RDC Research Proposal

General Information	
Date:	November 22, 2019;
Title of Project:	The Association Between Serum Vitamin D Levels and Childhood Obesity
NCHS Data System and Years:	NHANES 2003-2006
Non-NCHS Data Files:	N/A
Mode of Access:	[X] NCHS RDC, Hyattsville, MD [_] NCHS RDC, Rockville, MD [_] NCHS RDC, Washington, DC (Government Only) [_] NCHS RDC, Atlanta, GA [_] Federal Statistical RDC, specify: _____
Statistical Software: (Check all that apply)	[X] SAS/SUDAAN [_] Stata [_] Other, specify: _____
Proposed Start Date:	May 1, 2020
Funding Source:	Funded by the National Institute for Obesity Research, Grant No. 555
Billing Address: (include contact person)	Ima Business Manager University Department 1234 Research Way, Room 789 City, State, 12345 imabusiness@email.com 555-555-7890

List the name, institution, contact information, and role for anyone who will contribute to publications resulting from this project. Everyone listed must submit a C.V or resume. Add sections as needed.

Research Team		
Role:	**Principal Investigator**	**Co-Investigator**
Name	Lee Ann	Dolce Gab
Email	lann@email.com	dgab@email.com
Phone	555-678-2389	555-678-2214
Institution	University	University
Mailing Address	Department 1234 Research Way, Room 789 City, State 12345	Department 1234 Research Way, Room 789 City, State 12345
SSS Status?	N	N
US Citizen? Y or N	Y	Y
	Programmer	**Programmer**
Name	Jake Sterr	
Email	jsterr@email.com	
Phone	555-678-2894	
Institution	University	
Mailing Address	Department 1234 Research Way, Room 789 City, State 12345	
SSS Status?	Y	
US Citizen? Y or N	Y	
	Advisor (For Students and Post-Docs) [_] RDC-Student-Advisor Form	**Other, specify:**
Name		
Email		
Phone		
Institution		
Mailing Address		
SSS Status?		
US Citizen? Y or N		

Complete as applicable for your project. Address any "Yes" responses in the body of the proposal.

RDC Proposal Summary Information		
	YES	**NO**
Geographic variables		
Level of geography to be shown in **output** (check all that apply)		
National	X	
Regional		X
State		X
MSA		X
County		X
Urban/rural classification		X
Census tract		X
Latitude/Longitude		X
Other*	X	
Will geographic identifier(s) be removed after merge	X	
If yes, can true geographic identifiers be replaced with masked versions of these variables		X
Is GIS or mapping proposed		X
Dates and Temporal information		
Are exact dates requested other than to calculate time of follow-up		X
If more than 1 year/cycle, will years/cycles be presented separately		X
Merging of data with NCHS restricted data		
Are external data being merged with NCHS data		X
Linked Data Products		
Are you requesting linked Medicare/Medicaid files		X
If yes, are you using multiple years		X
Are you using public-use mortality data		X

***Explanation:** LAT = Location (latitude) of residence will be used to control for sun exposure. After the RDC analyst creates a derived variable for latitude of sun exposure with two values (North and South), the LAT will be removed from the analytic file.

A. Abstract: Please limit the project description abstract to 300 words.

Obesity has been linked to vitamin D deficiency in adults and adolescents. We aim to determine if an association exists between obesity and inadequate serum vitamin D levels among U.S. children. We will use serum 25-hydroxyvitamin D (vitamin D) and body measurement data from U.S. children aged 6–18 years examined in the National Health and Nutrition Examination Survey (NHANES) from 2003–2006 and evaluate the relationship between serum vitamin D levels and obesity, defined as a body mass index (BMI) ≥ 95th percentile. Vitamin D levels will be dichotomized as deficient (<15ng/ml) or not deficient in logistic regression models to assess odds of vitamin D deficiency accounting for age, sex, race/ethnicity, poverty status, and vitamin D-containing supplement use. We seek to adjust for two additional factors associated with serum vitamin D levels that may influence our results: latitude of residence and season of lab testing. These variables are restricted and only available through the Research Data Center.

B. Research Question: Describe study purpose, hypotheses, goals, or research questions.

- What is the relationship between vitamin D deficiency and obesity in US children aged 6-18 years?
- How does latitude of residence and season of lab testing influence this relationship?

C. **Background**: **Include a short literature review, no more than 2 pages, focusing on papers that discuss your topic and address the methodology that you plan to use. Please limit your reference list to 10 items or less.**

Vitamin D is a fat-soluble vitamin needed for promoting calcium absorption in the gut and ultimately enabling normal bone mineralization. It is also needed for bone growth and remodeling and has more recently been discovered to be involved in other physiologic processes, including modulation of neuromuscular and immune function, as well as reduction of inflammation. It may also play a role in modulating cancer cell proliferation. The growing evidence that vitamin D may help prevent several chronic diseases prompts the need to identify individuals at risk for vitamin D deficiency.

