



# **Adventist Health Study (AHS2)**

## **Pilot linkage study**

**US and Canada-wide**

**Special population with diet/cancer focus**

**Research opportunity for NAACCR members**

Susan Preston-Martin, PhD<sup>1</sup>

Gary Fraser, MD, PhD<sup>2</sup>, Larry Beeson, DrPH<sup>2</sup>

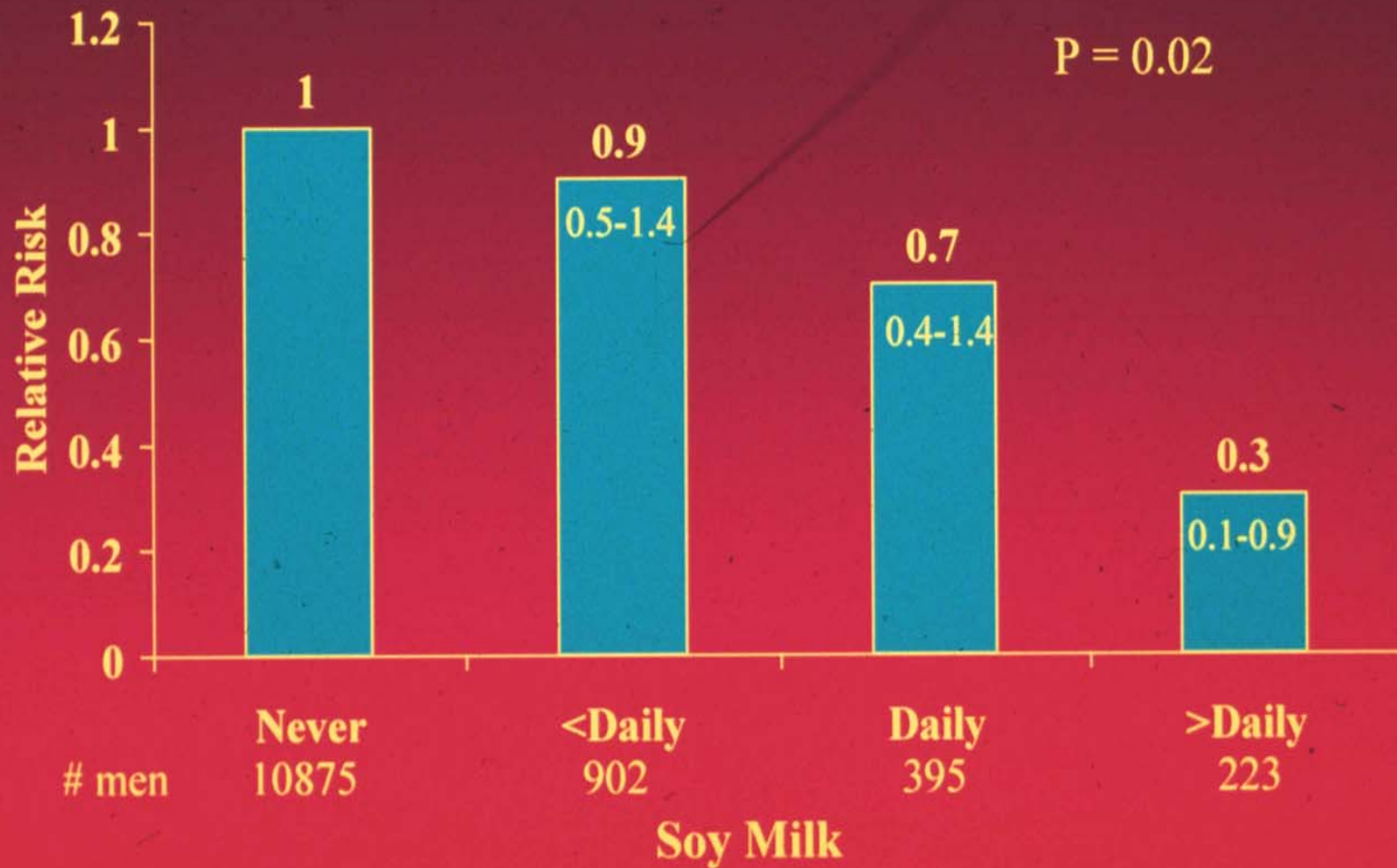
<sup>1</sup> University of Southern California, Los Angeles, CA

<sup>2</sup> Loma Linda University, Loma Linda, CA

# Goals of the Research

1. Effect of consuming **soy** products on risk of **colon, breast** and **prostate** cancers.
2. Effect of dietary and supplemental **calcium** on risk of these cancers.
3. Effect of **meat** and **fat** consumption on risk of these cancers.
4. Particular focus on **African-Americans**

# Soy Milk and Incidence of Prostate Cancer



# AHS 2 Methods

1. To enroll **70,000 White and 35,000 Black** Adventists across US and Canada.
2. Recruitment **church-by-church** (>4000 churches)
3. Also National promotional campaigns using church media.
4. Enrollment over a 6-year period.

