currently has a total of 43 categories. The distribution of this variable by jurisdiction is presented in the attached excel file. This variable can then be used to create new race variable(s) depending on the needs of the individual study or research questions. For example, we could create a new variable which will combine any mixed groups into that main ethnic groups, e.g., Mix_chuukese will be combined with Chuukese, mix_Hawaiian will be combined with Hawaiian, etc. In this way, there will be no mixed group. We can also think of creating a new category as Mixed, which combine all those different mixed groups into one.

IV. Hispanic. Whether the child is of Hispanic origin is not considered in the creation of new race variables. It is a separate group (ethnicity) intended to be in addition to the race variable group, per the OMB format. Nevertheless, there variable "hispanic" (coded 1=yes and 0=no) are there in the data set. In addition, currently there are 27 children that were described as of Hispanic origin but did not indicate any race group among our study participants. If in the future, someone is interested in study Hispanic in the Pacific, he/she can use those available variables to create new ones to address his/her research needs.

FAS Version -- Child and Household Demographics (Form 23-02)

We adapted for FAS Form 23-02 Information about your child and household, which includes the demographics questions. We changed, added, or dropped some items to tailor Form 23-02 for FAS. For example, the lower income response category was changed from under $10,000 to three categories: under $2500, from $2500 to less

than $5000, and from $5000 to less than $10000. We dropped questions about SNAP (food stamps) and WIC, since were not available in the FAS region. We added questions about resource availability and the source of water in the home as well as the main type of fuel their household uses for cooking. Questions about the parent/caregiver's and the household's betel nut, tobacco and alcohol use are part of the new questions. We added questions about the participating child's frequency of tooth brushing and preventative dental care, as these items are of particular interest to this region. Table 1 displays the variables that were changed or added for FAS or that offered different coding for responses.

Time 2 Version -- Child and Household Demographics (Form 23-02)

We adapted Form 23-02 Information about your child and household which includes the demographics questions for Time 2 data collection. We changed, added, or dropped some items to tailor Form 23-02 for Time 2. For example, we added questions about resource availability and the source of water in the home as well as the main type of fuel their household uses for cooking. We also added questions about child's specific health conditions and questions about the participating child's frequency of tooth brushing and preventative dental care. Table 2 displays the variables that were changed or added for Time 2 or that offered different coding for responses.

Culture (Form 23-03)

The "Culture" questionnaire" (Form 23-03) assesses native peoples' ethnic and American mainstream affiliation in line with a bi-dimensional assumption of ethnic identity in U.S dominated/controlled societies. Stronger empirical support for considering it as a bi-dimensional construct exists rather than unidimensional (Kaholokula, Grandinetti, Keller, Nacapoy, Kingi, and & Mau 2012; Kaholokula, Iwane, & Nacapoy 2010; Kaholokula, Grandinetti, Nacapoy, & Chang 2008).

Degree of participant's own group's cultural and US mainland cultural identifications were assessed using an acculturation questionnaire originally designed for use with Native Hawaiian (Kaholokula, Grandinetti, Nacapoy and Change, 2008). The questionnaire consisted of 2 subscales: a 4-item participant's native culture identify subscale (NCIS) and a 4-item US mainland culture identity subscale (USCIS). The two subscales assess the four same aspects of Native/Own group and US mainland cultural identifies: Knowledge about each cultural group, degree of involvement with, feeling towards and associations with each cultural group. A 5-point response scale, ranging from 1 (very knowledgeable; very involved; very positive; mostly of the time) to 5 (not at all knowledgeable; not at all involved; very negative; not at all associated) was used for each item.

The original 8 variables in the data set gives a “5” to very knowledgeable, very positive, or very involved and a “1” to not knowledgeable, very negative, or disinterested. Eight new variables were created to reserve the score so that the scoring pattern matches with the literature. A total score was created for each of those two subscale (variable names: NCISand USCIS) by summing the scores of the 4 items, which ranges from 4 to 20, with lower score indicated a stronger identity. Scores <=12 on each subscale (median score 12, range 4-20) indicated higher levels of affiliation. NCIS and USCIS scores were calculated only for subjects who had complete information for the two 4-items associated with that subscale.

A categorical variable, “ACCULTURATION” was created with four categories: Integrated, traditional, assimilated, and marginalized.

Sedentary Behavior (SB) / Screen Time (ST) (Form 23-04)

This “Lifestyle Behavior” tool was adapted to help measure one of CHL’s objectives -- to decrease sedentary behavior, that is recreational screen time, by 10 min/day. The measure was modified to include active and inactive video games. It was adapted from Buckworth, J., & Nigg, C. (2004); Nigg, C. R. (2005); Haas, S., & Nigg, C. R. (2009).

Sleep (Form 23-05)

The “Sleep Questionnaire” tool was modified from The Tayside children’s sleep questionnaire (McGreavey, Donnan, Pagliari, & Sullivan, 2005). One modification was to replace the word sleep “problem” with sleep “behavior” due to concerns and feedback in field testing that some of these behaviors have different meanings to some cultures in our jurisdictions. The demographic form asks about amount of sleep. Sleep quality and duration is considered a functional outcome or a secondary outcome of obesity.