Humans get vitamin D from their diet, in dietary supplements, and from exposure to sunlight. People living at higher latitudes have been shown to have lower levels of serum vitamin D compared with those living in lower latitudes, and levels of serum vitamin D are highest during the summer months when sun exposure is greater. Examination of the relationship between serum vitamin D levels and obesity is done using logistic regression analysis, with vitamin D deficiency as the binomial outcome and obesity as the binomial primary explanatory variable. Other important covariates we have adjusted for in our analyses thus far (using publicly accessible NHANES data) include: age, sex, race/ethnicity, poverty status, and the use of vitamin D-containing supplements. Because exposure to sunlight is also an important factor to account for when measuring serum vitamin D levels, we want to adjust for latitude of residence and month/season of lab draw, in addition to the other mentioned covariates. These variables are only available through the RDC.

D. Public Health Benefit: In one paragraph, how does your research benefit public health?

Our study seeks to examine the relationship between serum vitamin D levels, measured as 25- hydroxyvitamin D and dichotomized as vitamin D deficient or not, and obesity, defined as a BMI ≥ 95th percentile for age, in children aged 6–18 years. Prior research in adolescents and adults has shown a positive association between

vitamin D deficiency and obesity. By establishing an association between low serum vitamin D levels and obesity in children across a wider age range, we aim to identify an easy- to-obtain and objective measure with which to target children who may be at greater risk for vitamin D deficiency. Using this measure, children deficient in vitamin D may be more readily identified and started on supplementation to correct it. Because vitamin D may be involved in improving other health measures or preventing other chronic diseases or conditions, treating deficiency may have benefits that extend beyond improved bone health.

E. Data Requirements:
Please address the four items below and provide an explanation for "yes/checked" responses from the RDC Proposal Summary Information Table.

1. **Survey, Years, Files**:
   List survey name(s) and survey years you would like to access. For examples, NHIS 2005-2007 Household, Person and Sample Adult Files, NAMCS 2005-2006 Provider and Patient Visit Files, or NHANES 2005-2006 Examination and Demographic Files.

   NHANES 2003-2006
   Demographic variables and sample weights
   Physical examination measurements
   Lab component: Vitamin D
   Dietary supplements questionnaire

2. **Restricted-use Data:**
   List and describe the restricted-use variables that you will need. These variables must be listed in the Data Dictionary section of this proposal. Explain why each variable is needed and how you will include them in your analysis. Specify how geographic variables, if applicable, will be used to merge files, analyze the data and/or be presented in output.

   LAT = Location (latitude) of residence will be used to control for sun exposure.
   Here are categories for the derived variable:
   if LAT is missing then assign the value **.** to the new variable;
   else if LAT < X then assign the value 1 to the new variable;
   else if LAT is >= X then assign the value 2 to the new variable.
   LAT can be dropped after the derived variable for sun exposure is created.

   Month of MEC exam/lab draw will be used to control for season.
   Here are categories for the derived season variable:
   If month of exam in (1, 2, 12) then Season = Winter;
   else if month of exam in (3, 4, 5) then Season = Spring;
   else if month of exam in (6, 7, 8) then Season = Summer;
   else if month of exam in (9, 10, 11) then Season = Fall;
   else Season = .
   Month of MEC exam/lab draw can be dropped after the derived variable Season is created.

3. **Non-NCHS Data:**

If you plan to provide data from another source (such as Census or EPA) to merge to the restricted-use data, please describe the source, list the filename(s), and provide a general description of the non-NCHS data. The variables from the non-NCHS data must be listed in the Data Dictionary section.

N/A

4. **Merge Variables:**
   In detail, please describe the merge procedures needed to produce your analytic dataset(s). Highlight the variables used in the merge routine in the Data Dictionary. Leave blank if not applicable (e.g. NHDS, NAMCS/NHAMCS, Mortality, Natality, and DHHS Hosted Data Users). Note: The RDC Analyst will merge the data for you.

   SEQN will be used to merge the public and restricted data files.

F. Methodology:
We highly recommend you familiarize yourself with the analytic guidelines of the data you intend to use. Any deviations from the methodology suggested in the guidelines will require explanation as it may pose a disclosure risk. Please address these three items in your proposal:

1. **Unit or Level of Analysis and Subpopulation(s):**
   There can be many levels of analysis: be as detailed as possible in your description. A common example for an analysis using NHANES data is where the unit of analysis is the person while the subpopulation is adults ages 18-64. A common example involving geography is when the researcher aggregates persons to the state level to make comparisons between state policies.

