

SUPPORTING STATEMENT – Part B

Drug and Alcohol Warning Network (DAWN)

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Respondent Universe and Sampling Methods

Drug and Alcohol Warning Network (DAWN) is a nationwide public health surveillance system that improves emergency department (ED) monitoring of substance use crises. The proposed new DAWN data collection efforts expand the availability of early warning information, improves timeliness of data, increases frequent intervals of data availability, and covers a wide range of geographic area types.

DAWN eligible hospitals are defined as all non-Federal, short-stay, general surgical and medical hospitals located in the United States that operate at least one 24-hour ED with more than 100 annual ED visits. Hospital participation in DAWN is voluntary. SAMHSA anticipates review of all ED records within a participating hospital.

The DAWN eligible hospital population is determined by the most recently available American Hospital Association (AHA) Annual Survey Database at the time of sample selection. The AHA annual survey file includes both AHA member and non-member hospitals and takes into account changes in the hospital population, such as, the possibilities of openings, closures, mergers and demergers in the hospital population. Table B1 shows the total number of DAWN eligible hospitals in the 2021 AHA Annual Survey Database, as an example DAWN sampling frame.

Table B1: Number of hospitals and DAWN eligible hospitals in the 2021 AHA Annual Survey Database.

Number of hospitals or ED visits	2021 AHA
Total AHA records	6,201
Located within the U.S. (50 states and DC)	6,129
Total DAWN eligible hospitals	4,226

For hospital selection, DAWN uses a stratified systematic random sample combined with a purposive sample in a hybrid sample design. The three-part design allowed SAMHSA to identify sentinel hospitals after hospitals in Parts B and C were selected. Parts A, B, and C can work together to produce nationally representative estimates, or Parts B and C can stand alone to produce nationally representative estimates. The design of the hospital sample is outlined below in three parts:

- Part A – Sentinel hospitals are identified by SAMHSA based on highest priority sentinel areas with high potential substance use related ED visits and severity specific to all drug-related mortality.
- Part B – Probability sampling of hospitals in high priority suburban and rural areas with high potential DAWN case volume and severity specific to all drug-related overdose deaths.

- Part C – Probability sampling of hospitals in the remaining areas not included in Part B.

Types of substances misused often vary by geographic area or region. Hospitals are selected from each Census region and a mix of urbanity. Selection of hospitals also considers areas that have high prevalence of drug use and related morbidity and mortality. Tables B2 and Table B3 present counts of hospitals overall, by region, and by sampling stratum.

Table B2: Number of DAWN eligible hospitals by region

Census Region	2021 AHA
Overall	4,226
Midwest	1,273
Northeast	512
South	1,582
West	859

Table B3: Number of DAWN eligible hospitals by stratum

Stratum	2021 AHA¹
Overall	4,226
Part B-Midwest	17
Part B-Northeast	24
Part B-South	18
Part B-West	24
Part C-Midwest-HH	260
Part C-Midwest-HL	94
Part C-Midwest-LH	443
Part C-Midwest-LL	457
Part C-Northeast-HH	158
Part C-Northeast-HL	58
Part C-Northeast-LH	170
Part C-Northeast-LL	102
Part C-South-HH	464
Part C-South-HL	77
Part C-South-LH	576
Part C-South-LL	443
Part C-West-HH	200
Part C-West-HL	109
Part C-West-LH	278
Part C-West-LL	242

¹ Subgroups do not sum to overall total. Part A hospitals were part of the sampling frame for Part B and C selection and later designated as sentinel (Part A).

Sample Substitution to Account for Non Response

A target of 73 Part A (sentinel hospitals) and a total of 70 combined Part B and Part C hospitals will be recruited for DAWN data collection. In the case of originally sampled hospitals refused

or failed to engage or be responsive in the timeframe allowed for hospital recruitment, substitution would be used as a remedy for nonresponse in those cases.

As noted above, in DAWN, hospitals from the same stratum, and of similar size, characteristics, and predicted DAWN case volume could be identified to take the place of non-respondents, and yet provide real case-level data, which is key to DAWN's objectives. SAMHSA's approach for the treatment of nonresponding hospitals and the identification of their substitutes is as follows:

- Substitutes for nonresponding sampled hospitals will be identified by SAMHSA on an as-needed basis, usually as the hospital on the frame immediately above or below the originally sampled hospital and closest in the number of ED visits, which allows for control on hospital size and all explicit and (most) implicit stratification.
- In some instances, the substitute hospital above or below the nonresponding original would be identified based on community type, or state, with an eye to matching the original hospital in this regard, or control for the distribution of the resulting sample.
- In a few instances, a previously responsive hospital from the previous original sample, when close to the nonresponding hospital in the new sample in the frame on ED visits, would be identified as the substitute.

