

SAMHSA RDC Research Proposal

General Information

Project title	The Association Between Serum Vitamin D Levels and Childhood Obesity
Date of application	November 22, 2019.
Date of major amendments	01/20/2020; 05/01/2021; 11/01/2021
SAMHSA restricted data & years	Restricted NSDUH data, 2013-2017
Non-SAMHSA restricted data & years	NCHS: NHANES, 2013-2017;
Non-SAMHSA public data & years	Public Medical Expenditure Panel Survey, 2013-2017; Public Local Area Unemployment Statistics, 2013-2017.
RDC location selection	<input checked="" type="checkbox"/> <u>NCHS RDC site, specify:</u> Hyattsville, MD <input type="checkbox"/> <u>Census FSRDC site, specify:</u>
Preferred statistical software at RDC (Check all that apply)	<input checked="" type="checkbox"/> SAS/SUDAAN <input type="checkbox"/> Stata <input type="checkbox"/> Other, specify: _____
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

Research Team

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.

	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		
SSS Status?	Y	
US Citizen? Y or N	Y	
	Advisor (For Students and Post-Docs) <input type="checkbox"/> (Please sign Student Data User Acknowledgement form)	Other, specify: Graduate student (Please sign Student Data User Acknowledgement form)
Name		
Email		
Phone		
Institution		
Mailing Address		
SSS Status?		
US Citizen? Y or N		

RDC Proposal Summary Information

Check Yes or No as applicable for your project. Address any “Yes” responses in the body of the proposal.

	YES	NO
Level of geography to be shown in output (check all that apply)		
National	X	
Regional		X
State		X
MSA		X

	YES	NO
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

* 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: (Up 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 25hydroxyvitamin 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) \geq 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)

- The research objectives are:
 - To understand the behavior of opioid misuse driven by economic circumstances in the presence of illicit substitutes to the substance and supply control on prescription-based opioids
 - To predict what policies would effectively decrease the propensity to misuse opioids and thus the mortality rate induced by them.
- My hypotheses are:
 - In the presence of abundant substitutes to prescription-based opioids like illicit opioids and heroin, state-level supply control on opioid prescription does not affect the propensity to misuse opioids.
 - Socioeconomic circumstances like job displacement (and family-related shocks (e.g., divorce)) affect the propensity to opioid abuse.

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.

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 \geq 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. SAMHSA restricted data:

Restricted-use data of National Survey on Drug Use and Health (NSDUH), 2010-2017 will be analyzed. The following types of variables will be used:

- a) Identification and interview information: interview id, year, survey weight
- b) Demographic information: age, education, employment status, marital status, working environment, the composition of household, income, government assistance, health, and location.
- c) Geographical information: state, county, region, census tract 1. Counties, regions, and census tracts will not be revealed in the output. They are used to merge area-specific control variables, only when needed.
- d) Opioid use and misuse.

2. Non-SAMHSA restricted-use data:

- 1) NCHS restricted data of NHANES, 20013-2017. The following subcomponents will be accessed:
 - a. Demographic variables and sample weights
 - b. Physical examination measurements
 - c. Lab component: Vitamin D
 - d. Dietary supplements questionnaire
- 2) Public Medical Expenditure Panel Survey (MEPS) 2013-2017, Household Component and Prescribed Medicines Files.
 - a. The Household Component contains socioeconomic variables such as age, education, employment status, reasons for not working, marital status, income, and perceived health.
 - b. Prescribed Medicines Files contain respondent-specific prescriptions per round of panel. Specifically, this is used to produce variables for opioid prescriptions.

3. Non-SAMHSA public data:

- 1) Public Medical Expenditure Panel Survey (MEPS) 2013-2017, Household Component and Prescribed Medicines Files. I use this data set to produce marginal transition probabilities of state variables for receiving prescription opioids, employment status, perceived health, work decisions, etc. in the model and to provide expenditures that individuals spend for prescription opioids at the national level.
- 2) Local Area Unemployment Statistics (LAUS), American Consumer Survey (ACS), and Census data 2013-2017. These data sets provide 1-year estimates of population, single-year ages, employment rate, unemployment rate, income, marriage rate, and divorce rate at the state-level. They are used i) to compute aggregate mortality rates, opioid overdose mortality rates, and opioid misuse rates by year and age groups, ii) as control variables in the regression analysis, and iii) provide state-level aggregate statistics to fit the structural model.

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.]

F. Methodology:

Please address the following items in your proposal:

1. Unit or Level of Analysis and Subpopulation(s):

There can be many levels of analysis. For example, the unit of analysis is the person while the subpopulation is adults ages 18-64. For another example, the researcher aggregates persons to the state level to make comparisons between state policies, where the unit of analysis is the individual survey participants and the subpopulation is Children 6-18, excluding pregnant females.