   Unit of Analysis – individual
   Subpopulation – Children 6-18, excluding pregnant females, those with implausible BMIs, and those with missing covariate data.

2. **Analysis Plan: Provide an overall analysis plan that specifies what analytic procedures or models you will use, such as prevalence estimates, logistic regression, or log-linear modeling. Also, list the specific statistical package procedures you will use.**

   We have already performed logistic regression analyses using SAS-callable SUDAAN on the public use NHANES data for 2003–2006. The outcome is serum vitamin D level and is dichotomized into "deficient" (<15ng/ml or the 10^{th} percentile) or "not deficient" (≥ 15ng/ml). The primary explanatory variable is obesity (or BMI ≥ 95^{th} percentile for age) and is categorized as yes or no. The remaining covariates include:
   - age (as a continuous variable, in years, for one analysis),
   - age group (categorized into 6–9 years, 10–12 years, 13–15 years, and 16–18 years, for a separate analysis),
   - sex (male or female),
   - race/ethnicity (categorized as Non-Hispanic white, Non-Hispanic black, Mexican American, and other races, including multiracial),
   - poverty status (categorized as <2.0 PIR or ≥ 2.0 PIR), and
   - vitamin D-containing supplement use (yes or no),

- latitude of residence (North vs South);
- season of MEC lab draw (Winter, Spring, Summer, Fall).

3. **Complex Survey Design: Indicate how you will address sample weights, design variables, and other adjustments for the use of complex survey data, if applicable, using the statistical software listed in the General Information area. A detailed description per weight, design variables, and other adjustments are required and central to understanding the limitation of the data. This is a critical element to describe and is central to the proposal review process.**

   Our codes account for weighting and a complex sample design, where WTMEC4YR = 1/2 x WTMEC2YR for this 4-year sample. Example logistic regression code follows:

```
proc rlogist data = out.vitdobese_4 filetype=sas design=wr;
nest sdmvstra sdmvpsu/missunit;
weight wtmec4yr;
subpopn include = 1/name="6-18 year olds, no pregnant females";
class obese sex raceth fampir vitdsup/nofreqs;
reflevel obese=0 sex=1 raceth=1 fampir=2 vitdsup=2;
model vitd10 = obese sex raceth fampir vitdsup examageyr;
etc.
```

G. Output:

Please describe in full detail all the output you need to look at in order to ensure that the data file created for you is correct and complete. NCHS will only provide you with a SAS proc contents as the default if you do not specify any additional output needed.

Please describe in full detail all the output you would like to take out of the RDC that is needed for the final product you wish to create. Please be very specific in describing your output needs (i.e., state groupings, levels of output and how restricted-use data will be displayed) as this section is necessary for the Review Committee to assess disclosure risk of your output. Your examples should reflect the geographic variables you will use in your output. Note: Any output produced in the RDC that is not listed in your proposal will not be allowed out of the RDC.

1. **Output Needed to Confirm Accuracy of Analytic Dataset: Since your RDC Analyst will create your analytic dataset, please describe what output you need to review to ensure that your dataset was constructed accurately and is complete (e.g., need to review univariate frequencies of certain variables – please specify).**

   We request SAS code of the merge, log file, and PROC CONTENTS of the merged datasets. Also, we would like univariate frequencies of the following variables, all the years of the survey combined: SDDSRVYR, AGEGROUP, SEX, latitude of residence (North, South, missing), and season of MEC lab draw (Winter, Spring, Summer, Fall, missing).

2. **Table Shells: Include detailed examples of all table shells, models, and/or graphs with titles. Indicate the subsample and unit of analysis used in each type of table, model, or graphs. No graphical output is released by the RDC; you need to provide a table from which you can create graphs. Your proposal will not be approved without this information.**

Table 1: Weighted histogram of serum vitamin D levels in the sample population, NHANES 2003-2006. We will present 2 additional histograms by age group as well (e.g., 6–12 years old and 13–18 years old).
Total 3 tables.
Histograms will be created from output that looks like this:

BMI Percentile	5	10	25	50	75	85	90	95
Vitamin D >=15ng/ml								
Vitamin D < 15ng/ml								

Table 2: Summary statistics of the weighted percentage breakdown for each covariate and the outcome (vitamin D deficiency) in the sample population of 6–18-year-old, NHANES 2003-2006.

