

# **CHAD USER'S GUIDE**

## **Extracting Human Activity Information from CHAD on the PC**

June 30th 2000

Written for EPA National Exposure Research Laboratory

by

ManTech Environmental Technologies

**Casson Stallings**

**Jo Anne Tippet**

**Graham Glen**

**Luther Smith**

**Modified on March 22<sup>nd</sup>, 2002**

by

**Systems Development Center**

**Science Applications International Corporation**

# Table of Contents

List of Figures .....	v
List of Tables .....	vi
Abbreviations .....	vii
Introduction .....	1
CHAD Access: A Short Tutorial .....	2
The Basic Screens .....	2
Basic Respondent Info .....	2
Respondent Diary Info .....	4
Exiting Pages .....	5
Respondent METS Info .....	5
Filter Screen: Selecting a Cohort .....	5
Specifying Age .....	6
Clearing Queries .....	7
Specifying Gender .....	7
Specifying the State .....	7
Examining the Logical Expression .....	8
Saving and Restoring Logical Expressions .....	8
Viewing Selected Respondent Information .....	9
Creating Charts and Reports .....	9
Creating and Viewing Charts .....	9
Creating Formatted Reports for Printing .....	10
Exporting Data .....	11
Selecting the Variables to Export .....	11
File Formats .....	12
Moving On .....	12
What's Inside the CHAD Databases .....	14
Background on Individual Studies in CHAD .....	19
Activity Patterns in Cincinnati, Ohio (CIN) .....	19
Baltimore (BAL) .....	19
California Adults (CAA) and California Youth (CAY) .....	19
California Children (CAC) .....	19
Denver Study (DEN) .....	19
Los Angeles Studies (LAE,LAH) .....	20
University of Michigan (UMC) .....	20
National Human Activity Pattern Study (NHA, NHW) .....	20
Valdez Air Health Study (VAL) .....	20
Washington, DC (WAS) .....	20

Using Data from the Individual Studies .....	21
Finding the Correct Columns or Fields .....	22
Finding the Queries and the Coding .....	22
Defining Your Query .....	26
Caveats, Known Problems, Notes .....	28
Known Problems .....	28
CHAD Application .....	28
Nonrandom Sampling .....	28
California Adult (CAA) .....	28
NHAPS (NHA, NHW) .....	29
Data Notes .....	29
CHAD .....	29
California Adults (CAA) .....	29
California Youth (CAY) .....	29
Cincinnati (CIN) .....	29
NHAPS (HNA, NHW) .....	30
Washington DC (WAS) .....	31
Database Design and Development .....	32
Data Format and Content .....	32
Platform Selection .....	32
Development Process .....	33
Entering Data into CHAD Tables .....	33
Graphical User Interface Development .....	33
Querying CHAD Data .....	33
Code Consolidation .....	33
Activity and Location Codes .....	34
Standardization of Time .....	34
Questionnaire Data .....	35
Generation of body mass data from CHAD .....	35
MET Data .....	35
Quality Assurance Procedures .....	36
Completeness of Time of the Activity Diaries .....	36
Coding of Uncertain Activities .....	37
Coding of Missing Activities or Non-Sequential Activities .....	37
Descriptive Variables .....	37
RECCOUNT--The number of individual activity records in a diary day.	
.....	38
WRAPTIME--The time at which the 24-hour diary actually started.	
.....	38
Resolution of Quality Flags .....	38
Determination of Quality Flags .....	38
Quality Flag (QF) Variables .....	38
QFACTLOC--Indicates inconsistent activity-location pairs in the diary.	
.....	38
QFTRAVEL--Flag indicating possible travel time inconsistencies. . .	38
QFINFER--Indicates that a particular diary record was inferred (changed	
from the original). . . . .	39

QFMETAB--Indicates high (>3.0) metabolic rate factors. ....	39
Quality Count (QC) Variables .....	39
QCSLEEP--The amount of time spent sleeping. ....	39
QCMISS--Missing diary time (less than 24 hours). ....	39
QCMEALS--Number of meals in the diary day. ....	40
QCLONG--The duration of the longest activity in the diary. ....	40
Quality Flag Summaries .....	40
QCACTLOC--Total time (minutes) in diary with QFACTLOC flag set to "on". ....	40
QCEATIME--Total time (minutes) in diary spent eating. ....	40
QCINFER--Total time (minutes) in diary with QFINFER set to "on". ....	40
QCMETAB--Total time (minutes) in diary with QFMETAB set to "on". ....	40
QCHEAVY--Total time (minutes) in diary with HEAVYBR=1 (Heavy breathing). ....	40
References .....	42
Appendix 1: METS Distributions for CHAD Activities .....	44

# List of Figures

<b>Figure 1</b> Basic layout of CHAD Access screens and some of the buttons used to move from screen to screen..	3
<b>Figure 2</b> The <b>Record Navigation</b> widget.	4
<b>Figure 3</b> <b>Respondent Info</b> page (top only).	4
<b>Figure 4</b> The <b>Filter Screen</b> .	6
<b>Figure 5</b> The <b>Define New Filter</b> section of the <b>Filter Screen</b> .	6
<b>Figure 6</b> . The text lines that are used to input the minimum and maximum age range for a query.	6
<b>Figure 7</b> Pop-up form for selecting gender.	7
<b>Figure 8</b> Pop-up form for specification of states.	7
<b>Figure 9</b> The <b>Saved Filters</b> section.	8
<b>Figure 10</b> The <b>Filter Screen's Use Current Filter</b> section.	9
<b>Figure 11</b> The <b>Analysis Screen</b> .	9
<b>Figure 12</b> Sample page from generated report.	10
<b>Figure 13</b> The <b>Variable Selection</b> screen used when exporting diary data.	11
<b>Figure 14</b> The <b>Variable Selection</b> screen used when exporting questionnaire data.	11
<b>Figure 15</b> Box allowing the user to specify output file for the diary data.	12
<b>Figure 16</b> Box confirming that all of the data requested was written to the file.	12
<b>Figure 17</b> A schematic of the tables and relationships among the tables in the combined CHAD database.	14

# List of Tables

<b>Table 1</b> Examples of four text file formats that can be used when exporting questionnaire data. ....	13
<b>Table 2</b> CHAD Diary location code (LOC_CODE) definitions. ....	15
<b>Table 3</b> CHAD Diary activity code descriptions. ....	16
<b>Table 3</b> Questions relating to doors and windows in the two NHAPS studies. ....	21
<b>Table 4</b> Questions about doors and windows from the California, Cincinnati, and LA studies. ....	22
<b>Table 5</b> Location and coding of data relating to windows and doors in the individual studies' questionnaire data. ....	24
<b>Table 6</b> Location and coding of data relating to windows and doors in the individual studies' diary data. ....	25
<b>Table 7</b> The query c_qryNHA_Windows_Summary used to summarize the number of windows open in a household by season and temperature. ....	26
<b>Table 8</b> The query c_qryNHA_Windows_PreProcess that links several of the NHAPS data set tables, breaks the temperatures into 5 degree bins, extracts the month from the date, and selects those records where the window variables are defined; defined as being open, number opened, and time opened. ....	26
<b>Table 9</b> Results of the query c_qryNHA_Windows_Summary showing the average number of windows open by season and temperature. ....	27
<b>Table 10</b> Other descriptive CHAD variables. ....	37
<b>Table 11</b> Count of quality flags identifying activities or diaries with questionable or significant content. ....	41
<b>Table 12</b> METS distribution values used in CHAD. ....	44

## Abbreviations

APAI	Adjusted Physical Activity Index: a PAI calculated with MET values adjusted for oxygen debt and limits on maximum activity.
BAL	Baltimore Study
CAA	California Adults Study
CAC	California Children Study
CAY	California Youth Study
CARB	California Air Resources Board
CHAD	Consolidated Human Activity Database
CHADID	A unique identification number for each individual diary day in the CHAD database
CIN	Cincinnati Study
DC	Washington, DC Study
DDL	Data Definition Language
DEN	Denver Study
DNS	Domain Name Services
EPA	U.S. Environmental Protection Agency
EPOC	Excess Post-exercise Oxygen Consumption
GUI	Graphical User Interface
LAE	Los Angeles Elementary School Study
LAH	Los Angeles Highschool Study
MET	Ratio of an activity-specific metabolic rate to a person's resting metabolic rate
MM	Maximum MET value
NHAPS	National Human Activity Pattern Survey
NHEXAS	National Human Exposure Assessment Survey
PAI	Physical Activity Index: Time weighted average of MET values.
PEM	Personal exposure monitor
QC Variable	Quality count variable
UMC	University of Michigan Child Study
VAL	Valdez Alaska Study
WAS	Washington D.C. Study

# **Introduction**

The Consolidated Human Activity Database (CHAD) has been developed for the Environmental Protection Agency's National Exposure Research Laboratory. This database was created to support exposure, intake dose, and risk assessments. The overall design incorporates comments received from potential users knowledgeable about human activity pattern research and risk assessment.

CHAD is a relational database with a graphical user interface that facilitates queries and report generation. It contains databases from previously existing human activity pattern studies, which were incorporated in two forms: (1) as the original raw data and (2) as data modified according to predefined format requirements. The latter involved development of a common activity/location code system, compilation of background questionnaire information, and the application of data quality flags. To enhance the applicability to risk assessment, the ability to estimate metabolic activity ratios for respondents has been added. CHAD contains data obtained from pre-existing human activity studies that were collected at city, state, and national levels. The purpose of CHAD is to "provide a basis for conducting multi-route, multi-media exposure/intake dose/risk assessments" (McCurdy, 1997a). CHAD is typically used to provide input data for exposure simulation models such as HAPEM, SHEDS, SHEDSPM, and APEX.

Before we begin, a note of clarification. CHAD refers to two separate items: (1) the CHAD application, and (2) the CHAD database. The CHAD application is the software used to manipulate and provide a user interface to the data. The CHAD database is the one that contains data combined from several human activities studies. Sometimes a distinction will be made between these two entities, and other times it will not.



# CHAD Access: A Short Tutorial

## The Basic Screens

The major screens (or forms) in CHAD Access and their purposes are:

1. **Respondent Info:** display characteristics of an individual respondent.
2. **Diary Info:** display one day's worth of diary data for a respondent.
3. **METS Info:** display metabolic variables for one individual's day.
4. **Filter:** specify the query parameters to select groups of individuals.
5. **Analysis:** graph parameters of the query set selected group.
6. **Report:** display reports on individuals, their diary, or both.
7. **Variable Selection:** select the variables to be written to text files.

The relationship between these screens is shown in **Figure 1**. When Access starts you are put into the CHAD start screen. From here you can access the merged CHAD data by clicking on the central button with the blue and yellow CHAD logo. The buttons to the side of this take you to display screens for the individual studies.

**Note 1: Opening CHAD**  
If you hold the Shift key down while starting CHAD you will be put into the advanced mode. In this mode you can write your own queries against the database tables and use all of the Access facilities.

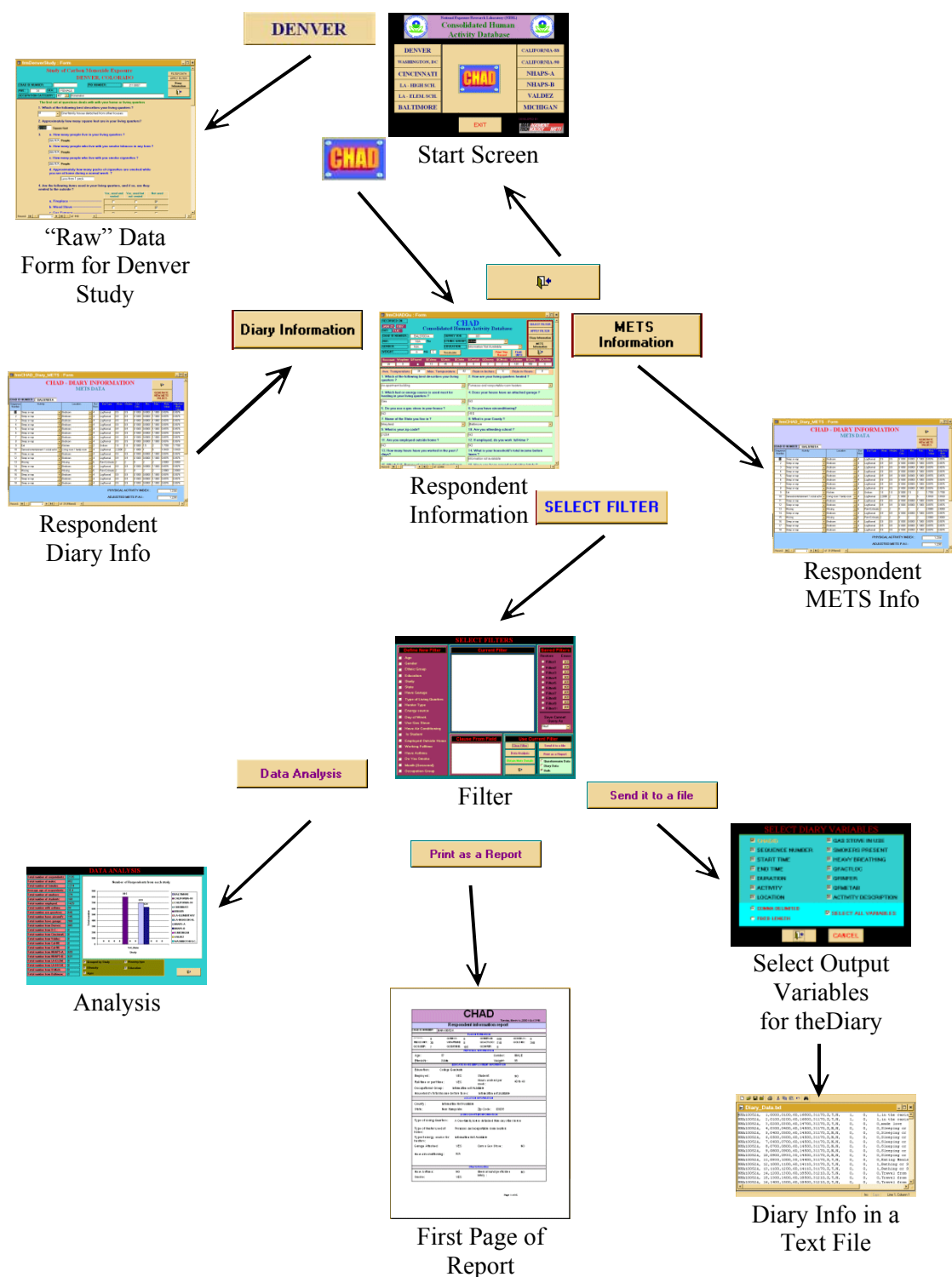
1. **Start CHAD Access.** Double click on the **mdb** file, or start Access and use the **File: Open** menu to open the **mdb** file.
2. Click on the CHAD logo to enter the CHAD Access application and view the combined database.

## Basic Respondent Info

Pressing the CHAD logo will bring you to the **Respondent Info** page. This page gives questionnaire information for one respondent. Actually, it gives information associated with one diary day for a respondent. In some cases, there may be several diary days associated with an individual. When this happens there are several records that can be displayed on this form.

