RDC Research Proposal

General Information	
Date:	February 25, 2009
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: (Check all that apply)	[X] NCHS RDC, Hyattsville, MD (Washington, DC-metro) [_] NCHS RDC, Atlanta, GA [_] Remote Access (ANDRE) [_] Census RDC, specify: ____________________
Statistical Software: (Check all that apply)	[X] SAS/Sudaan [_] Stata [_] Other, specify: _________________ * Remote access users can only use SAS/Sudaan
Proposed Start Date:	May 1, 2009
Funding Source:	Funded by the National Insitute 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

Complete as applicable for your project. Everyone listed in this section will need to submit a C.V. and if approved, must complete the Confidentiality Orientation. There can only be one ANDRE programmer.

Research Team		
	Primary Investigator	**Programmer** [X] On-site or [_] ANDRE (account holder)
Name	Ima Researcher	Ima Programmer
Email	imaresearcher@email.com	imaprogrammer@email.com
Phone	555-555-5555	555-555-1234
Institution	University	University
Mailing Address	Department 1234 Research Way, Room 789 City, State, 12345	Department 1234 Research Way, Room 789 City, State, 12345
US Citizen? Y or N	Y	Y

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

Research Proposal

A. Abstract:

Obesity has been linked to vitamin D deficiency in adults and adolescents. We aimed to determine if an association exists between obesity and inadequate serum vitamin D levels among U.S. children.

We used serum 25-hydroxyvitamin D (vitamin D) and body measurement data from 4,745 U.S. children aged 6–18 years examined in the National Health and Nutrition Examination Survey (NHANES) from 2003–2006, and evaluated the relationship between serum vitamin D levels and obesity, defined as a body mass index (BMI) ≥ 95th percentile. Vitamin D levels were 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 month/season of lab testing. These variables are restricted and only available through the Research Data Center.

B. Research Question:

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

C. Background:

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

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 easyto-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:

1. **NCHS Survey, Years, Files**
   NHANES
   2003-2006
   Demographic variables and sample weights
   Physical examination measurements
   Lab component: Vitamin D
   Dietary supplements questionnaire
   Dietary supplements questionnaire formats

2. **Restricted Variables**
   LAT = Location (latitude) of residence will be used to control for sun exposure.
   Month of MEC exam/lab draw will be used to control for season.

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

4. **Merge Variables**
   i) SEQN will be used to merge the public and restricted data files
   ii) N/A

F. Methodology:

1. **Unit or Level of Analysis and Subpopulation(s):**
   Unit of Analysis – individual
   Subpopulation – Children 6-18

2. **Analysis Plan:**
   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 10th percentile) or "not deficient" (≥ 15ng/ml). The primary explanatory variable is obesity (or BMI ≥ 95th 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). We intend on adjusting for two additional covariates obtained from the RDC: latitude of residence (specific latitude broken down into ranges, range size dependent on the variability of latitudes); season of MEC lab draw (determined by the month/date of the MEC exam/lab draw).

3. **Complex Survey Design:**
   Our codes already 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:

1. **Overview:**
   We will present a weighted histogram of serum vitamin D levels for the sample population of all children 6–18 years of age, excluding pregnant females, those with implausible BMIs, and those with missing covariate data. This will help establish our choice of cutpoint used to define vitamin D deficiency. Next, we will present proportions/percentages of the population for each covariate (age, in years, and age group, sex, race/ethnicity, latitude of residence (range), season of exam/lab draw, poverty status, vitamin D supplement use, obesity) and the outcome of vitamin D deficiency. We will then perform univariate analyses and present a proportional breakdown of each covariate with vitamin D deficiency. And finally, we will perform logistic regression analyses to determine odds ratios for vitamin D deficiency, adjusting for all of the above-mentioned covariates. Our primary focus will be on the odds ratio for vitamin D deficiency in obese vs. non-obese children, but we may present other significant odds ratios as well. We will perform one analysis with age as a continuous variable and another with age group as a categorical variable. If any significant interactions are found, we will also present those.

2. **Examples/Table Shells:** The desired output will consist of the following graphs/tables

Figure 1: Weighted histogram of serum vitamin D levels in the sample population, with labeled 5th, 10th, and 25th percentile cutpoints. We may present individual histograms by age group as well (e.g., 6–12 year olds and 13–18 year olds). It will be created from output that looks like this:

BMI Percentile	**5**	**10**	**25**	**50**	**75**	**85**	**90**	**95**	**100**
Vitamin D >=15ng/ml									
Vitamin D < 15ng/ml									

Table 1: Summary statistics of the weighted percentage breakdown for each covariate and the outcome (vitamin D deficiency) in the sample population of 6–18 year olds.

Table 2: Results of any significant univariate analyses for vitamin D deficiency status and each covariate, presented in odds ratios.

Table 1: % **Table 2: 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		
White		
African American		
Mexican American		
Other		
Ethnicity		
Hispanic		
Non-Hispanic		
Vitamin D Supplement Use		
Yes		
No		
Season		
Winter		
Spring		
Summer		
Fall		
Latitude		
North		
South		

Table 3: Results of logistic regression analyses for vitamin D deficiency, adjusted for age (in years, as a continuous variable) or age group (as a categorical variable), sex, race/ethnicity, poverty status, latitude of residence, season of exam, vitamin D supplement use presented as adjusted odds ratios. 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 * controlling for age, sex, race/ethnicity, poverty, latitude, season, supplement use		

3. **Presentation of Results:** Presentation to EIS officers and potential peer-review publication.

H. Data Dictionary:

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 Name	**Variable Description**
SEQN	Sequence Number – Used for Merging to Public Data
SDMVSTRA	Pseudo-stratum, used to identify segment in individual counties
SDMVPSU	Pseudo-primary sampling unit, used to identify households
SDDSRVYR	Survey year (3=2003–2004, 4=2005–2006)
WTMEC4YR	½ x WTMEC2YR, used to extrapolate sample data to entire population for the entire 4-year study period
RIDEXPRG	Pregnancy status of participant
VIT_D = LBXVID	Serum 25-OH vitamin D level, in ng/ml
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)
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
OBESE	Obesity status, categorized as yes (BMIPCT ≥ 95th percentile) or no (BMIPCT <95th percentile)
EXAMAGEYR	Age, in years (converted from RIDAGEEX or age, in months, at time of MEC exam, divided by 12)
AGEGROUP	Age, in years (EXAMAGEYR), categorized as 6–9 years, 10–12 years, 13–15 years, and 16–18 years
SEX	= RIAGENDR = Subject's sex

RACETH	Race/ethnicity (same as RIDRETH1, except for adding "Other Hispanic" into the "Other race, including multiracial" category)
FAMPIR	Poverty status (INDFMPIR categorized as a poverty income ratio <2.0 or $\geq$ 2.0)
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
INCLUDE	Sample selection variable, which includes only

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

I. References:
Please limit to 10

J. Other Authors:
Other Author One, University
Other Author Two, University

K. Resumes/CV:
Please limit each CV to 2 pages