Analyzing Audit Trails in the National Survey on Drug Use and Health (NSDUH)

Michael Anthony Penne & Jeanne Snodgrass, RTI International & Peggy Barker, SAMHSA

Abstract

The National Survey on Drug Use and Health (NSDUH) (formerly the National Household Survey on Drug Abuse) is one of the largest government-sponsored CAI household surveys conducted in the United States on a continuous basis. Beginning with the 2002 NSDUH, Blaise audit trail files were retained for each interview conducted. For each screen encountered during the interview, these files retain the date and time of entrance and exit, the response entered and the keystroke used to move to the next screen. Currently, we are using these audit trail files to examine the following data quality issues: 1) instances where respondents back up and change previously recorded answers and the potential effects on estimates 2) areas within an interview where respondent break offs occur and 3) the amount of time required to complete various sections and individual questions in the interview, focusing primarily on unusual response times and possible explanations for these outliers. We analyzed the data on a flow basis as a means to rectify processing problems as they occur and to ultimately streamline our audit trail processing and analysis methods for future years. During data processing and these analyses, we have discovered future areas of analysis that may be of interest to those examining data quality through audit trails.

The current paper will present the results of our analysis of the 2002 NSDUH audit trail files regarding impact on estimates of prevalence and recency of substance use; break offs and timing data; methods for streamlining both data processing and analysis; and potential for further data quality analysis.

1. Introduction

Early research on the National Household Survey on Drug Abuse (NHSDA) computer-assisted interview (CAI) demonstrated the value of using audit trail files to evaluate the ease with which respondents navigated through the audio computer-assisted self-interview (ACASI) portions of the questionnaire (Caspar 1997; RTI, 1997). Building on this initial work, in 2002 we used audit trail files from the first six months of the 2002 survey to further investigate data quality issues. In 2003, we conducted the same analyses on all 2002 NSDUH audit trail files to determine if increasing the sample size created differences in our data. In this paper, we briefly describe the early work, discuss possible methods for streamlining the data processing portion, then focus on use of audit trails in the 2002 survey to investigate three aspects of data quality: question timing, respondent break-offs, and changes in estimates of recency and prevalence of substance use.

1.1 Background and Preliminary Analysis

In 2002, the survey name was changed from the National Household Survey on Drug Abuse (NHSDA) to the National Survey on Drug Use and Health (NSDUH). The NSDUH, sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), is the federal government's primary source of information on the magnitude and correlates of substance use and abuse in the United States household population. The NSDUH produces annual estimates of

rates of use, numbers of users, and other measures related to alcohol, tobacco, illicit drugs and the non-medical use of psychotherapeutic drugs in the civilian, non-institutionalized population of the U.S. The NSDUH collects data by administering questionnaires to a representative sample of persons aged 12 and older at their place of residence. The questionnaire contains both respondent self-administered sections and interviewer-administered sections (SAMHSA, 2002). Research Triangle Institute (RTI) is under contract to SAMHSA to conduct the NSDUH.

In 1999, the NHSDA data collection method was changed from paper-and-pencil (PAPI) to CAI. The interview sections on substance use and other sensitive topics were changed from self-administered PAPI to ACASI. These changes were prompted by research which showed that CAI questionnaires reduced input errors. Research also showed that use of ACASI increased comprehension for less literate respondents, and, by increasing privacy, resulted in more honest reporting of illicit drug use and other sensitive behaviors (Lessler et al, 2000). At the same time, the sample size was increased almost three-fold, to approximately 70,000 persons per year.

In preparing for the CAI conversion, in the fall of 1996, RTI conducted a field test with 435 respondents to compare two CAI versions of the 1996 PAPI. As part of this experiment, an MS-DOS Blaise program was modified to capture each keystroke made during the course of the interview by both the respondent and the field interviewer (FI). Additionally, the program captured timing data for each screen encountered and an indicator of whether the audio or screen component was turned on or off. Analysis of these keystroke files centered on the respondents' interaction with the ACASI program. Keystroke file analysis from both CAI versions indicated that roughly 60% of all respondents utilized a function key at least once, with very little difficulty. Overall, the analyses suggested that respondents would not experience any difficulty while interacting with the NHSDA ACASI portion of the instrument (RTI, 1997).

With the conversion of the NHSDA questionnaire to CAI, the Blaise programming language was also changed from an MS_DOS based version to a Windows-based version, and the keystroke file application was no longer compatible. This version of Blaise did, however, contain code that would capture data similar to the previous keystroke file application. These new files were called "audit trails".

Interest in capturing this type of data was renewed after the first nationwide administration of the CAI NHSDA instrument in 1999. Issues surrounding changes in data collection mode prompted renewed interest in audit trail files analysis, and the decision was made that some portion of audit trail files should be retained for analyses. Since a sample size of approximately 25,000 audit trail files would be more than adequate to answer any proposed questions, one in every three transmitted audit trail files was retained starting with the 2000 survey year[1].

Analysis of the 2000 NHSDA CAI instrument audit trail files was done to validate the findings from the 1996 keystroke file analysis. Sample sizes for this analysis incorporated a little over 12,500 audit trail files (one third of the data from two calendar quarters). The results validated the 1996 findings that respondents were having little difficulty completing the ACASI portion of the questionnaire (Caspar, 2000).

[1] A single transmitted file does not imply either a single interview only or an interview in its entirety. Field Interviewers are required to transmit any completed work nightly for every day that they work. This implies that a single transmission may contain more than a single interview. Additionally, if an interview experienced a breakoff, the first portion of the interview would reside in one transmission file while, assuming that the interview was eventually completed at a later date, the remaining portion of the interview would reside in an entirely different transmission file.

Through the 2000 survey year analysis, we encountered problems with only retaining one-third of the transmissions. Most notably, when a breakoff occurred within an interview but was completed at a later date, there were instances where we retained the last portion of an interview, but not the initial portion, and vice versa. As a result, the level of partially completed interviews for audit trail analysis was high, which made it difficult or sometimes impossible, to use the data. This problem was large enough, and the interest for further data quality analysis was strong enough, to warrant retaining all transmitted audit trail files. Conclusions about the difficulties associated with retaining one in every three transmitted files were not made until well into the 2001 survey year, and since the questionnaire structure could not be changed to retain all transmissions at that time, the change was implemented at the beginning of the 2002 NSDUH survey year.

In 2002, attention turned toward an investigation of audit trail analysis to use for possible data quality measures. To assist us in preparing for this analysis, we used the 2001 audit trail data to plan, develop, and test both our analysis methods and programs. We then tested on the first six months of audit trail data from the 2002 NSDUH (Penne, Snodgrass and Barker, 2002).

2 Current Issues and Methods

All current analysis methods and results presented in this paper are from the entire 2002 NSDUH survey. This analysis includes all completed and partially completed interviews, regardless of final response or usable status. An interview is classified as "usable" if the respondent provided data on lifetime use of cigarettes and at least nine other substances (SAMHSA, 2002). Since our analysis issues deal only with respondent interaction with specific variables within the questionnaire, as opposed to the entire questionnaire, we felt that all data, no matter what its case status, should be used for each analysis. Prior to analyzing the 2002 NSDUH audit trail data, we outlined each method for examination.