Each respondent's diary day is given a unique identifier in the database field named **CHADID**. These unique IDs are often referred to as simply a CHADID. This identifier is shown on all pages that display information about a single diary day. On the **Respondent Info** page it is displayed on the left side about 4 lines from the top. The CHADID is composed of 3 characters, 5 numerals, and a final character. The first three characters indicate the study. The five numerals with the first three characters, provide a unique ID for each individual (not for an individual diary day). The final character indicates what day of the study a particular record or records (in the case of diary records) represent. Every CHADID representing respondents in studies with only 1 day of information contains an "A" as the final character. If a study collected 3 days of data, the records for an individual will contain the same 8 initial characters and end in "A", "B", and "C" depending on the day represented.

The **Respondent Info** page is separated into a non-scrolling top section, which displays some very basic information about the individual (e.g., CHADID, age), and a scrolling bottom section which gives weather (at top with peach colored background) and questionnaire answers for the respondent (actually the answers are for the respondent's diary



**Figure 1** Basic layout of CHAD Access screens and some of the buttons used to move from screen to screen..

day, but it will be easier to refer to them as being related to a respondent). The scroll bar to the right controls the scrolling of the bottom portion. Along the bottom border of the page is a record navigation widget shown in **Figure 2**.



**Figure 2** The **Record Navigation** widget.

The central window in this widget displays the current record number. The 5005<sup>th</sup> record is being displayed in the main **Respondent Info** page. The number on the right hand side displays the number of records that are currently in the selection set. All of CHAD's 22,968 respondent information records are selected. The arrows to the left and right of the window step through the currently selected records. The arrows with a line move to the beginning or end of the data base.

- 3 Click on the right arrow to move through the data one record (diary day) at a time.
- 4 Now click on the right arrow with the line to view the last record in the database.

This second image of the record navigation widget shows that only 29 respondents have been selected. The "(Filtered)" indicates that a selection filter has been applied.



The top of the **Respondent Info** page contains much of the basic demographic information on the respondent (**Figure 3**). The date sampled (in red background on the left side), age, gender, weight, ethnicity, education, and a number of quality flags (near bottom on red background). The buttons on this page also provide access to several other pages and some functionality. The **Diary Information** and **METS Information** buttons go to screens displaying activity and metabolic information for this respondent's diary day.

RECORDED ON :		CHAD		SELECT FILTER							
AUG 3, 1998				APPLY FILTER							
DAY: MON				Diary Information							
CHAD ID NUMBER: BAL98015F		SURVEY ID#: E15		METS Information							
AGE: 86 Yrs		ETHNIC GROUP: White									
GENDER: MALE		EDUCATION: Information Not Available									
WEIGHT: 98 KG C		Recalculate									
		Print This Record		FLAG INFO							
Recount	Wraptime	QFtravel	QCsleep	QCmiss	QCInfer	QCmetab	QCheavy	QCMeals	QCeatime	QClong	QCActloc
27	0	0	0	60	0	0	0	1	90	420	0

**Figure 3** **Respondent Info** page (top only).

## Respondent Diary Info

- 5 To go to the diary page, click on the **Diary Information** button in the upper right corner.



The CHADID is given in the upper left corner. The diary page contains a list of records representing the respondent's activities for a twenty four hour period. The period always starts at midnight. Each record or line of the data represents one activity or one hour, which ever took less time. Each record represents one period of time and contains the start and end times of the activity, the location of the person and additional information that varies from study to study such as whether or not the respondent or anyone else was smoking, was breathing hard, or was using a gas stove. A scroll bar to the right permits viewing additional records if they take up more than one page.

## Exiting Pages

Almost every page has an **Exit** button. Click on this to return to the page that you jumped from.

- 6 Click on the exit button to return to the **Respondent Info** page.



## Respondent METS Info

- 7 Click on the **METS Information** button to get to the **Respondent METS Info** page.



This page is similar to the **Respondent Diary Info** page. The CHADID is given in the upper left. The page uses the same diary records that the **Diary Page** uses. It also displays stochastically calculated metabolic rate information: the distribution type and parameters that were used to generate the values, the *METS Value*, the *Adjusted METS Value*, and at the bottom of the page-you may have to scroll down-the *Physical Activity Index* (PAI) and the *Adjusted Physical Activity Index* (APAI).

Since the METS values are stochastically calculated, a routine to recalculate them is available.

- 8 Click on the **Generate New METS Values** button.



Any METS values described by a distribution will be recalculated. Some of the METS values are described by a fixed value; these will not change.

- 9 Click on the **Exit** button to exit the **Respondent METS Info** page.

## Filter Screen: Selecting a Cohort

The **Filter Screen** is used to select a cohort of respondents (individuals) that fit specific criteria. The criteria used are often called a filter because they filter out the unwanted respondents.

We will go to the **Filter Screen** next to select a specific set of respondents.

### Note 2: SQL

The **Filter Screen** defines the WHERE clause of an SQL statement which is used to select records from the main respondents questionnaire database.

- 10 Click on the **Select Filter** button (upper right corner of screen) to bring the **Filter Screen** into view.

The **Filter Screen** has 5 sections. Each section is contained in a framed rectangle and contains a red title at the top (**Figure 4**).

**Define New Filter:** Use to add selection criteria to the filter.

**Clause From Field:** Displays a clause of the selection criteria for one field at a time.

**Current Filter:** Displays the entire filter.

**Saved Filters:** Use to save and restore filters.

**Use Current Filter:** Go to another page to display, analyze, or output the selected data.

Let's say we want to view all of the female respondents between 15 and 80 from California, Delaware, Hawaii, and Indiana. To retrieve these respondents we must enter a query that will select them. The query is a logical statement or expression. If it evaluates to True for a respondent, then the respondent is included in the selection set which can be viewed, printed, or exported. It is assumed that if a field is not included in the query, any value in that field is acceptable. So we only need to specify the acceptable values for age, gender, and state.

**Figure 4** The Filter Screen.

## Specifying Age

The acceptable age range can be set by clicking on the check box next to the “Age” label in the Define New Filter section (**Figure 5**) of the form. Once clicked, two boxes to enter the age ranges will appear to the right of the “Age” label (**Figure 6**). The minimum age value will be highlighted and will be replaced by simply typing a value. Hitting the tab or return keys will move the cursor to the maximum value and highlight it. After a value is entered hitting return will exit from the age range boxes. The age range selected is inclusive of the ages entered.

**Figure 6.** The text lines that are used to input the minimum and maximum age range for a query.

- 11 Click on the **Age** check box.
- 12 Type 15 and hit the return key.
- 13 Type 80 and hit the return key again.

**Figure 5** The Define New Filter section of the Filter Screen.

Notice at this point that the string “(Age Between 15 and 80)” is shown in both the **Current Filter** and **Clause From Field** boxes. The first shows the complete logical expression being built. The second shows only the part of the expression for the last entered field. In this case we have entered criteria for only one field, so they are identical.

## Clearing Queries

There are two methods to clear queries or clauses you want to take out. You can clear everything in the **Current Filter** window by clicking on the **Clear Filter** button at the bottom right of the **Filter Screen**. To clear a single clause, simply click on the check box which was used to produce that clause. For example, if you decide you do not want to use age in your query, click on its check box again. The clause will be removed from the current query and the check will be removed from the check box. Any other clauses will remain in the current query. Practice clearing errors? You will never need this skill. We only present it in case you need to teach others.

## Specifying Gender

To select gender:

- 14 Click on the **Gender** check box.

This will cause a pop-up menu to appear (**Figure 7**). This data entry form is typical of that for the remaining variables. The box on the left contains a list of the available codes for this field (M, F, X). Typically, “X”, indicates “N/A” or that no value is defined for a particular record. The meanings of these codes are given to the right of the codes. To

Please Select Gender

Select multiple items using Ctrl- or Shift- click: Selected Items: 1

M	Male
F	Female
X	Missing

[GENDER] = 'M'

Clear Selections OK

**Figure 7** Pop-up form for selecting gender.

select a code click on it using the left mouse button. You must select all codes for Gender that you want included. If any of the values are selected, only those selected will be kept in the selection set. The panel to the right will display the logical expression or clause that is generated by the current selection. The **Clear Selections** button will clear any selections currently set.

- 15 Click on “F” for female to select this value. Observe the clause defined to the right.
- 16 Click on the **Exit** button.

## Specifying the State

Now we will select the States that are to be included.

- 17 Click on the **State** check box.
- 18 Click on California.

To add additional selections, hold down the “Control” key and click on additional selections; this is sometimes written as “cntrl-Click”. Holding the control key down will maintain the existing selections and add any values that are clicked on. If you cntrl-Click on a selected value, it will be unselected.

Please Select Gender

Select multiple items using Ctrl- and Shift- click: Selected Items: 4

Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
Dist. Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana

[STATE] = 'California' Or [STATE] = 'Delaware' Or [STATE] = 'Hawaii' Or [STATE] = 'Indiana'

Clear Selections OK

**Figure 8** Pop-up form for specification of states.

- 19 Use cntrl-Click to select Delaware, Hawaii, and Indiana.

Another method of selecting several adjacent codes is to click on one, perhaps the top one; then hold down the shift key and click (“shift-Click”) on the bottom code. This will select the top, bottom, and all codes in between the two.

At this time notice that the clause for the states now includes several equalities joined by ORs. Additionally the top right of the form contains information on how many of the states have been selected. Respondents from all four of these states will be selected.

### Note 3: Query Limitations

In the current version of CHAD, query’s are limited to using the “=”, “AND”, and in some cases the “OR” logical operators in expressions. Expressions can not be formulated using NOT, XOR, or comparison operators (e.g., <, >, >=).

- 20 Click on the **Exit** button.

### Examining the Logical Expression

Examine the **Current Filter** box. This should now contain the complete logical expression being used to select respondents.

Examine the **Clause From Field** box. This contains the logical expression for the States field only. By clicking on any of the field names in the **Define New Filter** box, the clause specified for that field will be displayed in this box. You have to click on the label, not the check box itself.

- 21 Click on the **Study** label (not on the check box, but right over the text) and examine the **Define New Filter** box.
- 22 Click on the **Age** label and examine the box again.

The first click should have emptied the box since study is not used in our query. The second should have displayed just the age clause.

### Saving and Restoring Logical Expressions

Once you have created a satisfactory query, it can be saved for later recall. Up to 10 separate queries can be saved using the **Saved Filters** section of the **Filters Screen (Figure 9)**. With the query visible in the **Current Filter** section, simply click on the down arrow at the right of the pull down menu on the bottom of the **Saved Filters** section. This will bring down a list of 10 filters. Click on the one you want the current query saved in. Alternatively, double-clicking on the filter name displayed in the menu will save the current query to the filter named in the menu.

Restore a filter by clicking on any of the radio buttons next to the filter names. This will clear the current query and restore the specified query from memory.

The presence of the delete buttons to the right of the filter name indicates that a filter is stored by that name. Clicking on the delete button clears the filter from memory. The clear button itself also disappears. The filters are stored in a database and preserved between sessions.



**Figure 9** The **Saved Filters** section.

- 23 Click on the pull down menu once to expose the filter names. Click on **Filter 7** to save the current query.
- 24 Click on the radio button next to **Filter 1** to retrieve that existing query. Check the **Current Filter** box to see that a different query is being used.



- 25 Click on the radio button next to Filter 7 to retrieve your initial query.

## Viewing Selected Respondent Information

Now let's go back to the respondent info and see how many people we have selected and look at some of the questionnaire answers for them.

- 26 Exit the **Filter Screen** by clicking on the **Exit** button (lower right, **Figure 10**).
- 27 Examine the status bar at the bottom of the **Respondent Info** page. It should indicate that about 22,900 records are currently selected.
- 28 Apply the filter created to the data by clicking on the **Apply Filter** button.
- 29 After a brief wait, examine the status bar once again to see how many records are selected.



Figure 10 The Filter Screen's Use Current Filter section.

APPLY FILTER

In order to apply the filter to the records displayed on the **Respondent Info** page, you must click the **Apply Filter** button. Until then, all respondents are included and can be viewed. After clicking **Apply Filter**, all respondents that can be viewed on this or the diary pages are members of the set selected by the filter.

## Creating Charts and Reports

### Creating and Viewing Charts

The PC version of the CHAD application has the ability to display a limited number of pre-defined bar charts. The **Analysis Screen** is accessible from the **Filter Screen**. The current query as displayed in the **Current Query** section is used to select respondents before the information is displayed.

- 30 Click on the **Select Filter** button.
- 31 Define a query of interest.
- 32 Click on the **Analysis** button in the **Use Current Filter** section.

It will take a moment to perform the calculations necessary for the charts. An information screen will come up saying so.

- 33 Click the **OK** button on the information screen.
- 34 Minimize or move the **Filter Screen** window if necessary.

The **Filter Screen** usually blocks the initial view of the chart screen (**Figure 11**). It will take a moment for the computation necessary for the bar charts. When numbers appear in the boxes on the left, then you can proceed.

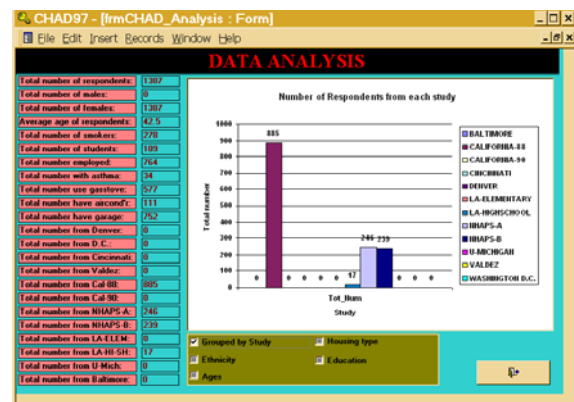


Figure 11 The Analysis Screen.



The **Analysis Screen** has three main sections.

**Left:** A list of values calculated for the current selection set.

**Main:** Area where the chart will be drawn.

**Bottom:** Check boxes to select the variable to be charted.

35 Examine the numbers on the left.

36 Click on the check box labeled "Grouped By Study". Your figure should look like the one in **Figure 11**.

37 Click on the other check boxes one at a time and examine the charts.

38 Click on the **Exit** button to leave the **Analysis Screen**.

## Creating Formatted Reports for Printing

The CHAD application can produce formatted reports (**Figure 12**) describing the respondents and their diaries. It works in a manner similar to exporting files, except that you cannot specify which variables to output. You should be back in the **Filter Screen**. Make sure the filter you are using returns a few, not many, respondents. One to four pages will be produced for each respondent.

Print as a Report

39 Click the radio button labeled "Questionnaire" (**Figure 10**).

40 Click the **Print as a Report** button.

A page will be produced inside an Access window. It may be behind the **Filter Screen**. At the bottom there is a record navigation widget, just like that used in the **Personal Info** and **Diary Info** pages. Clicking the forward arrow will advance the report one page. The widget should also indicate the total number of pages in the report.