# Enrollment simply requires completion of the questionnaire

- Past medical history
- **Diet**
- Physical activity
- Female questions
- Demographic questions
- **Tear-off section for personal information**

AHS-2



# ADVENTIST HEALTH STUDY - 2

## Connecting Lifestyle to Disease and Longevity

- a continuation of previous research on Adventists

Please return to:

Adventist Health Study - 2  
Loma Linda University  
Evans Hall - Room 203  
Loma Linda, CA 92350

**IT'S NOT AS LONG AS IT LOOKS!**  
See note on next page

7608201

AHS-2



# ADVENTIST HEALTH STUDY - 2

## Connecting Lifestyle to Disease and Longevity

- a continuation of previous research on Adventists

Please return to: Adventist Health Study - 2  
Loma Linda University  
Evans Hall - Room 207  
Loma Linda, CA 92350

**IT'S NOT AS LONG AS IT LOOKS!**  
See note on next page

Are you age 30 or older? If so, please complete this card now, tear it out, and give it to the Adventist Health Study coordinator. Thank you.

Name: \_\_\_\_\_  
first middle last

Address: \_\_\_\_\_

\_\_\_\_\_ city state zip

Telephone: ( \_\_\_\_\_ ) \_\_\_\_\_ area code Date of Birth: \_\_\_\_\_ month/day/year

Local church at which you hold membership: \_\_\_\_\_

DO NOT MARK BELOW THIS LINE



\* 0 1 1 5 5 1 \*

7608101

# A. MEDICAL HISTORY



For office use only

1. Would you say, in general, your health is:

- Excellent       Fair  
 Good             Poor

2. During the **past 12 months**, have you had influenza (flu)?

- Yes       No

3. During the **past 12 months**, how many times have you had:

Upper respiratory infection (cold, sore throat, sinusitis).

- None     1-2     3-4     5-6     7+

Gastroenteritis (stomach flu, food poisoning)

- None     1-2     3-4     5-6     7+

Bronchitis

- None     1-2     3-4     5-6     7+

4. Has a physician ever told you that you had any form of cancer (including leukemia, lymphoma, myeloma and skin cancer)?

No    ►► Go to Question 5, on the next page

Yes    ►► If YES, **write-in** site of cancer(s) (e.g. lung, colon, etc.) and approximate year it was first diagnosed in the space(s) below. **Also fill in the matching circles** below the year first diagnosed. Notice that there is space for up to three different cancers.

Please use CAPITAL LETTERS.

**A**

Write Site of cancer: Breast

Year first diagnosed: 1987

Fill in circles matching the numbers that you wrote above.