Child Anthropometry (Form 5901)

Weight, height, and waist circumference were measured by trained research staff based on standardized procedures and protocols (Lohman, Roche, & Martorell, 1988; Ikeda & Crawford, 2000; CDC, 2006). Child’s height, weight, and waist circumference measurements were taken three times at each visit. The results were reviewed right away to ensure that two of the three readings were within 0.2 units of each other (e.g., 0.2 kg for weight). If not, another series of three measurements were taken. This protocol ensured good data quality. Trained CHL staff measured each child’s height,weight and waist circumference, respectively, using a portable scale, stadiometer, and tape measure. A trained second person recorded the results using

the anthropometric recording sheet Form 59-01 and verified with the measurer that the written data were correct.

The Lohman, Roche, and Martorell (1988) Anthropometric Standardization Reference manual was also used in the form* and protocol development. This manual also informed us in the development of a internally-prepared CHL manual based on current standard practice in the field -- CHL Participant Measurement Training Guide. This protocol ensured good data quality. The Anthropometry form was developed using the Procedure Manual of the National Health and Nutrition Examination Survey (2002). The protocol for reading was adapted from the training module on technique to accurately weigh and measure infants, children and adolescents (Maternal and Child Health Bureau *) as well as The University of California at Berkeley's (2000) *Guidelines for collecting heights and weights on children and adolescents in school settings*. ?*

Zerfas informed our measurement standardization process. Zerfas criteria (*1986) were used to standardize research staff against the height, weight, and waist measurement of a certified anthropometrist, the CHL Principal Investigator, Dr. Novotny. No research staff assessed children for a measure for which they did not pass the Zerfas criteria. Zerfas did not provide any waist circumference criterion; however, Zerfas' criteria for assessments measured in cm (mm) units (height and arm circumference) was also applied to waist circumference.

Participants wore lightweight clothing and no shoes, and removed hair bands that added height. Height was measured to the nearest 0.1 cm using portable stadiometers (Perspective Enterprises, PE-AIM-101; Portage MI). Weight was measured to the nearest 0.1 kg using portable scales (Seca Model 876; Chino CA). Plastic tape (Seca Model 201; Chino CA) was used to measure waist circumference at the level of the umbilicus to the nearest 0.1 cm [Lohman et al., 1988). The parent/ caregiver often assisted to hold up the participant's shirt or by asking the child for permission to hold up their shirt. The child assumed a pose with arms crossed across the chest, holding the shoulders, with the shirt held up underneath the arms.

Time 1 and Time 2 anthropometry allow us to measure progress toward our objective to decrease the prevalence of young child overweight and obesity by 5%. These measures were used to compute Body Mass Index (BMI) as weight (kg) / height (m)2, waist (cm) to height (cm) ratios, and subsequently BMI z-score, waist circumference z-score, BMI-for-age-percentiles, and waist circumference-for-age percentiles (Barlow, 2007; Cook, Auinger, & Huang, 2009).

Anthropometry form was slightly modified for Time 2 data collection. Three new fields were added to the form: scale number, stadiometer number and tape number. Table 3 listed those additional variables.

Anthropometry Data Cleaning

CHL Protocol for data cleaning and data analysis for Anthropometry Data
The CHL protocol for anthropometry measurement states that each of the anthropometry components (height, weight, and waist circumference) must be taken 3 times regardless of the first 2 measurements being within 0.2 units (cm for height and waist and kg for weight). If no two of the 3 measures are within 0.2 units, the first 3 measures should be crossed out and the entire 3-measures process is repeated. That protocol was modified to continuing to take extra measures and crossing out (with initials of the recorder indicated) measures that were deemed out of range. The measurer should repeat the process until there were at least 2 measures within 0.2 units. The measurer was free to continue to obtain more than 3 measures if he/she believed that the first 3 measures were not consistent. A total of 9 variables were given to each of the 3 anthropometry components:

- For height, the 9 variables are:
  ht1 ht2 ht3 ht1_v2 ht2_v2 ht3_v2 ht1_v3 ht2_v3 ht3_v3;
- For weight, the 9 variables are:
  wt1 wt2 wt3 wt1_v2 wt2_v2 wt3_v2 wt1_v3 wt2_v3 wt3_v3;
- For waist circumference, the 9 variables are:
  waist1 waist2 waist3 waist1_v2 waist2_v2 waist3_v2 waist1_v3 waist2_v3 waist3_v3;

The suffix _v2 stands for the second measurement cycle for the 3 measures for any component and the suffix _v3 stands for the third measurement cycle for the 3 measures for any component when needed. Following this protocol, for each of the 3 anthropometry components, a child can have 3, or 6 or 9 measures, a number which is multiplied by 3 and the maximum total number of measure a component can have is 9.

CHL data revealed dozens of data patterns. For example, some study subjects had only 1 measure or 2 measures of one of those 3 components. There were study participants who had 4, or 5, or 7 or other numbers rather than a multiple of 3 measures for measurement components.

The following data cleaning and data analysis protocols were set for the anthropometry data, in particular, for the calculation of BMI, BMI Z score, BMI percentile or other BMI-related variables, and waist-related variables (e.g., waist height ratio) for different case scenarios.

- For study participants with no measure of height, weight, or waist
  BMI and BMI-related variables, or waist circumference and waist-related variables was set to missing.

- For study participant with 1 measure of height, weight, and waist
  The value of that single measure will be used as the value of that participant's height, weight, or waist circumference, respectively.