CHAD			
Tuesday, March 14, 2000 10:43 PM			
Respondent information report			
CONTACT NUMBER: NHA 10052A			
PLACED INFORMATION			
OF STUDY: 0	OCM 515: 0	OCMETAB: 660	OCMBVY: 0
RECODINT: 30	VIRAPRIME: 0	OCACCTLOC: 510	OCOLONO: 390
OCORREP: 7	OCERTIME: 105	OCORREP: 0	
PERSONAL INFORMATION			
Age: 57	Gender: MALE		
Ethnicity: White	Weight: 88		
EDUCATION AND EMPLOYMENT INFORMATION			
Education: College Graduate			
Employed: YES	Student: NO		
Full time or part time: YES	Hours worked per week: 40 to 49		
Occupational Group: Information Not Available			
Household's total income before taxes: Information Not Available			
LOCATION INFORMATION			
County: Information Not Available			
State: New Hampshire	Zip Code: 03055		
LIVING QUARTER INFORMATION			
Type of Living Quarters: A One-family house detached from any other house			
Type of Heater used at home: Finances and a portable room heater			
Type of energy source for heaters: Information Not Available			
Garage Attached: YES	Own a Gas stove: NO		
Have air conditioning: N/A			
Other Information			
Have Asthma: NO	Been around pesticides lately: NO		
Smoke: YES			

Page 1 of 16

**Figure 12** Sample page from generated report.

## Exporting Data

Now that you have decided that this is the set you would like to further analyze outside of Access, you will want to export the data in a format other programs can read. You do this from the **Filter Screen**. Like the **Analysis Screen**, the respondent info from the current selection set will be used. Two main data sets can be exported, the questionnaire data, and the diary data. To select one or both, click on a radio button in the **Use Current Filter** section (with the light green background, **Figure 10**). Then click on the **Send It To A File** button.

Send it to a file

- 41 Select the radio button labeled "Both" by clicking on the button or its label (see figure to right). This is in the same set of radio buttons used to determine which data sets were included in the report.
- 42 Click on the **Send It To A File** button.

## Selecting the Variables to Export

Two screens will pop up, one for the diary information (**Figure 13**) and beneath that, one for the questionnaire information (**Figure 14**). They each contain a number of similar options. You can accept the defaults by simply clicking on the **Exit** button at the bottom of the form. There are three sections on the screens. The top portion (blue-gray) lists all the variables that can be exported. Below this, there is a light blue strip where you can specify the format of the output and select to have all variables output (the default).

**Figure 13** The **Variable Selection** screen used when exporting diary data.

- 43 On the diary **Variable Selection Screen**, click on the check box labeled "Select All Variables".
- 44 Click on the check boxes for sequence number, start time, duration, activity, and location.
- 45 Click on the **Exit** button.

**Figure 14** The **Variable Selection** screen used when exporting questionnaire data.

To select individual variables for output you **MUST** manually deselect the "Select All Variables" check box. This is true for both **Variable Selection** screens. Then checking individual variables will cause those variables only to be written to the file. If you decide to quit the operation click on the **Cancel** button. Clicking on the **Exit** button proceeds with output. After proceeding, two boxes will pop up. The first (**Figure 15**) allows you to specify the output file; the default is c:\temp\Diary\_Data.txt. The second (**Figure 16**) confirms that the data was written to the file.

- 46 If necessary change the file name for output and click on the **OK** button. After the records have

- 47 been written to the file click on the **OK** button of the second box.
- 48 Now select the variables of interest for the questionnaire data.
- 49 Click on the radio button labeled "Comma Delimited".
- 49 Click on the **Exit** button and proceed as with the diary data.

## File Formats

The radio buttons at the bottom left of the screens control the output format of the data. There are four options (two for the diary data): comma delimited, tab delimited, fixed width, and last and least, the style from CHAD97, fixed width with commas (**Table 1**).

### Comma Delimited:

Values are output separated by commas with no leading or trailing spaces.

**Tab Delimited:** Values are output separated by a single tab character with no leading or trailing spaces. The values will not necessarily line up in a text or word processor.

**Fixed Width:** Values are output in a specific number of spaces. The values are padded with space characters if the value does not fill the specified number of spaces.

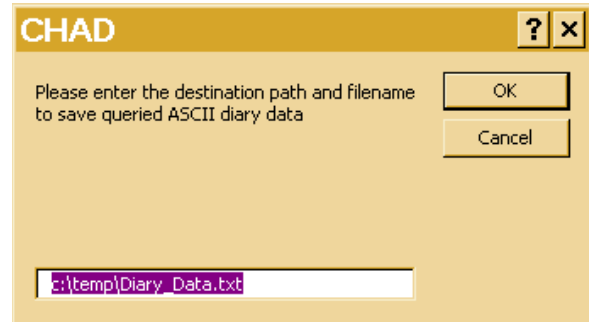
### Fixed Width With Commas:

This format is like the fixed width except that commas are placed between the values.

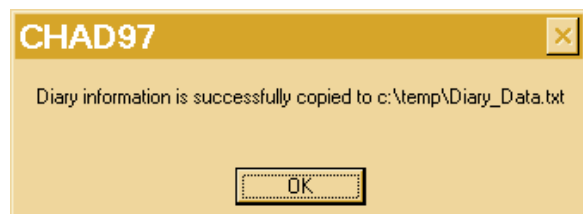
Each record is written on a separate line. For questionnaires this is a single day for one respondent. The diaries contain a single activity or time period on each line. The padding in the fixed width formats is placed on the left for numbers and on the right for strings. If some variables are not included the spaces or columns for those are not included.

## Moving On

That completes the coverage of CHAD PC for this tutorial. We suggest that you review the section on CHAD Oracle if appropriate. Your real analysis will undoubtedly start after you have extracted the data from CHAD.



**Figure 15** Box allowing the user to specify output file for the diary data.



**Figure 16** Box confirming that all of the data requested was written to the file.

**Table 1** Examples of four text file formats that can be used when exporting questionnaire data.

**Comma Separated**

NHA10052A,008427,1,1,4,20,1994,WED,SF,VB,X,N,Y  
NHA10053A,008437,1,1,5,2,1994,MON,AP,X,X,N,X  
NHA10054A,00844,1,1,5,18,1994,WED,SF,VB,OK,N,Y

**Tab Separated**

NHA10052A 008427 1 1 4 20 1994WED SF VB X N Y  
NHA10053A 008437 1 1 5 2 1994MON AP X X N X  
NHA10054A 008447 1 1 5 18 1994WED SF VB OK N Y

**Fixed Width\***

NHA10052A008427 1 1 4 201994WEDSFVBX NY  
NHA10053A008437 1 1 5 21994MONAPX X NX  
NHA10054A008447 1 1 5 181994WEDSFVBOK NY

**Fixed Width With Commas\***

NHA10052A,008427, 1, 1, 4, 20,1994,WED,SF,VB,X ,N,Y  
NHA10053A,008437, 1, 1, 5, 2,1994,MON,AP,X ,X ,N,X  
NHA10054A,008447, 1, 1, 5, 18,1994,WED,SF,VB,OK,N,Y

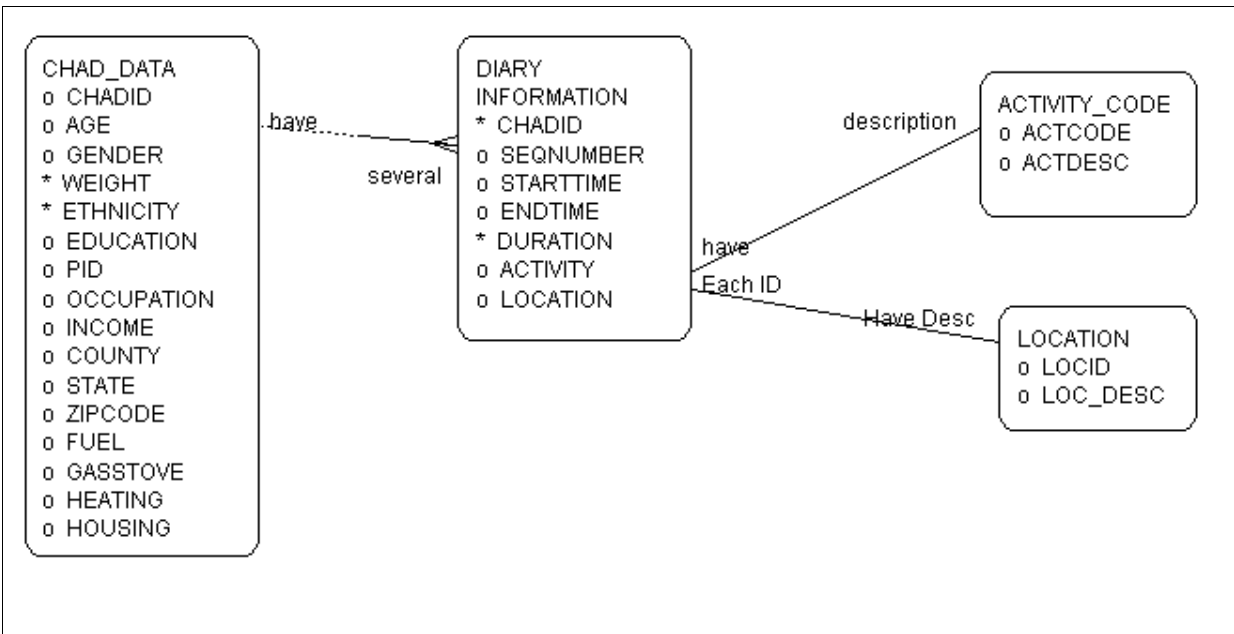
\* The spacing here does not represent actual spacing, but is shown as an example.

## What's Inside the CHAD Databases

The combined CHAD database, Access or Oracle, contains two main tables and several auxiliary tables. The most important are shown in **Figure 17**. The two main tables are CHAD\_DATA and CHAD\_DIARY (shown as DIARY). The CHAD\_DATA table contains one record for each person-day. Most of the fields are answers from the questionnaire. When an individual has more than one diary day in CHAD\_DIARY, there is more than one record in CHAD\_DATA representing the individual. Most of this data is usually redundant. Although some fields that may vary from day to day for an individual and fields based on the diary entries (e.g., quality flags) will most likely be different for each day.

The CHAD\_DIARY table contains one record for each activity during a 24 hour period. No record represents more than an hour or crosses the hour boundary. The minimal information in the diary data is the CHADID which links to the CHAD\_DATA, a sequence number, which indicates the order of the records in a day, start, end, and duration times, and an activity and location. Some of the studies have additional information on smoking and other variables.

Many of the values stored in the database are somehow coded; that is, rather than MALE, or Living Room, they might be coded as M or 1 for MALE and 30122 for the living room in the respondents house. The diary location and activity codes are given in tables that relate the coded value with a description (**Tables 2, 3**). Other codes can be deduced by looking at the queries and functions associated with a table (see the section on using individual data sets).



**Figure 17** A schematic of the tables and relationships among the tables in the combined CHAD database.

**Table 2** CHAD Diary location code (LOC\_CODE) definitions.in this table links with the Location field in tblCHAD\_Diary.

LOC_CODE	LOC_DESC
30000	Residence, General
30010	Your Residence
30020	Other's Residence
30100	Residence, indoor
30120	Your residence, indoor
30121	Kitchen
30122	Living room / family room
30123	Dining room
30124	Bathroom
30125	Bedroom
30126	Study / Office
30127	Basement
30128	Utility room / Laundry room
30129	Other indoor
30130	Other's residence, indoor
30131	Other's Kitchen
30132	Other's living room / family room
30133	Other's Dining room
30134	Other's Bathroom
30135	Other's Bedroom
30136	Other's Study / Office
30137	Other's Basement
30138	Other's utility room / laundry room
30139	Other indoor
30200	Residence, Outdoor
30210	Your residence, Outdoor
30211	Your residence - Pool, spa
30219	Your residence - Other outdoor
30220	Other's residence, outdoor
30221	Other's residence - Pool, spa
30229	Other's residence - Other outdoor
30300	Garage
30310	Indoor garage
30320	Outdoor garage
30330	Your garage
30331	Your indoor garage
30332	Your outdoor garage
30340	Other's garage
30341	Other's indoor garage
30342	Other's outdoor garage
30400	Other, residence
31000	Travel, general

31100	Motorized travel
31110	Travel by car
31120	Travel by truck
31121	Travel by truck(pick-up van)
31122	Travel by Truck (other than pick-up or van)
31130	Travel by Motorcycle / moped / motorized scooter
31140	Travel by bus
31150	Travel by Train / Subway / rapid transit
31160	Travel by airplane
31170	Travel by boat
31171	Travel by motorized boat
31172	Travel by unmotorized boat
31200	non-motorized travel
31210	Travel by walk
31220	Travel by bicycle / skateboard / roller-skates
31230	Travel in a stroller or carried by an adult
31300	Waiting
31310	Wait for bus, train, ride (at stop)
31320	Wait for travel, indoors
31900	Other travel
31910	Travel by other vehicle
32000	Other, indoor general
32100	Office building / bank / post office
32200	Industrial plant / factory / warehouse
32300	Grocery store / convenience store
32400	Shopping mall / non-grocery store
32500	Bar / night club / bowling alley
32510	Bar / Night Club
32520	Bowling alley
32600	Repair shop
32610	Auto repair shop /gas station
32620	Other repair shop
32700	Indoor gym / sports or health club
32800	Childcare facility
32810	Childcare facility, house
32820	Childcare facility, commercial
32900	Public building / library / museum / theater
32910	Auditorium, sport's arena / concert hall
32920	Library / courtroom / museum / theater
33100	Laundromat

**Table 2** CHAD Diary location code (LOC\_CODE) definitions.in this table links with the Location field in tblCHAD\_Diary.

LOC_CODE	LOC_DESC		
33200	Hospital / health care facility / doctor's office	35220	Parking lot
33300	Beauty parlor / barber shop / hair dresser's	35300	Service station / gas station
33400	At work : no specific location, moving among locat	35400	Construction site
33500	At School	35500	Amusement park
33600	At Restaurant	35600	School grounds / playgrounds
33700	At Church	35610	School grounds
33800	At Hotel /Motel	35620	playground
33900	At Dry cleaners	35700	Sports stadium and amphitheater
34100	Parking garage	35800	Park /golf course
34200	Laboratory	35810	Park
34300	Other, indoor	35820	Golf course
35000	Other outdoor, general	35900	Pool, river, lake
35100	Sidewalk / street / neighborhood	36100	Restaurant, picnic
35110	Within 10 yards of street	36200	Farm
35200	Public garage / parking lot	36300	Other outdoor
35210	Public garage	U	Uncertain
		X	Missing

**Table 3** CHAD Diary activity code descriptions. The Activity\_Sub\_Merge field in this table links with the Activity field in tblCHAD\_Diary.

