

# Secondary Data Analysis

Use of Data to Promote Awareness of Health Status in Haitians

National Haitian American Health Alliance



1<sup>st</sup> October 2005

Randolph, MA

**Linda Marc, ScD, MPH, MS**

Fellow of Epidemiology in Psychiatry

Cornell Medical School, New York

&

Lecturer in Practice

Yale School of Public Health, Connecticut

# Outline of Talk

- ✿ What is Primary Data?
- ✿ What is Secondary Data Analysis?
- ✿ Dataset
- ✿ Benefits of SDA
- ✿ Limitations of SDA
- ✿ Potential Data Sources
- ✿ Example: CCHER Datasets & Hospital Records
- ✿ Promoting Awareness of Health Status in Haitians

# Experience in SDA

- ✿ FDA Spontaneous Adverse Event Reporting System Data
  - ✿ Comparative Analysis of Adverse Events
    - ✿ Psychiatric Drugs (Anti-Depressive vs. Anti-Obsessive Compulsive Disorder)
    - ✿ Anti-Ulcer Drugs (H2 Antagonists vs. Proton Pump Inhibitors)
    - ✿ Calcium Channel Blockers
- ✿ NIH AIDS Clinical Trial Dataset
  - ✿ Investigate Social Determinants of HIV Treatment Effectiveness
  - ✿ Evaluate data from Health Survey Instruments
  - ✿ Compare health outcomes across race and social class groups
- ✿ Macro-Level Country Specific Socio/Economic/Political Data
  - ✿ Investigating relationship between socioeconomic indicators and availability of communication technologies
- ✿ Hospital Medical Records
  - ✿ Compare clinical characteristics of Haitians to African-Americans, admitted to an Acute Psychiatric Unit

# What is a Primary Data?

- ✿ The collection of original information to answer a research question about the relationship between two or more variables.
- ✿ Primary collection requires:
  - ✿ Conceptualizing the study
  - ✿ Identified type of data to be collected
  - ✿ Theoretically defining the cause-effect relationship between data
  - ✿ Designing the data collection method
  - ✿ Creating the **dataset**
  - ✿ Executing the analysis

# What is Secondary Data Analysis?

The use of an existing **dataset**  
to answer a new question.

# What is a Data Set?

- ✿ Defined as a collection of information

Statistical measures (quantitative)

Verbatim words (qualitative)

Including

Exposures (which cause) → Outcomes (effects)

- ✿ Usually stored in an electronic file (EXCELL, WORD, SAS, SPSS)

# What is the Goal of the SDA?

- ✿ To investigate the relationship between variables, not previously examined in the primary data analysis

# Who Performs a SDA & Why?

- Most often public health researchers investigating issues of:
  - health status
  - health behaviors, and
  - health policy
- Either in a cross-sectional or longitudinal dataset
- Why?
  - To generate useful **statistical conclusions** that **describe the health status** of population groups; because the findings can **influence change** at multiple levels of society (micro, mezzo, and macro).

# Hypothetical SDA

## Metro Boston Area Dataset (N=5000 Subjects)

Blood Samples – CD4 Counts (immune data)

Self-Reported Perception of Stress

Neighborhood-Specific Socioeconomic Data

*Average Income, Education, Home Value*

# Primary Study Question

1. Do individuals who live in **low socioeconomic neighborhoods** experience higher self-reported **stress**, when compared to individuals who live in middle and higher SES neighborhoods?

↓ SES neighborhood → ↑ Stress

# Primary vs. Secondary Question

1. Do individuals who live in **low socioeconomic neighborhoods** experience a higher self-reported **stress**, when compared to individuals who live in middle and higher SES neighborhoods?

↓ SES neighborhood → ↑ Stress

2. Do individuals who live in **low socioeconomic neighborhoods** have **lower immune functioning**, when compared to individuals who live in middle and higher SES neighborhoods?

