United States Food and Drug Administration

Healthcare Provider Perception of Boxed Warning Survey

OMB Control No. 0910-NEW

Part B Statistical Methods

1. <u>Respondent Universe and Sampling Methods</u>

Research Goals

Part A of the supporting statement described the rationale for conducting the study. The overall goal of the study is to better understand how HCPs perceive and information included in a product's boxed warning when making decisions about prescribing treatments.

The general research questions in this data collection are as follows:

1. What awareness, knowledge, and beliefs do HCPs have regarding boxed warning information for a prescription drug or class of drugs?
2. When making prescribing decisions, how do HCPs consider boxed warning information about a potential treatment? How does boxed warning information factor into their assessments of a drug's potential benefits and risks to their patients?
3. How do HCPs communicate with their patients about boxed warning information?
4. What factors (e.g., experience treating a condition) are associated with HCPs' awareness, knowledge, and beliefs about boxed warning information?

To explore a range of potential perceptions and uses of boxed warning information that may exist under different contexts, this study will include two medical product scenarios involving an FDA-approved medication or class of medications that include boxed warning information. The two scenarios are: (a) fixed-dose combination direct acting antiviral (DAA) medications indicated for the treatment of chronic hepatitis C virus (HCV; referred to as the *DAA* scenario); and (b) estrogen vaginal tablets indicated for the treatment of vulvar vaginal atrophy (VVA) associated with menopause (referred to as the *estrogen* scenario) The scenarios will include pertinent prescribing information from the FDA-approved labeling for these medications.

Sampling Frame

We will use SurveyHealthcareGlobus (SHG) U.S. healthcare professionals online panel as the frame for sample selection. The SHG U.S. panel comprises more than 640,000

HCPs, including over 540,000 physicians with MD/DO degrees, 100% of whom have prescribing authority. SHG regularly updates and validates panel members' credentials through the American Medical Association (AMA) and National Provider Identifier (NPI) databases for their ability to practice medicine, inclusive of prescribing authority. Additional verification sources include state license records, hospital directories, publicly available and verified healthcare portals, as well as the AMA database for physicians. In addition, SHG requests and collects copies of medical credentials and calls enrollees to verify them.

The SHG panel also includes nurses and other allied professionals, some with and some without prescribing authority. Because prescribing authority differs depending on state law, SHG does not maintain data on which of these HCPs have prescribing authority, Therefore, the study will depend on self-reported prescribing authority from the screener to determine eligibility for the subgroups of nurse practitioners and physician assistants.

Overall, the SHG offers a higher coverage of healthcare professionals compared to more traditional general population internet panels. SHG specializes in healthcare-related research recruitment and has one of the largest medical market research communities and samples from a population of over 2 million physicians and allied healthcare professionals.

Survey Population

The proposed data collection will include a diverse sample of HCPs reflective of a national population of U.S. HCPs. The following HCP groups have been selected inclusion in the study: physicians (primary care and specialists), nurse practitioners (NP), and physician assistants (PA) (Table 1). Eligible physicians include individuals with Doctor of Medicine (MD) or Doctor of Osteopathic Medicine (DO) degrees. To be eligible for this study, GPs must conduct direct patient care at least 50% of the time, and specialists must conduct direct patient care at least 20% of the time. All participants must write at least 30 prescriptions per week, report office-based or hospital-based practice as their major professional activity and have treated either VVA or chronic HCV. Specialists will be directed to the survey that corresponds with their specialty—OB/GYNs and geriatricians will see the estrogen survey whereas gastroenterologists, hepatologists, and infectious disease specialists will see the DAA survey. GPs will qualify for one or both surveys based on their experience treating the relevant conditions. GPs who qualify for both surveys will be randomly assigned to either survey, following their completion of the screener.

Table 1. Panel Representation of HCP Population

Specialty	Universe (U.S. Population)	% of Panel Reach[1]	Panel Size
General/Family Practitioners	300,161	28%	~84,845
Internal Medicine	186,936	57%	~106,553
Nurse Practitioners and Physician Assistants	~371,000	39%	~146,000
Infectious Diseases	8,280	34%	~2,815
Gastroenterology and Hepatology*	14,695	44%	~6,466
OB/GYNs	45,909	63%	~28,923
Geriatricians	6,061	42%	~2,546

*SHG does not track specific numbers of hepatologists within its panel book, but estimates that 20% of gastroenterologists would identify hepatology as a subspecialty (through the screener), which means the available panel sample size for hepatologists would be approximately $n = 1,293$. Based on the limited number of panelists, we will recruit self-identified gastroenterologists in addition to hepatologists and focusing on behavioral screener criterion (i.e., experience treating chronic HCV).