Activity_Sub_Merged_Code	Sub_Activity_Description		
10000	Work and other income producing activities, general	10117	Work for child / youth / family organizations
10100	Work, General	10118	Work for other organizations
10110	Work, general, for organizational activities	10120	Work, income-related only
10111	Work for professional/union organizations	10130	Work, secondary (income-related)
10112	Work for special interest identity organizations	10200	Unemployment
10113	Work for political party and civic participation	10300	Breaks
10114	Work for volunteer/ helping organizations	11000	General household activities
10115	Work off/for religious groups	11100	Prepare food
10116	Work for fraternal organizations	11110	Prepare and clean-up food
		11200	Indoor chores
		11210	Clean-up food
		11220	Clean house
		11300	Outdoor chores
		11310	Clean outdoors
		11400	Care of clothes
		11410	Wash clothes
		11500	Build a fire
		11600	Repair, general
		11610	Repair of boat
		11620	Paint home / room
		11630	Repair / maintain car
		11640	Home repairs

Activity _Sub _Merged _Code	Sub_Activity_Description
--------------------------------------	--------------------------

11650	Other repairs		training
11700	Care of plants	15200	Attend other classes
11800	Care for pets/animals	15300	Do homework
11900	Other household	15400	Use library
12000	Child care, general	15500	Other education
12100	Care of baby	16000	General entertainment / social activities
12200	Care of child		
12300	Help / teach	16100	Attend sports events
12400	Talk /read	16200	Participate in social, political, or religious activities
12500	Play indoors		
12600	Play outdoors	16210	Practice religion
12700	Medical care-child	16300	Watch movie
12800	Other child care	16400	Attend theater
13000	Obtain goods and services, general	16500	Visit museums
13100	Dry clean	16600	Visit
13200	Shop / run errands	16700	Attend a party
13210	Shop for food	16800	Go to bar / lounge
13220	Shop for clothes or household goods	16900	Other entertainment / social events
13230	Run errands	17000	Leisure, general
13300	Obtain personal care service	17100	Participate in sports and active leisure
13400	Obtain medical service	17110	Participate in sports
13500	Obtain government / financial services	17111	Hunting, fishing, hiking
		17112	Golf
13600	Obtain car services	17113	Bowling / pool / ping pong / pinball
13700	Other repairs	17114	Yoga
13800	Other services	17120	Participate in outdoor leisure
14000	Personal needs and care, general	17121	Play, unspecified
14100	Shower, bathe, personal hygiene	17122	Passive, sitting
14110	Shower, bathe	17130	Exercise
14120	Personal hygiene	17131	Walk, bike, or jog (not in transit)
14200	Medical care	17140	Create art, music, participate in hobbies
14300	Help and care		
14400	Eat	17141	Participate in hobbies
14500	Sleep or nap	17142	Create domestic crafts
14600	dress, groom	17143	Create art
14700	Other personal needs	17144	Perform music / drama / dance
15000	General education and professional training	17150	Play games
		17160	Use of computers
15100	Attend full-time school	17170	Participate in recess and physical education
15110	Attend day-care		
15120	Attend K-12	17180	Other sports and active leisure
15130	Attend college or trade school	17200	Participate in passive leisure
15140	Attend adult education and special	17210	Watch



Activity _Sub _Merged _Code	Sub_Activity_Description
--------------------------------------	--------------------------

17211	Watch adult at work
17212	Watch someone provide childcare
17213	Watch personal care
17214	Watch education
17215	Watch organizational activities
17216	Watch recreation
17220	Listen to radio / listen to recorded music / watch T.V.
17221	Listen to radio
17222	listen to recorded music
17223	Watch TV
17230	Read, general
17231	Read books
17232	Read magazines / not ascertained
17233	Read newspaper
17240	Converse / write
17241	Converse
17242	Write for leisure / pleasure / paperwork

17250	Think and relax
17260	Other passive leisure
17300	Other leisure
18000	Travel, general
18100	Travel during work
18200	Travel to/from work
18300	Travel for child care
18400	Travel for goods and services
18500	Travel for personal care
18600	Travel for education
18700	Travel for organizational activity
18800	Travel for event / social activity
18900	Travel for leisure
18910	Travel for active leisure
18920	Travel for passive leisure
U	Unknown
X	Missing

## **Background on Individual Studies in CHAD**

Data collection began with identification of data that would serve as the foundation of CHAD. Existing databases from several human activity studies were targeted because of their comprehensiveness and availability of documentation. Because of the passage of time, it was difficult to locate the original databases. However, the best known sources were contacted for each. If the primary source did not have access to it, then often a second or third potential source was contacted. In some cases, it was not possible to locate the original database; second generation databases were then used instead. At the end of this process, databases were obtained from the following studies:

### **Activity Patterns in Cincinnati, Ohio (CIN)**

Human Activity Patterns in Cincinnati, Ohio. This study was funded by the Electric Power Research Institute (EPRI) and was conducted in 1987 (Johnson, 1989). Activity diaries were carried by 973 randomly selected participants for three consecutive 24-hour periods in March or August 1985, resulting in 2800 subject days of diary data. Detailed background questionnaires were also completed by the participants. The electronic data were provided by James Capel.

### **Baltimore (BAL)**

This study was conducted in a retirement home in Baltimore County, Maryland during January-February, 1997 and July-August, 1998. All participants were Caucasian and at least sixty-five years of age; no individual was a part of both studies. For five days in each of three consecutive weeks, five people recorded diaries in 1997. In 1998, twenty-one people participated. All diaries were recorded in fifteen minute time segments.

### **California Adults (CAA) and California Youth (CAY)**

Activity Patterns of California Residents. This study was conducted from October 1987 to September 1988 by the California Air Resources Board (CARB; Wiley et al., 1991). It involved 1762 randomly selected participants aged 12 and above. Retrospective time diaries were collected via telephone for all activities occurring on the previous day. A total of 36,918 records were collected. Questionnaire data were collected on activity locations, presence of smokers, and housing and socioeconomic characteristics. Like NHAPS, this was a next-day telephone survey. The data for this study were also obtained from James Capel.

### **California Children (CAC)**

Activity Pattern Survey for California Children. This survey was conducted by CARB from 1989 to 1990 (Wiley, 1991). A total of 27,048 activity records were collected from 1200 children. The same methodology was used as described above, although the questionnaires and diary formats differed in certain details. James Capel also provided these data.

### **Denver Study (DEN)**

A Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. This study was conducted in 1983 with a total of 454 study participants (Johnson, 1984). Each participant carried a personal exposure monitor (PEM) and an activity diary for two consecutive 24-hour sampling periods. Approximately 900 person days of PEM and activity data were collected. Detailed background questionnaire data were also collected. The electronic data were obtained from James Capel. The questionnaire information was incomplete, and to date, another source for the remaining questionnaire information has not been found.

## **Los Angeles Studies (LAE, LAH)**

Los Angeles Area Studies: Development of Improved Methods to Measure Effective Doses of Ozone. These studies of outdoor workers exposed to oxidant pollution (Shamoo et al., 1991), elementary and high school students exposed to oxidant pollution (Spier et al., 1992), and ozone-exposed construction workers were conducted between 1989 and 1992. The activity records from these studies cover only a work-day or school-day period (approximately eight hours), as the purpose of these studies was to focus on exposures in specific microenvironments. The original data were obtained from Kenneth Clark. Because of the limited study sizes and the difficulty of extending the diaries to a standard 24-hour period, the construction and outdoor workers studies were not included incorporated into the CHAD database.

## **University of Michigan (UMC)**

The Survey Research Center of Institute for Social Research at the University of Michigan conducted this national study. Between March and December of 1997, researchers accumulated data related to 5,616 children twelve years of age or less. Questionnaires were administered either in person or by phone to primary care givers (usually the mother), other care givers, teachers, and others. Care givers completed twenty-four hour diaries for both a weekday and weekend day during a particular week.

## **National Human Activity Pattern Study (NHA, NHW)**

National Human Activity Pattern Study (NHAPS). This study was conducted between October 1992 and September 1994 (Klepeis et al., 1995), during which period 24-hour activity diaries were collected from 9386 participants. Questionnaire data were also collected. This survey was the largest and most geographically diverse (data from 48 states) of the studies going into CHAD. It was conducted as a next-day telephone interview. This method has the advantage that an interviewer can insist on filling in gaps in a diary, but suffers from the limitations of a participant's memory regarding the exact time and duration of activities. The original data for this study were obtained from William Nelson of EPA/NERL.

## **Valdez Air Health Study (VAL)**

Valdez Air Health Study. This study was conducted during 1990-91 (Goldstein et al., 1992). A total of 289 individuals from different households were interviewed as part of a continuous monitoring program that also measured criteria pollutants, volatile organic compounds, and weather conditions. This sample covered 23% of the residences in the Valdez area. The data for this study were provided by Ted Johnson.

## **Washington, DC (WAS)**

Study of Carbon Monoxide Exposure of Residents of Washington, DC. This EPA-sponsored investigation was carried out concurrently with the Denver study and covered 705 respondents for one day each (Settergren et al., 1984). The version of the data acquired for this project has incomplete questionnaire data. Also, the list of potential activity and location codes is fairly short, limiting the resolution in activity description that is possible in this study. James Capel was the source for these data.

## Using Data from the Individual Studies

The individual studies contain a wide variety of data not included in the combined CHAD database. These data can be more powerful in answering specific questions than the combined database. The CHAD application is not, however, set up to use this data easily. If you want to use the data from individual studies you will have to familiarize yourself with those studies, learn what individual fields mean, and write your own queries to produce the desired results. This section gives some pointers on that process.

The individual studies each had their own particular focus and collected some unique data. Some have more detailed medical data, smoking data, or data on when windows and doors were open either in the questionnaire or the diary. For our example, let's investigate how often people open their windows. The initial task is to look at the reference materials (some of which are hopefully available for download) and see which studies contain information about the opening of windows and doors.

The questions will not necessarily be in the forms display for the individual studies since answers are not displayed for all questions in each study. Examination of the survey sources shows the following questions were asked (Tables 3 and 4).

**Table 3** Questions relating to doors and windows in the two NHAPS studies.

### **NHAPS A**

- While [fill NAME] [fill WERE] at home yesterday, were any windows left open at all?
- How many windows?
- For about how long altogether were any windows left open?
- Does your home have any doors which open directly to the outside of your building or house?
- While [fill NAME] [fill WERE] at home yesterday, were any of these outside doors left open at all?
- For about how long altogether was any outside door left open?
- While [fill NAME] [fill WERE] at home yesterday, about how many times would you say that an outside door was opened?

### **NHAPS B**

Same as above questions plus these additional questions.

- When [fill name] showered, was the bathroom door closed?
- Was there a window open or an exhaust fan on?
- Was the bathroom door closed?
- Was there a window open or an exhaust fan on?

You will note a great variety of information, that is not all necessarily useful. Do we need to know how often the window was open when someone showered? Maybe, but this is going to account for very little of the time where someone might be near an open window. Some of the questions simply ask how many times a door was open, others for how much time, and still others keep track of it in the diary, allowing you to know when the door was opened.

You need to decide which of the answers for these questions will give the information desired? Are the studies chosen representative of the population of interest? Do the data have enough detail?

**Table 4** Questions about doors and windows from the California, Cincinnati, and LA studies.

**California 88**

- Were any doors or windows open for more than a minute or two at a time yesterday?

**California 90**

- Were any doors or windows in your home open for more than a minute or two at a time yesterday?
- For about how long during the day, that is, from 6am to 6pm (were they/was it) open?
- For about how long during evening or night hours, that is, from 6pm to 6 am, (were they/was it) open?

**Cincinnati**

- “Windows open” field is included in diary. Did not find this as a question in the main screen menu choices.

**LA Elementary Study and LA Highschool Study**

- “Use Windows” field is included in diary information, but not on primary screen.

## Finding the Correct Columns or Fields

Assuming you have found the appropriate questions, you must next find the corresponding answers in the data tables for that study and figure out how they are coded. The individual study data are typically stored in tables that use part of the study name in the table name. Looking around at the tables in CHAD Access you should be able to find those relatively easily.

Deciding which table and field among the tables for a study gives the answer is a little more complicated. Most of the fields do have a somewhat descriptive name, although often not quite descriptive enough. Additional information can be found by consulting the documentation from the original study. The field names are often named similarly to the variables used in the documentation. They are also often in the same order. Additional information about the data may be found in the definition of the forms and modules. The forms state which query and which fields they use. Many of the modules contain code to convert values from coded to understandable. If these fail, then look at the code associated with the form **frmTransferData**. The Visual Basic associated with this form was used to read in much of the data and gives additional information on which table and field corresponds to which part of the input data. The input data is often documented as part of the original studies. The information is contained in the code itself, there is not really any documentation *per se*.

## Finding the Queries and the Coding

Typically, in CHAD Access, the tables themselves are associated with a query which has easier to understand coding. In fact the only purpose of these queries is to convert the coding in the tables from the “native” coding to codes that can be put in the forms and understood. The query associated with a table usually has the same root name, but a different prefix: qry or qsel.

You can use the tables or the queries as sources for your own queries. The queries, however, tend to be more understandable. The queries can also be used to deduce the coding in the tables if you want to use the tables. In addition to the queries, you can look at the modules (e.g., mdl\_Cin\_Misc\_Func) that contain many transformation functions. For instance, the function CinHeatType converts the numeric table codes, 1, to text, “Steam or Hot Water System”. The second is much easier to interpret, but looking at the code tells you how the codes are defined.

The tables, fields or columns, queries, and coding of values relating to the window questions are given in the following tables.

**Table 5** Location and coding of data relating to windows and doors in the individual studies' questionnaire data.

<b>Study/Questions</b>	<b>Table/Column/Coding</b>	<b>Query/Column/Coding</b>
<b>NHAPS A</b>	<b>tblHHAPS_Dataset7</b>	
<i>Were any windows left open when respondent was at home? If yes:</i>	WINDY/ 0 (No), 1 (Yes), 8 (DK)	
<i>Number of windows left open</i>	NUM_WINDY/ 0 (None), 1 (1-2), 2 (3-5), 3 (6-9), 4 (10+), 8 (DK)	
<i>Number of minutes windows were open</i>	WINDMIN/ 0 (< 1 min), 1 to 119 (Minutes), 180 (2-4 hrs), 360 (4-8 hrs), 600 (8-12 hrs), 840 (12-16 hrs), 960 (16+ hrs), 888 (DK)	
<i>Do any home doors open directly to outside? If yes:</i>	DOOROUT/ 0 (No), 1 (Yes), 8 (DK)	
<i>Were any outside doors left open yesterday? If yes:</i>	DROPEN/ 0 (No), 1 (Yes), 8 (DK)	
<i>Number of minutes doors were open</i>	MIN_DROPEN/ 0 (< 1 min), 1 to 119 (Minutes), 180 (2-4 hrs), 360 (4-8 hrs), 600 (8-12 hrs), 840 (12-16 hrs), 960 (16+ hrs), 888 (DK)	
<b>NHAPS A</b>	<b>tblNHAPS_Dataset8</b>	
<i>How many times was outside door opened?</i>	DOOPENTI/ 1 (1-2 times), 2 (3-5), 3 (6-9), 4 (10-19), 5 (20+), 8 (DK)	
<b>NHAPS B</b>	<b>tblNHAPS_Dataset2</b>	<b>qselNHAPS_PersonalHlthNHyl nfo</b>
<i>Was there a window open or a fan on? [while respondent was in the shower.]</i>	SHWIND/ 0 (No), 1 (Yes), 8 (DK)	SH_WIN/ Yes, No, DK
<i>Was there a window open or a fan on? [while respondent was in the bath.]</i>	BAWIND/ 0 (No), 1 (Yes), 8 (DK)	BA_WIN/ Yes, No, DK
<b>California 90 (Children, CAC)</b>	<b>tblCali_Kids_HomeInfo</b>	<b>qselCali_Kids_HomeInfo</b>
<i>Were any doors or windows in your home open for more than a minute or two at a time yesterday?</i>	OPEN/ 1 (Yes), 5 (No), 7 (DK (this may be wrong))	DOOR_OPEN/ Yes, No, DK
<i>For about how long during the day, that is, from 6am to 6pm, (were they/was it) open?</i>	OPN1/ 0-30 (Minutes) 60-720 (Minutes, given as nearest (?) hour), except: 998 (? , now coded as ERR), blank or “.” (N/A)	MIN_OPEN/ 0-30 Minutes, 60, 120, ..., 720, 998 (unknown meaning), N/A, ERR. All given as 3 character string padded on left with zeros.

