Arrestee Drug Abuse Monitoring (ADAM II)

Request for OMB Review

Part B

Contract # TPD-NDC-10-K-00002

Revised Draft

December 11, 2012

Prepared for:
Fe Caces, Ph.D.
Executive Office of the President
Office of National Drug Control Policy (Planning and Budget)
Washington, DC 20503

Submitted by:
Dana Hunt, Ph.D.
Yuli Almozlino
Sarah Kuck Jalbert

Abt Associates Inc.
55 Wheeler Street
Cambridge, MA 02138

Request for OMB Review – Part B

Table of Contents

Part B.   Collection of Information Employing Statistical Methods

B.1   Respondent Universe and Sampling Methods

The Arrestee Drug Abuse Monitoring (ADAM) Program continues the seventh year of ONDCP's ADAM II program, a multi-year project designed to collect critical information about drug use among arrestees. The respondent universe is booked male arrestees in five counties designated as sites in the ADAM II program, which began in 2006. Data collection takes place in booking facilities in the sites from an annual, county-based representative sample of 350 arrestees (within 48 hours of arrest) per collection cycle, for an approximate total of 1750 respondent arrestees. Data collected include a voluntary and anonymous personal interview, a urine specimen, and data from arrestee booking sheet.

B.1.1   Sampling Sites and Facilities

The original 35 counties in the ADAM from 2000-2003 were selected through a competitive grant process sponsored by NIJ. Consequently, the counties did not constitute a probability-based sample of US counties, but were instead an intentional sample of grantee cities selected from major US urban areas.

In 2007 for ADAM II, ONDCP selected a **purposive sample** of 10 counties from the original 35 counties funded by NIJ to create 10 sentinel sites. Purposive sampling is not sampling by convenience, but instead is based on the broader interests of the study and selects sites with intent. The sites selected in 2007 were specifically identified based on geographic distribution (i.e., provide data from different sites across the country) and adequacy of prior data (complete quarters of collection from 2000-2003). Special consideration was given to sites east of the Mississippi to examine any possible spread of methamphetamine use from West to East in specific sites.

A site did not need to meet all of the above criteria to be considered, but had to meet at least the majority. The ten sites from 2007 continued into data collection for each year of 2008 through 2011. In 2012, federal budget reductions forced ONDCP to reduce the number of sites from 10 to 5, and to reduce data collection from two-week collections twice each year to one three-week collection period.

In 2012, ADAM II had to limit collection to 5 of those 10 counties, but continued to provide estimates and trend analysis for those 5 counties. Again the selection of the 5 sites in 2011 was purposive and based on specific criteria: case production and response rates, cost efficiency and sites in different geographic areas. For example, ONDCP wanted to retain at least one southern site (Atlanta or Charlotte) and one Western site (Sacramento or Portland), so that data on the unique drug use in those sites were collected. The final selection resulted in the following sites for the 2012 collection: New York, NY (Borough of Manhattan); Atlanta, GA (Fulton County); Chicago, IL (Cook County); Denver, CO (Denver County) and Sacramento, CA (Sacramento, County).

The five sites selected provide a *sentinel system.* Although in ADAM II the counties comprise a purposive sample and therefore a non-probability sample of counties, the sample of arrestees does constitute a probability-based sample of male arrestees booked into jails within those counties. The sites are not intended in any way to constitute a national sample nor to be generalized to represent

national trends. As sentinel sites, they do not represent the Nation as a whole and national estimates cannot be derived from the data; data are never combined in that fashion in analysis. However, they do represent the adult male booked arrestee population in each site, providing valuable information on variation in drug use in these sites and highlights important differences often masked in broader national estimates. It was ONDCP's interest in these site differences that guided the purposive selection of each ADAM II site.

Sampling within a County

The standard catchment area for each site is the county, although the sites are referred to by the primary city in that geographic region. Within each site, the number of booking facilities and the manner in which arrestees are moved from arrest to arraignment to holding varies.

In some cases, regardless of arresting agency, all bookings in the county take place in a single jail, while in other counties bookings may take place in multiple facilities across the county. Sampling plans are designed based on whether the site has a single or multiple booking facilities.

Two ADAM II counties have a single jail where all arrestees arrested in the county are brought to be booked pending further processing. The remaining three ADAM II counties, however, book in multiple jails. In these cases, each jail would constitute a stratum, and the result is a stratified random sample. However, resource constraints dictate that in some instances small booking facilities have to be excluded from the sample. For example, the Manhattan sample is restricted to the large central booking facility downtown (Manhattan House of Detention). The included jail, however, captures the overwhelming majority of the county bookings.[1] In Cook County (Chicago), the sample is limited to felony arrests and more serious misdemeanants who are brought from agencies throughout the city and county to be booked at the Cook County jail.[2]

ADAM II interviews booked male arrestees over 21 consecutive days in every sampled jail, with the exception of collection in Atlanta. In Atlanta (Fulton County and the City of Atlanta), there are two principal jails, one in Fulton County (Fulton County Jail) where all Fulton County felons and misdemeanants are booked. The second facility, the Atlanta Detention Center, books all misdemeanants arrested in the city proper by the Atlanta Police Department; all city felony arrests are taken to the Fulton County Jail. In 2012, ADAM II samples from the Atlanta Detention Center for the first 10 days and the Fulton County Jail for the second 11 days.

Sampling Method for Respondent Selection

To be eligible for inclusion in the ADAM II sample, in each county an arrestee must be: male, over 18, arrested no longer than 48 hours prior to the interview, coherent enough to answer questions and

---

[1]   It would have been possible to sample small jails and station interviewers in those facilities to provide representation for arrestees who do not appear in the included jails. However, so few arrestees are booked into the small jails that interviewers would spend most of their time waiting for arrivals. The resulting sample from the small jails would have a sampling variance that was so large that the small-jail estimate could not add appreciable information to a sample based exclusively on the large jail. A second jail in Manhattan was eliminated because it has a specialized caseload of public nuisance crimes and was excluded during 2002 and 2003 by ADAM.

[2]   A large proportion of minor misdemeanants is booked and released from over 100 small city precincts and suburban law enforcement facilities. It is impractical to sample from those facilities and, in any case, does not impact substantially estimates obtained from the facilities selected.

not in the jail as a hold for a federal agency. Field supervisors managing the sample (Lead Interviewers) and interviewers are trained to recognize when an arrestee is not sufficiently aware to provide consent or be interviewed. When such an individual is sampled and brought to the interviewing area by the assisting officer, if the interviewer sees the impairment, the officer is asked to return the arrestee to the cell and the interviewer will return to him, if possible, during the course of the interview shift or a replacement is selected and the reason for non-response is recorded on the facesheet. The sampled respondents who do not become cognizant enough to eventually be interviewed are recorded on the facesheet as the variables "violent or uncontrolled behavior" or "physically ill." The individual would be in one of these categories if his behavior continued to indicate serious drug or alcohol intoxication influencing his ability to give consent and be interviewed. This recording category on the facesheet would also be used for violent offenders who are deemed out of control and offenders who are brought in physically ill, not necessarily related to intoxication.

There are important practical challenges inherent in surveying the ADAM II target population and creating a representative sample of all males over 18 who are arrested over the course of a typical booking day. First, jails are chaotic, and law enforcement officials may not allow interviewers to be stationed within the jail during certain hours, particularly during hours when the booking process is most intense, due to security concerns and disruptions caused by our need to access booking records and arrestees. In contrast, there are also certain shifts during which so few arrestees are booked into the jail that interviewers stationed in the jail during those hours of relative quiescence could interview just one or two arrestees in an eight hour shift, compared to high-volume periods when many arrestees are available to be interviewed.

