Web-Enabled Coding Procedures in the National Survey on Drug Use and Health

Wafa Handley, Larry Kroutil, David Smarrella, Chuchun Chien

2005 International Field Directors and Technology Conference

Miami, Florida May 15 - 18, 2005

3040 Cornwallis Road ■ P.O. Box 12194 ■ Research Triangle Park, North Carolina, USA 27709

About the Project

- The National Survey on Drug Use and Health (NSDUH) – formerly NHSDA, is conducted by RTI, International for the Substance Abuse and Mental Health Services Administration (SAMHSA).
- Primary source of information on the prevalence and patterns of alcohol, tobacco, and illegal drug use and misuse of prescription medications.
- Administered to the general U.S. civilian non-institutionalized population, aged 12 or older.
- Data are collected quarterly and yield a final sample of approximately 67,500 persons each year.

Challenges

- Use of open-ended questions in NSDUH to capture information on
  - other drugs
  - other tobacco brands
  - other behaviors of interest (e.g., arrests, utilization of mental health services)
- Need to code responses numerically to facilitate data processing and analysis.
- High volume of responses for each survey year.

GOAL

To develop an automated system to

- reduce coder burden
- standardize coding procedure
- reduce turnaround time from receipt of data from the field to delivery of coded data sets.

How System Accomplishes Goals

- Allows for multiple analysts, with appropriate access, to code and review the data
- Utilizes responses that have been entered in previous years
  - Automatically assigns standardized codes for repeated responses
  - Maintains consistency and reduces coder burden
- Allows coding and updating the drug and tobacco brand data to proceed in a flow basis throughout the year
  - Reduces turnaround time for coding and processing

System Features

- We needed web pages that look like this (split screen)
- Grid design
- Log/tracking table
  - Populated each time entries are submitted
  - Stores date/time, user name and type of transaction
  - Can be used as a QC and a training tool (log-table).

Supporting Software

Intranet web site utilizes

- SAS V8.2 for initial processing of the data
- Microsoft Visual Basic 6 (MS VB6) for populating the SQL Server tables which store the data
- Cold Fusion CF software version 5
  - To display the data on the web
  - Perform other web utilities (e.g., wild card searches, navigation between tables).

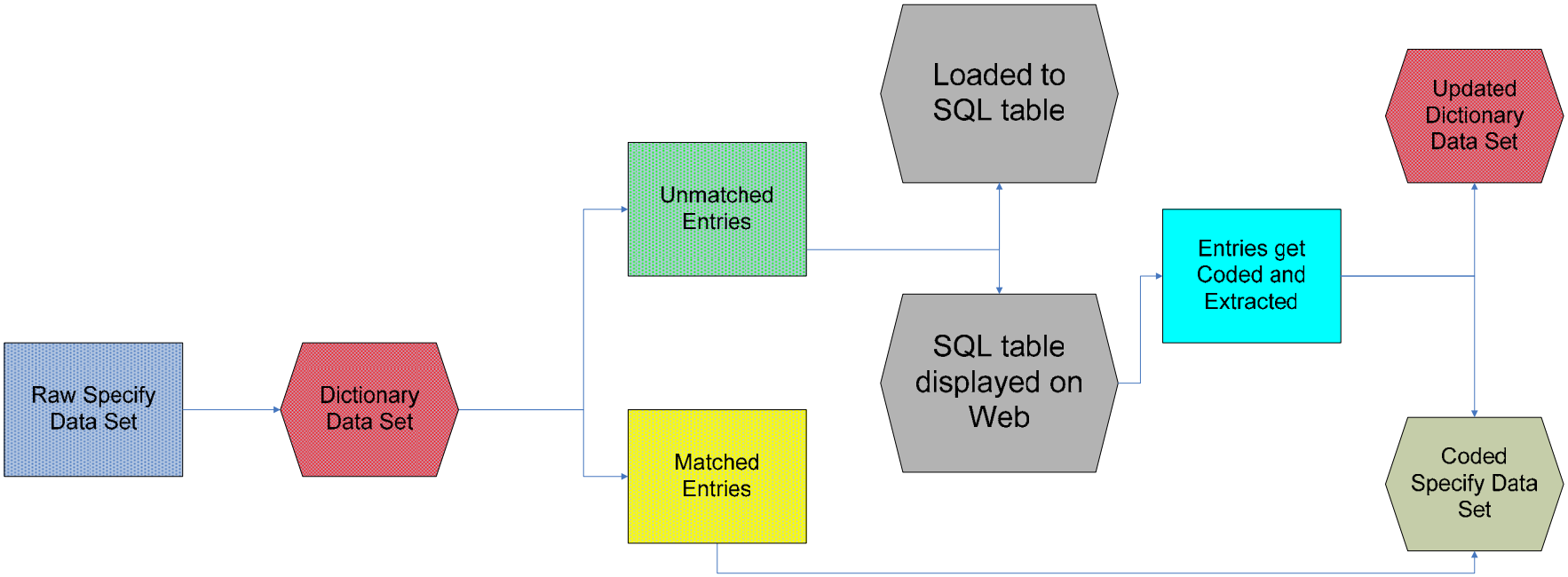
Data Flow

1. Daily transmitted data are processed and subset for cases where respondents have typed in specific drugs or tobacco brands.