Covariates	**Vitamin D >= 15 ng/ml**	**Vitamin D < 15 ng/ml**
Unweighted N		
Age 6 7 8 9 10 11 12 13 14 15 16 17 18		
Sex Male Female		
Race/Ethnicity Non-Hispanic white, Non-Hispanic black, Mexican American, other races, including multiracial		
Poverty status <2.0 PIR ≥ 2.0 PIR		
Vitamin D Supplement Use Yes No		
Season Winter Spring Summer Fall		
Latitude North South		

Table 3: Results of any significant univariate analyses for vitamin D deficiency status and each covariate, presented in odds ratios, NHANES 2003-2006.

OR	**Vitamin D >= 15 ng/ml**	**Vitamin D < 15 ng/ml**
Age 6 7 8 9 10 11 12 13 14 15 16 17 18		
Sex Male Female		
Race/Ethnicity Non-Hispanic white, Non-Hispanic black, Mexican American, other races, including multiracial		
Vitamin D Supplement Use Yes No		
Season Winter Spring Summer Fall		
Latitude North South		

Table 4: Results of logistic regression analyses for vitamin D deficiency, adjusted for age, sex, race/ethnicity, poverty status, latitude of residence, season of exam, vitamin D supplement use presented as adjusted odds ratios. NHANES 2003-2006.

If any significant interactions are found, they will also be presented, with corresponding p-values.

	Vitamin D >= 15 ng/ml	**Vitamin D < 15 ng/ml**
BMI >95th Percentile Interactions:		

3. **Presentation of Results:** Describe how you will present the results (in a report, publication in a peer-reviewed journal, presentation at a scientific meeting, or used for internal policy analysis).

   Presentation to EIS officers and potential peer-review publication.

H. Data Dictionary:

Include a data dictionary for each data source. Provide a public-use and restricted-use data dictionary for NCHS survey data. This should be a listing of variables you would like in your dataset. Also, provide a data dictionary for any non-NCHS data you want to use. See instructions and examples for creating the data dictionary. When asking for multiple years of data, make sure to reflect the public-use file layout for each year as variable names can change over years. Include all explanations in Section E. Data Requirements.

Note: NHDS, NAMCS/NHAMCS, Mortality, Natality, and DHHS Hosted Data Users: Provide a single data dictionary that includes all the variables (public-use and restricted-use) you would like extracted for your analytic data set.

1. **NCHS Restricted Data Dictionary**

Variable Name	**Variable Description**
SEQN	Sequence Number – Used for Merging to Public Data
LAT	Latitude of residence
EXAMDATE	Date of MEC exam/lab draw

2. **NCHS Public Use Data Dictionary**

Variable	**Variable Description**	Data availability, by cycle	
		2003-2004	2005-2006
SEQN	Sequence Number – Used for Merging to Restricted Data	X	X
SDMVSTRA	Pseudo-stratum, used to identify segment in individual counties	X	X
SDMVPSU	Pseudo-primary sampling unit, used to identify households	X	X
SDDSRVYR	Survey year (3=2003–2004, 4=2005–2006)	X	X
WTMEC4YR	½ x WTMEC2YR, used to extrapolate sample data to entire population for the entire 4-year study period	X	X
RIDEXPRG	Pregnancy status of participant	X	X
VIT_D	Serum 25-OH vitamin D level, in ng/ml	X	X

VITD10	Vitamin D deficiency: yes (serum 25-OH vitamin D <15ng/ml or <10th percentile), or no (serum 25-OH vitamin D ≥ 15ng/ml or ≥ 10th percentile)	X	X
BMIPCT	BMI percentile for age (in months), calculated with a pre-written program using height (BMXHT) and weight (BMXWT) variables measured on bmx data sets	X	X
OBESE	Obesity status, categorized as yes (BMIPCT ≥ 95th percentile) or no (BMIPCT<95th percentile)	X	X
EXAMAGEYR	Age, in years (converted from RIDAGEEX or age, in months, at time of MEC exam, divided by 12)	X	X
AGEGROUP	Age, in years (EXAMAGEYR), categorized as 6-12 years, 13-18 years	X	X
SEX	Gender, 1 - male, 2 - female	X	X
RACETH	Race/ethnicity (same as RIDRETH1, except for adding "Other Hispanic" into the "Other race, including multiracial" category)	X	X
FAMPIR	Poverty status (INDFMPIR categorized as a poverty income ratio <2.0 or ≥ 2.0)	X	X
VITDSUP	Vitamin D-containing supplement use (determined by finding any dietary supplements taken by each participant that contained an ingredient of vitamin D), categorized as yes or no	X	X
INCLUDE	Sample selection variable, excluding pregnant females, those with implausible BMIs, and those with missing covariate data	X	X

3. **Non-NCHS Data Dictionary:** N/A

I. Literature References: **Please provide a list of up to 10 references that are relevant to your project.**

J. Resumes/C.V.: Please include a 2-page C.V. for each member of the research team (not as attachments).