The sampling design in each facility divides the data collection day (and the interview cases) into periods of *stock* and *flow*. Interviewers arrive at the jail at a fixed time during the day. Call this H. They work a shift of length S. The *stock* comprises all arrestees booked between H-24+S and H, and the *flow* comprises all arrestees booked between H and H+S. For example, if interviewers start working at 4 PM and work for 8 hours, then the stock period runs from 12 PM to 4 PM, and the flow period runs from 4 PM to 12 PM. Cases are sampled from the stock and flow strata.

In the stock period, sampling is done from arrestees who have been arrested between H-24+S and H. This sampling begins at time H, and while arrestees identified as having been brought in during that time remain in the sample frame, interviewers can only interview those arrestees who remain in jail as of time H. In the flow period, sampling is done continuously for arrestees as they are booked between H and H+S.

To determine sampling rate, analysts estimate the number of bookings that occur during the stock and flow periods based on data for each facility reflecting the three-week period prior to the quarter's collection. Call the daily total N; call the number booked during the stock period N_S; and call the number booked during the flow period N_F. Then $N = N_S + N_F$. Supervisors set quotas from the stock and flow for each site equal to n_S and n_F, respectively, such that:

$$\frac{n_S}{n_F} = \frac{N_S}{N_F}$$

The actual sample size ($n = n_S + n_F$) depends on the number of interviewers and sometimes (for small jails) the number of bookings ($N = N_S + N_F$), since n cannot exceed N.

An analyst sorts arrestees based on booking time during the stock period and forms n_s equal sized strata based on that ordering. Sampling is systematic within each stratum: 1, n_S+1, n_S+2, etc. If the sampled arrestee is unavailable or unwilling to participate, the Lead Interviewer (LI) selects the nearest temporal neighbor—meaning the arrestee whose booking time occurs immediately after the arrestee who is unavailable or who declined. Replacement continues until the already established stock quota is filled. Because of administrative practices of jails and courts, arrestees are frequently unavailable to interviewers, i.e., they have been transferred to another facility, have already been released or are in court. The selection of the nearest neighbor is intended to reduce or eliminate any bias that otherwise would occur from apparently low response rates.

During the flow period, the sample manager selects the arrestee booked most recently and assigns an interviewer. If the arrestee is unavailable or unwilling to participate, the sample manager selects the next most recently booked arrestee as a substitute. This process continues until the workday ends at time H+S.

By addressing sample replacement as described above, we are replacing either an unavailable or a non-responding arrestee. This is a strategy well-known in the literature as "field substitution"[3]. Our specific strategy for field substitution, to select the nearest temporal neighbor for replacement, maintains sample balance. In ADAM II, we trade off a slightly less efficient variance estimate, to (1) maintain a self-balancing sample, and (2) considerably simplify the weighting procedure

This procedure produces a sample that is reasonably well balanced, meaning that arrestees have about the same probability of being included in the sample. If the sample were perfectly balanced, weighting would be unnecessary for unbiased estimates; and, in fact, estimates based on weighted and unweighted ADAM data are similar. The sample is not perfectly balanced, however, for several reasons.

First, while sample managers attempt to sample proportional to volume during the stock and flow periods based on recent data from the facility, achieving this proportionality requires information that is not available at the time that quotas are set. Analysts can only estimate N_S and N_F based on recent historical experience; furthermore, the LIs cannot know the length of time required to complete each interview because the length of the ADAM II interview depends on the extent of the arrestee's comprehension and cooperation level, as well as the extent of his reported drug use and market activity. Hence, the achieved value of n_F is variable.

Second, the number of bookings varies from day-to-day, but the number of interviewers arriving each day is constant. Days with a high number of bookings result in lower sampling probabilities than days with a low number of bookings. Furthermore, the number of bookings varies over the flow period, so that arrestees who are booked during periods with the most intensive booking activity have lower sampling rates than do arrestees who are booked during periods with the least intensive booking activity. Sampling rates do not vary as much across the stock period because of the way that the period is partitioned.

---

[3] Kish, L, (1965), *Survey Sampling*, New York: John Wiley & Sons, Inc. Lohr, S, (1999), *Sampling: Design and Analysis*, Pacific Grove, CA: Duxbury Press

Third, as noted above, arrestees can exit the jail during the stock period. The probability that an arrestee has been released prior to being sampled depends on both the time during the stock period when he is booked and his charge. The earlier that booking occurred during the stock period, the greater the opportunity he has had to be released. The more serious the charge, the lower the probability of being released, because serious offenders are more likely to be detained pending trial or require time-consuming checks for outstanding warrants. Neither factor plays an important role during the flow period because of the way that the sample is selected.

Cook County (Chicago) is unique to ADAM II sampling because ADAM II staff can only interview during more narrowly specified hours, precluding the use of an eight-hour flow period. In Chicago, the data collection window is 4-8 PM, the only time interviewers are allowed in the active booking area. Chicago is a flow only sample; that is, arrestees are brought in on transport buses in waves from over 100 precincts, and the sample is generated from paperwork arriving with each offender in the same manner as used with flow samples elsewhere. There is no access to those outside of the booking area, though cases are weighted using census data to represent those who were booked over the other 20 hour periods each day. By placing more interviewers in this high volume site during those hours, an adequate sample is developed. Eighty percent of the county's bookings are done at this jail.

The Illustration Box below shows site-specific modifications.

ILLUSTRATION OF HOW A SAMPLING PLAN IS DESIGNED

Based on county X's arrest data, a target of 350 cases each data collection would provide an adequate sample. In county X, all arrestees are booked into a single facility. Booking data for a seven-day period is reviewed to identify, the number of arrestees booked into the jail each hour on each day. This information is used to identify during what 8-hour period the highest proportion of arrestees are booked, and the proportion of arrestees booked during that 8-hour period versus the remaining 16-hour period. For this illustration, assume that 60% of arrestees are booked between 4:00 P.M. and 11:59 P.M., and the remaining 40% of arrestees are booked between 12:00 A.M. and 3:59 P.M. Assume also that the flow of arrestees is sufficient to produce the desired number of cases in a 21-day data collection period. This information provides the foundation for the development of the site's sampling plan.

The 350 cases would be distributed evenly across a 21-day period, resulting in a target of 17completed cases a day. The target of 17 would be divided between stock and flow, based on the percentage of bookings occurring during those time periods. In this case, interviewers would be looking to get seven interviews from stock and ten from the flow each day of data collection.

Sample Size Determination

Over 3,225 adult male booked arrestees will be sampled across all sites, an average of 646 cases sampled across the 3-week period per site. Not all sampled cases are available to be interviewed, however, as a number of sampled arrestees are physically unavailable, having been transferred to another facility, ill and in the medical unit, or isolated due to violent behavior (see above for

---

explanation of inclusion criteria). Of sampled cases in 2012, there were 2,107 adult male arrestees available for interview across all sites, with an average of 421 per site in 2012.

The comparisons made in ADAM II are only within a site and within each site over time. We do not make any statistical comparisons across sites. Therefore, the sample size of interest is one that provides adequate power for trend analyses (year to year and linear trends over the multiple years of data) *within a site*. From 2007-2011 ADAM II collected data for two 14 day periods in each site. In 2012, ADAM II collected data in one 21 day period, resulting in a lower sample size in each site, but one still sufficient to support our analyses. As noted in the OMB submission offered earlier this year, an estimated 350 cases per site per year given to the sites as a target is more than adequate for our analyses, and we describe the calculations that lead to that target below. Interviewed samples for 2012 ranged from 364 in Denver to 410 in Sacramento.