2. Procedures for the Collection of Information

This section describes the procedures for the collection of information of the three-part sample methodology, including information about the statistical methodology for sample selection, estimation procedure, and degree of accuracy. Data collection will be ongoing. Use of periodic data collection cycle to reduce burden is not suited for what this type of project will achieve.

2.1 Statistical methodology for stratification and sample selection

Part A (sentinel hospitals)

Past studies show a myriad of factors (such as social vulnerability index, racial and ethnic minorities, living situations, language, and geographic disparities) that may have relation with the overdose-related deaths or drug-related ED visits.

The Part A (sentinel hospital) selection strategy will be revisited every year to identify vulnerable geographic areas to recruit hospitals to inform outbreaks; to warrant representations from areas characterized by social vulnerabilities of high-risk populations with alcohol and drug overdose data, geography, and drug legalization; and to ensure that DAWN is considering shifts in drug-related overdoses across the United States.

The design of the sentinel hospitals will implement a robust methodology and utilize publicly available local-level data to ensure the DAWN system can identify counties with high risk of outbreaks and the emergence of new drugs of use, which can be used for epidemiologic investigations, better allocation of resources for prevention, treatment, and recovery for policymakers at national level while assisting public health officials at local level. Data sources include but are not limited to the following and underlying data will continuously be updated to reflect current trends in drug and social vulnerability:

- US Census data;
- Center for Disease Control and Prevention's (CDC) Wide Ranging Online Data for Epidemiological Research (WONDER) Multiple Causes of Death (MCOB) database;
- National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties;
- The CDC's Agency for Toxic Substances and Disease Registry (ATSDR) Social Vulnerability Index (SVI) measures overall vulnerability and vulnerability across four specific themes: (1) socioeconomic status, (2) household composition and disability, (3) racial and ethnic minority status and language, and (4) housing type and transportation. The components of the SVI will be analyzed on an ongoing basis for model refinement of composite score to best reflect current trends in alcohol and drug use; and
- AHA aggregate number of hospitals, aggregate number of buprenorphine providers and aggregate number substance use treatment facilities in each county.

The analysis will be two stepped and compared for best outcomes. The first step of the analysis will be to map the variables across the counties. When original variables are continuous, simple averaging or weighted averaging will be considered. In the first step, simple averaging will be used to creating a composite variable based on computation of a Z-score for each geographic unit. In the second phase, based on the mapping of z-score, the analysis will potentially implement a linear index to compare the results from Z-score analysis. Finally potential spatial statistical test will be conducted to identify global and local clusters of identify spatial patterns of the index (Moran's I or Getis and Ord statistics), followed by the implementation of a regression methodology (spatial and non-spatial) to evaluate the association of overdose-related deaths with the SVI, AHA hospitals, aggregate number of buprenorphine providers and aggregated number of substances use treatment facilities in each county.

Part B

The Part B hospital population is a sample of hospitals in high priority suburban and rural areas. High priority areas are the top 15 suburban and rural counties within each Census region, ranked by drug-induced-death rates for 2012-2016. The sample of hospitals is explicitly stratified by region and implicitly stratified (i.e, sorted) by community type and rank and selected via systematic sampling given that sorted list.

The following data sources would be used to identify the highest priority suburban and rural counties, as an example:

- CDC's WONDER MCOB database;
- NCHS Urban-Rural Classification Scheme for Counties; and
- The AHA annual survey dataset.

Part C

Part C consists of a traditional probability sample of hospitals in the balance (i.e., areas not included in Part B above) of the U.S., stratified and distributed as follows: explicitly stratified by region (4 levels), drug-induced-death counts (high/low), and ED visits (high/low). Hospitals with drug-induced-death counts larger than the median within the hospital's region were defined as high. Hospitals with ED visit counts larger than the median within the hospital's region were defined as high. The sample was implicitly stratified (i.e., sorted by, prior to sample selection,

within explicit strata) by urbanicity, and 16 strata resulted. Hospitals are selected with equal probability per stratum via systematic sampling. Sample selection was independent stratum to stratum, and a reasonable distribution by the implicit stratification variables were the result.