<b><i>Study/Questions</i></b>	<b><i>Table/Column/Coding</i></b>	<b><i>Query/Column/Coding</i></b>
<i>For about how long during evening or night hours, that is, from 6pm to 6am, (were they/was it) open?</i>	OPN2/ see OPN1	MIN_OPEN_NIGHT/ see MIN_OPEN
<b><i>California 88 (Adults CAA)</i></b>	<b>tblCali_Dataset1</b>	<b>qselCali_ExpToChem</b>
<i>Were any doors or windows in your home open for more than a minute or two at a time yesterday?</i>	OPEN/ 1 (Yes), 5 (No), 7 (DK (this may be wrong))	DR_OPEN/ Yes, No, DK
<i>For about how long during the day, that is, from 6am to 6pm, (were they/was it) open?</i>	OPN1/ 0-30 (Minutes) 60-720 (Minutes, given as nearest (?) hour), except: 998 (? , now coded as ERR), blank or “.” (N/A)	MIN_OPEN/ 0-30 Minutes, 60, 120, ..., 720, 998 (unknown meaning), N/A, ERR. All given as 3 character string padded on left with zeros.
<i>For about how long during evening or night hours, that is, from 6pm to 6am, (were they/was it) open?</i>	OPN2/ see OPN1	MIN_OPEN1/ see MIN_OPEN

**Table 6** Location and coding of data relating to windows and doors in the individual studies’ diary data.

<b><i>Study/Questions</i></b>	<b><i>Table/Column/Coding</i></b>	<b><i>Query/Column/Coding</i></b>
<b><i>Cincinnati (CIN)</i></b>	<b>tbl_Diary</b>	<b>qryCin_Diary</b>
<i>“Windows Open” field in diary form.</i>	WINDOWS/ 0 (N/A), 24 (Yes), 25 (No), 26 (Uncertain)	WINDOW/ Yes, No, Uncertain, N/A
<b><i>LA Highschool (LAH)</i></b>	<b>tblLAHI_Diary</b>	<b>qselLAHI_Diary</b>
<i>“Use Windows” field in diary</i>	WIN_OPEN/ 24 (Yes), 25 (No), -9 (N/A)	WIN_OP/ Yes, No, N/A
<b><i>LA Elementary (LAE)</i></b>	<b>tblLAEM_Diary</b>	<b>qselLAEM_Diary</b>
<i>“Use Windows” field in diary</i>	WIN_OPEN/ 24 (Yes), 25 (No), -9 (N/A)	WIN_OP/ Yes, No, N/A



## Defining Your Query

Now that you have located and understand the coding of your data you need to extract or summarize it as desired. The basic method of doing this is to use an SQL query which manipulates the data. Tables 8 and 7 show examples of such queries. The first selects the records which have valid data out of the table, joins the data from 3 tables, performs a few transformations (e.g., derives the month from the date), and redefines some column names to be clearer. The second query then uses the results of the first. It derives the season from the month, derives the temperature category (5 degree bins) from the temperature, creates a table using the first as column headers and the second as row headers, and then averages the number of windows open by season and temperature with which it fills the table (**Table 8**).

Although you can type in an SQL statement directly, it will probably be easier to use the query builder which is much more intuitive. The table produced can be cut and pasted in to a word processing or spread sheet document.

**Table 8** The query c\_qryNHA\_Windows\_PreProcess that links several of the NHAPS data set tables, breaks the temperatures into 5 degree bins, extracts the month from the date, and selects those records where the window variables are defined; defined as being open, number opened, and time opened.

```
SELECT Val(Left([DATE],2)) AS Month, CInt(5*(Int([AVGTEMP]/5))) AS Temp,
IIf([WINDY]="1","Yes","X") AS WindowOpen, cNHA_Num_Windy([num_windy])
AS Windows, cNHA_Windmin_hours([windmin]) AS Hours,
tblCHAD_Data.STATE, tblCHAD_Data.AVGTEMP
FROM (((tblNHAPS_Dataset7 INNER JOIN tblNHAPS_Dataset8 ON
(tblNHAPS_Dataset7.QUARTER = tblNHAPS_Dataset8.QUARTER) AND
(tblNHAPS_Dataset7.RESPID = tblNHAPS_Dataset8.RESPID)) INNER JOIN
tblNHAPS_Dataset1 ON (tblNHAPS_Dataset8.QUARTER =
tblNHAPS_Dataset1.QUARTER) AND (tblNHAPS_Dataset8.RESPID =
tblNHAPS_Dataset1.RESPID)) INNER JOIN c_tblChad_To_NHAPS ON
(tblNHAPS_Dataset1.QUARTER = c_tblChad_To_NHAPS.Quarter) AND
(tblNHAPS_Dataset1.RESPID = c_tblChad_To_NHAPS.Respid)) INNER JOIN
tblCHAD_Data ON c_tblChad_To_NHAPS.CHADID = tblCHAD_Data.CHADID
WHERE (((IIf([WINDY]="1","Yes","X"))="Yes") AND
((cNHA_Num_Windy([num_windy]))>=0) AND
((cNHA_Windmin_hours([windmin]))>=0));
```

**Table 7** The query c\_qryNHA\_Windows\_Summary used to summarize the number of windows open in a household by season and temperature.

```
TRANSFORM Avg(c_qryNHA_Windows_PreProcess.Windows) AS AvgOfWindows
SELECT CInt(5*(Int([AVGTEMP]/5))) AS AveTemp
FROM c_qryNHA_Windows_PreProcess
WHERE (((c_qryNHA_Windows_PreProcess.WindowOpen)="Yes") AND ((IIf([month]
Between 1 And 2,"4 Winter",IIf([month] Between 3 And 5,"1
Spring",IIf([month] Between 6 And 8,"2 Summer",IIf([month] Between 9
And 11,"3 Fall",IIf([month]=12,"4 Winter","NA")))))<>"NA") AND
((CInt(5*(Int([AVGTEMP]/5))))>=-20))
GROUP BY CInt(5*(Int([AVGTEMP]/5)))
PIVOT IIf([month] Between 1 And 2,"4 Winter",IIf([month] Between 3 And 5,"1
Spring",IIf([month] Between 6 And 8,"2 Summer",IIf([month] Between 9
And 11,"3 Fall",IIf([month]=12,"4 Winter","NA"))));
```

When creating a table such as this you will want to be careful to also create a table of counts for the number of records used to calculate the average in each cell of the table.

**Table 9** Results of the query  
c\_qryNHA\_Windows\_Summary showing the average  
number of windows open by season and temperature.

<i>AveTem</i>	<i>1 Spring</i>	<i>2 Summer</i>	<i>3 Fall</i>	<i>4 Winter</i>
0				4.00
5				1.50
10	1.50			1.50
15	1.50			1.50
20	1.50			1.78
25	1.50			1.97
30	2.00		1.50	2.03
35	1.78		2.13	2.04
40	1.83		2.65	1.98
45	2.81		2.06	2.35
50	2.92	2.92	2.73	2.56
55	3.36	3.13	3.19	2.56
60	4.03	4.21	3.58	2.93
65	4.32	5.14	4.32	2.44
70	4.34	5.49	4.02	4.63
75	5.11	4.67	5.45	8.75
80	4.90	5.05	4.27	
85	2.75	4.70	4.50	
90		2.88		
95		4.25		

# **Caveats, Known Problems, Notes**

## **Known Problems**

### **CHAD Application**

While not a problem, please be aware that the District of Columbia, or Washington, D.C. is represented as "DC" in the database's "State" field. To search for records from DC in CHAD Oracle, you must enter "DC" without the quotes.

An earlier version of the Oracle application sometimes lost screen elements in some browsers when one was scrolling up and down on a web page generated by Oracle. If this happens, press your browser's refresh or reload button.

All calculated weights should be the same for an individual across days when the individual has multiple diary days. This is implemented correctly in Oracle, not in Access.

Calculated weights should have a flag indicating such in exported files.

There is no easy method of exporting METS values. The **Obtain METS Details** button has not been implemented.

Adjusted METS functions occasionally return a zero to the CHAD Diary screen.

The entire table of METS should be recalculated when recalc is requested. Currently only those joined to an activity in the diary are calculated.

### **Nonrandom Sampling**

Users may wish to obtain "random" activity diaries from CHAD that meet certain conditions. However, the respondents sampled in CHAD are not randomly distributed. Some cities or states are oversampled because entire studies were carried out in those places. Other studies disqualified large groups of people such as smokers, or non-English speakers, or people without telephones. Many surveys were age-restricted, or they preferentially sampled certain target groups (such as asthmatics in Cincinnati). As a result, users should be cautioned against using CHAD to draw random individuals who represent Americans as a whole. One solution to this would be to create a set of statistical weights for CHAD to reflect the probability of sampling. The least biased study is NHAPS which was conducted using a random national phone survey. The study documentation talks more about how to pull unbiased samples from the raw data.

### **California Adult (CAA)**

The version of the California adult survey (1988) that was obtained by ManTech contains certain activity codes that are not included in the study documentation. Telephone conversations with Susan Lum of the California Air Resources Board have revealed that our documentation is identical to the "official" version, but that our electronic data are not. Several draft versions of the database were developed by various researchers, and evidently we have one of these. Our version nevertheless contains a finer breakdown (more codes) than does the final version. Use of these extra codes would require mapping the 100 undocumented codes into the nearest CHAD codes; some cases would require new CHAD codes. An alternative would be to request the official version of the California data; the CHAD user could then view the text description corresponding to each activity to determine any finer breakdown of activity categories.

### **NHAPS (NHA, NHW)**

NHAPS contains 9386 diary days. However, the RESPIDS field only contains about 9088 unique values. In order to uniquely identify an individual you need to combine the RESPIDS and QUARTER fields. Most of the NHAPS data tables have both of these items in them. When joining tables make sure that there is a link between both the RESPIDS and QUARTER fields of the two tables.

## Data Notes

### CHAD

All children age 4 and under with a student value of “X” were set to non-students (N=1026).

All children less than 10 (0-9 years) years old with an “X” for SMOKER were set to “N” for SMOKER (N=1999, studies effected: CAC, NHA, NHW).

### California Adults (CAA)

The table c\_tblCali\_To\_Chad was created to link CAA observations in tblCali\_DataSet1 (and others) and tblCHAD\_Data.

The CHADID is calculated from the original data as CHADID = “CAA” & right(“00000”&trim([PID],5)&”A”.

### California Youth (CAY)

The California Youth are really part of the California Adults Study (CAA). The data contain information on the youth that lived in households interviewed in the CAA study. The tables contain the full 1579 records, but only 183 of them have any information. The CASEID is the same as the adult in the same household.

The data is stored in tblCali\_DataSet3 and 5. Most item names start with “y”. There is no diary information for the youth.

The CHADIDs for the youth and the IDs in the California diary are 2 greater than the corresponding adult. The CHADID is calculated from original data as CHADID = “CAY” & right(“00000”&str(val(trim([PID]))+2),5)&”A”.

The table c\_tblCHAD\_To\_CA links the main California data tables with CHAD tables. The diaries use different IDs.

### Cincinnati (CIN)

Contains 2608 diary days. The typical individual, of the 887 was interviewed 3 times.

The table c\_tblCHAD\_ToCinByPerson links summaries made by person to CHAD.

CHAD values for SMOKER were set based on diary entries. If a person had smoked at all during any DIARY day, then SMOKER=“Y” was assigned for all of his diary days. If he never smoked then SMOKER=“N” was set.

Values for passive smoke exposure, SMOKER2, were set for a diary day (not an individual) to “Y” if any passive smoke exposure was recorded in the diary. Otherwise the value was set to “N”

## NHAPS (HNA, NHW)

NHAPS contains 9386 diary days and 9386 unique individuals.

Diary Days	Study
4723	A
4663	B

The Primary key for the original data tables (tblNHAPS\_Dataset1..20) are really two keys combined: RESPID and QUARTER. Together they form a unique ID. RESPID alone is not unique (there are only 9088 unique RESPID values). Join tables in these sets by linking both the RESPID and QUARTER fields together.

The table c\_tblCHAD\_To\_NHAPS can be used to link the raw tables with the consolidated database.

The NHAPS data contain a set of primary and secondary tables. The primary tables created by reading the NHAPS data are named tblNHAPS\_DatasetXX where XX=1, 2, 3, ..., 20. Then there are a set of update queries (e.g., qupd\_ExpToSmoke) that create subject specific tables (tbl\_ExpToSmoke) from the somewhat randomly organized data set tables. Once created, these subject specific tables, with matching select queries to transform some of the coding are used to display values in the forms. There are values in the data set tables that are not represented in the subject specific tables.

The individuals exposure to second hand smoke (SMOKER2) was re-coded in CHAD for NHAPS with the following logic.

Start with all SMOKER2 = "X" or missing.

If "Someone else smoked at house" (WHOSMHM),

OR

("Was at Bar/Night Club/Or Bowling is True

AND environmental tobacco smoke (ETS) is unknown")

then SMOKER2="Y".

If the respondent was a non-smoker AND the diary indicated that someone was smoking,  
then SMOKER2="Y"

If the respondent was a smoker

AND other smokers lived in the house

AND someone smoked at the house

then SMOKER2="Y"

If NOT someone smoked at their house

(No one smoked in house.)

AND NOT someone smoked in their diaries,

(No smoking in diary.)

then SMOKER2="N"

Otherwise, SMOKER2="N/A"

## **Washington DC (WAS)**

The Washington DC data set contains many incomplete responses for questionnaire data. This causes many variables to contain missing data.

## **Database Design and Development**

This process began in September 1997. Much of the initial database structure was based on comments received during contacts with potential users and was further modified after an exchange of ideas during demonstrations and presentations given in October and November 1997. The intervening time frame has seen improvements based primarily on users' experiences and the availability of data from additional studies.

### **Data Format and Content**

One of the most common concerns from the viewpoint of potential users was that a consolidated database might not retain information of interest that was collected in only one or two of the component studies. Also, the recoding of variables into a standard format inevitably results in some information loss, as the exact phrasing of the questions (and the allowed responses) differ somewhat from study to study. To answer these concerns, the CHAD database system contains (1) the CHAD consolidated database itself and (2) a section consisting of the original (or raw) data values obtained from the various component studies. The earliest "clean" versions of the original questionnaire and diary data and the modified data were entered into CHAD. Queries can be made of either the original data or the modified (according to CHAD quality standards) data. The original data contain many questions or details that pertain to only one or two of the studies and, therefore were not common enough to be part of a single, consolidated database. In the consolidated portion, a user can select a single search variable (e.g., respondent's age) and apply it to all of the studies in CHAD. Also, interpretation of the variables (e.g., activity codes) is identical for all studies, which is not the case for the original data.