Using a prequalification screener, sampled HCPs in SHG's network will be screened to determine their eligibility for the study. (See Appendix B for screener.) A pretest will be conducted prior to the main study to test data collection process.

SHG employs quality control processes to help ensure the identity of survey respondents, for example, using IP address information to verify panelists and flag respondents whose IP information is outside of the expected location.

Sample Size

The sample size for the pretest is $n=50$ and the main study is $n=1156$, divided into the follow subgroups (Table 2). The pretest sample ($n=50$) will not overlap with the main study sample ($n=1156$).

---

[1] Percentage of universe/population represented by panel membership.

Table 2 *Survey sample breakdown*

Population	Survey 1: Atrophic Vaginitis/Vagifem	Survey 2: Hepatitis C Infection/DAAs
GPs (PCPs, Internists, Family Medicine, NPs, PAs)	347	347
OB/GYNs and Geriatricians	231	0
Gastroenterologists/Hepatologists and Infectious Disease Specialists	0	231
TOTAL	578	578

Power

Power analysis is not applicable for this data collection as there are no experimental manipulations. Rather than conducting typical power analyses, we arrived at the sample sizes based on a target margin of error (MOE) under a probability-based sampling scenario, using the formula below.

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{MOE} \right)^2 p(1-p)UWE$$

Notation	Description
n	Total number of survey completes
p	Estimate of a proportion; assume 0.5 for a conservative variance estimate
MOE	Margin of error
$z_{1-\frac{\alpha}{2}}$	Quantile from the standard normal distribution; assume a confidence level of 95%
UWE	Kish's unequal weighting effect[2]

For the physician population, we assumed a 5% MOE with an unequal weighting effect (UWE) of 1.5, which means that 578 completed surveys would be needed for each of the two surveys ($n = 1,156$) in a probability-based context. Note that the term "margin of error" typically invokes design-based principles of inference, which are not directly relevant in this particular survey context since it entails the use of nonprobability data, and implicitly relies on model-based assumptions. However, in the absence of well-established and accepted principles of sample design for nonprobability surveys, we have

---

[2] Normally, we adjust the sample size by an estimated correction factor $(1 + L)$ for weighting variability and its effect on precision, in which $L = n^{-1} \sum_s \frac{\left(w_i - \overline{w} \right)^2}{\overline{w}^2}$ and is the squared coefficient of variation of the sample weight w_i. This $1 + L$, termed the relative loss due to weighting by Kish (1992) and commonly referred to as the unequal weighting effect (UWE), is a reasonable approximation for the design effect (DEFF) when the weights are unrelated to the outcome of interest (e.g., see Spencer, 2000).

borrowed upon this terminology as to inform the sample size calculations, while recognizing that the term "margin of error" is not, strictly speaking, appropriate in this context. We also note that the term "unequal weighting effect" is not directly applicable in this study, since inferences will be unweighted. However, we have incorporated the UWE in the above calculation to mitigate against inflation in variances that could arise due to inefficiencies in the sample.

For each of the two survey cohorts (VAA and HCV), we have set a quota for relative proportions of GPs (60%) and specialists (40%). We arrived at these quotas based on balancing several considerations specific to both prescribing scenarios (VAA and HCV):

1. The size of GP universe population is much larger compared to specialists in the U.S. (See Table 1). However, specialists prescribe at a greater volume than general practitioners.[3] According to an internal FDA analysis, approximately 40% of the total number of prescriptions for estrogen vaginal tablets written in the U.S. in 2017 were prescribed by general practitioners of the type included in our general practitioner cohort for the estrogen scenario. Approximately 55% of total prescriptions for estrogen vaginal tablets were prescribed by the specialists included in our specialist cohort. The remaining 5% of prescriptions were written by a variety of other types of specialties. The prescribing volumes for the hepatitis C DAAs had similar percentages.
2. Balancing subgroup precision and overall precision. Overrepresenting a subgroup (in this case, specialists) that has a small universe population relative to the total universe population share improves domain precision for that subgroup at a cost to overall precision.
3. Availability of sufficient sample from within the research panel, as there are smaller populations of specialists compared to GPs.