↓ SES neighborhood → ↓ Immune Functioning

**Same exposure, but different outcome  
for secondary data analysis**

# Primary vs. ~~Secondary~~ Question

**THIRD**

**New exposure, Previous outcome**

# Primary vs. ~~Secondary~~ Question

THIRD

New exposure, Previous outcome

3. Do individuals with higher self-reported **stress** have lower immune functioning?

↑ **Stress** → ↓ Immune Functioning

# Primary vs. ~~Secondary~~ Question

THIRD

New exposure, Previous outcome

3. Do individuals with higher self-reported **stress** have lower immune functioning? Will this relationship vary by neighborhood socioeconomic status?

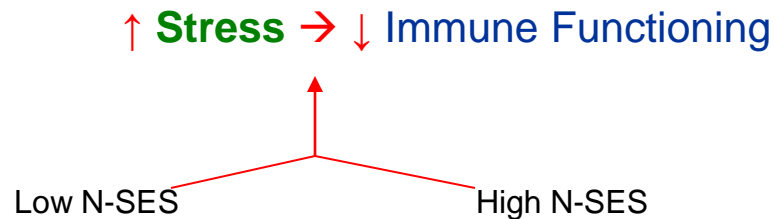
↑ **Stress** → ↓ Immune Functioning

# Primary vs. ~~Secondary~~ Question

THIRD

New exposure, Previous outcome

3. Do individuals with higher self-reported **stress** have lower immune functioning? Will this relationship vary by neighborhood socioeconomic status?



# Three Research Questions

↓ SES neighborhood → ↑ **Stress**

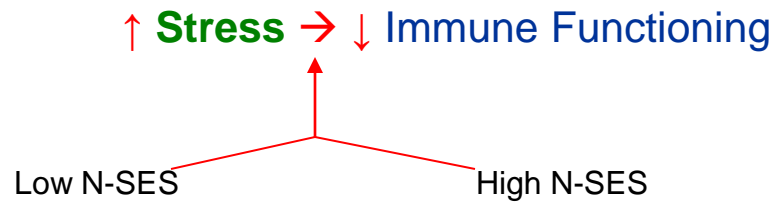
↓ SES neighborhood → ↓ Immune Functioning

↑ **Stress** → ↓ Immune Functioning

# Three Research Questions

↓ SES neighborhood → ↑ **Stress**

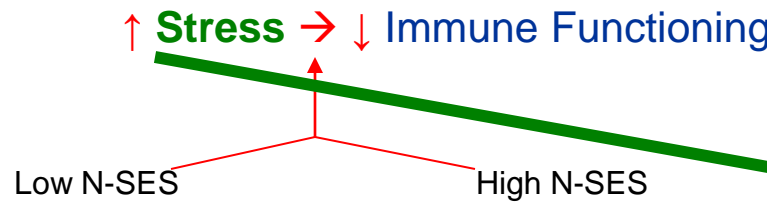
↓ SES neighborhood → ↓ Immune Functioning



# Three Research Questions

↓ SES neighborhood → ↑ **Stress**

↓ SES neighborhood → ↓ Immune Functioning



# What are the Benefits of SDA?

- ✿ Popular method among researchers who lack personnel and financial resources to collect original data
  - ✿ Accessible, and easier than original data collection
  - ✿ No need to identify study subjects (often difficult)
  - ✿ Requires less money for personnel and subject identification
  - ✿ In a timely fashion, research questions can be answered  
*provided the variables of interest are available*

# What are the Benefits of SDA?

- ✿ Many datasets provide the benefits of a representative sample, standard measures, questions or health instruments
  - ✿ Unobtrusive measures (e.g., height, weight, blood pressure)
  - ✿ Biological measures (e.g., serum blood for cholesterol, diabetes)
  - ✿ Psychological Measures (Depression Scale, Quality of Life Scale)
  - ✿ Disease Specific Measures (e.g., HIV Symptom Distress Questionnaire)

# What are the Limitations of SDA?

- ✿ UnPopular because:

- ✿ Data is NOT accessible

- ✿ Stored in electronic file but no manpower to access information

- ✿ Sometimes EXPENSIVE

- ✿ To create new electronic file including specific variables for SDA

- ✿ Takes too LONG

- ✿ Original collected at multiple sites (i.e., local health centers) and merging data is time consuming

# What are the **Limitations** of SDA?

## Coding Limitations

- ✿ Want to generate average age of individuals in the dataset
  - ✿ **Dataset A** has individual ages
  - ✿ **Dataset B** has month, day, and year of birth
  - ✿ **Dataset C** has age coded (0-10)
    - Reflects groupings for every 10-year category (ages 0 to <100).