<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**B**

Write Site of cancer: Colon

Year first diagnosed: 2000

<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**C**

Write Site of cancer: Liver

Year first diagnosed: 2001

<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1

2

3

4

5

6

7

8

9

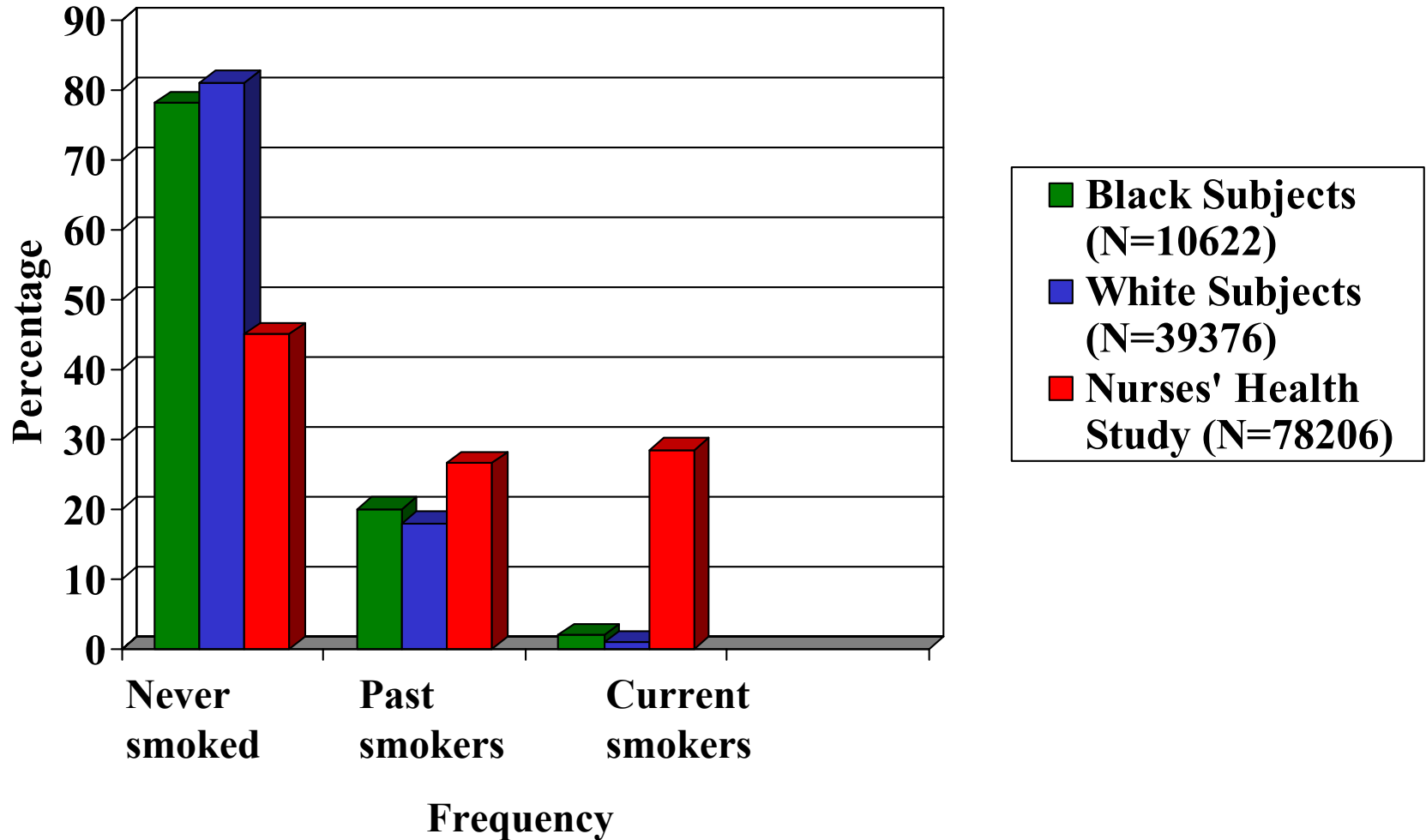
10

11

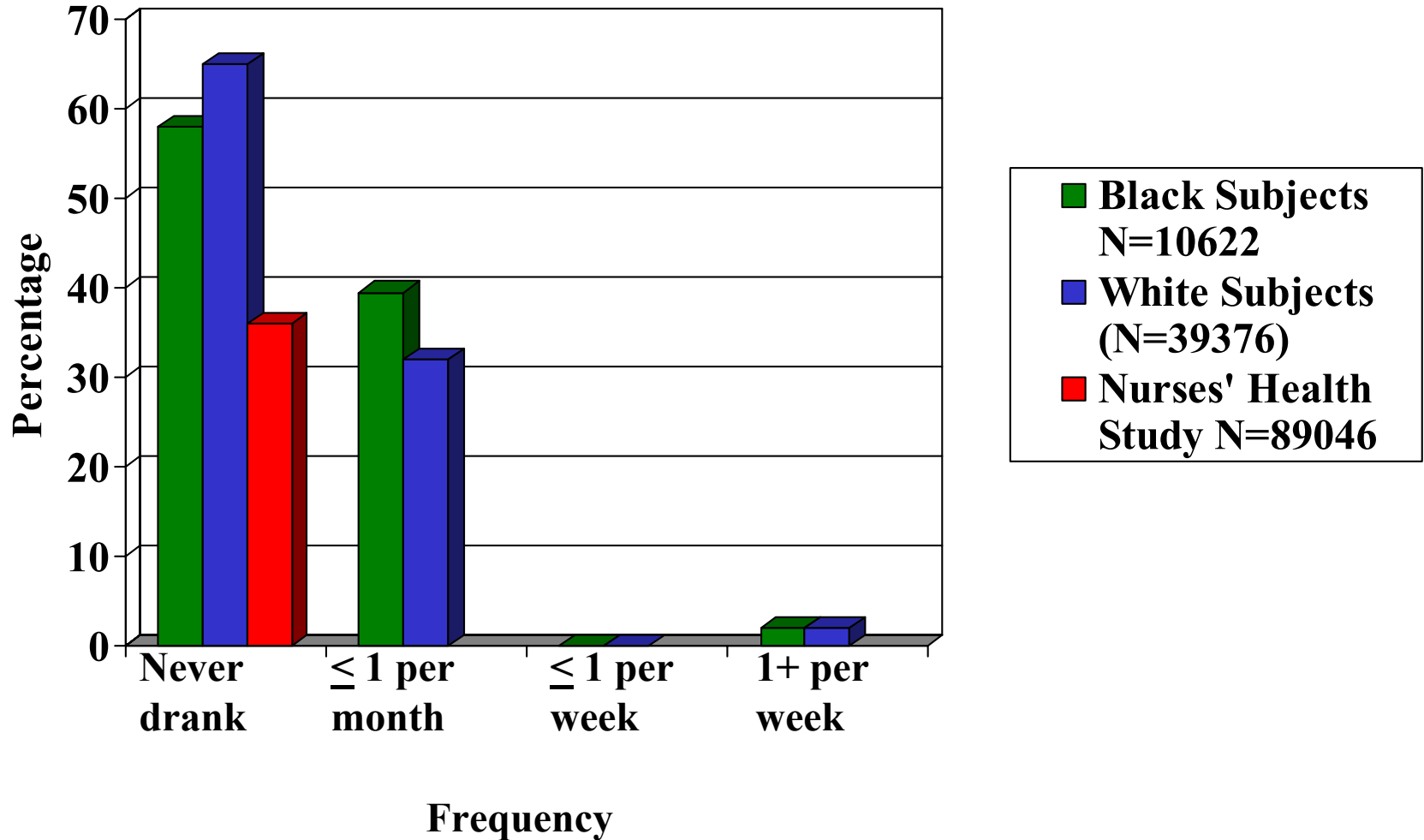