2. Data are run through existing dictionary to code matched entries.

3. Data that do not match an existing entry are output for coding.

Data Flow


Raw Specify Data Set
Dictionary Data Set
Unmatched Entries
Matched Entries
Loaded to SQL table
SQL table displayed on Web
Entries get Coded and Extracted
Updated Dictionary Data Set
Coded Specify Data Set

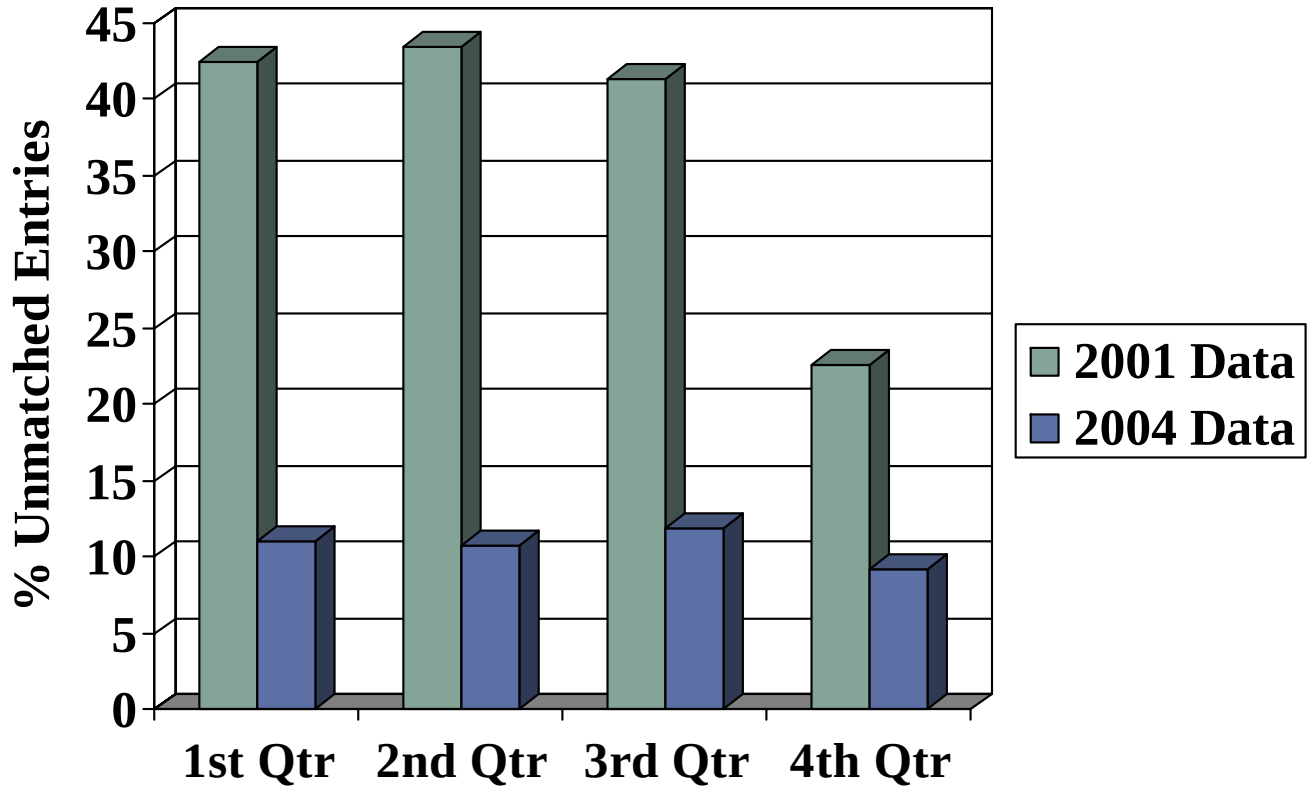
Results for 2004

	Number	Percent
Interview records processed	67,760	--
Open-ended drug responses	4,416	100.0
▪ Coded through the dictionary	3,940	89.2
▪ Coded through website	476	10.8
Open-ended tobacco brand responses	2,526	100.0
▪ Coded through the dictionary	2,375	94.0
▪ Coded through website	151	6.0