# What are the Limitations of SDA?

## Poor Data Quality

- ✿ Missing Data

- ✿ Unknown vs. Refused to Answer

- ✿ How do you code this as the primary researcher?
    - ✿ How do you interpret this in a secondary dataset?
    - ✿ Will you be informed that the code “99” in the secondary dataset reflects both?

# What are the **Limitations** of SDA?

## Poor Data Quality

### ✿ No Codebook

- ✿ Research Director fails to create a codebook for research staff
- ✿ Staff create their own codes to record information
- ✿ EXAMPLE:
  - High School education verbally reported in years completed.
  - ✿ Staff A records data as (HS=12);
  - ✿ Staff B records it as (0 or 1; 1=completed HS)

# What are the Limitations of SDA?

## Uncooperative Collaborators

- ✿ Lack of trust to share
  - ✿ Data
  - ✿ Ideas
- ✿ Issues of Fame & Reputation
  - ✿ Publish or Perish (hording)

# What are the **Limitations** of SDA?

## Using Data from Multi-Site Studies

- ✿ European & US Combined Dataset
  - ✿ Never told that data was captured using difference methodologies across countries, or that non-validated questionnaires were administered in multiple languages.

# What are the Limitations of SDA?

- ✿ Has the data been standardized?
- ✿ Standardizing measures, methods and capture tools
  - ✿ Unobtrusive measures

Height: Metric vs. Inches

Weight: kgs vs. lbs

BP: Sitting, standing, averaged over 3 times  
Calibration of Pressure Gauges

Body Temp: Anal, Mouth, Ear  
Glass Thermometer vs. Electronic Device

# What are the **Limitations** of SDA?

- **Standardizing** measures, methods and capture tools
  - ✿ Biological Measures → Calibration of Instruments?
  - ✿ Depression Instrument → Same Language, Version, Validated?

# What are the Challenges of SDA?

- ✿ Computer Skills
- ✿ Data Management
- ✿ Logical Reasoning
- ✿ Programming
- ✿ Statistics
- ✿ Substantive Knowledge of Content Area
  
- ✿ **AND, availability of data, which includes desirable information that helps to ask research questions not previously investigated**

Lets reflect on what we are  
really talking about



Are their available datasets  
including health status on Haitians  
that we can analyze?



There are FEW

# Public Health Data Sources

## Is Country of Birth Collected?

### Mortality

- Vital Records
- Specific Diseases – AIDS
- Infant & Fetal Mortality
- Accidents
- Homicide
- Occupational Fatalities

### Morbidity

- Hospital Discharge Data
- Surveillance Systems
- Infant & Fetal Mortality
- Accidents
- Specific Diseases-AIDS
- Occupational Morbidity
- Blood Disorders