12



# Smoking

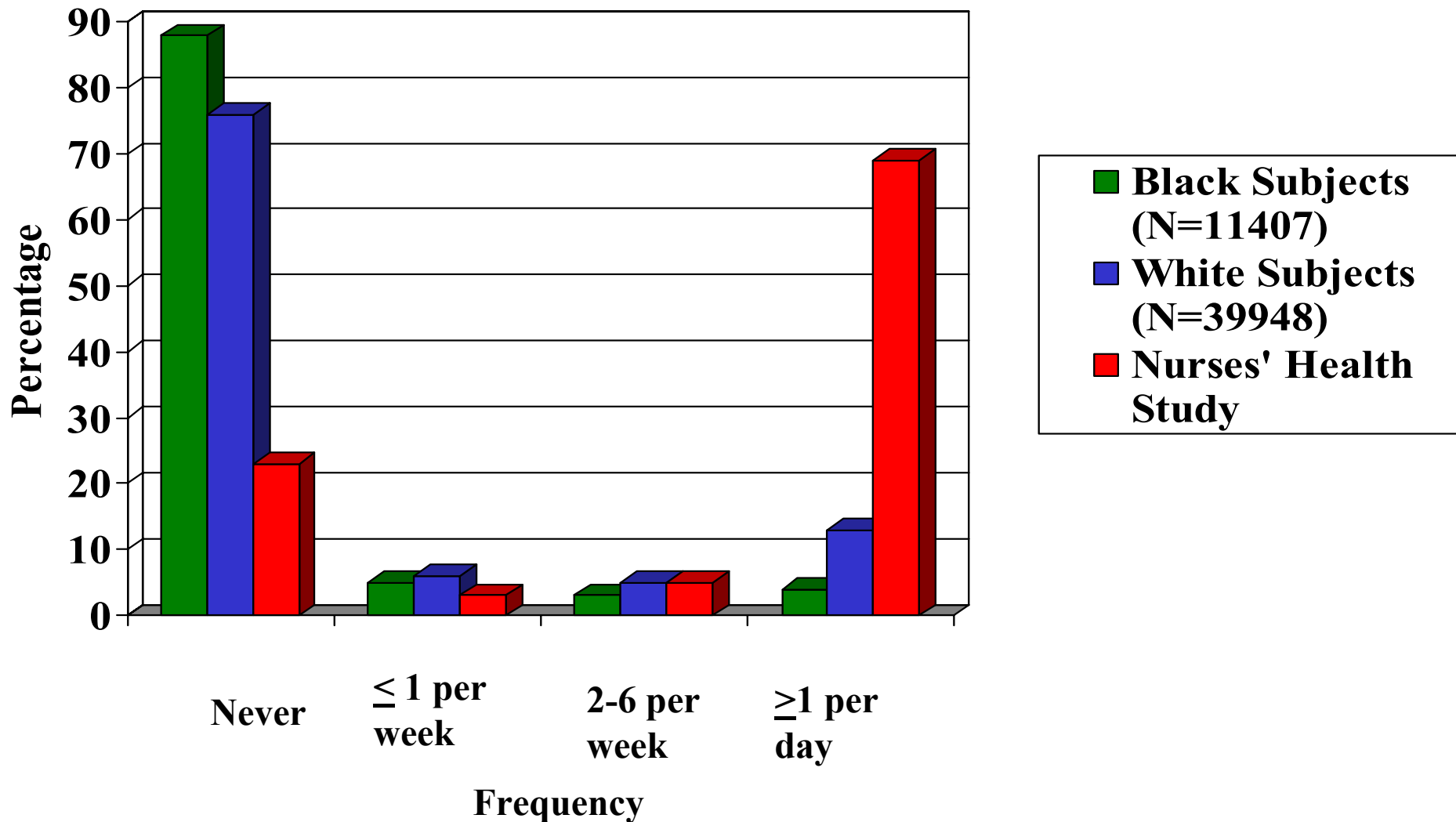


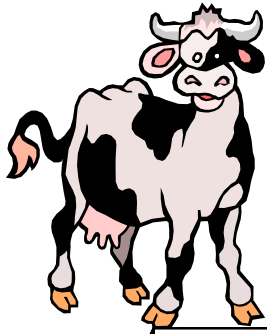
# Alcohol Use



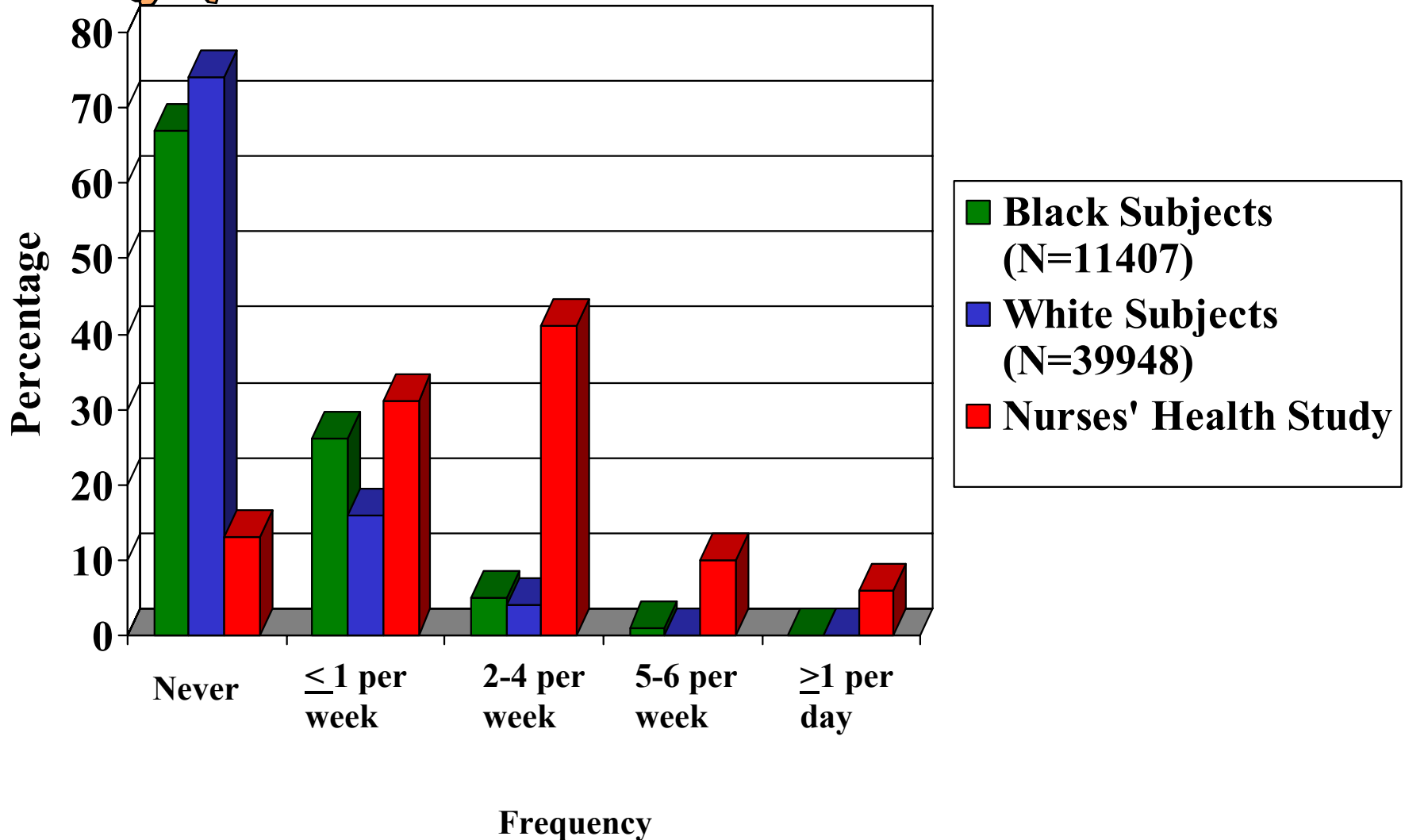
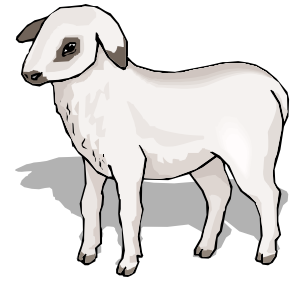