2.1 Timing

As was the focus of the timing analysis in our earlier studies of keystroke/audit trail data, our initial purpose for this analysis was to monitor respondent difficulties in utilizing ACASI. Though not presented in this paper, our results were the same as before: respondents were not having any more than expected difficulties in navigating the questionnaire. This reassurance allowed us to turn our focus towards FI performance and what possible ramifications it might have on respondent behavior.

With audit trail data, timing can be calculated for any screen or group of screens within the interview. It can also be calculated up to a hundredth of a second. Timing data collected with timestamp data within the Blaise program is limited to entire sections or modules, and is only calculated to a tenth of a second. We chose 6 measures within the NSDUH instrument to examine FI behavior with regard to timing:

Title	Screen Name	Description
Introduction to CAI	INTROCAI	The interviewer begins rapport with the respondent, and reads them the introduction to the study, if applicable
Calendar Set-up	CALENDAR	The FI sets up the reference calendar for the respondent to use throughout the interview.
ACASI Set-up	INTROACASI	The FI explains the Audio CASI set up to the respondent, adjusts the headphones and starts them on the tutorial.

End of ACASI	ENDAUDIO	The respondent finishes the ACASI portion, and turns the laptop over to the FI, who enters a 3-letter code to continue.
Verification Form	TOALLR3I	The FI interacts with the respondent to complete the verification Completion form.
Ending Interview with Respondent	INCENT01	The FI ends the interview with the respondent, and gives the respondent a cash incentive for their participation.

These single question measures were present in every interview, as opposed to all other questions in the CAI that are asked dependent upon skip logic, and we felt it would give us insight into how long FIs are spending on these important aspects of the questionnaire. Our analysis entailed calculating the amount of time spent on a question deemed appropriate for a respondent to fully comprehend the information on that particular screen, and then comparing the percentages of FIs falling below this amount of time. This allowed us to get a sense of the magnitude of shortcutting. The Gold Standard (GS) timing was developed for each of these six measures, at RTI, with 3 NSDUH staff members administering the NSDUH CAI instrument to 3 non-NSDUH staff members who ranged in age from 27 to 32. The interview files were sent to a Blaise programmer, who calculated the timing for each of the 3 interviews. In all six measures the highest timing across the three interviews was taken as the GS.

2.2 Breakoffs

In the breakoff analysis, we wanted to explore, 1) magnitude of breakoffs overall, 2) if there is any particular section of the interview where more than the average number of breakoffs are occurring, possibly indicating that respondents are having trouble or are being offended by the questions, and 3) if there are any interviewers who account for a large percentage of the overall number of breakoffs. We looked at all instances where a breakoff occurred within the interview. We examined the overall number of occurrences (total interviews with at least one breakoff), the number of breakoffs associated with individual FIs, and the location within the interview of the breakoffs. With regard to location, we classified breakoffs into three main categories: ACASI, FI administered sections and lastly, instances where the CAI system itself crashed. The ACASI sections were further categorized as: Tutorial, Main Core Drugs, Psychotherapeutic Drugs and remaining Non-Core sections. FI-Administered sections were categorized as those with some respondent interaction (instances where the FI is actually administering the questionnaire to the respondent) versus no respondent interaction (e.g., confirmation of FI identification number, FI debriefing questions, etc).

2.3 Changes in Prevalence and Recency of Use

When the NHSDA was conducted in PAPI, respondents were asked every follow-up question in the core drug modules; even if they indicated they had never used a particular drug. Data editing procedures used answers from all follow-up questions to classify a respondent as a drug user or non-drug user[2]. With the switch to CAI, routing logic was implemented that skipped the respondent out of a drug module if they indicated at the beginning of the module that they never used a particular drug. With this implementation, the previous editing procedure was no longer possible. Though not a direct correlation with the previous editing procedures, the audit trail data does allow us the benefit of observing instances where a respondent

[2] Since skip patterns did not exist within the PAPI, a respondent was directed through all follow-up questions about their use of a particular drug, regardless of the respondent's actual use. Any indication of use of a particular drug classified the respondent as at least a lifetime user of that substance.

backs-up to a previously answered question and changes their answer. Our first step was to determine if in fact there were a sufficient number of instances where this occurred to have an impact on drug use estimates. Since true reasons for a respondent changing their answers are speculative at best, we took a worst case scenario approach to our analysis. In other words, we presumed that all changes in answers were reflective of the respondent's desire that no one know their true drug usage. With this in mind, we examined the audit trails to find instances where a respondent at any time indicated a positive response to using a particular drug. We wanted to know how much drug use estimates would increase when assuming that if a respondent ever said 'Yes' to using a drug, they were in fact a drug user.

We focused on two aspects of interest. First, we analyzed all individual lifetime use gate questions[3] to view the overall effect on lifetime drug use prevalence. Second, we examined the effects these changes might have on recency of use of a drug. In this analysis, we sought to determine the most recent period ever that a respondent indicated they had last used a drug regardless of their final response. For example, if a respondent initially reported past 30-day use of a drug, but later changed their response to past year, this analysis would finalize that respondent as a past 30-day user.

2.4 Data Management

As noted earlier, starting with the 2002 survey year, all transmitted audit trail files were retained. Unfortunately this did not alleviate all of the data management difficulties. Specifically, the potential for the same FIs to replicate the same unique identification (ID) number for different individual interviews, or for two separate FIs to use the same ID, still existed. Additionally, since the audit trail analysis is currently separate from the questionnaire data analysis, corrections of incorrect ID numbers and subsequent linking of follow-up portions of an interview from an earlier interview breakoff were not applied to their respective audit trail files. Ultimately, this resulted in approximately 6.2% (~4,457 records) of our 2002 records having at least one duplicate ID represented on our analysis file. For processing expedience, and since we still retained a sufficient sample size to produce reliable estimates, we decided to remove these records with duplicate IDs from the timing, lifetime and recency analyses (n = 63,811). However, all records were considered within the breakoff analysis (n=68,268).

3. Results and Discussion

The timing data in Table A1 shows that when measured against a Gold Standard (GS) time, FIs are spending approximately the correct amount of time with the very beginning of the interview, at the introduction to the CAI instrument screen. However, once past this point, they spend less time than the GS on several important aspects of the questionnaire, such as setting up the calendar, setting up the ACASI tutorial, completing the verification form and ending the interview with the respondent. Conversely, they are taking longer than the GS in ending the ACASI portion of the interview. There are at least two reasons for these results. The first might be that the FIs are not performing their jobs correctly. The shorter times might indicate that FIs are reading through the questions too quickly for the respondent to understand or get the full meaning of the question.