Considering all of these factors together, we concluded that the survey sample should include more GPs than specialists, with enough representation from both groups to ensure sufficient specificity of survey estimates. We determined that a 60% (general practitioner) /40% (specialist) split was an appropriate balance for each cohort (VVA and HCV). Overrepresenting the specialist subgroups above should allow for reasonable precision for all subgroups without introducing too great a reduction in overall precision should one group be overrepresented to a moderate extent.

Weighting

All analysis will be conducted on unweighted data. Weighting is best suited when: (a) statements about the generalizability of the analysis to the national population of healthcare providers is desired; and (b) when appropriate benchmarks can be obtained and determined. Neither conditions apply in this case. This study does not attempt to generalize its findings applicable to the broader U.S. population of HCPs. Rather, the intent is to understand the perceptions, attitudes and behaviors of a set of prescribers who prescribe medications to treat the specific condition for each scenario, Hepatitis C Virus

---

[3] We recognize that prescribing volume reflects a different unit of analysis than our survey population (i.e., number of prescriptions rather than number of prescribers.)

and VVA, respectively. As a subset of the broader population of HCPs with prescribing authority, this set of prescribers is defined by their prescribing behavior. We are not aware of any well-defined external benchmarks for a survey population defined by whether they prescribe medicines for HCV and VVA respectively. In this case, applying weights calibrated instead to benchmarks for a broader population (all general practitioners) to this particular survey population would be difficult to interpret and may increase, rather than reduce, the bias of estimates.

We believe an appropriate and more efficient means of adjusting for potential selection bias is by including potential confounders as covariates in our analysis (see Nonresponse Bias Analysis).

2. <u>Procedures for the Collection of Information</u>

For both the pretest and main study, the online panel vender SHG will initially send a recruitment email (Appendix A) that links to a prequalifying screener (Appendix B) to identify eligible HCPs. The screener will include questions about the amount of time spent seeing patients, demographic questions (age, race/ethnicity, and gender) and question to confirm the respondent's specialty and experience treating the relevant scenario condition. All respondents that meet eligibility requirements will be invited to participate in the survey within 24 hours of completing the screener. In accordance with their screening criteria met, respondents will be assigned to one of two nearly identical survey instruments (Appendix C) that correspond to the two prescribing scenarios. Respondents will be paid incentives for completing the survey.

Analysis Plan

We will compute frequencies for all survey questions (see Appendix C for a listing of the questionnaire items). Means and standard deviations will be provided for scale items. To compare responses between different subgroups of interest, we will carry through chi-square tests for categorical variables (using t-tests to compare frequencies of specific priority response options), and ordinal logistic regression for ordinal variables. Chi-square and ordinal logistic regression tests will test the null hypothesis that subgroup membership is independent of responses to each question. To assess relationships among key variables of interest, we will conduct correlation and regression tests that test the null hypothesis of the correlation/regression coefficient not being significantly different from 0. Analyses will be conducted using Stata statistical software. For hypothesis testing for subgroup comparisons and correlations, a 95% confidence level and $p < 0.05$ will be used as a standard for statistical significance. These will enable us to answer the key research questions relating to descriptive assessment of HCPs' prescription decision-making, as well as how the key variables of prescribing scenario and participant variables (GPs vs. specialists) are related to outcome variables such as understanding of boxed warning risks, assessment of drug efficacy, and likelihood of prescribing.

Analysis of Open-ended Questions

Responses to open-ended questions will be coded by trained research staff. We will develop the coding schema using the responses from the pretest surveys (n=50) as well as a sample of responses from the main survey. At least two coders first complete at least 10% of the responses—if an intercoder reliability of .75 (kappa) is obtained, then the remaining responses are divided among the coders. Any differences between coders will be discussed and adjudicated by a third reviewer. All coded responses will be transformed into categorical numeric proxy measures for analysis.

3. <u>Methods to Maximize Response Rates and Deal with Non-response</u>

    Because this study utilizes a non-probability sampling approach and a web-based panel, we will calculate a participation rate rather than a response rate, in accordance with AAPOR's Standard Definitions (2016)[4]. For each survey, we will calculate a participation rate as the number of individuals who provide usable data divided by the estimated number of panelists who were invited to take the survey.