Rationale for Using 5-Year Drug-induced Death Indicator for Parts B and C

A combined 5-year estimate of drug-induced death rates for each county was generated for this analysis for the following reasons:

- DAWN focuses on multiple substances. Drug-induced deaths can be better predictors of DAWN case volume as opposed to looking at individual drug types separately.
- Looking at drug-induced deaths will eliminate double-counting. For example, if both opioids and cocaine played a role in drug overdose deaths, it will be listed in both opioids and cocaine-related overdose categories leading to some overestimation of drug-related deaths.
- Focusing on death rates as opposed to numbers will lead to better representation from suburban and rural counties as opposed to death counts, where most of the top counties will be urban given their population.
- Looking at the 5-year combined drug-induced deaths gives us death data for 65% of counties (2,031 out of 3,147 total counties) as opposed to single year rates where data is suppressed for over 68% counties (estimates for only 1,000 out of 3,147 counties can be generated for 2016 drug induced death rates).
- Looking at the 5-year rate as opposed to the 2016 rate will identify counties where the rates have been consistently higher, and will take into account spikes in death rates especially for smaller/low prevalence counties.

2.2 Estimation Procedure

DAWN employed a multi-step weighting process to produce nationally representative estimates given DAWN's hybrid sentinel surveillance and probability sample design. The multi-step weighting process involved (1) calculating initial base weights for each sampling part/stratum, (2) adjusting the initial base weights for changes in the sample design and sampling frame, (3) adjusting for hospital non-response, and (4) post-stratification to adjust DAWN estimates of ED visit totals to AHA ED visits for the given stratum.

2.3 Degree of Accuracy Needed for the Purpose Described in the Justification

DAWN hospital samples will be drawn sequentially instead of selecting all hospitals in the first year. This approach benefits DAWN by providing an opportunity to adjust in response to recent DAWN data or changes in the hospital population; and maintaining a representative sample of the population that can be used to create national estimates and for sentinel hospital surveillance with increasing precision each year. A final total sample of 143 hospitals will be able to support national estimates as well as estimates for smaller domains of interest, such as region, type of geography (urban, suburban, rural), specific geography (state, county, city), and priority subpopulations (e.g., race/ethnicity). Estimates by state, county, and city

may be difficult to achieve with reasonable precision from a design-based perspective with a responding sample size of fewer than 143 hospitals.

With the sample size of 143 hospitals, SAMHSA anticipates the relative standard errors (RSEs) for key measures will decrease. As part of continuous improvement of DAWN approach and procedures, SAMHSA will evaluate the sample design's precision performance (via RSEs), noting outlier values in data, weights, adjustments to weights, and their causes, and develop improved data quality control, imputation, weighting, estimation, and variance estimation procedures.

New data abstraction fields of Gender Identity and Sexual Orientation (SOGI) are proposed to be added in the DAWN Case Report Form (Attachment A), following the Department of Health and Human Services (HHS) and SAMHSA's commitment to meeting the administration's call to reduce behavioral health inequities faced by LGBTQ+ children, adolescents, and their families. The primary purpose of this addition is to gather information that can help researchers, policymakers, service providers, and other stakeholders understand diverse populations and create policies, programs, and budgets that meet these populations' needs. These efforts aim to reflect the identities and experiences of people and communities that deserve to be heard and respected. It is also important to consider that data collection is not an end in itself and attention is required to the analytical aspects as well. Methodological imprecision can lead to mismeasurement of the relevant concept or misuse of the data, and this can have negative repercussions for the community, as well as for the overall data quality. There is an increasing concern about the potential damages that can occur from mismeasurement or misuse of measures of sex and gender, particularly in health care. DAWN proposes the following methodology to pursue the most scientific methodology in categorizing and analyzing the SOGI data to be collected in DAWN. Data collected will be assessed for missing values before being categorized into the HHS/SAMHSA selected categories. Missing data may cause bias and will cause a reduction in efficiency as well as data quality. DAWN will analyze the missing for reasons for missingness, primarily to identify whether the data is missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). Thresholds will be identified for reporting the SOGI data to protect the quality of the data and avoid mismeasurement or misuse of the data. DAWN will abstract the data, code, and implement the above-mentioned methods to evaluate missing data. Post analysis DAWN will decide the threshold of missing for reporting to ensure the scientific integrity of the data which aligns with CBHSQ's role as a Federal Statistical Unit (FSU).

3. Methods to Maximize Response Rates

This section describes the methods and activities that are implemented to maximize response rates of hospital respondents that were selected to participate in DAWN and how DAWN accounts for nonresponse.