The second possible reason could lie with the method of the GS calculation itself. Since GS timing was calculated using simulated interviews at RTI, it did not capture the situations that affect timing in many real field situations. The GS

[3] Hallucinogens, inhalants and psychotherapeutic drugs (pills) each contain multiple gate questions. Each question focuses on particular substances classified as a hallucinogen, inhalant or pill. For instance, hallucinogens ask questions on LSD, PCP, peyote, angel dust, mescaline, psilocybin, "Ecstasy" and any other hallucinogen not already asked about.

calculation, then, could be seen as a little higher than what is realistic. Another limitation to the GS calculation is that it was not calculated by age. The ages of the mock respondents (27, 29, and 32) were not representative of the entire NSDUH survey population and may not serve as a realistic benchmark measure across respondent age groups. Means for determining the validity of these GS measures would entail either modifying the measuring criteria for each aspect or conducting field tests, with an adequate sample and appropriate age range. On another note, the longer time taken to end the ACASI portion of the instrument could simply be accounted for by rapport between the interviewer and the respondent. At this point in the interview, the computer is handed back to the interviewer, and the interviewer may stay on this screen (ENDAUDIO) while talking to the respondent, or answering any of the respondent's questions about the ACASI.

Though the reasons for these timing results are somewhat speculative and GS times require additional validation, we did notice a trend across our analysis age groups. As the age group of interest goes up, we tend to see a decrease in the percentage of FIs below the GS time. This indicates that FIs are taking more time with older respondents, who tend to be less comfortable with the CAI.

To further assist us in drawing proper conclusions about the interviewers' time to complete, we produced graphical distributions of the audit trail timing data for the 12 and older population[4]. These graphs show detail which Table A1 does not. For instance, in Table A1, it appears that the interviewers are equally distributed around the GS for the introduction to the CAI, but from the graph we see that the distribution is in fact bimodal. Further, we observe that roughly 1/3 of the sample is taking less than 2 seconds to complete this screen (1/3 of the GS time). We see these same patterns in later screens also (Verification Form Completion and Ending the Interview With the Respondent). These results indicate that short-cutting by the interviewer is a possibility. The data in Table A1 will be investigated further by the NSDUH data quality staff, along with many other indicators of data quality, for flagging interviewers whose work may warrant a closer look.

Tables A2.1 and A2.2 illustrate that interview breakoffs within the ACASI are minimal. It does not appear that there are specific areas in the questionnaire where respondents are breaking off, and the majority of the breakoffs are within FI administered portions of the interview. This suggests that while breakoffs do happen, we can't identify any particular questions that generate a significant number of breakoffs. Within the ACASI portion, the breakoffs are clustered in the tutorial and non-core sections, suggesting that if a breakoff does happen during the ACASI, it probably will occur right at the beginning or towards the end where respondents may be getting tired of answering questions. Respondents' reluctance to continue with a long interview may also explain why a majority of breakoffs are occurring in the FI administered sections, which are at the beginning and the end of the entire interview. It should be noted that as a result of not being able to resolve all issues of duplicate ID numbers, and retaining all files for the breakoff analysis, the potential for double counting of interviews and instances per FI does exist.

An interesting finding is in Table A2.3. This shows that there are 90 interviewers with 5 or more instances of a breakoff, and 21 interviewers with 10 or more instances, which might suggest possible interviewer problems. In this instance, the audit trail files could be used as an FI monitoring tool, to alert project staff of

[4] Initially, we produced graphs for each of the six timing measures for each age group (including an overall). Upon careful inspection, we noticed only negligible differences in the shapes of the graphs among the different age groups. The only noticeable differences occurred in the magnitude of the distribution, which was expected as a result of the varying sample sizes in each age group. Hence, only results for the 12 and older population are presented.

interviewers who may have problems interacting with the CAI instrument. In the NSDUH, this breakoff data from Tables A2.1-3, along with the timing data, will be used by NSDUH data quality staff in monitoring FI adherence to study protocols.

Table A1. 2002 NSDUH Audit Trail Timing Analysis: 12+ Year-olds

Modules of Interest	Gold Standard Time (Minutes)	Respondents 12 + (n = 63,811)				Respondents 12-17 Years-old (n = 22,221)				Respondents 18-49 Years-old (n = 36,645)				Respondents 50+ Years-old (n = 4,945)			
		Median	% Below	% Equal	% Above	Median	% Below	% Equal	% Above	Median	% Below	% Equal	% Above	Median	% Below	% Equal	% Above
Introduction to CAI	0.10	0.10	49.81	2.00	48.19	0.08	53.32	2.07	44.61	0.10	48.47	1.92	49.61	0.17	43.88	2.29	53.83
Calendar Set-up	1.55	1.22	67.93	0.00	32.07	1.23	67.34	0.00	32.66	1.20	68.89	0.00	31.11	1.27	63.42	0.00	36.58
ACASI Set-up	2.23	1.48	74.62	0.00	25.38	1.52	74.21	0.00	25.79	1.45	75.95	0.00	24.05	1.70	66.63	0.00	33.37
End of ACASI	0.20	0.48	3.97	0.00	96.03	0.50	2.78	0.00	97.22	0.47	4.30	0.00	95.70	0.48	6.86	0.00	93.14
Verification Form Completion	0.42	0.08	75.87	0.00	24.13	0.07	75.61	0.00	24.39	0.10	76.26	0.00	23.74	0.10	74.07	0.00	25.93
Ending Interview with Respondent	0.52	0.05	84.98	0.00	15.02	0.05	84.42	0.00	15.58	0.05	85.44	0.00	14.56	0.05	84.11	0.00	15.89

Table A2.1 Distribution Summary Of Breakoff Analysis

Total Audit Trail Files	68,268
Total Breakoff Instances	1,563 (2.3%)
Total Respondents	1,411
Single Instance	1,293 (91.6%)
Two Instances	92 (6.5%)
Three Instance	20 (1.4%)
Four Instances	4 (0.3%)
Five Instance	2 (0.1%)
Total Field Interviewers	523