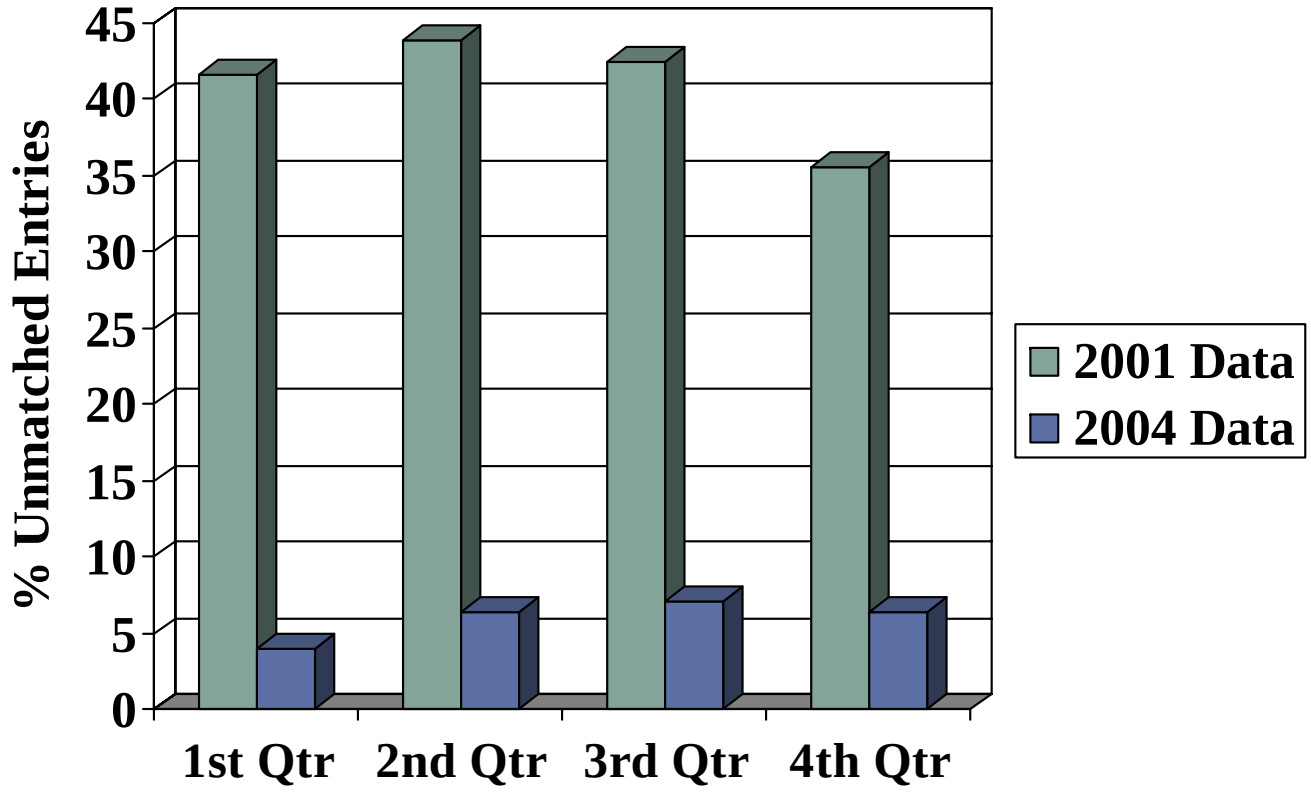
Comparative Results for 2001

	Number in 2001	Percent in 2001	Percent in 2004
Interview records processed	68,929	--	--
Drug responses	2,973	100.0	100.0
▪ Coded through dictionary	1,852	62.3	**89.2**
▪ Manually coded	1,121	37.7	**10.8**
Tobacco brand responses	2,093	100.0	100.0
▪ Coded through dictionary	1,236	59.0	**94.0**
▪ Manually coded	857	41.0	**6.0**

Drugs
2001 Compared to 2004



Tobacco
2001 Compared to 2004



Summary

Meeting our Goals

- ✓ Reduce coder burden

  number of cases that coders need to review has been reduced by 26.9% for drugs and 35% for tobacco brands.

- ✓ Standardize coding procedure

  coding is more consistent due to frequent updates of data dictionaries.

- ✓ Reduce turnaround time from receipt of data from the field to delivery of coded data sets.

  coding is completed and data are processed at the end of each collection period.

Questions or Comments?