    To help ensure that the participation rate is as high as possible, FDA and the contractor will:

    - Design a protocol and cognitively-tested survey instrument that minimizes burden (reasonable in length, clearly written, and with appealing graphics).

    - Use incentive rates that are as close to industry standards as possible. In addition to offsetting respondent burden, using market-rate incentives tends to increase participation rates.[5]

    - Use government sponsorship on the survey invite to increase response rate. An experiment conducted by FDA and RTI[6] found that among endocrinologists, response rates were 6 percentage points higher when FDA was disclosed as the sponsor in the survey invitation than when no sponsor was listed. All study materials will reference the U.S. Food and Drug Administration.

    - Use multiple email reminder invitations to increase opportunities for invited panelists to participate, including more custom personal reminders for specific targets of interest, based on observed participation rates while in the field.

    Nonresponse Bias Analysis

    Given the use of a nonprobability panel, for each of the two surveys, we will apply benchmarking methods to assess how the demographics of respondents differ from identified external benchmarks. For physicians, these benchmarks will be based on

---

[4] The American Association for Public Opinion Research. 2016. Standard definitions: Final dispositions of case codes and outcome rates for surveys. 9th edition. AAPOR.

[5] Dykema, J., Stevenson, J., Day, B., Sellers, SL, Bonham VL (2011). Effects of Incentives and Prenotification on Response Rates and Costs in a National Web Survey of Physicians. *Eval Health Prof*, 34(4): 434–447.

[6] Aikin, KJ, Betts, K, Boudewyns, V, Stine, A, Southwell, B. (2016). Physician responsiveness to survey incentives and sponsorship in prescription drug advertising research. *Annals of Behavioral Medicine, 50*(Suppl.), s251.

comprehensive statistics from the American Medical Association (AMA) Physician Masterfile; analogous sources will be selected for NPs and PAs based on availability, quality, and relevance. We will ensure that the benchmark analyses reflect comparable populations, to the extent possible (e.g., sample physicians [regardless of prescribing activities] compared to AMA physicians).

We anticipate that this analysis will reflect the most important sources of potential selection bias, since it will reflect potential selection bias arising from panel recruitment and retention as well as from study-specific sampling and screener nonresponse. However, an HCP who completes the screener and qualifies for the study may still fail to complete the entire survey. Therefore, the second analysis analyzes the likelihood of completing the survey, conditional upon screener completion and eligibility. This entails estimating a logistic regression model to predict survey completion as a function of demographic information collected during the screener. This model will be restricted to individuals who completed the screener and meet all screening eligibility criteria.

Item Nonresponse

We expect low levels of item-missing data in this survey, due to the methodologies employed and the data completeness criteria that will be imposed. For each survey respondent, we will compute the proportion of the survey that was completed, while accounting for skip logic. We plan to classify eligible units as complete respondents only if they complete at least 75% of the survey.

After classifying sample members based on survey eligibility and completion, we will analyze item nonresponse for respondents and tabulate item nonresponse rates for every question. Additional examination will be employed for questionnaire items that register unusually high item nonresponse rates (e.g., > 5%). The specific actions we will take in response to item nonresponse will depend on the amount of missing data, the practical importance of the specific analysis, and the potential mechanisms associated with the missing data, following practices established by Little & Rubin (2002). [7]

4. Test of Procedures or Methods to be Undertaken

   We conducted five cognitive interviews with healthcare providers to assess questionnaire flow and wording and received written pre-test feedback from an additional three healthcare providers. Before proceeding with the main survey, we will conduct a pre-test survey with a sample of 50 respondents from the web panel to ensure that the main study will run smoothly.

5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

   The contractor, Fors Marsh Group, will collect and analyze the data on behalf of FDA as a task order under Contract HHSF223201510001B. Elise Bui, Ph.D., 571-444-1131, is

---

[7] Little, R. J., & Rubin, D. (2002). Statistical analysis with missing data. 2nd Edition, John Wiley and Sons, 389 pp.

the Project Director for this project. Data analysis will be overseen by the Office of Program and Strategic Analysis, Office of Strategic Programs, CDER, FDA, and coordinated by Sara Eggers, Ph.D., 301-796-4904.