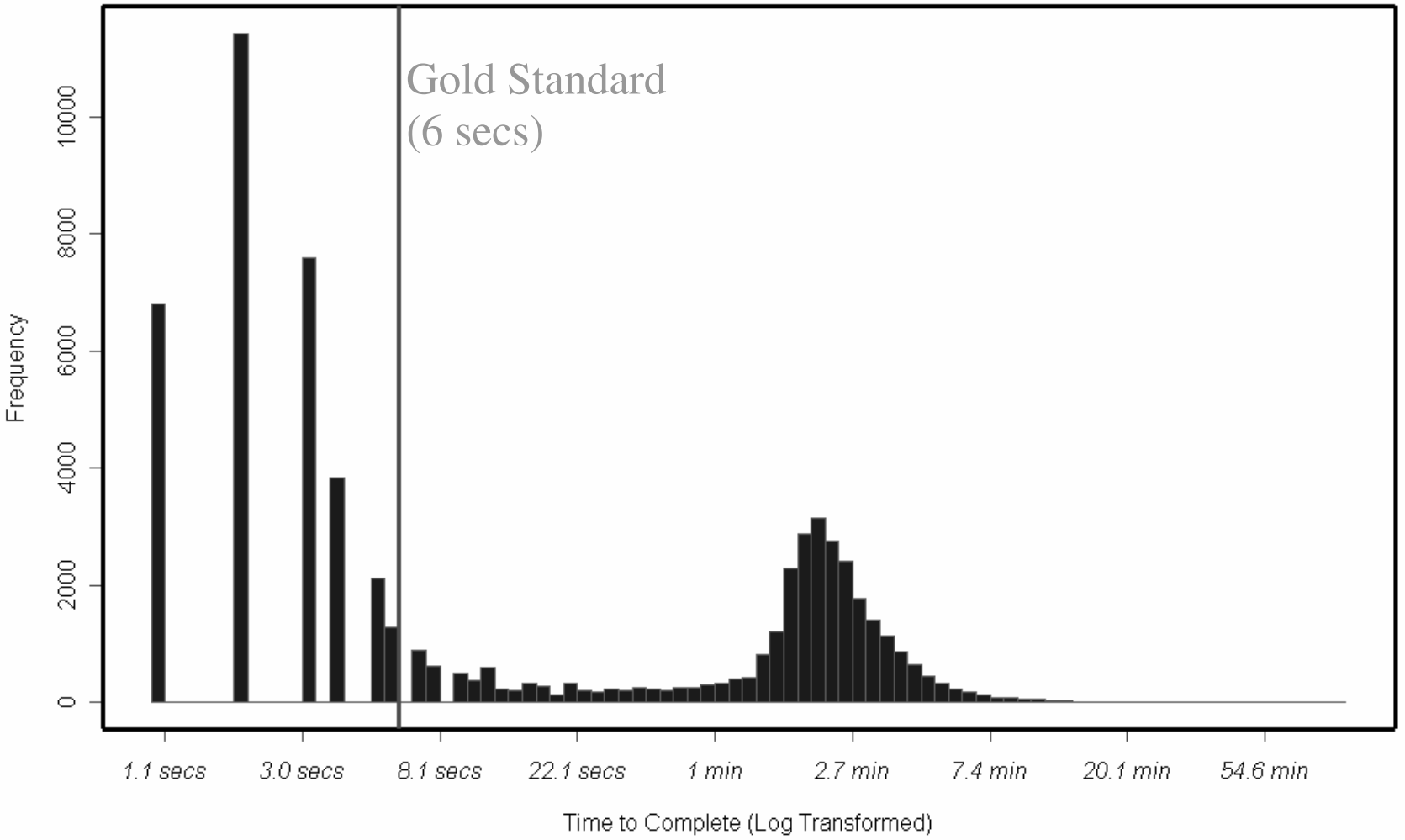
Table A2.2 Distribution of Breakoffs by Field Interviewer Interaction

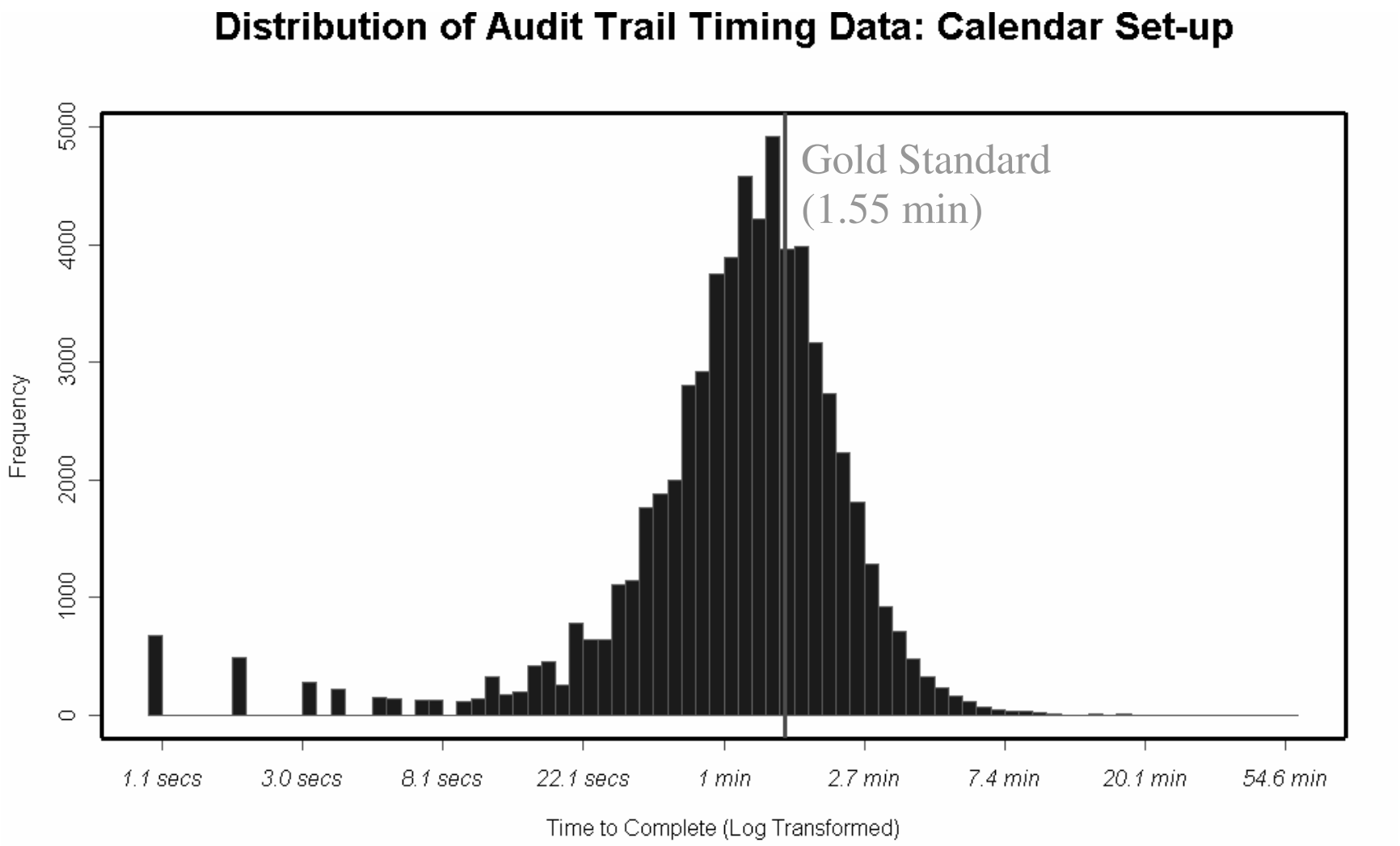
Domain of Interest	N	% of Total Breakoffs	% Within Domain
Field Interviewer Administered	926	59.25	
Respondent Interaction	543		58.64
No Respondent Interaction	383		41.36
ACASI	631	40.37	
Tutorial	189		29.95
Main Core Drugs	102		16.16
Psychotherapeutic Pills	52		8.24
Non-Core Sections	288		45.64
Systems Crash	6	0.38	