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.

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 on page 1. A detailed description per weight, design variables, and other adjustments are required and central to understanding the limitation of the data. For example:

Our codes account for weighting and a complex sample design, where $WTMEC4YR = 1/2 \times WTMEC2YR$ 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;
```

Note: SAMHSA RDC allows only the weighted estimates/numbers to be published, with unweighted estimates and unweighted sample sizes being considered in the output data tables only on a case-by-case basis. Please provide a paragraph of justifications here if you intend to include unweighted results and/or sample sizes in the output.

G. Output:

1. General description of the output data you intend to take out from your RDC analyses.

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. **Any output produced in the RDC that is not listed in your proposal will not be allowed out of the RDC.**

2. Provide the RDC output data 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. All tables must have unique sequential ID and properly labels. The table IDs, titles, column headers, and row headers you provided here must match exactly with your future RDC output data tables. You may provide one table shell to cover a group of tables of exactly the same pattern, but all table IDs and titles must be listed out. **Your proposal will not be approved without this information.**

Three table shell examples are provided below.

Table 1: Weighted histogram of serum vitamin D levels in the sample population, NHANES 2003-2006.

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 groups 6 - 10 11 - 18		
x Male Female		
Race/Ethnicity Non-Hispanic white, Non-Hispanic black, Mexican American, other races, including multiracial		
Poverty status <2.0 PIR ≥ 2.0 PIR		

Covariates	Vitamin D >= 15 ng/ml	Vitamin D < 15 ng/ml
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		
x Male Female		
Race/Ethnicity Non-Hispanic white, Non-Hispanic black, Mexican American, other races, including multiracial		
Vitamin D Supplement Use Yes No		

OR	Vitamin D >= 15 ng/ml	Vitamin D < 15 ng/ml
Season Winter Spring Summer Fall		
Latitude North South		

H. 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).

I. Data Dictionary:

Include a data dictionary for each data source described in section “E. Data Requirement”. Please refer to “Example of Data Dictionary” on our [webpage](#). When asking for multiple years of data, make sure your data dictionary reflects the variable name changes over the years. If possible, provide a single data dictionary table for each survey or database you are requesting to access at the RDC.

1. SAMHSA’s NSDUH Restricted Data Dictionary

Variable Name	Variable Description	Data Availability by Year								
		2011	2012	2013	2014	2015	2016	2017	2018	2019
STATE	State FIPS code (numeric)	√	√	√	√	√	√	√	√	√
STATENAME	State name	√	√	√	√	√	√	√	√	√
COUNTY	County FIPS code (numeric) - for merging with restricted NSDUH.	√	√	√	√	√	√	√	√	√
STUSAB	State US abbreviation (string)	√	√	√	√	√	√	√	√	√
ZIPCODE	Postal zip code	√	√	√	√	√	√	√	√	√
ZCTA	Zip Code Tabulation Area	√	√	√	√	√	√	√	√	√
MCDIMD	State Waiver Status on Medicaid Institution of Mental Disease Payment Restriction.	√	√	√	√	√	√	√	√	√
PR_MCO_ALL	Penetration rate of Medicaid comprehensive	√	√	√	√	√	√	√	√	√

Variable Name	Variable Description	Data Availability by Year								
		2011	2012	2013	2014	2015	2016	2017	2018	2019
	managed care in a state among all Medicaid enrollees									
PR_MMC_ALL	Penetration rate of any type of Medicaid managed care program in a state among all Medicaid enrollees	√	√	√	√	√	√	√	√	√
PR_MMC_C1214	Penetration rate of any type of Medicaid managed care program in a state among Medicaid children aged 12-14	√	√	√	√	√	√	√	√	√
PR_MMC_C1518	Penetration rate of any type of Medicaid managed care program in a state among Medicaid children aged 15-18	√	√	√	√	√	√	√	√	√