We present data on two issues regarding statistical power: 1) power calculations related to minimum sample size numbers for an individual site to detect a change in a year to year trend, and 2) power calculations for looking at the significance of the trends in drug use over the entire ADAM II time period,

Testing for year-to-year changes can be relatively uncertain not only because sample sizes are small, but also because short-term idiosyncrasies existing at the time of data collection can affect estimates. Therefore, we prefer to rely on trends estimated over a longer period. Nevertheless, Figure 1 reports power calculations for detecting a change in drug use for six drugs (or combinations of drugs) given the current ADAM II sampling rate. These estimates are based on the average sample accumulation over two years (2011 and 2012) of a total of 487 arrestees per site or roughly 244 arrestees per site per year. Our power analysis has conservatively powered detectable effect sizes. The power curves we present depend upon an overall sample across the two years of 447 observations. The sampling protocol for ADAM II sets targets above this (i.e., 350 cases per year, for a total of 700 across two years), and historically, the ADAM survey has exceeded its sample targets.

The yearly ADAM samples are well within this target (running from 364 to 410 last year). As noted in the OMB package submitted earlier this year (Part A, Part B and the 60 day notice) an estimated 350 cases per site is more than adequate power for these analyses.

Figure 1 -- Power of Detecting a Change in Positive Drug Tests for Six Drugs given ADAM's Current Sample Sizes
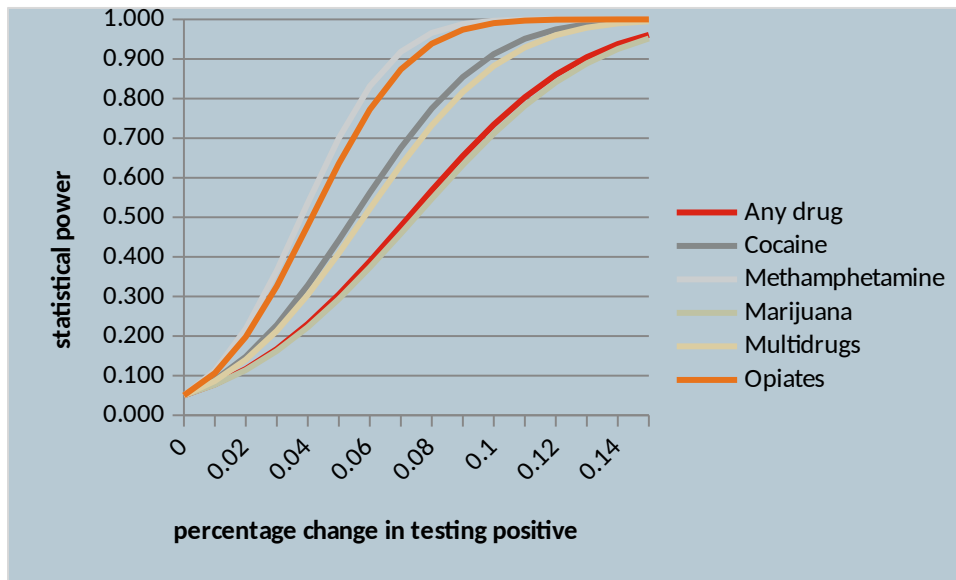
Testing for trends over time. Table 2 below provides power calculations for looking at the significance of the trends in drug use over the ADAM II time period, that is, the ability to reject the null hypothesis that there is no change in the proportion of arrestees testing positive for specified drugs in each of the five ADAM sites that reported urine test results between 2007 and 2012. We have used the average sample size across these years in each site in these calculations. The table lists six types of drugs, including the categories "any drug" and the category "multiple drugs". The second column shows the average rate of testing positive between 2007 and 2012 where each year and each ADAM site receives the same weight. The rest of the table reports the power of detecting changes of 0.00, 0.01, 0.025, 0.05, 0.075 and 0.10.

Table 1   Power Calculations for Linear Trends is Positive Tests of Drug Use across Five ADAM Sites between 2007 and 2012

		Linear Change in Testing Positive between 2007 and 2012 (±)					
1	2	3	4	5	6	7	8
Drug Type	rate of testing positive	0.00	0.01	0.025	0.05	0.075	0.10
Any drug	69.4%	0.050	0.147	0.439	0.910	0.998	1.000
Cocaine	25.8%	0.050	0.164	0.509	0.955	1.000	1.000
Marijuana	46.6%	0.050	0.141	0.412	0.885	0.996	1.000
Opiates	8.4%	0.050	0.285	0.853	1.000	1.000	1.000
Methamphetamine	8.2%	0.050	0.480	0.990	1.000	1.000	1.000
Multiple drugs	22.2%	0.050	0.182	0.577	0.979	1.000	1.000

As indicated by the third column, all power calculations were based on type-1 error of 0.05. Using 0.80 as the criterion for "sufficient power", ADAM II lacks sufficient power to detect a very small linear change in illegal drug use of 0.01 regardless of the drug (Column 4). This means, for example, that ADAM II could not reliably detect a one percentage point change in cocaine use that went from 25.8% to 24.8 %. But ADAM II samples have acceptable power to detect a linear trend of 0.025 (Column 5) or a 2.5 percentage point change for drugs that are used relatively infrequently (opiates and methamphetamine). However, at .05 (Column 6), current ADAM sample sizes provide more than sufficient power to detect a 5 percentage point change trend over time for each of the six categories of illegal drugs.

B.1.2 Estimating Trends for 2008 and Beyond

From 2000 through 2003, ADAM used post-sampling stratification methods to estimate sampling probabilities and to calculate weights. Data were stratified by jail, stock and flow, and day of the week. Within each stratum, the sampling probability was estimated as the number sampled per number booked. Although conceptually simple, the approach was operationally difficult. The principal difficulty was that strata sometimes had no or few members of the sample. This meant that strata had to be merged, and it often resulted in heterogeneous strata being combined.

To avoid these complications, ADAM II adopted propensity scores as an alternative device for estimating sampling probabilities and computing weights. The propensity score approach does not require stratification, because the sampling probability can be modeled as a continuous function of factors that affect the sampling rate (time of day, day of the week, charge). Because 2000 and 2001 ADAM data provided the necessary census data, the survey team replaced the original weights for the 2000 and 2001 ADAM data with new weights based on propensity scores. That is, the survey team replicated the ADAM II weighting procedure using the 2000 and 2001 ADAM data.

This replication was not possible for the 2002 and 2003 ADAM data because the national ADAM contractor during that period did not retain the census data for those years. Thus, for purposes of reporting trend statistics, the ADAM II survey team:

- Uses the reweighted ADAM data for 2000 and 2001;
- Uses the ADAM data for 2002 and 2003 without changing the weights; and
- Uses the propensity score weights for the ADAM II data.

It is important to note that there was nothing wrong with the original ADAM weights. They simply led to sampling variances that were larger than necessary, so the ADAM II study team improved the weights when possible. Because there was nothing wrong with the original sampling weights, there is nothing misleading about mixing the reweighted data for 2000-2001, the 2002-2003 data with their original weights, and the new ADAM II data in producing trend estimates.

However, reweighting has two consequences. The first is that the 2000- 2001 estimates changed slightly from those reported earlier. The second is that estimates from year-to-year in reweighted years are no longer independent. Consequently, to test for trends, an analyst requires an estimate of the parameter covariance matrix.

As anticipated, this has the result of potentially slightly changing the prior years' estimates that appeared first in the 2007 report. Although this approach improves the efficiency of the estimates, there is concern that yearly revisions going forward, regardless how slight, would be confusing. Consequently, 2008-2012 estimates are developed holding earlier estimates at their previously reported levels. In theory, we could update the propensity weights at each data collection. In practice, we freeze the weights for each year at the end of the year, and use those weights in any analyses that include older data. We do this for three reasons. One, while the propensity weights are made more precise by adding new data, the weights themselves do not change appreciably. Two, the ADAM data need to be archived at the end of the year, and we do not want to re-archive data if possible. Three, ONDCP requires that past estimates may be re-created, which may only be achieved by freezing the weights at the end of each year.