Hospital Recruitment

The DAWN contractor has a hospital recruiting expert that has significant experience in recruiting hospitals for large hospital-based public health projects. DAWN's recruitment team

developed a recruitment plan and materials to use to inform hospitals about the project and aid in all recruitment-related activities.

- **Research Hospitals:** First, before contacting any hospitals, the hospital recruiter researched each hospital to collect any information that could aid the initial recruitment process, for example a search of the hospital website to learn more about the hospital's mission, trauma level, and general population served. Confirmation of hospital eligibility was also obtained by the hospital recruiter's preliminary research.
- **Identify Hospital Contact:** The next step in the recruitment process was to identify the appropriate hospital contact. The recruiter researched the hospitals to identify the appropriate person(s) to receive the invitation.
- **Sent Invitation Package:** Once the recruiter identified the principal contact she arranged to send the invitation package to the contact by her preferred method; i.e., email, mail, or in person.
- **Follow-up:** Approximately 3 days after the initial invitation packet, the recruiter sent one follow-up email or make one follow-up phone call, depending on how the original invitation was sent. A second attempt to contact the individual was made the next day. Continued follow-up alternating between phone and email was completed, as appropriate. If the recruiter was unable to reach the contact after two weeks, the recruiter may have identified an alternate contact and requested a mailing or send an email to the new contact. Given the relatively small number of hospitals being recruited, there was no maximum number of contact attempts. The recruiter used her best judgment to determine when to put recruitment on hold for a particular hospital. However, if contact wasn't established within 30 days, the recruiter notified SAMHSA and requested a replacement hospital.
- **Schedule Call:** Once the recruiter confirmed that the primary contact has received the initial invitation packet, the recruiter arranged for a time to discuss DAWN and the steps necessary to begin abstraction.
- **Participation Persuasion:** If the primary contact was not interested in being part of DAWN, the recruiter will emphasize the importance of their participation. That is, the hospital was selected to be representative of communities affected by drug use; their participation is integral in addressing new and emerging drug use trends to improve public health; the data collection protocol is intended to have minimal impact on their day-to-day operations. If the primary contact expressed interest in participating, but may not have had the infrastructure to support participation, the recruiter should discuss the specifics of the concerns, including the abstraction options available.
- **Refusal Conversion:** In order to aid refusal conversion attempts, the recruiter documented the following information when a hospital refused to participate:
 - Was a reason for the refusal given?
 - Who refused?

- o When did the refusal occur?
- o What was the degree of the refusal (i.e., hard or soft refusal)?

Answers to these questions varied by hospital, however, the answers helped to assess the hospital-specific refusal and allowed the recruiter to engage in the refusal conversion process.

Hospital Retention Activities

Maintaining a strong rapport with the participating hospitals is essential to the success of DAWN. In order to retain participating hospitals, we continue to identify and develop materials to aid in maintaining the relationships established with key contacts. Once data collection has started, weekly calls with the field managers responsible for managing the abstraction efforts will include time to discuss new facility relationship-building approaches. Retention activities may include, but are not limited to, thank you notes and DAWN update letters.

4. Tests of Procedures

Many years of experience with Legacy DAWN has informed the data collection elements and the processes. Throughout data abstraction, SAMHSA assesses the data and the processes and make any modifications as necessary to ensure that SAMHSA is receiving the high-quality data that is desired. Data from the legacy DAWN was used to test the Machine Learning Process and the Statistical process control (SPC) system. The SPC provides a well-established and robust framework to identify anomalies by applying statistical outlier detection rules to an ongoing time series of observations.

5. Consultants

The following individuals were consulted on statistical aspects of the DAWN sample design.

Name	Organization	Email Address
Matt Gladden	CDC/DDNID/NCIPC/DOP	gkv7@cdc.gov
Alana Vivolo-Kantor	CDC/DDNID/NCIPC/DOP	goz4@cdc.gov
Michael Coletta	CDC/DDNID/NCIPC/OD	mac0@cdc.gov
Michael Cala	OND CP	Michael.A.Cala@ondcp.eop.gov
James Green	Westat	JimGreen@westat.com
Rick Valliant, PhD	Westat	valliant@umich.edu
Suparna Das, PhD	SAMHSA/CBHSQ/OTS	suparna.das@samhsa.hhs.gov
Nathan Donnelly	SAMHSA/CBHSQ/OTS	nathan.donnelly@samhsa.hhs.gov
Brittany Wilbourn, PhD	SAMHSA/CBHSQ/OTS	brittany.wilbourn@samhsa.hhs.gov