PR_MMC_C1218	Penetration rate of any type of Medicaid managed care program in a state among Medicaid children aged 12-18	√	√	√	√	√	√	√	√	√
PR_MMC_ADULT	Penetration rate of any type of Medicaid managed care program in a state among adult Medicaid enrollees	√	√	√	√	√	√	√	√	√
ENROLL_ADULT	Mandatory/voluntary enrollment in a MMC program of a Medicaid eligibility group: Adult				√	√	√	√	√	√
STNMPSYL	Estimated number of psychologists in a state	√	√	√	√	√	√	√	√	√
STNMMHCO	Estimated number of MH counselors in a state	√	√	√	√	√	√	√	√	√
STNMMH1	Number of MH treatment facilities in a state	√	√	√	√	√	√	√	√	√
STMHSAT1	Number of MH treatment facilities offering substance abuse treatment in a state	√	√	√	√	√	√	√	√	√
STOPMH1	Number of outpatient MH treatment facilities in a state	√	√	√	√	√	√	√	√	√
STOPSAT1	Number of outpatient MH treatment facilities offering substance abuse treatment in a state	√	√	√	√	√	√	√	√	√

Variable Name	Variable Description	Data Availability by Year								
		2011	2012	2013	2014	2015	2016	2017	2018	2019
STIPMH1	Number of inpatient MH treatment facilities in a state	√	√	√	√	√	√	√	√	√
STIPSAT1	Number of inpatient MH treatment facilities offering substance abuse treatment in a state	√	√	√	√	√	√	√	√	√
STRSMH1	Number of residential MH treatment facilities in a state	√	√	√	√	√	√	√	√	√
STRSSAT1	Number of residential MH treatment facilities offering substance abuse treatment in a state	√	√	√	√	√	√	√	√	√
STNMSA2	Number of substance abuse treatment facilities in a state	√	√	√	√	√	√	√	√	√
STSAMH2	Number of substance abuse treatment facilities offering MH treatment in a state	√	√	√	√	√	√	√	√	√
STOOSA2	Number of outpatient substance abuse treatment facilities in a state	√	√	√	√	√	√	√	√	√
STOPMH2	Number of outpatient substance abuse treatment facilities offering MH treatment in a state	√	√	√	√	√	√	√	√	√
STIPSA2	Number of inpatient substance abuse treatment facilities in a state	√	√	√	√	√	√	√	√	√
STIPMH2	Number of inpatient substance abuse treatment facilities offering MH treatment in a state	√	√	√	√	√	√	√	√	√
STRSSA2	Number of residential substance abuse treatment facilities in a state	√	√	√	√	√	√	√	√	√

2. Non-SAMHSA Restricted Use Data Dictionary

NCHS restricted NHANES 2003-2006 data dictionary.

Variable	Variable Description	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

Variable	Variable Description	2003-2004	2005-2006
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 prewritten 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-SAMHSA public data dictionary:

Data dictionary of public use MEPS-prescription file from AHRQ

No.	Name	Variable Description	2013	2014	2015	2016	2017	2018	2019
1	DUID	PANEL # + ENCRYPTED DU IDENTIFIER	X	X	X	X	X	X	X
2	PID	PERSON NUMBER	X	X	X	X	X	X	X
3	DUPERSID	PERSON ID (DUID + PID)	X	X	X	X	X	X	X
4	DRUGIDX	DRUG ID (DUPERSID + COUNTER)	X	X	X	X	X	X	X
5	RXRECIDX	UNIQUE RX/PREScribed MEDICINE IDENTIFIER	X	X	X	X	X	X	X

No.	Name	Variable Description	2013	2014	2015	2016	2017	2018	2019
6	LINKIDX	ID FOR LINKAGE TO COND/OTH EVENT FILES	X	X	X	X	X	X	X
7	PANEL	PANEL NUMBER	X	X	X	X	X	X	X
8	PURCHRD	ROUND RX/PRESCR MED OBTAINED/PURCHASED	X	X	X	X	X	X	X
9	RXBEGMM	MONTH PERSON STARTED TAKING MEDICINE	X	X	X	X	X	X	X
10	RXBEGYRX	YEAR PERSON STARTED TAKING MEDICINE	X	X	X	X	X	X	X
11	RXNAME	MEDICINE NAME (IMPUTED)	X	X	X	X	X	X	X
12	RXDRGNAM	MULTUM MEDICINE NAME (IMPUTED)	X	X	X	X	X	X	X
13	RXNDC	NDC (IMPUTED)	X	X	X	X	X	X	X
14	RXQUANTY	QUANTITY OF RX/PRESCR MED (IMPUTED)	X	X	X	X	X	X	X
15	RXFORM	DOSAGE FORM (IMPUTED)	X	X	X	X	X	X	X
16	RXFRMUNT	QUANTITY UNIT OF MEDICATION (IMPUTED)	X	X	X	X	X	X	X
17	RXSTRENG	STRENGTH OF MEDICATION (IMPUTED)	X	X	X	X	X	X	X
18	RXSTRUNT	UNIT OF MEDICATION (IMPUTED)	X	X	X	X	X	X	X
19	RXDAYSUP	DAYS SUPPLIED OF PRESCRIBED MED (IMPUTED)	X	X	X	X	X	X	X
27	RXFLG	NDC IMPUTATION SOURCE ON PC DONOR REC	X	X	X	X	X	X	X

J. Literature References:

Please provide a list of major references that are relevant to your project.

K. Resumes/C.V.:

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