Graphical user interface (GUI) screens were developed for the CHAD data and most of the raw data sets. For the raw data, part of this process involved examining the questionnaire record sheets and grouping the variables that covered similar topics. For instance, all variables relating to bathing or showering (i.e., BATH, BATH#) were grouped. This ensured that all of the necessary questionnaire data were entered into the database correctly and that access to the data would be convenient for future users. The modified data are augmented by a set of quality and METS variables. Some data, such as body weight and daily weather measurements were not part of the original surveys but were estimated or obtained from other sources and statistically modified for use.

### **Platform Selection**

Microsoft Access was chosen as the front-end interface because it is a relational database. A relational database is a linked collection of tables, each containing data about one subject. Because the tables are related, information from more than one table can be used at a time. A data value is stored just once, and this reduces required disk storage space and reduces time required for updating and retrieving data. For example, personal data on a respondent does not need to be repeated on every line of the activity diary, because it is constant for a diary day. In this relational database, the personal data file is linked to the diary file via CHADID, which uniquely identifies a single diary-day (further details on CHADIDs are given in Section 5.3.). Files and reports can then be generated from both the diary file and the personal (questionnaire) data at the same time.

A runtime version of the Access database can be created for distribution to those who do not have MS Access on their computer.

The database accessible over the Internet was developed in Oracle. At the beginning of the project that was the accepted standard for Internet databases at EPA.

### **Development Process**

## **Entering Data into CHAD Tables**

The earliest versions of the original questionnaire and diary data, available in ASCII format, have been read into tables contained in the database. Additional tables have been created to link each field to associated text. Questionnaire and diary information is stored in two different tables and is linked through a unique CHADID. Some fields in a main table is linked to associated data tables; this link enables access to textual information for the specified ASCII code.

## **Graphical User Interface Development**

The GUI of CHAD allows a user to browse through all records of a respondent's information. The GUI design includes buttons and pull-down menus for customizing queries. A user may access diary information for each respondent through the designated GUI screen. Also, a user may view explanations of the data codes through pull-down menus.

The main screen for each study depicts information for one individual. Personal information such as age, gender, ethnicity, and education may be viewed at the top portion of the screen. A user may view other questionnaire information by scrolling the bottom portion of the screen--clicking on the "Diary Information" button brings up the diary information for that person. Quality flag information for the questionnaire and diary information of that person may also be viewed in the same manner.

## **Querying CHAD Data**

A main menu screen was created so that the original and CHAD-formatted data could be accessed from one screen. Data for all questions common to most studies have been incorporated into CHAD. Available data from the studies may be read into CHAD tables.

Each record in the CHAD tables is identified by a unique CHADID, which consists of nine characters (e.g., DEN10807A). The first three characters represent the study source (e.g., DEN for Denver). The next five characters represent the person's ID used in the study, and the last character may be A, B, C, ...: A to denote the first person-day, B the second person-day, and so on for multi-day studies.

Data may be queried by clicking the Filter Data button. This button leads to a screen where the user can query data from options available on the screen. The user can save the selected query in one of the boxes available on the right portion of the screen, which can then be accessed later or used for batch queries. The queried data can be displayed on the screen, output in the form of an ASCII file, or printed as a report. A summary of queried data can be viewed by clicking on the Data Analysis button on the Filter screen. The Data Analysis screen displays the number of respondents belonging to the selected criteria. This set of queried data can be further analyzed through various charts available to the user.

## **Code Consolidation**

To conflate the selected databases and incorporate them into CHAD, a single code system was created to account for all activity and location codes across studies. This was accomplished after examining the code systems of each database and then comparing them for common categories.

## **Activity and Location Codes**

The NHAPS and California diaries had approximately 100 different activity codes and another 100 location codes, in contrast with earlier studies such as the Denver and Washington studies, which had only 10 to 20 codes each for activities and locations. In the NHAPS and California systems, a nesting format was used in which main activity categories were divided into



subcategories. In some cases, these subcategories were further divided into more specific categories. This structure accommodated a specific and comprehensive documentation of participants' activities over time. Because of the strong similarity between the coding systems of the NHAPS and the California studies, use of one of these as the CHAD coding system would mean that the studies with the greatest number of diary entries would need relatively little modification. The NHAPS system is slightly more succinct than the California code system. For these reasons, the NHAPS coding system was selected as the foundation for the CHAD system.

The NHAPS system has been refined and regrouped as necessary to incorporate coding systems from other studies. In some cases, these other systems had finer categories than did NHAPS, although more often they used broader categories. Generally, all the NHAPS coding categories are retained, along with several extra ones reflecting the categories used in other studies. Thus there are more than 100 each of CHAD activity and location codes.

In the CHAD coding system, activities and locations are each designated by five digits. For activities, the first digit is always 1, and for locations, it is 3; this prevents activity and location codes from being confused. Subsequent digits define finer distinctions. With a five-digit system, there is room to add new codes as they are needed. The resulting CHAD activity and location code systems are presented in the section on what is inside CHAD.

In retrospect, certain difficulties are evident with the NHAPS system. For instance, both indoor and outdoor (e.g., back yard) locations are parts of the primary location category of "at home". For exposure purposes, the primary distinction is the indoor/outdoor one, while the type of building is secondary.

## **Standardization of Time**

The original studies used a variety of protocols regarding diary start time, stop time, and resolution. All of the studies incorporated into CHAD so far have nominal 24-hour (or multiples thereof) durations and theoretical resolutions of one minute. In practice, the studies using written diaries did not start at a standard time, and they ended only approximately 24 hours later. The telephone recall surveys generally enforced a standard midnight-to-midnight time frame, but seldom were able to pinpoint activity start and stop times to the nearest minute, so that five-minute or in some cases only 15-minute resolution was achieved.

Diaries were standardized to an exact 24-hour length for CHAD. If the original diaries were longer, then the first 24 hours were retained. If the diaries were shorter, then extra records with missing activities and locations were added to fill the missing time. Activities that cross hourly boundaries (i.e., that start before the top of the hour but finish after) are divided at the hour. This was motivated by the standard practice of hourly measurement of concentrations of air pollutants. A single value can be associated with each activity record in CHAD, assuming that concentration remains fixed within the hour. This assumption eases exposure modeling calculations.

Although each diary is standardized to 24 hours in length, there is no common start time that could be applied to all diaries. This has been resolved by "wrapping" the diary around so that in CHAD, all diaries begin and end at midnight. The true starting time of the original diary is indicated by the variable WRAPTIME, which is discussed later.

The CHAD variables have been re-coded with SAS programs. Quality flags were applied to the data during this process. A detailed discussion of the quality flags is presented later.

## Questionnaire Data

Each component study in CHAD was accompanied by a survey or questionnaire that recorded personal data from the respondent, such as age, gender, occupation, and house type. The list of such questions varied significantly in both length and focus, depending on the purposes of the original study. This made it difficult to consolidate the questions into a single set applicable across all studies.

There are two main philosophies on which a consolidated list can be based either one may include those questions common to most studies (effectively, the intersection set of questions), or include all questions asked in any study (the union of all questions). In the case of the studies in CHAD, the union would contain nearly 1000 questions per person-day, with the vast majority of these responses missing for any individual person. This would make CHAD unwieldy and would not effectively consolidate the data, as most questions would apply to only one of the component studies. Therefore, a small set of fairly common questions have been selected for CHAD, for which several (but not necessarily all) of the original studies provided data. In addition, a total of about 15 new variables have been composed and added to the CHAD data to assist the user and indicate data quality. These variables are described in detail later.

## Generation of body mass data from CHAD

Where not collected as part of the survey, body masses are generated from lognormal distributions based on their parameters as estimated by Burmaster and Crouch, (1997) from their exploratory method. The distributions are functions of age and gender.

## MET Data

A MET is defined as the ratio of an activity-specific metabolic rate to a person's resting metabolic rate. The resting metabolic rate can be estimated from personal data such as gender, age, weight, and occupation. Metabolic rates are important in the estimation of intake dose, pharmacokinetic modeling, and risk assessment. MET values are utilized in modeling efforts that seek to quantify the intake of substances and relate this to health effects.

The MET values for specific activities are not always single point estimates, but in many cases are provided as ranges or distributions. CHAD describes MET values in the form of uniform, exponential, normal, lognormal, and triangular distributions, as well as point estimates. The distributions and their associated parameters are estimated from data reported in the literature (values are given in Appendix 1).

The MET Data screen, which can be accessed by clicking the METS Data button on the Diary Information Activity Monitor screen lets the user conduct stochastic sampling of MET values from the specified distributions, and the resultant activity-specific and daily total energy expenditures will be calculated. At the press of a button, a new set of randomly based MET values can be produced. To assist users who wish to emphasize large MET values in their analyses, CHAD contains a flag (QCMETAB) that indicates cases for which the median MET value is greater than or equal to 3.

## Quality Assurance Procedures

Data quality flags are used to indicate problems with data content. There are three classes of such problems that can arise in the activity database. The first is missing or blank data. Missing data may result from unreported gaps in the diary data, diaries that are shorter than 24 hours, or missing questionnaire data. The second type of problem occurs if data are present but outside the

allowable range. If "clean" data sets are obtained, this type of error should occur very seldom, but such errors will occasionally be found. When these errors are detected, either they are corrected (if possible), or the values are deleted. The third and most difficult type of problem to resolve occurs when the data are present and the values are within allowable range, but are deemed to be of poor quality by a researcher or analyst.

In CHAD, missing diary data are filled by dummy activity records that have activity='X' and location='X'. No activities have been extended or inferred to fill in these gaps., although this may be done in a future version (in which case a flag would be set to indicate that the data were changed). The second type of problem occurs infrequently and can be eliminated with careful quality control. The third type of problem is by far the most extensive, as there are literally thousands of cases of unusual or unlikely events in the database. Some of these problems can be identified by using the data quality flags added to the CHAD database.

Interviews with prospective users made clear that most users would like to have access to the unadjusted (original) data, even if of poor quality, with the option of deleting or replacing data according to their needs. Therefore, quality flags are available to identify suspect records, but it is left to the user to determine which flags are important. For example, some diaries do not report any sleep time or any meals in a 24 hour period. These can be identified by using the QCSLEEP and QCMEALS flags. The user must decide whether or not the diaries are acceptable to them.

## **Completeness of Time of the Activity Diaries**

The activity data in each study should reflect activities for a 24-hour period. However, it was found that records may have missing activity data such that a complete 24-hour period was not represented. Also, in three major studies (Denver, Washington, and Cincinnati) the start and end times were not consistent from person to person. In CHAD, all diaries are exactly 24 hours long and start at midnight. As explained above, the variable WRAPTIME indicates the true diary starting time. Two sequencing numbers are provided: RECNUM indicates the new midnight-to-midnight order, while SEQ indicates the original chronological sequence.

The Denver study collected two diary days from each participant, that is, 48 continuous hours. The two diary days from each respondent were first recombined into one 48-hour diary. If data were available for less than 45 hours, then only one diary day was retained. The number of consecutive diary days from the same respondent is indicated by the variable NDAYS. In most cases of the original data, there was some time overlap between the first and second days. This was due to the extending or adding of records, often with missing activity codes, so that the diary days either started or ended on an hour. These records were removed from the data, and a stop time was set exactly 24 hours after the first record started. The second diary day then starts immediately and continues another 24 hours. If the second day was too short, then additional records were added with the activity and location codes set to missing (i.e., 'X'). The different diary days from the same individual are indicated in two ways: (1) The leftmost eight characters in the CHADID are the same, and the ninth (rightmost) is A for the first diary day, B for the second, and so on; and (2) the DAYNUM variable on the personal summary screen indicates the day number numerically.

## **Coding of Uncertain Activities**

In a number of diary records in the target databases, the applicable activity code was uncertain. These are coded as "U". Too many instances of the "U" code would result in a record with a large amount of uncertain activity time. Activities coded as "U" are distinct from activities not recorded at all, since other data may have been provided (such as the location, the presence of smokers,

etc.). Often, “U” signifies a combination of activities (such as eating and TV watching) for which the respondent was uncertain of the proper coding method.

## Coding of Missing Activities or Non-Sequential Activities

In a number of instances (e.g., Denver), blank spaces were found in some of the diary data, indicating missing data. All such time gaps were filled by new activities using "X" for both the activity and location code. In some cases there were time overlaps, usually because two diary days extended over each other. In most cases this was easy to resolve, but in 35 cases a specific decision was made on the best method of removing the overlap. Most studies did not have any gaps or overlaps in their diaries.

## Descriptive Variables

There are several additional descriptive variables that have been added to the CHAD database that can be used for quality control purposes, even though their names do not start with the letter “Q” (**Table 9**). For example, diaries with a small number of records (i.e., small RECCOUNT) might be eliminated from consideration on the basis that they do not provide a detailed enough description of daily activities.

All of the following quality flags were added to the data by the SAS programs that converted the original data into CHAD form. There is one SAS program for each of the studies incorporated into CHAD.

**Table 10** Other descriptive CHAD variables.

<i>Name</i>	<i>File</i>	<i>Type or Range</i>	<i>Description</i>
CHADID	Both	9-Char	Unique 9 character ID identifying each person-day
NDAYS	Summary	1,2,3,...	Total # of consecutive diary days from this person
DAYNUM	Summary	1-NDAYS	Day number (1=first 24 hours, 2=second 24 hrs...)
DAYTYPE	Summary	WD,WE	WD=Weekday, WE=Weekend (Sat or Sun)
RECCOUNT	Summary	24,25,26,...	Number of records in diary
WRAPTIME	Summary	0000-2359	Time at which a 24-hour diary day actually started

### **RECCOUNT--The number of individual activity records in a diary day.**

A small number of reported activities over a 24-hour period may be a sign of poor or unreliable diaries. Some activity databases delete diaries that contain fewer than some specified number of entries (records), but any specific number of activities would represent an arbitrary standard. Using the RECCOUNT variable as a filter, the CHAD user can select any desired minimum number. Note that all diaries have at least 24 records because there is one or more for every hour.

### **WRAPTIME--The time at which the 24-hour diary actually started.**

For consistency, all of the CHAD diaries are organized into a midnight-to-midnight form. However, some studies (e.g., Denver) did not have a standard starting time. In such cases, a consecutive 24-hour period was taken, starting at the actual start time, and the part until midnight was "chopped off" and added back at the other end. The WRAPTIME variable indicates if this

has been done, so the end user could "unwrap" the diary into its original form if desired. WRAPTIME indicates the actual start time of the diary in military time, so it has the value '0000' for diaries that were never wrapped (always midnight to midnight) and can range to a value of 2359. For wrapped diaries, the date and the weekday given in CHAD represent the day that contained more than 12 hours of time.

## Resolution of Quality Flags

The flagged diaries have not been modified. To resolve the quality issues being flagged, an analyst would have to examine each flagged diary in detail to determine whether changes should be made. This is a very labor intensive process since so many diaries are involved. It will be left to the user to "correct" any flagged records in the database.