The ADAM II 2012 technical documentation report details the estimation procedures for 2008-2012. That report is attached to this request; previous year's reports are available at the University of Michigan's Inter-university Consortium for Political and Social Research:

(http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/27221?q=ADAM&permit%5B0%5D=AVAILABLE)

Table 2 reports the standard errors for the reweighted 2000-2003 trend estimates and the standard errors achieved for 2007-2012 ADAM II data collection. The trends were estimated as percentage changes. Standard errors for linear trends from 2007-2012 are available in the Annual Reports published on the ONDCP website.[4]

Table 2		2000-2012 Standard errors for urine test trends

---

[4] http://www.whitehouse.gov/sites/default/files/ondcp/policy-and-research/adam2011.pdf

	Site	2000	2001	2002	2003	2007	2008	2009	2010	2011	2012
Any drug positive	Atlanta			0.04	0.04	0.04	0.05	0.05	0.06	0.05	0.06
	Chicago	0.04	0.05	0.01	0.01	0.03	0.03	0.04	0.04	0.04	0.03
	Denver	0.02	0.02	0.02	0.02	0.03	0.03	0.02	0.03	0.03	0.03
	New York	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03
	Sacramento	0.02	0.02	0.02	0.02	0.03	0.02	0.03	0.03	0.03	0.03
Cocaine	Atlanta			0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	Chicago	0.09	0.08	0.02	0.02	0.04	0.04	0.05	0.04	0.03	0.04
	Denver	0.02	0.02	0.02	0.03	0.03	0.03	0.02	0.02	0.03	0.03
	New York	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03
	Sacramento	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
Methamphetamine	Atlanta			0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.00
	Chicago	0.00	0.02	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01
	Denver	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03
	New York	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sacramento	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04
Marijuana	Atlanta			0.04	0.04	0.04	0.04	0.05	0.05	0.05	0.06
	Chicago	0.08	0.08	0.02	0.02	0.04	0.04	0.05	0.05	0.04	0.05
	Denver	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03
	New York	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04
	Sacramento	0.03	0.03	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04
Multiple drugs positive	Atlanta			0.04	0.03	0.03	0.03	0.03	0.03	0.04	0.03
	Chicago	0.08	0.07	0.02	0.02	0.04	0.04	0.05	0.04	0.04	0.04
	Denver	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03
	New York	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03
	Sacramento	0.03	0.02	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.04
Opiates	Atlanta			0.02	0.01	0.01	0.01	0.02	0.04	0.04	0.03
	Chicago	0.09	0.07	0.02	0.02	0.03	0.04	0.04	0.03	0.03	0.04
	Denver	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.02
	New York	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02
	Sacramento	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.02	0.02

B.1.3 Annualizing Point Prevalence Estimates

Most of the statistics appearing in the ADAM II reports are point prevalence estimates. A point prevalence estimate is straightforward, because it only requires weighting the desired variable by the propensity score weights. The statistics reported in the 2007-2012 ADAM II reports use this estimator.

In preparation for the 2007 ADAM II report, the team determined that the prevalence estimates should be annualized to account for the fact that the ADAM sample was collected at different times during the year (three or four quarters versus two quarters in ADAM II or one quarter in 2012 and beyond). This complicates the estimation explained in the previous subsection. The statistical procedure for producing annualized estimates is detailed in the attached technical documentation report.

B.2   Information Collection Procedures

Data collection protocols are described in detail in the *ADAM II 2012 Annual Report* available through ONDCP's website. The protocols are briefly summarized here to provide some context for the discussion of weighting and estimation methodologies.

B.2.1   Selecting Study Subjects

Interviewers work in teams in each jail. The sample manager samples from the stock and flow. Sampling from the stock requires a list of all individuals who were booked since the interviewer's last work period. Not all arrestees are still in the facility, but the sample manager does not know that. He or she seeks the sampled arrestee, and, if that arrestee is unavailable or unwilling to be interviewed, the sample manager records the reason and seeks a replacement. Sampling from the flow requires a list of individuals as they are booked into the jail. The sample manager continuously compiles a list of incoming arrestees and seeks the most recently booked arrestee. If that arrestee is unavailable or unwilling to be interviewed, the sample manager records the reason and seeks the closest temporal replacement.

When any arrestee is sampled (regardless of their availability), the sample manager completes a facesheet. The facesheet contains sufficient identifying information that the arrestee can be matched with census data (that is, a census or records representing all bookings into the jail in each of the fourteen data collection days) that are collected long after sampling. The role of the census data is described in Section B.2.2. The sample managers use the facesheet to record that an interview occurred, and if it did not, the reasons why it did not. Analysts use the facesheet to compute response rates. Bar-coded labels are attached to the facesheet, the interview form and the urine specimen bottle, tying all data together. All arrestees sampled have a facesheet, but not all have the other components of the collection (interview, urine specimen).

Eligible arrestees who consent to an interview are administered the ADAM series questionnaire. The request for a urine sample is made at the beginning of the interview and repeated at its completion. If the arrestee consents, he is given a specimen bottle which he takes to a nearby lavatory to produce a sample. The bottle is returned to the interviewer, bagged and sent at the end of the shift to a national laboratory for urinalysis. In most sites over 85% of arrestees consent to provide a urine specimen. The urine specimen is linked to the facesheet and the interview through common bar-coded labels.

B.2.2   The Role of Census Data

Developing propensity scores for case weighting requires complete data on all bookings (a census) that occurred in each ADAM II facility during the 21-day period of data collection. These data are provided by each law enforcement agency participating in ADAM II after their data collection is completed. Site law enforcement partners submit census data in a variety of forms: electronic files listing each case, PDF, or other text files of cases and paper format listing all cases. The Abt Data Center staff transforms each into site and facility specific data sets containing the following data elements for each arrestee:

- Date of Birth and or Age
- ID (computer generated number)
- Charges

- Time of arrest

- Time of booking

- Day of arrest

- Race

Whether the census data are transmitted electronically, as a PDF file, or a paper file, the data are transformed into a SAS dataset. The census data become the sampling frame. As noted, ADAM II interviewers complete a facesheet that includes the above variables for every arrestee sampled for the study, records whether the arrestee answered the interview and whether he provided a urine specimen.

Figure 1 represents the steps included in the manipulation of the raw census data done in preparation for matching with the ADAM II facesheet data. The raw census data received from booking facilities are cleaned to correct invalid data and reformatted for compatibility with the other data components. The census data typically have one row of data per charge and must be converted to single records identifying arrestees with multiple charges. First, arrestees are excluded in the census data who are ineligible for the ADAM II survey: juveniles, women and people booked on days other than those when ADAM II surveys were conducted. Second, charges recorded in the census data are converted into a set of standardized ADAM II charges. Additionally, the top severity, top charge and top charge category (violent, property, drug, other) are determined for each individual.

Figure 1.　　　First Step in Matching Process
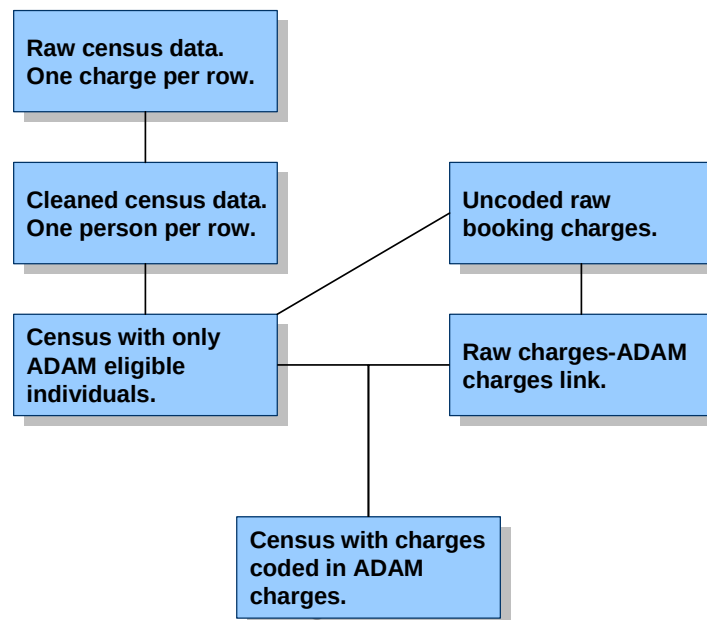
Figure 2 shows the process of matching the census records to the ADAM II facesheet records. The variables common to both the facesheet and the census data that are used to match the records are: booking date/booking time, date of birth, arrest date/arrest time, charges and race. Potential matches are outputted if records match on any single key variables; they are then ranked into tiers based on the goodness of the fit. For example, a facesheet record that matches a census record on just booking

date/booking time and charges will be superseded in rank by a facesheet-census match that links on booking date/booking time, charges and date of birth. Out of all the potential matches the best census match is selected for each facesheet. If, in fact, multiple census records match the same facesheet, and these duplicate matches have equivalent rankings, booking date/time is used as a tiebreaker. The output dataset from this process is a one-to-one match between each facesheet record and census records.