Table A2.3 Distribution of Breakoffs Attributable to FIs

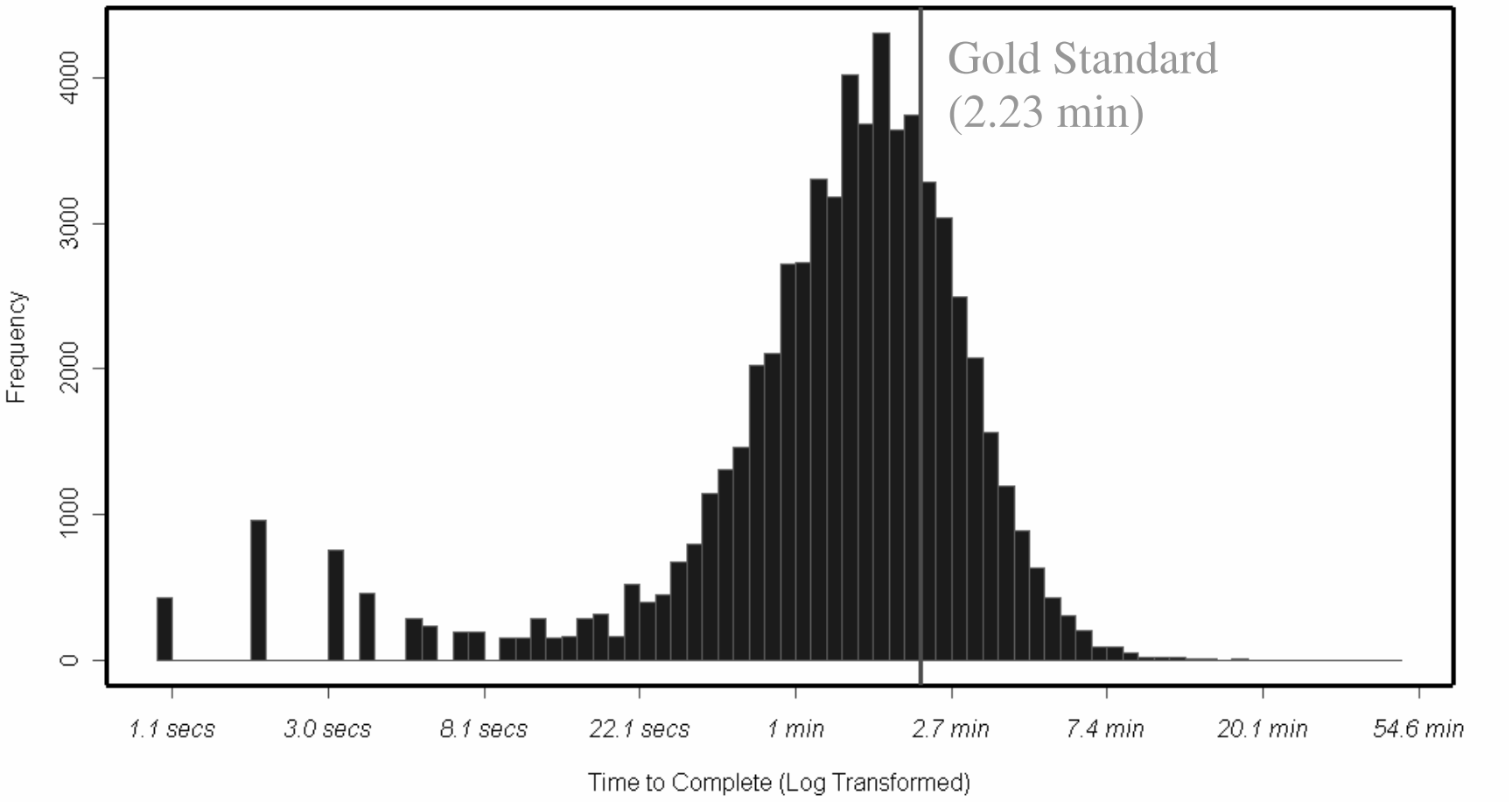
# of Breakoff Instances	Accountable Field Interviewers (n = 523)	
	N	%
1	222	42.45
2	107	20.46
3	68	13.00
4	36	6.88
5	30	5.74
6	16	3.06
7	14	2.68
8	6	1.15
9	3	0.57
10	7	1.34
11	3	0.57
12-45	11	2.10

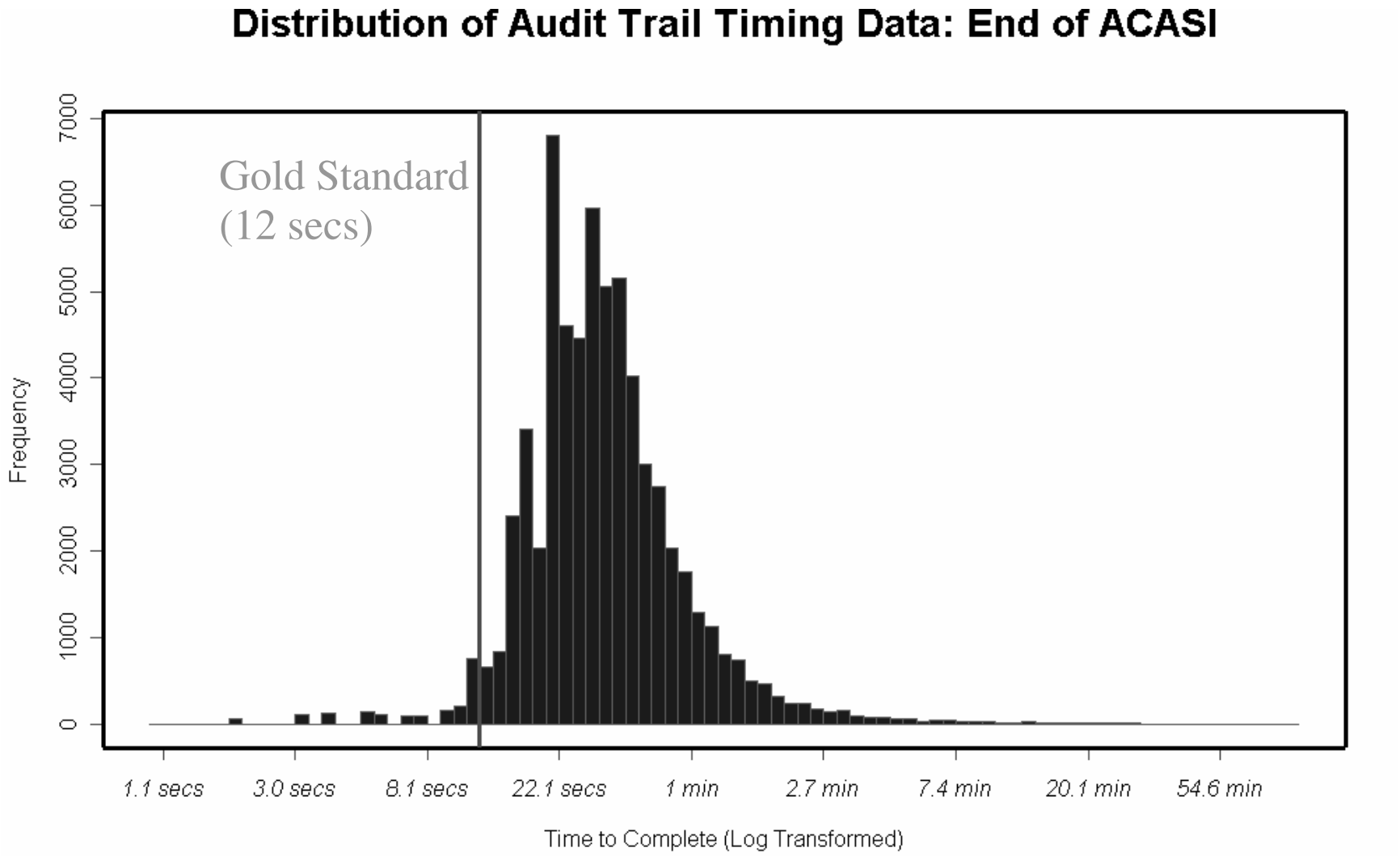
Distribution of Audit Trail Timing Data: Introduction to CAI