- For study participants with 2 or more measures (usually up to 9 measures) of height, weight, and waist
  Mean value of all available measures of each of the 3 components was used as the value of that component and be used in the later calculation of BMI, BMI-related variables, waist circumference and waist-related variables. The rationale of this decision is from the experience of anthropometric standardization where low variability in measures was found to lead to more biased results than measures with larger variability.

- In addition, to capture the number of measures each child has for the 3 components and whether among all available measures of any component, there are at least 2 measures within 0.2 units, the following variables are created:
  - ***validity_ht, validity_wt, validit_waist***: all those 3 variables are coded 1 for yes to reflect the fact that there are at least two measures within 0.2 units and coded 0 for no to reflect that fact that no two measures are within 0.2 units among all available measures. Those 3 variables can be used in the future to determine whether an analysis will use all available data or only those whose data is valid (at least 2 measures are within 0.2 units). They can be used in combination with the following variables, ***numbermeasures_ht, numbermeasures_wt, numbermeasures_waist***, in future analysis for specific investigations.

  - ***numbermeasures_ht, numbermeasures_wt, numbermeasures_waist*** : to reflect the number of measures each child has for the 3 components. This variable can be used in the future to determine whether an analysis will include all data, or only data from children with certain number of measures (e.g., n=3). It can be used in combination

with the above three variables (***validity_ht, validity_wt, validit_waist)***, in future analysis for specific investigation.

Calculation of BMI and BMI-related variables
Once the value of height and weight are determined for a participant, BMI and BMI-related variables were calculated using CDC's SAS program which uses the 2000 CDC growth charts for ages 0 to <20 years (available at: http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm).

For BMI-related variables, including BMIZ, BMI percentile to be calculated, the dataset must include the following variable with the same exact name and the given data format before to run this SAS program:
AGEMOS: child's age in months. This numeric variable may need to be calculated from the date of anthropometry measurement and the date of birth.
SEX: child's gender with 1 for male and 2 for female.
HEIGHT: child's recumbent length or standing height in centimeters. It is a numeric variable.
RECUMBENT: Indicator of child's height measurement with 1 for recumbent length and 0 for standing height. It is a numeric * variable.
WEIGHT: Child's weight in kilograms. It is a numeric variable.

In some cases, study participants may miss information on one or more of those variables. As a result, the total number of study participants may be different from the total number of study participants with a calculated BMI, or other BMIrelated variables.

Acanthosis Nigricans Screening (Form 52-09)

Acanthosis Nigricans is a skin condition characterized by dark, velvety skin in the body folds and creases such as the armpits, groin, and neck. (Burke, Hale, Hazuda, & Stern, 1999). It is often associated with conditions that raise insulin levels such as obesity, Polycystic Ovarian Syndrome, and Cushing's syndrome; it is a risk factor for type 2 diabetes. Acanthosis nigricans is considered a functional and secondary outcome of obesity.

CHL measurement staff were trained to identify acanthosis nigricans, an indication of insulin resistance.. Each child's neck will be examined at baseline and post-intervention period. Two trained research staff examined child participants' necks for the presence of acanthosis nigricans (AN). Using Burke's quantitative scale for AN, staff rated each child on a scale for Acanthosis Nigricans severity: 0 to 4 [68]. Staff marked their rating on the recording sheet Form

59-02. Participants with a score of one or higher were considered to have AN. AN is independently associated with hyperinsulinemia, an important risk factor for type 2 diabetes [69].

These measurements will allow us to measure progress toward our objective of decreasing Acanthosis Nigricans by 5%.

Acanthosis Nigricans referral. Parents/caregivers of participants with a positive screen for AN were provided a referral to follow-up with their children's health care providers or a public health service provider. If staff observes that a child has acanthosis nigricans, they will talk to the parent in private and present a written referral for their child to see a medical professional. The referral form was used to refer positive children to local medical attention. Each jurisdiction identified the most appropriate referral options for their community.

Diet of Children - Food and Activity Log -

To collect information about food and beverage intakes of the children participating in CHL, the dietary record method was used. This method is useful for collecting detailed quantities of food consumed. The dietary record administered at time one for the intervention jurisdictions and at time one for the prevalence jurisdictions included collection of physical activity as activities while eating and daily physical activity; thus, the tool was referred to as the "food and activity log" or FAL. The FAL was based on the Food and Physical Activity record used for Dr. Novotny's study, Pacific Kids DASH for Health (Novotny R, Nigg C, Li F, Wilkens L, 2015) and the dietary records used in the dietary data collection conducted by Dr. Boushey as part of the CoASTAL cohort (Fialkowski MK, McCrory MA, Roberts SM, Tracy JK, Grattan LM, Boushey CJ, 2010). Parents/caregivers were asked to complete the FAL for their children on two randomly assigned non-consecutive days, which included weekdays and weekend days, between visit 1 and 2. Assignment of recording days was based on the day of the child's first visit (Sunday – Saturday). Standard techniques were used to improve accuracy of information recorded in the FAL. Parents/caregivers were instructed in record keeping techniques with the aid of food models, service ware, and utensils and were provided a tool kit of calibrated utensils (i.e., measuring cups and spoons); the FAL; and a zip-top bag in which to place food wrappers, labels, and packages (WLP). CHL staff followed-up with reminder telephone calls. During visit two, research staff reviewed the FAL with the parents (e.g., for completeness of food entries, portion size estimation, food preparation methods, accuracy of recording data) (Yonemori KM, Ennis T, Novotny R, Fialkowski MK, Errienne R, Wilkens L, Leon Guerrero RT, Bersamin A, Coleman P, Li F, Boushey CJ, 2017). The FAL were used to estimate dietary intake over the two randomly

assigned days. The information gathered in "real time" was used to measure progress toward the CHL objectives of increasing fruits and vegetables and water; and reducing sugar-sweetened beverages. From two 24-hour periods, the data estimated foods and beverages and amounts each child consumed.