Rarely, a facesheet fails to match any booking record. When this happens, a pseudo-booking sheet is created and inserted into the booking data. This process is represented by the right-hand flow in Figure 2.

Figure 2.   Matching Census with Facesheet Data
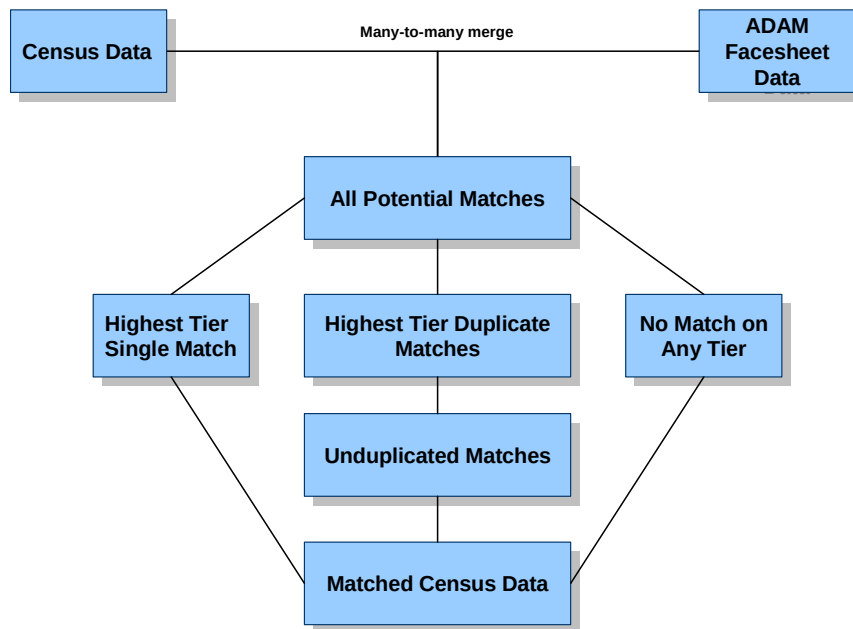
Figure 3 demonstrates the last step in the construction of the analysis file for each site and each data collection quarter. The linked census-facesheet data are merged with the appropriate urinalysis and survey record using unique identification numbers recorded in barcoded labels on the facesheet, interview and urine specimen. The result is the final analysis dataset for each quarter for each particular ADAM II site.

Figure 3.   Creation of Final Analysis File
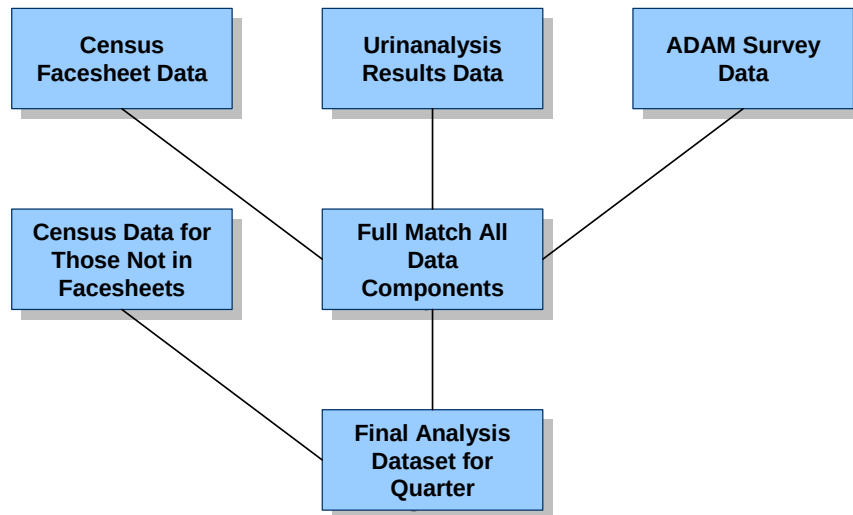
B.3   Methods to Maximize Response Rates

In ADAM II we report two response rates (Table 3): the *overall response rate* (identical to the OMB designated "unit response rate" or RRU) that indicates the number of eligible booked arrestees who were interviewed, relative to all of those eligible booked arrestees[5] sampled, (regardless of whether they were physically available); and the *conditional response rate* that indicates the number of booked arrestees who were interviewed, relative to the total number of eligible arrestees sampled (and available for interview; that is, physically in the facility.) For 2012, the overall response rate for interviews for the program was 60%, ranging from 38% in New York to 92% in Chicago, and the conditional response rate was 92%, ranging from 84% in Denver to 96% in Sacramento. A key to conducting a successful study is ensuring that response rates remain at acceptable levels throughout the study. Abt Associates' rigorous procedures for interviewer selection, training, and quality control have resulted in steadily improving response rates for ADAM II sites since 2006.

The overall response rate in ADAM II for 2012 (60%) reflects the difficulty that the interview setting presents.   Because we are surveying booked arrestees within 48 hours of arrest, we have to station ourselves in the active booking areas of jails. Consequently, offenders are continuously being brought in, processed, moved to court or housing, or released.

Prior to discussing the actual response rates, it is important to remember that the most critical part of the ADAM II sampling and weighting strategy is to provide the basis for making inferences about booked arrestees given the idiosyncrasies imposed on ADAM II sample due to the setting (booking facilities). The sampling strategy balances the sample, and the propensity score weights control for things correlated to testing positive for drugs, such as day and time of booking and severity of

---

[5] Again, to be eligible for interview, in each county an arrestee must be: male, over 18, arrested no longer than 48 hours prior to the interview, coherent enough to answer questions and not in the jail as a hold for a federal agency.

offense. This sampling and weighting strategy, rather than simply pure response rates, justifies the ADAM II sample as a valid indicator of the adult male booked population.

The *overall response rate* is computed as the number of arrestees completing interviews divided by the sum of the number of arrestees completing interviews and the number of sampled eligible arrestees not completing interviews. We partition the eligible arrestees not completing interviews into two subgroups: arrestees not available for interview (e.g. taken to court) and arrestees available for interview but refusing or unable to take the interview (e.g. a language barrier) or who agree to the interview but do not complete it. For any ADAM II site *i*, this may be written as:

$$\textbf{ResponseRate}_i = \frac{\textbf{Resp}_i}{\textbf{Resp}_i + \textbf{EligUnavailable}_i + AvailableNon\,\textbf{Re}sp_i} \tag{B.2}$$

Where

ResponseRate   The response rate to the interview

Resp   The number of *eligible and availabl*e arrestees responding to the interview

EligUnavailable   The number of eligible but unavailable arrestees

AvailableNonResp   The number of *eligible and available* arrestees not completing an interview

The *conditional response rate* is nested within the overall response rate, and is written as the number of arrestees completing interviews divided by the sum of the number of arrestees completing interviews and the number of sampled eligible and available arrestees not completing interviews. For any ADAM II site *i*, this may be written as:

$$\textbf{CondResponseRate}_i = \frac{\textbf{Resp}_i}{\textbf{Resp}_i + AvailableNon\,\textbf{Re}sp_i} \tag{B.3}$$

Overall response rates for the interview may be computed according to Equation (B.2), and conditional response rates may be computed according to Equation (B.3). For each ADAM II site, Table 2 reports the number of arrestees eligible to be interviewed, eligible and available for the interview, completing the interview, and providing a urine specimen in 2012.