# Coffee





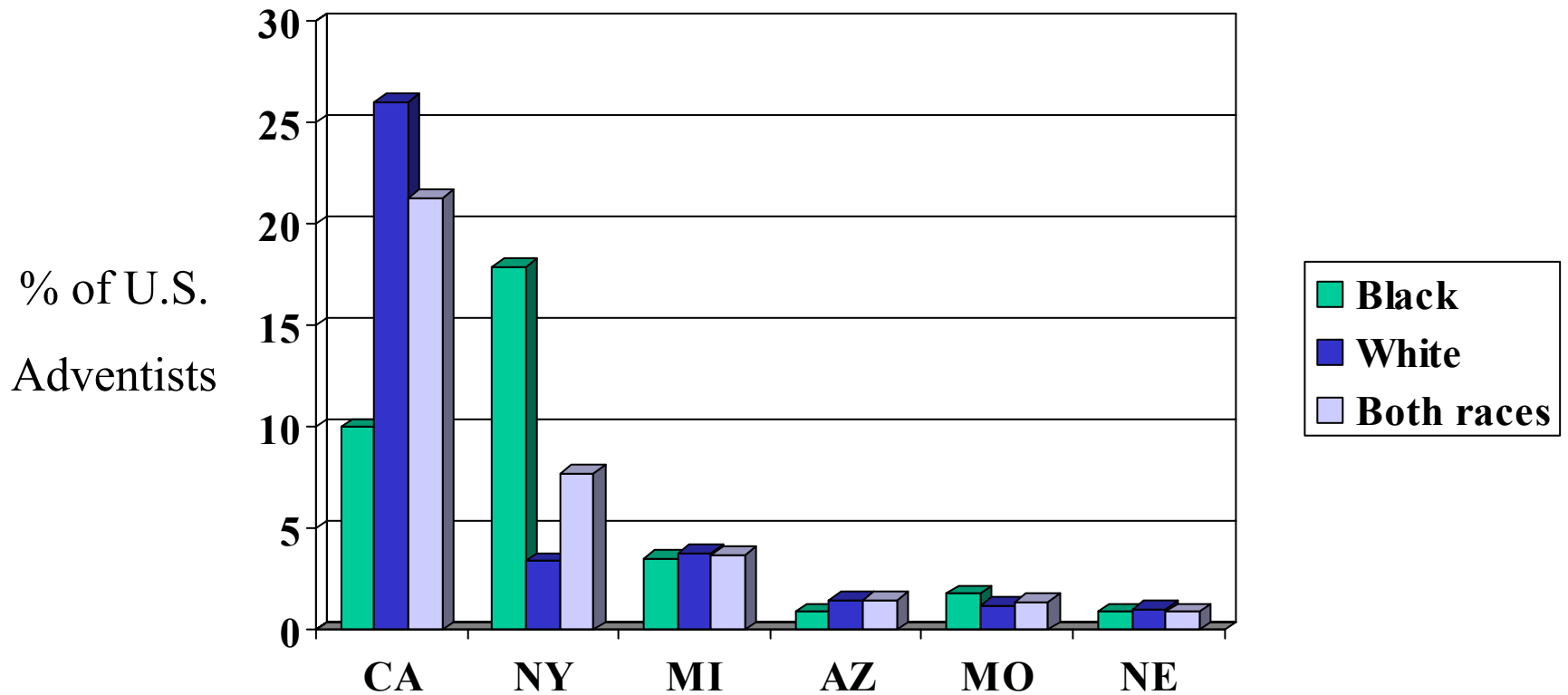
# Beef, Lamb as Main Dish



# Pilot study of linkage to registry files

- **Why do?**
  - To **work out details** of linking AHS2 to registry files – develop common procedures
  - **Confirm self-reported prevalent cancers**
  - Show we can do **for grant renewal** application
- **Why these six states?**
  - **Recruitment finished** by 2004; **#s of both** black and white Adventists
  - **Range of states** by size, region, age of registry

# Percent of U.S. Adventists by state and race, AHS2 pilot linkage to state cancer registries planned for late 2004; completed late May 2005



# Obtaining approval at pilot registries

- Time (first contact to final) - **4-16 months**
- **Longer than expected** in all 6 states
- **Reasons for unexpected delay**
  - IRB form changed
  - Registry director changed
  - Additional approvals needed after IRB
  - Staff change
  - Application forgotten in drawer
  - Additional info requested after initial review

# Traveling Linkage Programmers

- **AHS-2 programmers** work at each registry
- On-site linkage on **AHS laptop** PC
  - using Link Plus software
  - AHS-2 files on PC
  - **PC left for registry use**
- Only cancer data on matches taken to LLU
- Linkage done with or without supervision by registry personnel