Additional data collected in the FAL for Freely Associated States (FAS)

The FAL used for the FAS jurisdictions differed from the FAL used in the intervention jurisdictions study. Unique to the FALs used in the FAS with regard to data collection was the request to provide source information for all foods and water used as single items and as ingredients in prepared dishes. Examples of sources were provided in the FAL and shown below in Table 11. The FAL for the FAS jurisdictions included an additional portion size estimation aid for fish. Pictures of fish were included in the FAL representing three portion sizes by whole fish and cut pieces of fish. The additional aids for estimating fish intake were provided due to the importance of fish in the diet among the FAS jurisdictions.

Table 11: Examples of Source of Food on FAL in FAS

Purchase	Communal/gift/donation	Local labor or self-labor
Supermarket	Food bank / food pantry	Fishing
Restaurant	Field trip	Hunting
Road side stand / stall	Church gathering	Home garden
Convenience store	Government assisted	Personal farm
Grocery store	Gift from friend/relative	Community garden
Farmers' market	USDA Commodities	Commercial farm
Lunch wagon / food wagon	Funeral	Ocean gathering
Fish markets	Traditional event	Animal husbandry
Merchant/Cargo		Specify: non-purchase

PacTrac3

The Pacific Tracker 3 (PacTac3) database and web application is a modification of the MyPyramid Tracker developed by the U.S. Department of Agriculture's (USDA) Center for Nutrition Policy and Promotion and the PacTrac2 modification by the UH Cancer Center and the Human Nutrition, Food and Animal Science (HNFAS) department (Murphy S, Blitz C, & Novotny R, 2006). PacTrac2 modified the MyPyramid Tracker for collection of dietary data in the Pacific islands. Two modifications had been made to the existing MyPyramid Tracker: 1. The addition of a function to save entered data and allow data to be accessed at a later date; aAnd 2. The addition of foods specific to the diets of the Pacific Islands' populations. PacTrac2 was modified for use in CHL and was designated as PacTrac3. This tool was used to input and analyze data collected using the food and activity log (FAL).

Pac Trac 3 generates two data tables that can be used for data analysis. The "heh" table includes derived food groupings, energy, and nutrients based on information recorded on the Food and Activity Log (FAL) and entered by CHL staff into PacTrac3. The heh file has one or two record days along with the dates of each record day per CHL ID. The "hei" table contains the names of the foods and beverages recorded on the FAL by the parents or caretakers. There is one data row per food/beverage entered associated with the user ID, record date, record time, and other relevant variables. The University of Hawaii Cancer Center's Nutrition Support Shared Resource (NSSR) Food Composition Table (FCT) was used in PacTrac3.

Most participants have records over two days at – baseline and/or post-intervention period. The index d1 or d2 at the end of each variable indicates the record is for day 1 or day 2 of each visit. Variable ***heh_visit_no*** tells which visit the data were associated with.

Average 2 day Variables for Nutrients. If only data from a certain visit are needed, use variable *heh_visit_no* (values 1 to 2) to the particular visit number of interest.

Weighted Variables for Nutrients. If only data from a certain visit are needed, use variable *heh_visit_no* (values 1 to 2) to the particular visit number of interest.

Quality Control (QC) Training

Each staff member entering dietary data needed to demonstrate adequate skills related to identifying foods, amounts, and best translation of this information as data entry into PacTrac. These skills were evaluated prior to entering any participant data. QC FALs were created to help standardize the data entry of the diet and physical activity (PA) logs. Staff completed data entry offour of the standardized diet logs. The completed data entry data using PacTrac and the entries were reviewed and scored by the staff of the NSSR. To be eligible to perform data entry, individuals needed to obtain a score within an acceptable range to begin entering participant data. If a score was outside the acceptable range, staff were asked to enter two additional QC FALs until an acceptable score was attained.

NSSR Review Summary

After entry of the FALs into PacTrac 3 by staff at the local level, the FALs were sent to the NSSR for review. For the first 6 months, the NSSR thoroughly reviewed all FALs. However, this procedure was deemed too tedious based on the large number of FALs to be reviewed. Therefore, the procedure was altered to a more focused review using systematic selection of FALs within jurisdictions and searching for specific errors using data output files from PacTrac.

Checks completed by the NSSR on the data entered into PacTrac 3 included examining:

1. Extra days

    a. FALs with more than 2 days entered were checked to see that only the correct days were entered.
2. Missing days
    a. FALs with only 1 day entered were checked to ensure there was no other day of recording.
3. Unlisted foods (i.e., foods recorded in FAL and entered into PacTrac as “Unknown”)
    a. Foods that weren’t truly unknown were checked and modified as needed.
        i. These included food items with a match in PacTrac, (e.g., Sour Cream and onion chips). For the Sour Cream and Onion chips, the best match in PacTrac is “Chips, Potato, Salted”.
    b. Unknown foods were flagged, investigated, and modified as needed.
        i. Post resolution foods (no food in PacTrac, but food exists in USDA Standard Reference).
        ii. School meals
            1. Guam did not collect any school menus, so inferences were made from Hawaii menus.
        iii. Recipes were investigated and modified as needed.
            1. If a recipe matched closely enough with an item in PacTrac 3, that food was used in place of entering each ingredient in the recipe.
        iv. Items not adequately identified were investigated and deleted as a PacTrac entry.
            1. Using the food occasion, the food, e.g., “snack” (that wasn’t consumed at school), these occurences were deleted.
            2. Printed entries that were illegible were deleted.
4. Missing detailed descriptions
    a. Missing detailed descriptions were entered into PacTrac 3.