In 2012, amongst those eligible and available to be interviewed, ADAM II sites had an overall conditional response rate of 92% for the questionnaire. Amongst those completing interviews, ADAM II sites had a 90% agreement rate for providing the urine specimen. The study team anticipates an even higher conditional response rate in 2013.

Table 3   Sample Sizes and Response Rates for Interview and Urine Specimen

	Atlanta	Chicago	Denver	New York	Sacramento	Overall

Sample Sizes						
Provided Urine Specimen	323	374	324	351	364	1,736
Completed Interviews	367	395	364	402	410	1,938
Eligible and Available to be Interviewed	395	418	432	433	429	2,107
Eligible to be Interviewed	528	430	597	1,056	618	3,229
Interview Response Rates						
Conditional Response Rate	0.929	0.945	0.843	0.928	0.956	0.920
Overall Response Rate	0.695	0.919	0.610	0.381	0.663	0.600
Urine Response Rates						
Urine Agreement Rate	0.880	0.947	0.890	0.873	0.888	0.896
Conditional Response Rate	0.818	0.895	0.750	0.811	0.848	0.824
Overall Response Rate	0.612	0.870	0.543	0.332	0.589	0.538

Table 3 shows the final disposition of the sample, based on facesheet information.

Table 4   Final Disposition of Completed Facesheets

	Atlanta	Chicago	Denver	New York	Sacramento	Overall
Ineligible for the Interview						
Arrested More than 48 Hours Ago	0	0	0	0	0	0
Eligible but Unavailable for the Interview						
Taken to Court	0	0	1	118	6	125
Released	88	1	121	0	142	352
Transferred	1	0	5	499	1	506
Medical Unit	6	1	4	0	7	18
Violent or Uncontrolled Behavior	23	2	20	0	21	66
Physically Ill	0	8	4	3	11	26
Shift Ended	4	0	0	0	0	4
Other/Missing	11	0	10	3	1	25
Eligible and Available for the Interview						
Did Not Want to Answer Interview	25	20	65	23	19	152
Could Not Answer Interview Due to Language Barrier	1	1	0	3	0	5

Other/Missing	1	0	0	2	0	3
Agreed, Did not Complete Interview	1	2	3	3	0	9
Completed Interview						
No Urine Sample	44	21	40	51	46	202
Provided Urine Sample	323	374	324	351	364	1,736

Table 5 indicates the factors that can affect the response rates to the survey.

Table 5   Characteristics of Response to the Survey

	Atlanta	Chicago	Denver	New York	Sacramento
Day of Week					
Monday	78%	92%	55%	29%	65%
Tuesday	67%	87%	62%	46%	69%
Wednesday	65%	92%	62%	48%	66%
Thursday	70%	93%	57%	42%	71%
Friday	78%	92%	63%	38%	73%
Saturday	69%	96%	65%	31%	70%
Sunday	61%	91%	65%	37%	55%
Total N (non-missing)	527	430	597	1056	618
Chi-Square	8.47	3.60	2.95	18.10	9.85
p-value	0.206	0.730	0.815	0.006	0.131
Booking Time					
12:00am-8:59am	50%	67%	63%	25%	50%
9:00am-3:59pm	86%	90%	67%	25%	70%
4:00pm-11:59pm	80%	100%	56%	80%	85%
Total N (non-missing)	525	132	596	1054	618
Chi-Square	64.09	4.35	4.93	244.34	65.24
p-value	<0.001	0.114	0.085	<0.001	<0.001
Sample Type					
Stock	63%	n/a	58%	23%	56%
Flow	82%	92%	69%	82%	85%
Total N (non-missing)	527	430	597	1054	618
Chi-Square	19.72	n/a	5.47	289.22	52.36
p-value	<0.001	n/a	0.019	<0.001	<0.001
Age					
18-23	69%	93%	60%	41%	69%
24-29	70%	96%	65%	39%	60%
30-35	70%	90%	63%	39%	59%
36-44	69%	88%	59%	31%	69%
45+	70%	88%	59%	39%	74%
Total N (non-missing)	524	429	596	1054	612
Chi-Square	0.08	5.30	1.32	5.38	8.51
p-value	0.999	0.258	0.859	0.250	0.075
Race					
Black	71%	91%	60%	37%	74%
Hispanic	100%	93%	67%	46%	66%
White	57%	95%	56%	28%	61%
Other	50%	0%	50%	34%	65%
Total N (non-missing)	528	430	597	1056	618
Chi-Square	10.66	0.97	5.00	13.75	7.96
p-value	0.014	0.615	0.172	0.003	0.047
Top Severity					

	Atlanta	Chicago	Denver	New York	Sacramento
Felony	82%	94%	71%	41%	80%
Misdemeanor	70%	90%	59%	37%	39%
Other	58%	95%	53%	40%	0%
Total N (non-missing)	528	430	597	1056	618
Chi-Square	15.24	2.70	12.15	1.90	102.52
p-value	<0.001	0.259	0.002	0.388	<0.001
Top Charge Type					
Violent	82%	93%	68%	36%	78%
Drug	60%	90%	64%	33%	51%
Property	76%	90%	63%	44%	72%
Other	67%	95%	56%	35%	73%
Total N (non-missing)	519	419	593	1028	606
Chi-Square	15.40	2.08	6.11	9.60	37.26
p-value	0.002	0.555	0.106	0.022	<0.001

Low response rates are often linked to worries about low data quality, but it is the bias that non-response can introduce —which is independent of response rate—that is the concern for considerations about the quality of the data. ADAM II analysts are also concerned with the issue of non-response as it relates to bias. This is why we take critical diagnostic steps to determine bias due to non-response and incorporate the information gained from those steps into our weighting scheme.

We have a distinct advantage in dealing with non-response bias in ADAM II in that we can utilize two rich sources of data on persons who are sampled and not available, those sampled and refuse and those in the universe of all arrested and booked during the data collection periods: 1) the official booking sheets on all arrestees and, 2) each sites' data on all persons arrested in the time period in which we are collecting data (the census data). Data from the booking sheets on each sampled arrestee are recorded on the study facesheet and include arrest date and time, arresting agency and arrest location, arrestee birthdate, race/ethnicity, three most serious charges, and booking date and time. Interviewers also record on this sheet the status of the interview (agreed, declined) and the reason for non-response (did not want to, taken to court, released, transferred, in medical unit, violent or uncontrolled behavior, physically ill, language difficulty, other). The census data (an electronic record of all males booked on the data collection days with the same data as found on the booking sheet) includes all but the reason for non-response data cited above, allowing a direct comparison of observables on who , once sampled, was interviewed.

Using these data sources, we take three steps in adjusting for potential bias in the sample due to non-response:

- Diagnostic analyses to determine the source of non-response, using data on all arrested (census data) and all sampled (booking data and reason for non-response recorded on facesheets)

- Based on diagnostic analysis results, adjusting bias through weighting with propensity scores

- Adjusting bias related to non-response to urine test (missing urine test data)through data imputation

Diagnostic analysis to determine the source of non-response.

For each year and each site in the ADAM II sample, we look for statistically-significant correlates of non-response with characteristics we are able to observe on all sampled cases and available in what we term *the census data*. The census data are data available on all arrests occurring in each site on all days on which data collection occurred for ADAM II. The comparison of ADAM II facesheet data and the census data allow a direct examination of the representativeness of the final ADAM II interview cases. We use these data to create the variables linked to the probability of selection (booking day and time,[6] the booking charge and severity, the age of respondent, sample type [stock or flow] and the race/ethnicity of the respondent), the stratifying variables that are the basis of the propensity score weights.