# Link Plus record linkage software

---

- Developed at **CDC**
- Computes **probabilistic linkage scores** ( < 5; 5-24; **25+**)
- Facilitates a simple and efficient **“OR” blocking mechanism** by matching pairs on at least one key variable
  - AHS2 used **SS#, surname, and birthdate**
- **Output** is match +/- text files **easily imported into Excel**
- **AHS2 supplementary program** run on all files with scores **5-24** and **displays linked pairs** on the screen for review:
  - **Definite** match or non-match
  - **Still undecided** (i.e. query) – often missing data

# Link Plus Comparisons

---

- **Frequency-based sets weights** for matching score :
  - **Frequent** value → **low** weight
  - **Rare** value → **high** weight
- **Last and first name** comparison uses **both partial and frequency-based matching** - typos, misspellings, hyphens
- **Middle name** either full name or initial OK
- **SSN** comparison **allows partial matching** to account for typos and transposition of digits
- **Date** comparison **allows partial matching** if missing values

# AHS2 Custom Matching Code

The matching code is composed of 12 digits.

Each digit represents result from comparing a pair of fields.

Example Code: 210191117891

## Codes

0 = No match

1 = Good match

2 = Probable match

7 = Missing CR data

8 = Missing AHS data

9 = Missing both

## Position / Description

- 1) First name
- 2) Last name
- 3) Middle name
- 4) Gender
- 5) Date of birth
- 6) Social security number
- 7) Address number
- 8) Address street name
- 9) City
- 10) Zip
- 11) Phone area code
- 12) Phone local code

# Clerical review of uncertain matches

**Manual Resolution** \_ □ ×

**ID's**   **Link Plus Score**  **Matching Code**  **1 of 1 records**

**FirstName**   **LastName**   **MiddleName**   **Sex**

**DOB**   **SSN**   **Address**

**City**   **Zip**   **Phone**

**Query**

# **Years linked at six pilot registries**

Range: all years **1981-2003**

(only MI link used all these years)

**first year (1981-1995)**

**most recent year (2001-2004)**

# Possible matches still unresolved

- **Range** [possible/(definite+possible)]  
across states (<5% - >80%)
- **Key missing data** (>50% of unresolved)
  - Address** (3 states)
  - Phone #** (3 states)
  - SS#** (AHS2 and 2 states)\*
  - Middle name** or initial (3 states)

\* 4 states if missing in AHS2, didn't record if also missing in registry file

# **Cancers that would have been missed if only matched at one state**

- **132 Adventists** with cancer (**8.5%**) total  
(range: 1-82; 3-14%)
- **144 Cancers** missed (**8.4%**) across 6 states
- **Need to match with entire AHS2 file**

# How good is self report of cancer?

Cancers identified by **registry only** – **268/1657 (16%)**  
(46 breast, 31 colon, 50 prostate, 141 other)

## **Accuracy** of confirmed self reported cancer **1657**

<b>Site</b>	same	1389	<b>84%</b>
	different	56	
	cancer not self-reported	212	

## **Year** of diagnosis

	exact	978	<b>59%</b>
	+/- 1 year	278	
	> 1 year	203	
	cancer not self-reported	198	



# Sensitivity by cancer site

(#self reports / #registry matches)

• Breast	<b>91.0%</b>	<b>(466/512)</b>
• Colon, rectum, anus	80.3%	(126/157)
• Prostate	88.6%	(387/437)
• All other (except skin)*	72.6%	(321/442)
• Total cancers	<b>83.8%</b>	<b>(1389/1657)</b>