Final Data Cleaning Steps

After the preliminary checks of the specific food entries, a final check was made that included looking at 25 different foods that were thought to be more traditional, foods that were commonly entered incorrectly, or foods for which a new food code was created in the NSSR FCT. These foods included: 1% chocolate milk, apple juice concentrate, banana ketchup, breast milk, chia seeds, chorchorus leaves/saluyot leaves, cranberry juice, Greek yogurt, gum, koko, laupele, mango juice, nori seaweed, palolo worm, Paradise Sun tea, powdered milk, Selecta Moo, dry saimin, sugar cane, tapioca, tea, tocino meat, and Yoo Hoo Chocolate Drink.

In addition, several key variables comprised of energy, nutrients, food groups, and water were examined including the following:

Outlier (variable name)	Cut-off value for examination
Energy (kcal)	≥3755 kcal

Energy (kcal)	≤1 kcal
Fruits (hei3)	≥5 cups
Vegetables (hei2)	≥3 cups
Vitamin A (vitamina)	≥2005 ug RAE
Water	≥5 cups
Quantity (e.g., household measure by food or beverage)	≥20 (e.g., cups of milk, number of doughnuts)

Entries were checked for missing detailed descriptions, serving sizes, and quantities as were ID's to ensure they were entered correctly. Records were checked for more than 2 days of entry and unlisted foods were checked and resolved as needed.

The following variable tables from the FAL are from the UH Nutrition Support Shared Resource (NSSR).

CHL Targets: SSB and Water

The analytical program used to analyze the foods/beverages entered into PacTrac 3 does not calculate total water or sugar sweetened beverages (SSB). To address the CHL target of SSBs, a classification scheme was created referencing WWEIA/NHANES and approved by the CHL data leads team. All beverages in the NSSR FCT were manually assigned one of the codes in the following table. Using the NSSR FCT foodcode for water (which includes only plain water consumed alone and not as part of a recipe) and those for SSB, the output data file from PacTrac 3 (named "hei") was used to compute each participant's intake.

The beverages in Table X. below were used to create the variables in Table X.

Table X. Beverage Coding Scheme and Definition

Code	BeverageGroup	Definition	Notes
1	Regular soft/other drinks	Includes regular (i.e., calorically sweetened) versions of the following: Soft drinks (pop, soda), fruit drinks (fruit flavored or containing less than 100 percent juice), sports drinks, and energy drinks.	SSB
2	Tea	Includes tea and tea-based drinks, such as ready-to-drink sweet tea and tea made from presweetened mix.	
2.1	Tea (sweetened)		SSB
2.2	Tea (unsweetened)		
3	Coffee	Includes coffee and coffee-based drinks, such as latte and coffee made from presweetened mix	
3.1	Coffee (sweetened)		SSB
3.2	Coffee (unsweetened)		
4	100% juice	Includes all fruit and vegetable juices that are 100 percent juice	
5	Milk	Includes plain milk (whole, reduced fat, lowfat, skim), soy milk	
6	Milk drinks	Includes milk-based drinks such as chocolate milk, milk shakes, and hot cocoa; and sweetened soy milk	

7	Diet soft/other drinks	Includes low-calorie versions of the following: Soft drinks (pop, soda), fruit drinks (fruit flavored or containing less than 100 percent juice). Also includes carbonated water	
8	Do not include/Other	Do not include meal replacement beverages, electrolyte replacement (e.g., Pedialyte) beverages, shave ice, powders	

The variables in the table below represent those created to address the specific diet related CHL targets of SSB and water. The standard nutrients and components in the CHL dataset can be found in the CHL Data Dictionary.

Table X. Sugar Sweetened Beverages (SSB) and Water CHL Target Variables in Data Set

Variable	Type	Length	Descriptive Label
SSB_PortionInCups	Number	8	Intake per day: SSB Consumption In Cups, weighted for weekday/weekend days
SSB_PortionInCups_adjforVar	Number	8	Intake for day: SSB Consumption In Cups, weighted for weekday/weekend days and adjusted for within person variance
SSB_PortionInGrams	Number	8	Intake per day: SSB Consumption In Grams, weighted for weekday/weekend days
SSB_PortionInGrams_adjforVar	Number	8	Intake for day: SSB Consumption In Grams, weighted for weekday/weekend days and adjusted for within person variance
water_PortionInCups	Number	8	Intake for day: Water Consumption In Cups, weighted for weekday/weekend days
water_PortionInCups_adjforVar	Number	8	Intake for day: Water Consumption In Cups, weighted for weekday/weekend days and adjusted for within person variance
water_PortionInGrams	Number	8	Intake for day: Water Consumption In Grams, weighted for weekday/weekend days
water_PortionInGrams_adjforVar	Number	8	Intake for day: Water Consumption In Grams, weighted for weekday/weekend days and adjusted for within person variance

CHL Targets: Fruit and Vegetable Dietary Components

The hei and pyr variables represent the dietary components available in the CHL dataset. When evaluating outcomes for the vegetable and fruit targets, the variables used were hei2 (for vegetables) and hei3 (for fruits). The other dietary components can be used to create different dietary patterns based on specific research questions. The pyr components represent individual components which can be used at the users own discretion.