## Determination of Quality Flags

In CHAD, there are two main types of quality indicators. Variables whose names start with the prefix "QF" are binary quality flags of the pass/fail type. These are set to "on" to indicate questionable data. The other type of quality indicators use the prefix "QC". These variables are counts (usually of time), which indicate how often within a diary some condition was met. For example, the variable 'QCSLEEP' records the number of hours of sleep time in each diary day. This variable can then be used to select diaries that contain sleep durations within a specified range.

### Quality Flag (QF) Variables

#### **QFACTLOC--Indicates inconsistent activity-location pairs in the diary.**

Each record in the diary portion of the database contains both an activity code and a location code. Certain activities are not compatible with certain locations. For example, travel activities (those with codes over 18000) should not take place in one's home (location codes less than 31000). A list of incompatible combinations was developed and implemented. Frequency tables of flagged data were prepared for certain studies (Denver, NHAPS, and California adults) and were used to revise the flagging criteria. Overall, about 1.8% of the activity records are flagged. This still amounts to several thousand records. Since certain activity-location combinations are unlikely but not impossible, the identification of suspect combinations will always be partly subjective. The total amount of time that QFACTLOC is "on" in each diary is summarized in the variable QCACTLOC.

#### **QFTRAVEL--Flag indicating possible travel time inconsistencies.**

One of the logical constraints on a valid activity diary is that travel to a remote location (such as a workplace) should be roughly matched in duration by a later return trip. There are complicating factors such as traffic jams, running errands, which may result in differing travel times. Another difficulty arises from simply trying to identify trips that qualify as out-and-back pairs. Without examining all the diaries individually, an attempt at flagging potential travel inconsistencies was made. The QFTRAVEL variable flags those person-days that meet four conditions: (1) the person is employed outside the home, (2) the sampled day is a weekday (Monday-Friday), (3) the total travel time in the morning rush hours (6-9 am) differs by more than a factor of 2 from the evening rush hour (4-7 pm) travel time, and (4) either the morning or evening travel time was at least 30 minutes. The travel time used is the travel time to/from work (Activity=18200, if recorded) plus the time for unspecified travel (Activity=18000). Only the NHAPS and California studies recorded the purpose of the travel.

Even with these restrictive conditions, a significant percentage of diaries are flagged (roughly 20% for studies with only general travel, and 5% of diaries with commuting identified as such). Most of these diaries are likely to be valid, but individual examination would be necessary to

make specific decisions. A refinement of the flag definition might be possible to better restrict the flagging to only work commuting trips, but it is not clear how best to do this since the purpose of a trip must be inferred from the diary as a whole.

**QFINFER--Indicates that a particular diary record was inferred (changed from the original).**

All data in CHAD is faithful to the original raw data. If the original data contained gaps in time, or missing activity or location codes, then these show up as CHAD records with missing codes (ACTIVITY=X and/or LOCATION=X). At some point it may be desirable to assign inferred activities in place of the missing data. If this is done, then the QFINFER flag will be set to indicate this, and the user will have the option of "turning off" the changes. Similarly, diaries with no missing time might be altered (to add meals or sleep time, for instance), and the QFINFER flag would then be used to indicate this condition. Presently, QFINFER=0 for all data since no changes have yet been made. The cumulative variable QCINFER indicates the total time per diary with QFINFER=1; this is also currently zero in all cases.

**QFMETAB--Indicates high (>3.0) metabolic rate factors.**

A distribution of possible METS values is associated with each activity. The QFMETAB flag indicates those activities for which the median value in the distribution is greater than or equal to 3. These activities will have an effect on the breathing rates of subsequent activities through the phenomenon of "excess post-exercise oxygen consumption" (EPOC). The total amount of time per diary with QFMETAB set to "on" is recorded in the count variable QCMETAB. QFMETAB is on in 134,001 or 15.3% of the diary records.

## **Quality Count (QC) Variables**

**QCSLEEP--The amount of time spent sleeping.**

The only two activities that are almost essential each day are sleeping and eating. For sleeping, the daily total time (rounded to the nearest whole number of hours) is recorded in the variable QCSLEEP, which appears with the questionnaire (summary) data. This flag does not necessarily indicate poor quality data, but may point to unusual behavior. People may spend little or no time sleeping on a given day. Poorly recorded diaries do not necessarily underestimate sleep time. However, diary days with little or no reported sleep time should probably be examined more closely to allow the analyst to decide on their reliability. QCSLEEP is equal to zero in 2.7% of the diary days.

**QCMISS--Missing diary time (less than 24 hours).**

The CHAD database is divided into person-days that contain exactly 24 hours of consecutive diary data. In some cases, the original data did not span a full 24 hours, or else it contained one or more gaps (time intervals) with no recorded activities. The total amount (in minutes) of such time is recorded in the QCMISS variable, so that incomplete diaries can be eliminated by filtering.

QCMISS is greater than zero indicating that the diaries are incomplete in 8.9% of the diary days.

**QCMEALS--Number of meals in the diary day.**

A meal is defined as one or more consecutive records of eating activity. Two meals must be separated by at least one non-eating activity. Roughly 8% of the diaries do not report any eating on that particular day. Examination of the diaries shows that in some cases eating is not reported because it is subsumed under some other activity (e.g., visiting a friend or relative for a very long period of time). Eating may also take place while watching TV or traveling, for example. There are a few diaries that report excessive eating (more than six hours per day). The diaries are flagged but are not excluded because of unusual eating patterns. The analyst must decide on the appropriateness of using any of these diaries.

**QCLONG--The duration of the longest activity in the diary.**

Some diaries report a single activity (meaning that both the ACTIVITY and LOCATION codes do not change) lasting 10 hours (600 min) or more. The duration in minutes of the longest such

activity is recorded in the variable QCLONG for each diary. Diaries with large values (say, >600) for QCLONG could be examined to determine whether or not the diary appears complete.

## Quality Flag Summaries

The remaining QC variables provide time summaries of other flags. This allows diaries to be excluded on the basis of thresholds of time flagged in each category.

**QCACTLOC--Total time (minutes) in diary with QFACTLOC flag set to "on".**

**QCEATIME--Total time (minutes) in diary spent eating.**

**QCINFER--Total time (minutes) in diary with QFINFER set to "on".**

**QCMETAB--Total time (minutes) in diary with QFMETAB set to "on".**

**QCHEAVY--Total time (minutes) in diary with HEAVYBR=1 (Heavy breathing).**

The NHAPS and Cincinnati studies asked the respondents to identify those activities accompanied by heavy breathing. The definition of heavy breathing is therefore subjective and not consistent across individuals or across studies. Nevertheless, this flag is provided so that the user can quickly find those diaries that are likely to represent active individuals. The fields QFMETAB and QCMETAB give similar information.

**Table 11** Count of quality flags identifying activities or diaries with questionable or significant content.

Quality Flag or Count	Condition/Count of	Count	Percent
<b>CHAD Diary Records</b>		<b>875,338</b>	<b>100.0%</b>
QFACTLOC	Inconsistent activity-location pairs	15,703	1.8%
QFMETAB	The median METS value for an activity is greater than 3	134,001	15.3%
QFINFER	Where diary data is filled in based on assumptions	0	0%
<b>CHAD Questionnaire Records</b>		<b>22968</b>	<b>100.0%</b>

<b>Quality Flag or Count</b>	<b>Condition/Count of</b>	<b>Count</b>	<b>Percent</b>
QCACTLOC	Diary days with QFACTLOC set in one or more diary records (QCACTLOC > 0).	5033	21.9%
QCMETAB	Diary days in which the QFMETAB flag is set (QCMETAB > 0).	20899	91.0%
QCINFER	Diary days with QFINFER set for one or more records (QCINFER > 0).	0	0.0%
QFTRAVEL	Diary days with with possible travel conflicts (QFTRAVEL = 1).	1084	4.7%
QCSLEEP	Diary days listing no sleep time (QCSLEEP = 0).	612	2.7%
QCMISS	Diary days with missing time (QCMISS > 0).	2051	8.9%
QCMEALS	Diary days with no meals reported (QCMEALS = 0).	1849	8.0%
QCEATTIME	Diary days with no eating time recorded (QCEATTIME = 0).	1849	8.0%
QCLONG	Diaries with the longest activity being greater than 10 hours (QCLONG > 600 min).	2020	8.8%
QCHEAVY	Diary days containing activities causing “heavy breathing” (QCHEAVY > 0).	3426	14.9%



## References

- Burmaster, DE, and AEC Crouch (1997) Lognormal distribution for body weight as a function of age for males and females in the United States, 1976-1980. *Risk Analysis* 17:497-505.
- Deyling, K. 1997. Development of a Consistent & Comprehensive Data Base for Assessing Multipathway/Intake Dose in Children & Others: Part I. Consolidated Report, prepared under Work Assignment 39 of Contract 68-D5-0049 for the U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC.
- Deyling, K. (1997). *Report of the development of a consistent & consolidated human activity data base for assessing multipathway/intake dose in children & others*. Research Triangle Park NC: ManTech Environmental Technology, Inc.
- Glen, G. (1999). *Revisions to the metabolic rate (MET) values and excess post-exercise oxygen consumption (EPOC) adjustments in the CHAD database*. Research Triangle Park NC: ManTech Environmental Technology, Inc.
- Glen, G., Lakkadi, Y., Tippet, J.A., and del Valle-Torres, M. (1997). *Development of NERL/CHAD: The National Exposure Research Laboratory Consolidated Human Activity Database*. Research Triangle Park NC: ManTech Environmental Technology, Inc.
- Goldstein, B., R. Tardiff, G. Hoffnagle, and R. Kester. 1992. Valdez Air Health Study: Summary Report. Prepared for Alyeska Pipeline Service Company, Anchorage, AK.
- Johnson, T. 1984. Study of Personal Exposure to Carbon Monoxide in Denver, Colorado. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.
- Johnson, T., J. Capel, and L. Wijnberg. 1986. Selected Data Analyses Relating to Studies of Personal Carbon Monoxide Exposure in Denver and Washington, DC. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.
- Johnson, T. 1989. Human Activity Patterns in Cincinnati, Ohio. Final Report. Prepared for Electric Power Research Institute, Health Studies Program, Palo Alto, CA.
- Klepeis, N., A. Tsang, and J. Behar. 1995. Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment. Final Report. Prepared for U.S. Environmental Protection Agency, National Exposure Research Laboratory, Las Vegas, NV.
- Lakkadi, Y., Glen, G., Smith, L., Tippet, J.A., and del Valle-Torres, M. (1998). *Development of the NERL/CHAD: enhancements to PC CHAD and creation of an internet application*. Research Triangle Park NC: ManTech Environmental Technology, Inc.
- Lakkadi, Y. and McCurdy, T. (1998). *CHAD user manual*. Research Triangle Park NC: ManTech Environmental Technology, Inc.

- Linn, W., C. Spier, and J. Hackney. 1993. Activity Patterns in Ozone-Exposed Construction Workers. *Journal of Occupational Medicine and Toxicology*, 2:1. pp. 1-14.
- McCurdy, T. 1997a. Work Assignment 39 of Contract 68-D5-0049. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC.
- McCurdy, T. 1997b. Modeling Multi-route Intake Dose in an Internally Consistent Manner. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC.
- Paul, R., J. Yoder, T. Johnson, and J. Capel. 1991. Documentation for the Multi-City Activity Data Base. Prepared for U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.
- Settergren, S.K., T.D. Hartwell, and C.A. Clayton. 1984. Study of Carbon Monoxide Exposure of Residents of Washington, DC.: Additional Analyses. Prepared for U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Research Triangle Park, NC.
- Shamoo, D., T. Johnson, S. Trim, D. Little, W. Linn, and J. Hackney. 1991. Activity Patterns in a Panel of Outdoor Workers Exposed to Oxidant Pollution. *Journal of Exposure Analysis and Environmental Epidemiology*, 1:4. pp. 423-438.
- Spier, C., D. Little, S. Trim, T. Johnson, W. Linn, and J. Hackney. 1992. Activity Patterns in Elementary and High School Students Exposed to Oxidant Pollution. *Journal of Exposure Analysis and Environmental Epidemiology*, 2:3. pp. 277-293.
- Wiley, J., J. Robinson, T. Piazza, K. Garrett, K. Cirkse, Y. Cheng, and G. Martin. 1991. Activity Patterns of California Residents. Final Report. Prepared for California Air Resources Board, Research Division, Sacramento, CA.
- Wiley, J. 1991. The Study of Children's Activity Patterns. Final Report. Prepared for California Air Resources Board, Research Division, Sacramento, CA.

## Appendix 1: METS Distributions for CHAD Activities

Key for activity distribution table.

Act = activity code (activity 10... refers to a set of activity values, rather than a single value)

Age = age group (1= <25; 2= 25-39; 3= >40; X = missing)

Occ. = occupation (X=missing)

Dist = indicator for distribution type (1 - normal; 2 - uniform; 3 - exponential; 4 - lognormal; 5 - triangular; 6 - point)

DistType = indicator for distribution type (first letter of distribution name)

Mean = mean value

Med = median value

StdDev = standard deviation

Min = lower bound

Max = upper bound

LTail = percent of the distributions left tail to cut off.

RTail = percent of the distribution's right tail to cut off.

The parameters are displayed with the precision that is used inside CHAD. The values may not be represented to full precision though. All parameters were calculated outside of CHAD and are simply recalled during calculations (e.g., the standard deviation is not calculated internally).

The parameters actually used to calculate the METS are given below for each distribution type.

- a. Normal Mean, Standard Deviation
- b. Uniform Minimum, Maximum
- c. Exponential Minimum, Standard Deviation
- d. Log Normal Maximum, Median
- e. Triangle Minimum, Mean, Maximum
- f. Point Mean
- g. Not Used
- h. Uses 1, 2, or 4 above.
- i. Uses 4 or 5 above.

In practice, the minimum and maximum are used to limit the range of the values. The LTail and RTail values reflect the minimum and maximum, but are not used in any calculations.

The distribution parameters reflect the underlying distribution and not the distribution as modified by RTail and LTail. The truncated distributions, those with non-zero LTails and RTails, force values to the minimum or maximum values so there tends to be more observations at these values.