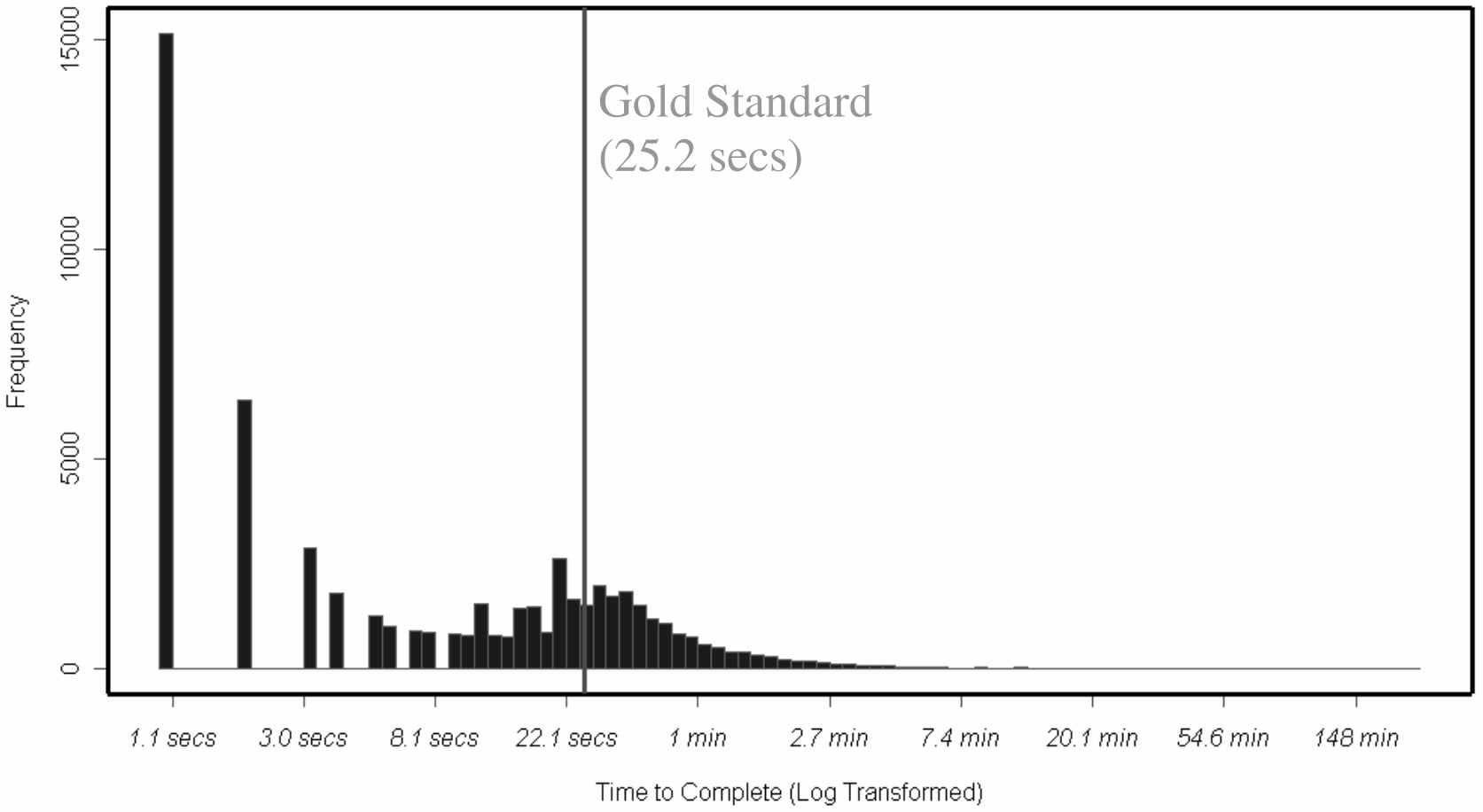

Distribution of Audit Trail Timing Data: Calendar Set-up
Gold Standard
(1.55 min)
Frequency
0
1000
2000
3000
4000
5000
1.1 secs
3.0 secs
8.1 secs
22.1 secs
1 min
2.7 min
7.4 min
20.1 min
54.6 min
Time to Complete (Log Transformed)

Distribution of Audit Trail Timing Data: ACASI Set-up




Distribution of Audit Trail Timing Data: End of ACASI
Gold Standard
(12 secs)
Frequency
0
1000
2000
3000
4000
5000
6000
7000
1.1 secs
3.0 secs
8.1 secs
22.1 secs
1 min
2.7 min
7.4 min
20.1 min
54.6 min
Time to Complete (Log Transformed)