References:
Friday, J.E., and Bowman, S.A. (2006). MyPyramid Equivalents Database for USDA Survey Food Codes, 1994-2002 Version 1.0. [Online]. Beltsville, MD: USDA, Agricultural Research Service, Beltsville Human Nutrition Research Center, Community Nutrition Research Group. Available at: http://www.barc.usda.gov/bhnrc/cnrg.

Bowman SA, Friday JE, Moshfegh A. (2008). MyPyramid Equivalents Database, 2.0 for USDA Survey Foods, 2003-2004 [Online] Food Surveys Research Group.
Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD. Available at: http://www.ars.usda.gov/ba/bhnrc/fsrg

CHL dietary variables: weighting for weekday/weekend days

Overall, weekdays and weekends are represented across the sample of FALS at time 1 and time 2. However, a single child could have only dietary records for weekdays, weekend days, or both. In order to best reflect the weekly dietary intakes per child, a weighted average of the daily dietary records is taken, using the weighting scheme given in the table below. The basic idea is that weekday intakes represent the majority of the week (5 days out of 7) and that weekend intakes represent the minority of the week (2 days out of 7). A general formula for the weighted average is:

$$mean_{wt} = \frac{w_{weekday} \times mean_{weekday} + w_{weekend} \times mean_{weekend}}{w_{weekday} + w_{weekend}}$$

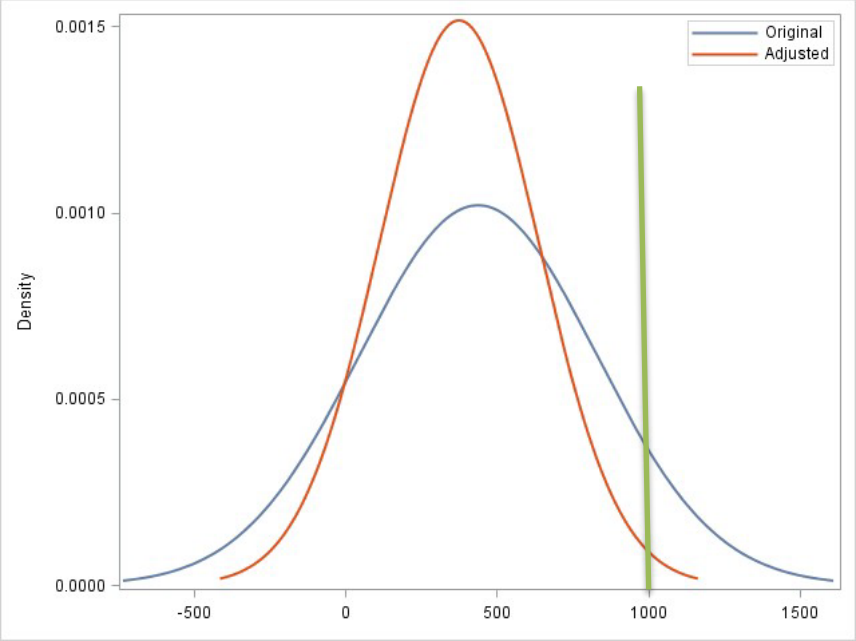
where w_{weekday} is the weight for weekdays and w_{weekend} is the weight for weekend days, and $mean_{\text{weekday}}$ is the mean of intakes for dietary records in weekdays and $mean_{\text{weekend}}$ is the mean for dietary records in weekend days. Note that the weights are in the range 0-1, inclusive, and the sum of w_{weekday} and w_{weekend} is 1.

Number of valid days of dietary records per child	w_{weekday}	w_{weekend}
1 weekday dietary record	1	0
1 weekend day dietary record	0	1
2 weekday dietary records	1	0
2 weekend day dietary records	0	1
1 weekday dietary record and 1 weekend day dietary record	5/7	2/7

CHL dietary variables: adjustment for within person variance

Long-term diet is difficult to measure, and measures of diet are subject to measurement error. Daily dietary intakes from dietary records or recalls are subject to random (rather than systematic) measurement error. While random measurement error reduces power, a comparison of means between groups, such as in the CHL analysis, with random error is unbiased and valid. However, an analysis comparing the proportion that are less than or greater than a cutpoint, such as a dietary recommendation, will not provide valid results.

The variance for a single day of dietary intake has two components: between-person variance (V_B) and within-person variance (V_W). V_B represents the difference between the long-term average diets of different individuals. V_W represents the day-to-day fluctuation in the diet, which is expected to vary randomly about the long-term diet for the person. This indicating that the mean of a large number of daily intakes would well estimate the long-term diet. The variance of a dietary intake for a single dietary record is $V_{total} = V_B + V_W$, while the variance of a mean of m dietary records is $V_{total} = V_B + V_W/m$. This (original) inflated variance creates a distribution that is more spread out than the (adjusted) distribution based only on V_B, as shown in the graph below.



Note that the mean values are quite similar between the distributions. However, the original distribution has a wider range. The percentage above the cutpoint represented at 1000 by the green line for the original distribution is 7.8% and for the adjusted distribution is 2.4%. Therefore the within-person variance overestimates the percentage above or below a cutpoint.