Not every arrestee sampled answers a survey, but facesheet information is collected on all sampled cases and includes the reasons arrestees do not respond to the interview. As Table 3 shows, for 2012 the overwhelming numbers of arrestees who are eligible, but were not interviewed, were those not physically available to be interviewed. Most frequently, this was due to early release, taken to court or transferred out (30% of those sampled). In Chicago in 2012, there were very few unavailable arrestees and most of these were physically ill. The examination of differences between those interviewed and those not interviewed (both through refusal and non-availability) form the basis of the ADAM II propensity scores (discussed below).

 For eligible arrestees *physically in the facility at the time of interview*, in every site the most frequent reason for non-response is the arrestee not wanting to participate. Of the 3229 arrestees eligible and sampled, 2107 arrestees were available to be interviewed; of those available, 7% (152) did not want to participate. Because we use Spanish speaking interviewers in all sites, there were few refusals due to language difficulties – only 5 across all 5 sites.

We examine each year for each site potential sources of bias due to differences in response rates among subpopulations of the eligible arrestees. For each of the stratifying variables described above, Tables 5 reports the number of facesheets with non-missing values for the set of stratifying variables, the percentage of arrestees among the subpopulations with facesheets that agree to the interview, and a Chi-square test of significance that assesses whether the response percentages are statistically different across the subpopulations. We consider a difference statistically significant if its p-value is less than or equal to 0.05.

For example, for eligible arrestees in Atlanta, New York, and Sacramento, the *time when an arrestee is booked* differentiates respondents from non-respondents. In all three sites arrestees booked earlier in the day are interviewed at a lower rate, as the lowest rate is always from 12:00 AM – 8:59 AM. For Atlanta, the highest rate is in the middle of the day (9:00 AM – 3:59 PM), while in New York and Sacramento the highest interview response rates are late in the day (4:00 PM – 11:59 PM). For all four sites where there is both a stock and flow sample (Chicago is a flow-only sample), the highest response rates come from those respondents entering during the flow period.

---

[6]   The probability of selection is lower on busier days and at busier times of the day.

Race/ethnicity is a factor that differentiates responders from non-responders in Atlanta, New York, and Sacramento. In Atlanta and New York, Hispanics are more likely to be interviewed, while in Sacramento, blacks are more likely to be respondents.

The *severity of the most serious charge* at the time of arrest differentiates responders from non-responders in Atlanta, Denver, and Sacramento. In all three sites, those with felony charges were more likely to be interviewed.

Responders and non-responders differ in terms of the *type of arrest* for the most serious charge in Atlanta, New York, and Sacramento. In all three sites, those with drug charges were less likely to be respondents than those with other charges.

We conduct the same analysis to look at response bias in providing a urine sample (Table 6). There are no statistically significant associations between facesheet variables and the agreement to provide a urine sample. Therefore, we can conclude that based on observable characteristics there is no bias related to willingness to provide a urine specimen for testing.

Table 6   Characteristics of Non-Response to the Urine Test

	Atlanta	Chicago	Denver	New York	Sacramento
Day of Week					
Monday	93%	91%	98%	85%	95%
Tuesday	88%	89%	87%	89%	81%
Wednesday	94%	100%	87%	81%	90%
Thursday	84%	95%	90%	90%	87%
Friday	89%	93%	86%	94%	96%
Saturday	84%	98%	91%	85%	86%
Sunday	84%	97%	84%	86%	85%
Total N (non-missing)	366	395	364	402	410
Chi-Square	5.44	9.76	6.50	5.38	10.46
p-value	0.489	0.135	0.369	0.496	0.107
Booking Time					
12:00am-8:59am	91%	75%	90%	89%	92%
9:00am-3:59pm	89%	94%	86%	86%	90%
4:00pm-11:59pm	85%	89%	90%	88%	86%
Total N (non-missing)	364	118	363	401	410
Chi-Square	2.49	2.56	0.96	0.39	3.16
p-value	0.288	0.278	0.619	0.824	0.206
Sample Type					
Stock	90%	n/a	90%	86%	91%
Flow	85%	95%	87%	88%	86%
Total N (non-missing)	366	395	364	401	410
Chi-Square	1.90	n/a	0.51	0.50	3.22
p-value	0.168	n/a	0.477	0.481	0.073
Age					
18-23	91%	93%	87%	84%	94%
24-29	83%	97%	92%	86%	85%
30-35	84%	92%	90%	90%	91%
36-44	88%	93%	85%	89%	86%
45+	90%	98%	91%	91%	88%
Total N (non-missing)	366	394	363	400	406
Chi-Square	3.19	4.21	2.36	3.04	5.45
p-value	0.526	0.379	0.669	0.551	0.244
Race/ethnicity					
Black	87%	95%	89%	90%	88%
Hispanic	100%	98%	89%	87%	88%
White	95%	93%	89%	75%	89%
Other	100%	0%	100%	91%	94%
Total N (non-missing)	367	395	364	402	410
Chi-Square	3.72	1.01	0.41	6.19	1.07
p-value	0.294	0.604	0.937	0.103	0.783
Top Severity					
Felony	91%	95%	87%	87%	89%
Misdemeanor	87%	95%	92%	87%	86%
Other	88%	94%	87%	88%	0%
Total N (non-missing)	367	395	364	402	410
Chi-Square	0.64	0.04	2.18	0.03	0.72
p-value	0.726	0.981	0.336	0.985	0.397
Top Charge Type					
Violent	89%	93%	91%	90%	87%
Drug	83%	92%	90%	87%	90%
Property	92%	100%	84%	83%	87%
Other	88%	97%	89%	89%	90%
Total N (non-missing)	363	387	361	384	402
Chi-Square	3.19	6.24	1.92	2.20	0.84
p-value	0.363	0.101	0.589	0.533	0.839

Based on diagnostic analyses, adjusting bias through weighting with propensity scores.

As is clear from the discussion above, there is sampling rate variance that occurs during data collection based primarily on releases during the non-interview hours. Originally, ADAM assigned weights by assigning all arrestees to strata based on offenses and the time they were booked. This approach was not altogether satisfactory because samples were often small or even missing within a stratum, so that strata had to be merged. Merging required considerable manual manipulation of the data, and too frequently disparate strata were merged. As a result ADAM II statisticians turned to another method that has been developed to adjust for sampling bias, the use of propensity scores.

Since 2007, ADAM II has developed **propensity scores** to weight the data to address the sampling bias inherent in working in this dynamic setting. A propensity score is the estimated probability that a member of the population of arrestees is included in the sample and ultimately interviewed. The use of propensity scores dates to work such as Rosenbaum and Rosenbaum and Rubin[7]. Rotnitzky and Robins [8], among others, proposed using "inverse probability weighting" as a solution for missing data problems, of which non-response is one. Wooldridge[9] proposed a generalized two-step estimation method, which produces consistent and asymptotically normal estimates. This method estimates propensity scores (i.e., probabilities of being sampled) in the first step, and uses inverses of the estimated propensity scores as weights when estimating the parameters of interest in the second step.

Using this method, we calculate the propensity score for each arrestee using a logistic regression, where the explanatory variables are those things we can observe that impact the probability of an arrestee being sampled and interviewed: time of day of the arrest, type and severity of the offense, day of the week, age, and sample type (stock or flow). The inverse of the propensity score is the ADAM II case weight.

 Again, we know these sources of bias from year to year and site by site through examination of the sample data and the census data.