**Table 12** METS distribution values used in CHAD.

Description	Act	Age	Occ	Dist	DistType	Mean	Med.	StdDev	Min	Max	LTail	RTail
Work, general	10...		ADMIN	9	LogNormal	1.7	1.7	0.3	1.4	2.7	0.16	0.01
Work, general	10...		ADMSUP	9	LogNormal	1.7	1.7	0.3	1.4	2.7	0.16	0.01
Work, general	10...		FARM	9	LogNormal	7.5	7.0	3.0	3.6	17.0	0.04	0.01
Work, general	10...		HSHLD	9	LogNormal	3.6	3.5	0.8	2.5	6.0	0.07	0.01
Work, general	10...		MACH	9	Uniform	5.3	5.3	0.7	4.0	6.5		
Work, general	10...		PREC	9	Triangle	3.3	3.3	0.4	2.5	4.5	0.00	0.01
Work, general	10...		PROF	9	Triangle	2.9	2.7	1.0	1.2	5.6		
Work, general	10...		PROTECT	9	Triangle	2.9	2.7	1.0	1.2	5.6		
Work, general	10...		SALE	9	Triangle	2.9	2.7	1.0	1.2	5.6		
Work, general	10...		SERV	9	Triangle	5.2	5.3	1.4	1.6	8.4		
Work, general	10...		TECH	9	Triangle	3.3	3.3	0.4	2.5	4.5		

Description	Act	Age	Occ	Dist	DistType	Mean	Med.	StdDev	Min	Max	LTail	RTail
Work, general	10...		TRANS	9	LogNormal	3.3	3.0	1.5	1.3	8.4	0.03	0.01
Breaks	10300			2	Uniform	1.8	1.8	0.4	1.0	2.5		
General household activities	11000			5	Triangle	4.7	4.6	1.3	1.5	8.0		
Prepare food	11100			4	LogNormal	2.6	2.5	0.5	2.0	4.0	0.13	0.01
Prepare and clean-up food	11110			3	Exponential	2.8	2.5	0.9	1.9	4.0		0.02
Indoor chores	11200			3	Exponential	3.4	3.0	1.4	2.0	5.0		0.01
Clean-up food	11210			2	Uniform	2.5	2.5	0.1	2.3	2.7		
Clean house	11220			3	Exponential	4.1	3.5	1.9	2.2	5.0		0.01
Outdoor chores	11300			1	Normal	5.0	5.0	1.0	2.0	7.0	0.00	0.02
Clean outdoors	11310			3	Exponential	5.3	4.5	2.7	2.6	6.0		0.00
Care of clothes	11400			3	Exponential	2.2	2.0	0.7	1.5	4.0		0.02
Wash clothes	11410			6	Point Est.	2.0	2.0		2.0	2.0		
Build a fire	11500			6	Point Est.	2.0	2.0		2.0	2.0		
Repair, general	11600			1	Normal	4.5	4.5	1.5	2.0	8.0	0.05	0.01
Repair of boat	11610			6	Point Est.	4.5	4.5		4.5	4.5		
Paint home / room	11620			3	Exponential	4.9	4.5	1.4	3.5	6.0		0.00
Repair / maintain car	11630			5	Triangle	3.5	3.4	0.4	3.0	4.5		
Home repairs	11640			3	Exponential	4.7	4.5	0.7	4.0	6.0		0.00
Other repairs	11650			2	Uniform	4.5	4.5	1.4	2.0	7.0		
Care of plants	11700			2	Uniform	3.5	3.5	0.9	2.0	5.0		
Care for pets/animals	11800			2	Uniform	3.3	3.3	0.1	3.0	3.5		
Other household	11900			3	Exponential	6.6	5.5	3.6	3.0	9.0		0.00
Child care, general	12000			4	LogNormal	3.1	3.0	0.7	2.5	5.0	0.20	0.01
Care of baby	12100			2	Uniform	3.3	3.3	0.1	3.0	3.5		
Care of child	12200			2	Uniform	3.3	3.3	0.1	3.0	3.5		
Help / teach	12300			2	Uniform	2.8	2.8	0.1	2.5	3.0		
Talk /read	12400			2	Uniform	2.8	2.8	0.1	2.5	3.0		
Play indoors	12500			2	Uniform	2.8	2.8	0.1	2.5	3.0		
Play outdoors	12600			2	Uniform	4.5	4.5	0.3	4.0	5.0		
Medical care-child	12700			2	Uniform	3.2	3.2	0.1	3.0	3.3		
Other child care	12800			2	Uniform	3.0	3.0	0.3	2.5	3.5		
Obtain goods and services, general	13000			5	Triangle	3.8	3.7	0.8	2.0	6.0		
Dry clean	13100			2	Uniform	3.3	3.3	0.4	2.5	4.0		
Shop / run errands	13200			5	Triangle	3.7	3.6	0.8	2.0	6.0		
Shop for food	13210			5	Triangle	3.9	3.8	0.8	2.2	6.0		
Shop for clothes or household goods	13220			2	Uniform	3.4	3.4	0.6	2.3	4.5		
Run errands	13230			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Obtain personal care service	13300			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Obtain medical service	13400			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Obtain govern't / financial services	13500			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Obtain car services	13600			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Other repairs	13700			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Other services	13800			2	Uniform	3.5	3.5	0.6	2.5	4.5		
Personal needs and care, general	14000			2	Uniform	2.0	2.0	0.6	1.0	3.0		
Shower, bathe, pers. hygiene	14100			1	Normal	2.0	2.0	0.3	1.0	4.0	0.00	0.00
Shower, bathe	14110			2	Uniform	3.0	3.0	0.6	2.0	4.0		
Personal hygiene	14120			2	Uniform	1.8	1.8	0.4	1.0	2.5		
Medical care	14200			2	Uniform	1.8	1.8	0.4	1.0	2.5		
Help and care	14300			4	LogNormal	3.1	3.0	0.7	2.5	5.0	0.20	0.01
Eat	14400			2	Uniform	1.8	1.8	0.1	1.5	2.0		
Sleep or nap	14500			4	LogNormal	0.9	0.9	0.1	0.8	1.1	0.09	0.01
dress, groom	14600			6	Point Est.	2.5	2.5		2.5	2.5		
Other personal needs	14700			5	Triangle	2.0	2.0	0.4	1.0	2.9		
General educ. and pro. training	15000			4	LogNormal	1.9	1.8	0.7	1.4	4.0	0.23	0.01
Attend full-time school	15100			2	Uniform	2.1	2.1	0.4	1.4	2.8		
Attend day-care	15110			2	Uniform	2.3	2.3	0.4	1.5	3.0		
Attend K-12	15120			2	Uniform	2.1	2.1	0.4	1.4	2.8		
Attend college or trade school	15130			2	Uniform	2.0	2.0	0.3	1.4	2.5		

Description	Act	Age	Occ	Dist	DistType	Mean	Med.	StdDev	Min	Max	LTail	RTail
Adult education and special training	15140			2	Uniform	1.8	1.8	0.2	1.4	2.2		
Attend other classes	15200			2	Uniform	2.2	2.2	0.5	1.4	3.0		
Do homework	15300			6	Point Est.	1.8	1.8		1.8	1.8		
Use library	15400			2	Uniform	2.3	2.3	0.4	1.5	3.0		
Other education	15500			2	Uniform	2.8	2.8	0.7	1.5	4.0		
General entertainment / social activities	16000			4	LogNormal	2.2	2.0	1.1	1.0	6.0	0.07	0.01
Attend sports events	16100			2	Uniform	2.7	2.7	0.8	1.4	4.0		
Participate in social, political, or religious activities	16200			2	Uniform	1.7	1.7	0.2	1.4	2.0		
Practice religion	16210			2	Uniform	1.7	1.7	0.2	1.4	2.0		
Watch movie	16300			2	Uniform	1.3	1.3	0.2	1.0	1.6		
Attend theater	16400			2	Uniform	1.7	1.7	0.4	1.0	2.3		
Visit museums	16500			2	Uniform	2.5	2.5	0.3	2.0	2.9		
Visit	16600			2	Uniform	1.5	1.5	0.3	1.0	1.9		
Attend a party	16700			4	LogNormal	3.3	3.0	1.4	1.5	8.0	0.05	0.01
Go to bar / lounge	16800			4	LogNormal	3.3	3.0	1.4	1.5	8.0	0.05	0.01
Other entertainment / social events	16900			2	Uniform	3.8	3.8	1.3	1.5	6.0		
Leisure, general	17000	20		8	LogNormal	5.7	5.0	3.0	1.4	16.0	0.01	0.01
Leisure, general	17000	30		8	Normal	5.0	5.0	2.0	1.0	9.0	0.02	0.02
Leisure, general	17000	40		8	Normal	4.5	4.5	1.4	1.7	7.3	0.02	0.02
Sports and active leisure	17100	20		8	LogNormal	5.7	5.0	3.0	1.4	16.0	0.01	0.01
Sports and active leisure	17100	30		8	Normal	5.0	5.0	2.0	1.0	9.0	0.02	0.02
Sports and active leisure	17100	40		8	Normal	4.5	4.5	1.4	1.7	7.3	0.02	0.02
Participate in sports	17110	20		8	LogNormal	3.6	3.2	1.9	1.4	10.0	0.05	0.01
Participate in sports	17110	30		8	LogNormal	3.6	3.2	1.9	1.4	10.0	0.05	0.01
Participate in sports	17110	40		8	LogNormal	3.4	3.0	1.7	1.4	9.0	0.05	0.01
Hunting, fishing, hiking	17111	20		8	Normal	5.6	5.6	2.1	1.4	9.8	0.02	0.02
Hunting, fishing, hiking	17111	30		8	Normal	5.8	5.8	2.4	1.0	10.6	0.02	0.02
Hunting, fishing, hiking	17111	40		8	Normal	4.7	4.7	1.8	1.1	8.3	0.02	0.02
Golf	17112	20		8	Uniform	3.8	3.8	1.0	2.0	5.5		
Golf	17112	30		8	Uniform	3.8	3.8	1.0	2.0	5.5		
Golf	17112	40		8	Uniform	3.5	3.5	0.9	2.0	5.0		
Bowling / pool / ping pong / pinball	17113			2	Uniform	3.0	3.0	0.6	2.0	4.0		
Yoga	17114			5	Triangle	3.1	3.2	0.6	1.4	4.0		
Participate in outdoor leisure	17120	20		8	LogNormal	4.2	3.9	1.5	2.0	9.0	0.03	0.01
Participate in outdoor leisure	17120	30		8	LogNormal	4.2	3.9	1.5	2.0	9.0	0.03	0.01
Participate in outdoor leisure	17120	40		8	Point Est.	3.5	3.5		0.0	0.0		
Play, unspecified	17121	20		8	LogNormal	4.2	3.9	1.5	2.0	9.0	0.03	0.01
Play, unspecified	17121	30		8	LogNormal	4.2	3.9	1.5	2.0	9.0	0.03	0.01
Play, unspecified	17121	40		8	Point Est.	3.5	3.5		0.0	0.0		
Passive, sitting	17122			2	Uniform	1.5	1.5	0.2	1.2	1.8		
Exercise	17130	20		8	LogNormal	5.8	5.5	1.8	1.8	11.3	0.00	0.01
Exercise	17130	30		8	Normal	5.7	5.7	1.8	2.1	9.3	0.02	0.02
Exercise	17130	40		8	Normal	4.7	4.7	1.2	2.3	7.1	0.02	0.02
Walk, bike, or jog (not in transit)	17131	20		8	LogNormal	5.8	5.5	1.8	1.8	11.3	0.00	0.01
Walk, bike, or jog (not in transit)	17131	30		8	Normal	5.7	5.7	1.8	2.1	9.3	0.02	0.02
Walk, bike, or jog (not in transit)	17131	40		8	Normal	4.7	4.7	1.2	2.3	7.1	0.02	0.02
Create art, music, work on hobbies	17140	20		8	Normal	5.3	5.3	1.8	1.7	8.9	0.02	0.02
Create art, music, work on hobbies	17140	30		8	Normal	5.2	5.2	1.7	1.7	8.9	0.02	0.01
Create art, music, work on hobbies	17140	40		8	Normal	3.8	3.8	1.0	1.8	5.8	0.02	0.02
Participate in hobbies	17141			5	Triangle	2.8	2.7	0.8	1.5	5.0		
Create domestic crafts	17142			5	Triangle	2.0	1.9	0.4	1.5	3.0		
Create art	17143			2	Uniform	2.5	2.5	0.3	2.0	3.0		
Perform music / drama / dance	17144	20		8	Normal	5.3	5.3	1.8	1.7	8.9	0.02	0.02
Perform music / drama / dance	17144	30		8	Normal	5.2	5.2	1.7	1.7	8.9	0.02	0.01
Perform music / drama / dance	17144	40		8	Normal	3.8	3.8	1.0	1.8	5.8	0.02	0.02
Play games	17150			5	Triangle	3.3	3.2	0.6	2.4	5.0		
Use of computers	17160			2	Uniform	1.6	1.6	0.2	1.2	2.0		

Description	Act	Age	Occ	Dist	DistType	Mean	Med.	StdDev	Min	Max	LTail	RTail
Recess and physical education	17170			2	Uniform	5.0	5.0	1.7	2.0	8.0		
Other sports and active leisure	17180	20		8	LogNormal	6.6	5.9	3.2	2.0	17.4	0.01	0.01
Other sports and active leisure	17180	30		8	Normal	6.0	6.0	2.0	2.0	10.0	0.02	0.02
Other sports and active leisure	17180	40		8	Normal	4.8	4.8	1.4	2.0	7.6	0.02	0.02
Participate in passive leisure	17200			4	LogNormal	1.3	1.3	0.3	1.0	2.3	0.14	0.01
Watch	17210			2	Uniform	1.5	1.5	0.2	1.2	1.8		
Watch adult at work	17211			2	Uniform	0.0	0.0	0.0	1.2	0.0		
Watch someone provide childcare	17212			2	Uniform	0.0	0.0	0.0	1.2	0.0		
Watch personal care	17213			2	Uniform	0.0	0.0	0.0	1.2	0.0		
Watch education	17214			2	Uniform	0.0	0.0	0.0	1.2	0.0		
Watch organizational activities	17215			2	Uniform	0.0	0.0	0.0	1.2	0.0		
Watch recreation	17216			2	Uniform	2.7	2.7	0.8	1.4	4.0		
Listen to radio / recorded music / watch T.V.	17220			4	LogNormal	1.2	1.2	0.4	0.9	2.3	0.15	0.01
Listen to radio	17221			2	Uniform	1.2	1.2	0.1	1.0	1.3		
listen to recorded music	17222			2	Uniform	1.9	1.9	0.2	1.5	2.3		
Watch TV	17223			6	Point Est.	1.0	1.0		1.0	1.0		
Read, general	17230			2	Uniform	1.3	1.3	0.2	1.0	1.6		
Read books	17231			2	Uniform	1.3	1.3	0.2	1.0	1.6		
Read magazines / not ascertained	17232			2	Uniform	1.3	1.3	0.2	1.0	1.6		
Read newspaper	17233			2	Uniform	1.3	1.3	0.2	1.0	1.6		
Converse / write	17240			2	Uniform	1.4	1.4	0.2	1.0	1.8		
Converse	17241			2	Uniform	1.4	1.4	0.2	1.0	1.8		
Write for leisure / pleasure / paperwork	17242			2	Uniform	1.4	1.4	0.2	1.0	1.8		
Think and relax	17250			2	Uniform	1.2	1.2	0.1	1.0	1.3		
Other passive leisure	17260			2	Uniform	1.9	1.9	0.2	1.5	2.3		
Other leisure	17300			2	Uniform	1.5	1.5	0.2	1.2	1.8		
Travel, general	18000			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel during work	18100			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel to/from work	18200			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for child care	18300			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for goods and services	18400			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for personal care	18500			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for education	18600			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for organ. activity	18700			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for event / social act	18800			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for leisure	18900			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for active leisure	18910			4	LogNormal	2.3	2.0	1.3	1.0	7.0	0.10	0.01
Travel for passive leisure	18920			4	LogNormal	2.3	2.0	1.3	1.8	7.0	0.42	0.01