We adjusted the dietary variables to the distribution removing the within-subject variance. This is done by estimating V_W in order to adjust the distribution to have variance $V_B = V_{total} - V_W$. The

adjusted variables have the suffix _adjForVar. These adjusted variables can be computed with the programs PC-SIDE or IMAPP (http://www.side.stat.iastate.edu/imapp.php) or the R package SPADE. If the distribution contains many zero values (zero-inflated), SPADE should be used.

Physical Activity

Physical Activity – Food and Activity Logs

At Time 1, parents were asked to complete a record of their child's activities for two days. The activity log in conjunction with the use of the accelerometer provided more specific information about the type of activity the child was doing. At Time 2, parents were only asked to complete a record of activities done while eating.

PacTrac3 was developed from MyPyramid Tracker (cnpp.usda.gov/MyPyramidTracker.htm) with added metabolic equivalents (METS) for children (Ridley, Ainsworth, & Olds, 2008), research functionality to manage output, and local/traditional activities, such as surfing. Thus, we can quantify child physical activity in MET-adjusted minutes per day, from two days of physical activity logs per child.

Quality Control (QC) Training

Each staff member entering physical activity data needed to be cleared to do so prior to entering any participant data. The QC FALs were created to help standardize the data entry of the diet and physical activity (PA) logs. Each potential data entry staff member, entered four standardized PA logs. The completed PA logs were reviewed and scored by the NSSR staff. Data entry staff needed to obtain a score within an acceptable range to begin entering participant data. If a score was outside the acceptable range, individuals were asked to enter two additional QC PA logs until an acceptable score was attained.

NSSR Review Summary

The FALs were sent to the NSSR for review after entry of the FAL into PacTrac 3 by staff at the local level. Focus was placed on PA logs with accelerometer data.

Checks completed by the NSSR on the PA data entered into PacTrac 3 included examining:

1. Extra days
    a. FALs with more than 2 days entered were checked to see that only the correct days were entered.
2. Missing days
    a. FALs with only 1 day entered were checked to ensure there was no other day of recording.
3. Unlisted activities
    a. Activities that weren't truly unknown were checked and modified as needed.
    b. Unknown activities were flagged, investigated, and modified as needed.
4. Missing detailed descriptions
    a. Missing detailed descriptions were entered into PacTrac 3.

Final Data Cleaning Steps

After resolving the unlisted activities and inputting missing detailed descriptions, a final check was made that included the following:

Outlier (variable name)	**Cut-off value for examination**
Minutes (minutes)	>1200 minutes (or 20 hours) of activity
Minutes (minutes)	<120 minutes (or 2 hours) of activity

Physical activity – Accelerometry

Objective Measurement of Physical Activity – Accelerometry. Accelerometers have become an important tool to objectively monitor physical activity in community-dwelling conditions. Triaxial accelerometers measure vertical, horizontal and lateral acceleration and raw data can be analyzed to provide an objective measure of the intensity, duration and frequency of physical activity throughout the day. The detected accelerations are filtered, converted to a numerical value and summed over a specified time interval or epoch. The recorded counts for each epoch can be used to represent the intensity of the physical activity. Children's physical activity tends to be intermittent and characterized by rapid changes from rest to vigorous physical activity. In order to accurately quantify physical activity intensity and duration, short (<30 secs.) epoch durations are necessary (Baquet, 2006; Nilsson, 2002).

These monitors store data over long periods, allowing analysis of patterns of physical activity in free-living subjects over the course of several days to weeks. The small size of the device is unobtrusive and allows monitoring of subjects without interfering with normal movement. Several studies including our own unpublished CHL Physical Activity Pilot have concluded that accelerometers can be effectively used in free-living children to measure levels of physical activity (Trost, 2002; Freedson et al., 2005; Hoos et al., 2003). In addition, relatively short periods of monitoring (4-7 days) have been found to be reliable (Trost, 2000).

Accelerometers measured physical activity in young children. Accelerometers provide data of different intensities, including inactivity during waking hours, making them ideal for lifestyle interventions or interventions not specific to a location. This approach allowed us to obtain objective measures for sedentary behavior and physical activity at different intensities. Each participant was asked to wear an accelerometer for six days. We collected accelerometry for about 100 participants per each intervention and comparison community as well as at all FAS sites in a subset of the study population (50%; n of around 100 per jurisdiction). The accelerometer data were processed using the manufacturer's software and its output is in activity in counts/minute.

Actical Accelerometer Data

CHL participants were expected to wear the Actical accelerometer for six consecutive days. In general, the Actical accelerometer was placed on the child on day 1 and removed on day 7. For each day, a child's activity was recorded from 0:00:00 a.m. in the morning to 23:59:59 p.m., which sums to 1440 minutes. Occasionally, a child might have had his/her accelerometer removed on some day(s) during the week and later have had the accelerometer placed back on. For these children, their accelerometer data may show as non-consecutive days.

Counts per minute (cpm) were first calculated by summing the counts per second within that minute as recorded by the accelerometer. Accelerometer data were then summarized every minute into 4 activity levels using the following rules (as recommended by Respironics for the Actical):

- sedentary, if cpm ≤40
- light, if 41≤ cpm ≤ 2295
- moderate, if 2296≤ cpm ≤ 6815
- vigorous, if cpm ≥ 6816

The number of minutes on each day spent in each of the 4 activities is then calculated as the sum of minutes from 0:00:00 am to 23:59:59 pm assigned to that category.