### Nativity & Maternal & Child Health

- Nativity
- Fertility

### Census

- Census Track Data
- Employment
- Demographics

### Nutritional Surveillance

- Nutritional evaluations
- Food Consumption

**Can Scientists in the Haitian Community  
Collaborate to Advance  
the Health Status of US Haitians?**



**National Haitian American Health Alliance**

# Overcoming Historical and Institutional Distrust in the Boston Haitian Community

REACH 2010 for HIV

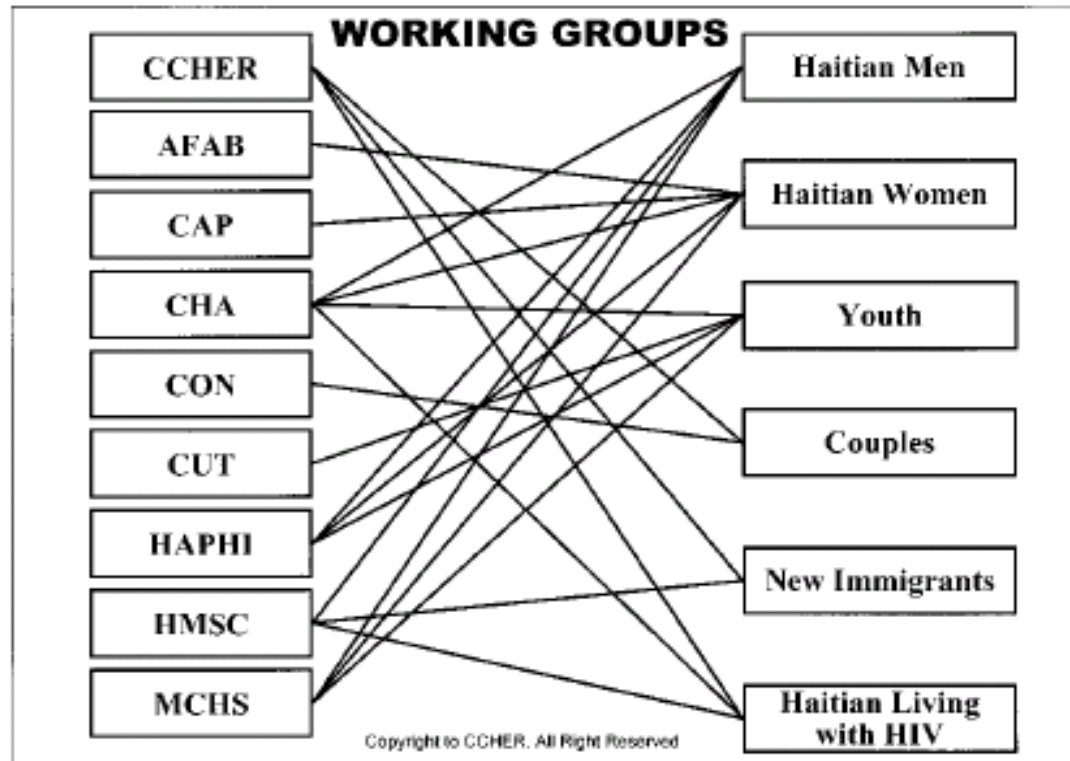


Fig 1. Coalition members

AUTHORS:

Metayer, Jean-Louis, Madson in Ethnicity & Disease, 2004

# What are the Barriers to Secondary Data Analysis on US Based Haitians

- ◆ Only 6% of U.S. Haitian households speak English exclusively.
- ◆ English-only Health Questionnaires
- ◆ Low Reading Ability
- ◆ Low Health Awareness (misinformation)
- ◆ Low Computer Literacy
- ◆ Distrust of the Healthcare System

## Solution

Steps	Products
I. Translation	Survey form
II. Scale construction	Scoring algorithms
III. Validations and norming	Interpretation

# What are the Barriers to Secondary Data Analysis on US Based Haitians

Funders think Creole is a difficult language barrier and are reluctant to fund grants.

- English-only Health Questionnaires

- Low Reading Ability

- Low Health Awareness (misinformation)

- Low Computer Literacy

- Distrust of the Healthcare System

## Solution

Steps	Products
I. Translation	Survey form
II. Scale construction	Scoring algorithms
III. Validations and norming	Interpretation

# What are the Barriers to Secondary Data Analysis on US Based Haitians

- ◆ Only 6% of U.S. Haitian households speak English exclusively.

- ◆ English-only Health Questionnaires

- ◆ Low Reading Ability

- ◆ Low Health Awareness (misinformation)

- ◆ Low Computer Literacy

- ◆ Distrust of the Healthcare System

- ◆ Translate Health Questionnaires

- ◆ Increase Reading Ability

- ◆ Assess Health Awareness (misinformation)

- ◆ Develop Computer Skills

- ◆ Distrust of the Healthcare System