\* i.e., other cancers excludes malignant melanoma  
and other skin cancer

# Positive Predictive Value by Site

(#registry matches / # self reports)\*

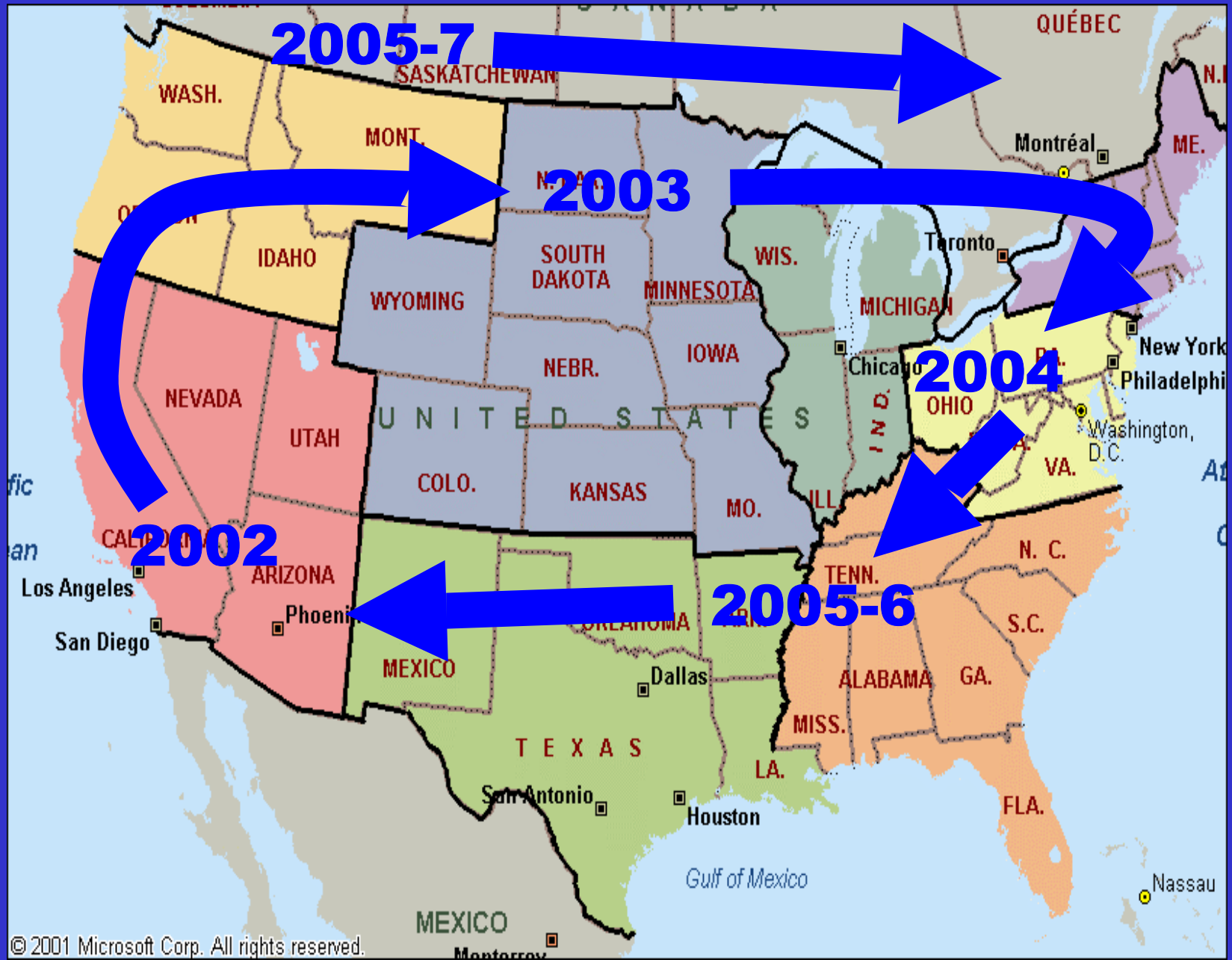
- Breast
- Colon
- Prostate
- All other (except skin)
- Total cancers

\* among those who lived within 10 miles of current residence during match period

# Variation by Demographics

(race, age, sex, year, and state)

# Schedule for Recruitment



# Years we hope to do links

- First link
  - already done in 2005 in 6 pilot states
  - 2006-2007 in rest of US and Canada
- Link to identify incident cancers
  - 2008-2010

# Measures to assure Confidentiality during the linkage process

- Only **personal identifiers** we'll take away will be ones we brought on AHS2 file
- Will bring **AHS2 back-up file to resolve dubious matches**
- **Portable AHS2 PC** used for match will be **left at registry as payment for assistance**
- Disk made will only have cancer data on matches and AHS id #s; reg. staff can check.

# Confidentiality of cancer data taken to LLU

- **No personal identifiers on any electronic file** that also has lifestyle or cancer data
- Study **ID # can be linked** to name and address **only by 2** senior data managers.
- When **junior staff** work with data only out of range or unlikely responses show on PC.
- Thousands of subjects contacted about this; none have expressed concern

# What have we learned?

- 1) IRB approval takes longer than expected
- 2) Traveling programmer model works
- 3) Missing data is key issue
- 4) Level of missing registry data varies widely
- 5) Cancer self-reports need registry details
- 6) PPV
- 7) Registry and IRB staff most helpful

**THANK YOU ALL!**



# Summary

- A **diet- cancer study** in a special population whose results can be generalized.
- Participants **across the U.S. and Canada.**
- **Opportunity to use NAACCR member registry data in a uniform manner** to greatly enhance validity and efficiency of this important research.
- **A new era for cohort studies** of NAACCR member collaboration continent- wide.

# Acknowledgments

**AZ** Georgia Yee  
Chris Newton  
Keith Laubham

**CA** Bill Wright  
Mark Allen  
David Harris

**MI** Glenn Copeland

**LLU** Larry Beeson  
Lars Sveen

**MO** Jeanette Jackson-Thompson  
Mary Jane King  
Sue Vest  
Saba Yemane

**NE** Kurt Weiss  
Judy Paradies

**NY** Maria Schymura  
Amy Kahn