The number of minutes per activity level and some combined levels (sedentary/light, moderate/vigorous) within 5 minute bouts was also calculated. This was done as it was found that young children often had periodic activity spikes for several seconds. For some children, the majority of their activity came from these isolated spiking events. It is unclear whether these events have the same effect on health as sustained activity. The US recommendations for adults were to perform physical activity within intervals of least 10 minutes in duration. This recommendation was changed in 2018 to perform activity of any intensity and any duration; the purpose of the change was not based on an change in understanding of the value of sustained activity, but to further encourage adults to engage in activity (https://www.advisory.com/daily-briefing/2018/11/13/exercise). The information on activity in bouts is more important for moderate and vigorous activities than sedentary and light activities.

Below is an example of how bouts would be computed for a 40 minute period. The bouts highlighted in blue are those that have duration of 5 or more minutes.

Time	Intensity	Sedentary Bout	Light Bout	Moderate Bout	Vigorous Bout	Sedentary / Light Bout	Moderate / Vigorous Bout
12:00:00 PM	Sedentary	1				1	
12:01:00 PM	Sedentary	1				1	
12:02:00 PM	Sedentary	1				1	
12:03:00 PM	Sedentary	1				1	
12:04:00 PM	Sedentary	1				1	
12:05:00 PM	Sedentary	1				1	
12:06:00 PM	Sedentary	1				1	
12:07:00 PM	Sedentary	1				1	
12:08:00 PM	Sedentary	1				1	
12:09:00 PM	Sedentary	1				1	
12:10:00 PM	Sedentary	1				1	
12:11:00 PM	Light		1			1	
12:12:00 PM	Light		1			1	
12:13:00 PM	Sedentary	2				1	
12:14:00 PM	Sedentary	2				1	
12:15:00 PM	Sedentary	2				1	
12:16:00 PM	Sedentary	2				1	
12:17:00 PM	Vigorous				1		1
12:18:00 PM	Sedentary	3				2	
12:19:00 PM	Sedentary	3				2	
12:20:00 PM	Sedentary	3				2	
12:21:00 PM	Moderate			1			2
12:22:00 PM	Moderate			1			2
12:23:00 PM	Moderate			1			2
12:24:00 PM	Moderate			1			2
12:25:00 PM	Moderate			1			2
12:26:00 PM	Vigorous				1		2
12:27:00 PM	Moderate			1			2
12:28:00 PM	Moderate			1			2
12:29:00 PM	Moderate			1			2
12:30:00 PM	Light		2			3	
12:31:00 PM	Light		2			3	
12:32:00 PM	Light		2			3	
12:33:00 PM	Light		2			3	
12:34:00 PM	Light		2			3	
12:35:00 PM	Light		2			3	
12:36:00 PM	Sedentary	4				3	
12:37:00 PM	Moderate			2			3
12:38:00 PM	Sedentary	5				4	
12:39:00 PM	Sedentary	5				4	
12:40:00 PM	Sedentary	5				4	
Total Minutes		**22**	**8**	**9**	**2**	**30**	**11**
Minutes in 5+ minutes bouts		**11**	**6**	**5**	**0**	**24**	**9**

The CHL protocol for Actical accelerometer data cleaning and data analysis includes the following 3 rules:

1. The Actical accelerometer data for each day was assessed as valid or not valid. The criteria for valid data for each day are:
    a. The total number of minutes for any one single type of activity must not exceed 1300 minutes (90% of the time).
    b. The number of minutes of moderate/vigorous activity in 5+ minute bouts must not exceed 720 minutes (12 hours).
    c. The total number of minutes of sedentary plus light activity must not be less 720 minutes (12 hours).
    d. The total number of minutes of sedentary activity must not be less than 300 minutes (5 hours).
    e. The total number of minutes for the day must be 1440 minutes (24 hours).

    Data for any day that was not valid was excluded from data analysis.

2. Data from the first day a child wore the Actical was excluded from data analysis in order to avoid potential reactivity effects of children wearing the Acticals. Data from the last day a child wore the Actical was also excluded because it may not have complete data (e.g., 1440 minutes) for that day. As a result, to be included in this data analysis, the participants need to have at least 3 days of valid Actical accelerometer data after excluding the first and last days.

The major statistic of interest from the Actical Data is the weighted average minutes per day at the various activity levels (e.g., sedentary, light, moderate, vigorous, sedentary/light, modeate/vigorous). The steps that were followed to compute this statistic are below.

A weighted average for each of the 4 activities was then calculated for each child based on the number of valid days of Actical data available and the number of weekdays or weekend days included. The averages are computed only for children who have 3 or more valid days.

If a child has only weekdays or only weekend days, the simple average is taken. If there is at least one weekday and at least one weekend day, then the formula for the weighted average is:

$$mean_{wt} = (5/7) \times mean_{weekday} + (2/7) \times mean_{weekend}$$

where $mean_{\text{weekday}}$ is the mean of intakes for dietary records in weekdays and $mean_{\text{weekend}}$ is the mean for dietary records in weekend days.

Means were created using all valid days and only using days 4-6. It was found that activity was higher on Day 1 and steadily declined over the week of wearing the Actical device, likely due to reactivity. Days 4-6 were selected as the last half of the week to reduce the influence of reactivity.