Adjusting bias related to non-response to urine test (missing urine test data) through data imputation

Another possible source of bias in the ADAM II samples comes from those interviewed who refuse to (or cannot) produce a urine specimen for testing. Approximately 11% of those interviewed fail to provide a urine specimen for testing. A common way of dealing with missing data is to discard data that are missing and only work with data that are not missing. However, this approach for ADAM II introduces a possible bias because those arrestees who fail to provide a urine specimen may differ

---

[7] Rosenbaum, P. (1984) "From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment," *Journal of the American Statistical Association* Vol 79 (385): 41-48; Rosenbaum, P. and Rubin, D (1984) "Reducing bias in observational studies using sub classification on the propensity score," *Journal of the American Statistical Association*, Vol 79 (387): 516-524.

[8] Rotnitzky, A. and Robins, J. (1995) "Semi parametric regression estimation in the presence of dependent censoring," *Biometrica* Vol 82 (4): 805-820.

[9] Wooldridge, J (2003) Cluster sample methods in applied econometrics," *The American Economic Review* Vol 93 (2): 133-138.

systematically from those who provide a urine specimen, and the propensity score may fail to control for those differences.

We explored several statistical approaches to dealing with missing data in 2007 with the reinstatement of ADAM data collection and the following method successfully used in other studies[10] to address issues of bias was determined the most robust. There is a high correlation between self-reported drug use during the last three days and the results of the urine test. When an arrestee admits to using a drug during the last three days, it comes as no surprise that his drug test will usually be positive. Not all arrestees tell the truth, however, and many deny recent drug use. Nevertheless, self-reports of recent drug use are highly correlated with urine test results, so self-reports (which are known provided the arrestees answered the survey) can be used to impute urine test results when the latter were missing.

Deriving an imputation is as follows. The probability that a urine test result is positive when an arrestee said that he had used a drug during the last three days can be estimated. In fact, the probability is close to 1. The probability that a urine test result is positive when an arrestee said that he had not used a drug during the last three days can be estimated. In fact, the probability is positive, but much closer to 0. Basically, the approach is to estimate this probability, draw a random sample from a Bernoulli distribution, and thereby assign a value of 1 or 0 to replace the missing value.

Although the basic approach to the imputation is simple in theory, using the imputation when estimating the proportion of arrestees who tested positive for drug use is more complicated. Although a value of 1 or 0 based on the above procedure can be imputed, subsequent statistical analysis would not reflect two forms of sampling error without additional steps. First, the estimates of the probability of testing positive conditional on a self-report of recent drug use is in fact an estimate with its own sampling variance. Second, the random draw from the Bernoulli distribution is only one possible realization of a random process. Estimation takes that additional sampling variation into account. A step-by-step explanation is provided below:

1. The probability of testing positive conditional on admission of use in the last three days does not vary much over time. Conditional on the respondent saying "YES" to the three day use question, the probability of testing positive when the urine test is known is estimated. Call this P1. The same can be done when the respondent says "NO" to the three day use question. Call this P2.

2. P1 and P2 are estimates, but the distribution of the estimates is known—they are asymptotically normal with estimated variances of P1(1-P1)/N1 and P2(1-P2)/N2 respectively, where N1 and N2 are the number of observations with self-reports of "YES" and "NO" that have corresponding urine test results.

3. Estimates of P1 and P2 are drawn from these distributions. Given these estimates, first round imputations for missing urine test results are done. Using the resulting sample (which now includes the imputed responses), P1 and P2 are re-estimated.

---

[10] Rubin, D. (1987) Multiple Imputation for Nonresponse in Surveys. New York, John Wiley & Sons; Schaefer, J. (1997) Analysis of Incomplete Multivariate Data, New York, Chapman and Hall.

4. Assuming that the posterior distribution is normal, a new P1 and P2 are sampled from the posterior distribution.

5. The approach runs the algorithm through a burn period of 1000 iterations; then imputes the missing values. This is repeated until there are 20 simulated data sets.

6. When the urine test results are reported, they do not change from data set to data set. Only the imputed data vary across the 20 simulated data sets.

7. Each of the 20 data sets is used to estimate the weighted probability of testing positive in each of the 20 data sets.

8. Each of these analyses yields parameter estimates and a parameter covariance matrix.

    a. These are used to compute 20 point estimates for the probability of testing positive conditional on the offense and year. These estimates are averaged to produce the grand estimate. This is reported as the estimate.

    b. Twenty variance estimates are computed for each of the 20 point estimates. These are averaged to produce a grand estimate of the variance. Call this V1.

    c. The variance of the 20 point estimates is computed. Call this V2.

    d. The variance estimates used for reporting is V=V1+V2.

The three steps outlined above (diagnostics, propensity score weighting and imputation of missing test data) form the basis of ADAM II's comprehensive approach to addressing the important issue of non-response and the bias that non-response may introduce into results. We are fortunate to have access to official records (booking sheets on all sampled and a complete census of all arrested) that can be used to compare respondents from non-respondents on observable characteristics. Case weighting can then be developed for each arrestee in each site on those observables that impact the probability that he is interviewed. This is accomplished through the use of propensity scoring. Lastly, the imputation of missing urine test data is conducted to address the potential that those who refuse or cannot supply a urine sample for testing are appreciably different from those who are tested.

B.4   Tests of Procedures

The utility of the core ADAM II instrument for analysis and ease of administration has been demonstrated in the over 115,000 interviews collected since 2000. This instrument was developed beginning with de novo question creation in a series of focus groups of drug users and sellers in sites throughout the country. Final sets of questions underwent cognitive testing with heavy drug users. Additional validity testing of specific sections[11] was undertaken during development (the calendar portion and the dependency screener). The entire instrument and sampling protocols underwent beta testing and revisions at two large ADAM sites (New York and San Antonio) in the fall of 1999. The

---

[11]   Hoffman, N., Hunt, D., Rhodes, W. and Riley, J. (2003). UNCOPE: A brief substance abuse screener for use with arrestees, *Journal of Drug Issues*, Winter, 29–44.

modifications (see below) are minimal and designed to capture updated information on prescription drug prices, combat veteran status and eligibility for Veteran's Administration benefits, and better understand the changing legal climate for marijuana purchasers.

ADAM II has the advantage of being able to assess the validity of one of its most important pieces of data, self-report of drug use, through a separate source, the urine test results. While there is variation in the reliability of time frames for particular drugs in urine tests, the rate of truth telling can be approximated. Specifically, ignoring a small rate of false positive tests, arrestees who test positive for a drug have used that drug within the last two or three days (30 days for marijuana). Across 4 drugs and all 5 ADAM II sites, the proportion telling the truth was extremely high in 2012. For marijuana, 82 percent of arrestees were consistent in their response to self-reported use and the results of the testing of their urine specimen. A similar percent of congruence was identified for cocaine (88 percent) and even higher rates for heroin (95 percent) and methamphetamine (96 percent). This proportion is driven, in large part, by the high number of arrestees who do not report any drug use and do not test positive for any drugs.

Protocols for sampling in ADAM II are identical to those used and approved in ADAM. For ADAM II, revisions to the original ADAM instrument and protocols are minimal. As discussed in Section A, due to budgetary constraints, ADAM II will have fewer sentinel sites than past ADAM data collections; ADAM II will study five sites. Other procedural changes are limited to a shorter total number of days for the data collection period (21 days instead of 28 days) and the elimination of one of the collection cycles (one quarter of collection instead of two quarters).The ONDCP ADAM II team has also added two questions designed to correct an omission in the original ADAM instrument —namely regarding community supervision and veteran status. Please refer to Section A for a more detailed discussion.

No additional tests of procedures will be necessary for these changes.

B.5   Individuals Consulted on Statistical Aspects of the Design

William Rhodes, Ph.D.	Dana Hunt, Ph.D.
Fe Caces, Ph.D.	Terry Zobeck, Ph.D.
Ryan Kling, M.A.	Richard Kulka, Ph.D.
Michael Battaglia	