Distribution of Audit Trail Timing Data: Verification Form Completion



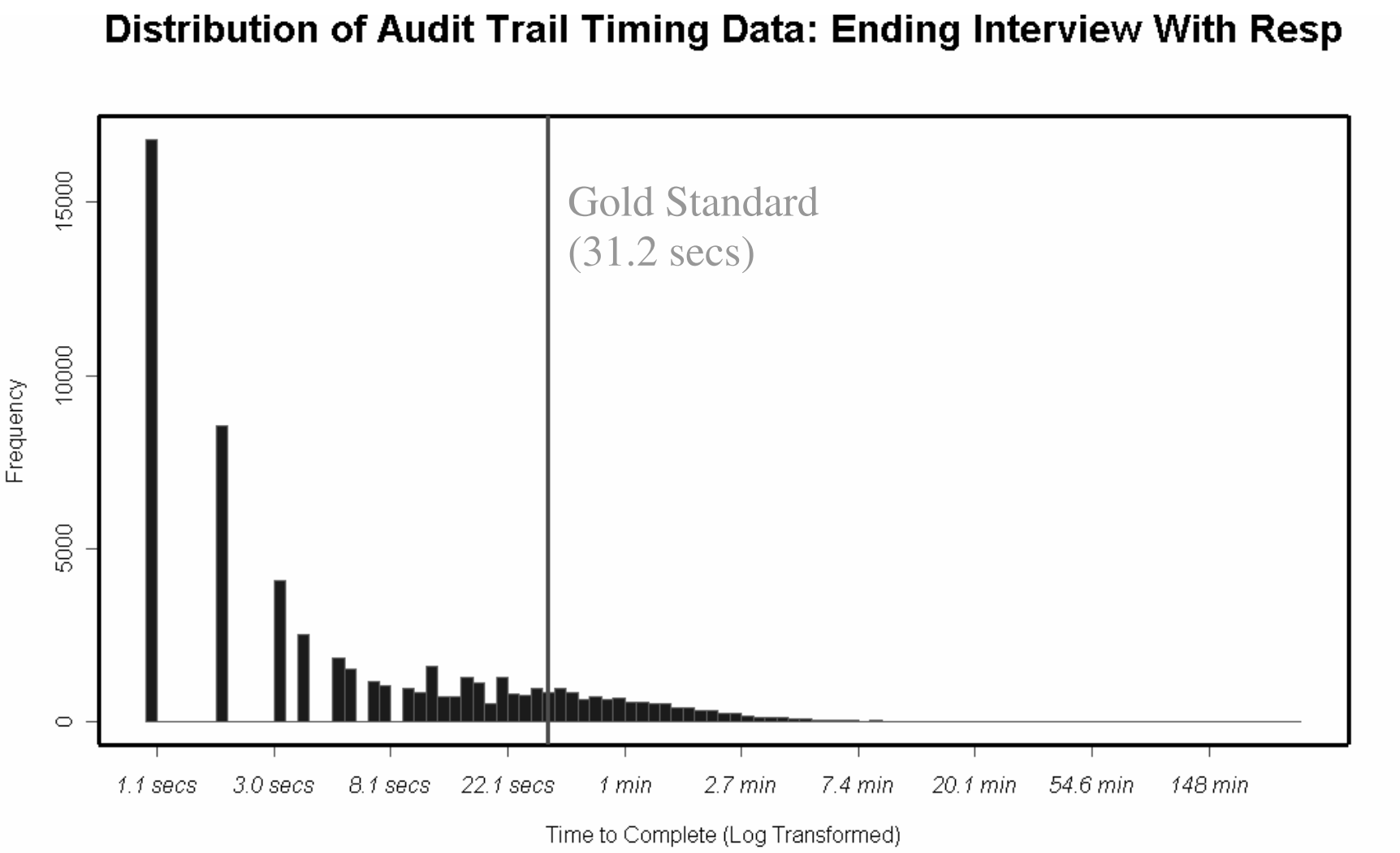

Distribution of Audit Trail Timing Data: Ending Interview With Resp
Gold Standard
(31.2 secs)
Frequency
0
5000
10000
15000
1.1 secs
3.0 secs
8.1 secs
22.1 secs
1 min
2.7 min
7.4 min
20.1 min
54.6 min
148 min
Time to Complete (Log Transformed)

Table A3. Relative Percentage Change in Estimated Prevalence for Lifetime Drug Use Measures: Raw Questionnaire Data Vs. Using Additional Audit Trail Information

Drug of Interest	Relative Percentage Change			
	12+ Years-old	12-17 Years-old	18-49 Years-old	50 + Years-old
Cigarettes	1.06	3.15	0.54	0.63
Alcohol	0.94	3.55	0.22	0.43
Marijuana	1.69	3.73	1.03	5.18
Cocaine	4.06	16.24	2.87	8.30
Heroin	10.48	27.59	7.69	23.68
Any Hallucinogen	2.46	9.77	1.24	6.67
Other Hallucinogen	30.41	51.25	23.15	23.52
Any Inhalant	5.62	11.67	2.78	15.04
Other Inhalant	32.30	38.90	23.42	*
Any Pain Reliever	11.72	17.51	9.11	30.00
Other Pain Reliever	34.29	39.07	31.15	52.77
Any Tranquilizer	5.69	13.84	3.72	18.13
Other Tranquilizer	28.20	38.32	24.35	*
Any Stimulant	6.13	14.26	4.26	6.03
Other Stimulant	27.09	40.13	22.61	*
Any Sedative	11.78	44.50	6.84	12.60
Other Sedative	27.07	38.37	20.93	*

* Low Precision, no estimate reported (initial sample < 40).

Tables A3 and A4 present the results of our analysis on the impact of using audit trail data in estimating both lifetime drug use prevalence and recency of use. With regard to the lifetime prevalence measures, we see a good deal of variation across various drugs and age groups. In some instances, particularly for more prevalent substances such as cigarettes and alcohol, these slight increases in prevalence would have little or no effect. However, for some of the rarer drugs, such as heroin or cocaine, even the smallest increases could have large effects. For instance, the 0.14 percentage increase in heroin accounts for a 10.48% relative increase from the estimate based on raw questionnaire data alone.

Additionally, some drug categories (Hallucinogens, Inhalants, and all psychotherapeutic drugs, or "pills") contain multiple gate questions with a last, "Other, specify" question. These questions prompt the respondent to indicate any additional drugs used within a particular category besides those specific drugs indicated in the preceding multiple gate questions. These questions, alone, account for more than a 25% relative increase in their respective prevalence rates. Though the "Other, specify" category represents a large relative increase when considered

by itself, it's impact on its respective overall drug category must be calculated with regard to all the other respective gate questions before any specific conclusions can be drawn. Though not shown here, the impact of the "Other, Specify" category only on the overall drug category, on average, accounts for about 26 percent of the lifetime prevalence increase (high of 41.8% for Hallucinogens and a low of 14.7% for Inhalants). These results direct the focus of the "Other, specify" analysis towards questionnaire methodology, in an attempt to discover why a large number of respondents are backing up and changing their answers to these specific questions.

Table A4. Relative Percentage Change in Recency of Drug Use Measures: Raw Questionnaire Data Vs. using Additional Audit Trail Information

Drug of Interest	Usage Classification	Relative Percentage Change			
		12+ Years-old	12-17 Years-old	18-49 Years-old	50 + Years-old
Cigarettes	Past 30 Day User	0.95	1.74	0.65	2.99
	Past Year User	0.07	0.12	0.09	0.92
	Used 1-3 Years Ago	0.29	1.10	0.21	1.02
	Confirmed Lifetime	0.76	0.70	0.67	1.06
	At least Lifetime User	712.11	657.64	*	*
Alcohol	Past 30 Day User	0.30	1.12	0.15	0.36
	Past Year User	0.78	1.22	0.41	1.80
	Confirmed Lifetime	1.19	1.45	1.00	1.38
	At least Lifetime User	509.70	537.10	*	*
Marijuana	Past 30 Day User	0.74	0.59	0.68	11.32
	Past Year User	0.65	0.66	0.65	0.00
	Confirmed Lifetime	0.37	1.43	0.26	0.40
	At least Lifetime User	1150.35	*	895.29	*
Cocaine	Past 30 Day User	3.02	1.72	3.28	*
	Past Year User	0.80	1.37	0.58	*
	Confirmed Lifetime	0.11	0.00	0.14	0.40
	At least Lifetime User	1236.39	*	*	*
Heroin	Past 30 Day User	7.45	*	5.30	*
	Past Year User	1.73	3.03	1.16	*
	Confirmed Lifetime	0.15	0.00	0.00	2.85
	At least Lifetime User	*	*	*	*
Hallucinogens	Past 30 Day User	2.73	1.63	3.29	*
	Past Year User	1.58	2.63	1.15	*
	Confirmed Lifetime	0.25	1.36	.18	0.00
	At least Lifetime User	766.30	*	*	*
Inhalants	Past 30 Day User	3.88	3.80	4.02	*
	Past Year User	2.00	1.77	2.37	*
	Confirmed Lifetime	0.26	0.29	0.26	0.00
	At least Lifetime User	419.14	355.43	566.48	*

* Low Precision, no estimate reported (initial sample size < 20).

For the recency results displayed in Table A4, we see relatively the same results as for lifetime prevalence; however, the magnitude of the change is now dispersed across the various recency periods. In the recency of use categories, "Confirmed Lifetime User" means that all of the respondent's questionnaire responses indicated only use prior to the past year, or in the situation of cigarettes, prior to the past three years. "At least a Lifetime User" stems from situations where a respondent has indicated some lifetime use within their audit trail data, but their questionnaire data indicates no use, so they were not subsequently ever routed though any of the respective recency of use questions. In short, there is neither questionnaire data nor audit trail data available on recency of use for these respondents. Though the prevalence and recency analyses have shown some interesting results, they have not shown any conclusive evidence that there are any problems with the reliability of the NSDUH estimates.

All data in Tables A1-4 were compared to the results from the six month 2002 NSDUH analyses. There were no significant differences at all, nor was there any shift in any of the final analyses. This suggests that, for future exploration with any large audit trails data set, half of the data set would be sufficient from which to draw conclusions.

Some discussion should be noted on our findings concerning data management of the Audit Trail files. Though the size of the NSDUH study is fairly large and at first glance appears to be burdensome, it was relatively easy to manage the data. With today's advances in computer processing abilities and relative low cost of storage space, we were able to store and process all the files with minimal down time of any analyses (usually on average about 8 hours to reprocess all the 68,000 files depending on the extent of the modifications requested and about 1.5 GB of storage space). This should provide some assurances to anyone who is considering performing a study of this size and wishes to utilize audit trail data.

Before using audit trail data on any large study, however, a significant change needs to be made to the processing methods used in this analysis. This entails processing the incoming transmitted files on either a daily or at most weekly basis. Though some automated machine editing procedures can be implemented to capture and resolve a bulk of the audit trail idiosyncrasies, it will never be able to resolve all of them. This may be relieved by implementing some automated process that will at least flag certain files that need further investigation or flag situations that indicate a problem. Some examples would be: 1) identifying files with duplicate IDs and resolving them; 2) linking breakoff interviews with subsequent follow-ups; 3) detecting instances where the FI is having technical difficulties and taking appropriate action to correct them.

Ultimately, this hands-on processing would enable project staff to immediately deflect any problems that they see occurring in the field and would produce a fairly clean data file that could be processed quickly. We suggest having an analyst clean and process the data as it is coming in. It is the strong opinion of the authors that when the bulk of audit trail data reaches the magnitude of the NHSDA/NSDUH and covers such a lengthy time-span (i.e., 1 calendar year, divided into 4 quarters) that the data files be streamline processed (e.g., as close to real-time processing as is feasible). There are several reasons for this. First, in post-processing after the end of a quarter of data collection, the data is manageable, but when dealing with the sheer number of records, detection and resolution of inconsistencies in file management become difficult. Hence, the probability of not being able to recover a complete audit trail increases the longer the duration between transmission and processing. More importantly, the closer to real-time that processing can occur, the better opportunity to utilize the data in a manner effective to resolving problems or to developing an enhancement/modification to continue to improve data quality.

As an example, if a new FI is experiencing technical difficulty at the beginning of the data collection period, it would be better to rectify this situation early on, instead of discovering this situation after numerous interviews have been conducted.

4. Future Analyses

The authors' experiences working with the 2002 NSDUH audit trail data has been an enlightening one. Although time and cost constraints limited the amount of analysis we could do with the 2002 data, we discovered many potential analyses that we could perform. One further analysis would be to utilize the timing information on the initial routing through a prevalence or recency question of interest and determine the number of subsequent questions answered prior to the respondent backing-up and changing their answers. For instance, if a respondent takes two seconds to respond to a drug lifetime use question with 'Yes" and then after going through one or two more questions, backs-up and changes that answer to 'No', we might be more inclined to concede that the respondent had rushed through the initial question and did not realize their initial response was incorrect. Taking an opposite approach, if a respondent takes a relatively long time to answer the initial question and then proceeds through five or more questions before returning to change their answer, we might be more inclined to believe that the respondent is hiding the truth of their drug use. Furthermore, a comparison could be made between the outlying timing data of respondents who change their answers within the interview, with those who do not. This approach is still speculative, but it does provide another means to utilize additional data from the audit trail files, and based on the criteria used to determine what is a long time or a sufficient number of questions to go through prior to backing-up, may provide a more realistic assessment of the effect of changing answers.

Also, with advances in data mining techniques and audit trail file data management procedures, this would be an excellent opportunity to model fraudulent cases from the audit trail data of cases already proven to be fraudulent. Though the NSDUH currently conducts a random 15% verification of all interviews, it might be useful to calculate a predicted probability of being a fraudulent interview and specifically designate these cases to be verified within that FI's 15%. This method would require 1) identifying fraudulent cases and using their audit trail data to establish response patterns and 2) continually updating the predictive model to include new data. We plan to continue to utilize audit trail data to monitor our survey methodology and search for additional avenues of research for which audit trails would be an indispensable aspect.

5. References

Penne, M., Snodgrass, J. and Barker, P. (2002)
Analyzing Audit Trails in the National Survey on Drug Use and Health: Means for Maintaining and Improving Data Quality. Presented at the International Conference on Questionnaire Development, Evaluation, and Testing Methods. Charleston, South Carolina. Nov. 14-17, 2002.

Caspar, R. and M. Couper (1997)
Using Keystroke Files to Assess Respondent Difficulties with an Audio-Casi Instrument. Survey Research Methods – Proceedings of the American Statistical Association, pp 239-244.

Caspar, R. (2000)
Using Keystroke Files to Identify Respondent Difficulties With an ACASI Questionnaire. Presented at Fifth International Conference on Social Science Methodology. Cologne, Germany. Oct. 3-6, 2000.

Lessler, J.T., R. Caspar, M. Penne and P. Barker (2000)
Developing Computer Assisted Interviewing (CAI) for the National Household Survey on Drug Abuse. Journal of Drug Issues 30(1), 9-34, 2000.

Research Triangle Institute (1997)
1997 National Household Survey on Drug Abuse: Results from Analyses of the Keystroke File for Field Test No.1. Prepared for SAMHSA. Contract: 283-96-0001. October 9, 1997

Substance Abuse and Mental Health Services Administration (SAMHSA). (2000)
Development of Computer-Assisted Interviewing Procedures for the National Household Survey on Drug Abuse (Office of Applied Studies, Methodology Series M-3, DHHS Publication No. SMA 01- 3514). Rockville, MD.

Results from the 2001 National Household Survey on Drug Abuse: Volume I. Summary of National Findings (Office of Applied Studies, NHSDA Series H-17, DHHS Publication No. SMA 02-3758). Rockville, MD. (2002)