

Data and Methods for Evaluating the Impact of Opioid Formulations with Properties Designed to Deter Abuse in the Postmarket Setting

July 10-11, 2017
Docket No. FDA-2017-N-2903

Center for Drug Evaluation and Research
Food and Drug Administration
Department of Health and Human Services

Issues Paper

Table of Contents

1.	Overview	2
2.	Background	2
3.	Overview of Current Data Sources Used in Postmarket Evaluation of Opioid Formulations with Abuse-Deterrent Properties.....	3
3.1	Poison Control Center Call Data.....	3
3.2	Data Collected from Individuals Entering or Being Assessed for Substance Use Disorder Treatment	4
3.3	Electronic Health Care Data, Including Administrative Claims Data	6
3.4	Population-Based Surveys.....	7
3.5	Other Potential Sources of Information	8
3.5.1	Spontaneous Adverse Events	9
3.5.2	Drug Diversion Data	9
3.5.3	Web Monitoring Programs	10
3.5.4	Street Price	10
3.5.5	Prescription Dispensing Data: Doctor and/or Pharmacy Shopping.....	10
4.	Current Methodological Approaches and Challenges	11
4.1	Outcomes.....	11
4.2	Current Study Designs	13
4.3	Metrics and Denominators.....	13
4.4	Misclassification/Ascertainment of Products	14
4.5	Sampling and Selection Bias	15
4.6	Confounding and Secular Trends.....	16
4.7	Statistical Models.....	17
5.	Organization of Meeting and Discussion Topics.....	18
6.	References.....	21

1. OVERVIEW

Determining the real-world impact of opioid formulations with properties designed to deter abuse¹ is a key next step in understanding the utility of these products in helping to curb the opioid abuse epidemic. Ascertaining the extent of prescription drug abuse and related safety outcomes presents unique challenges because the data sources and methods typically used to study other drug safety issues may not provide relevant information. Studies of prescription drug abuse differ from traditional pharmacoepidemiologic investigations in a number of ways, for example:

- Product-specific exposure can be problematic to determine because it often occurs outside the health care system; outcomes commonly occur in individuals not prescribed opioid products.
- Many factors that can affect both the ability to assess and the overall levels and trends in prescription drug abuse are not captured in clinical information (e.g., state and Federal law enforcement and policy changes, regional trends).
- No national-level data resource can provide estimates of prescription drug abuse, at all levels of severity, and link those data to relevant clinical and other information needed to form a comprehensive assessment of the problem.
- Available data resources generally capture one aspect of interest (abuse, clinical, or mortality data) without the ability to link to other relevant datasets.
- Outcomes that come to medical attention cannot generally be linked to a specific product or products.

Although this area of research has advanced since the first opioid formulation with properties designed to deter abuse was approved,² significant challenges remain at all stages of study design, execution, and interpretation. This document represents a high-level summary of the often-used data sources and methodological concerns currently facing investigators and regulators in this continually evolving area.

2. BACKGROUND

Prescription opioid products are an important component of modern pain management. However, abuse and misuse of these products have created a serious and growing public health problem. One of the highest priorities of the Administration, the Department of Health and Human Services, and the U.S. Food and Drug Administration (FDA or Agency) is addressing the abuse and misuse of opioids. The FDA is currently working to

¹ This term refers to opioid products that incorporate technologies designed to deter abuse, regardless of whether the product is labeled as such.

² http://www.accessdata.fda.gov/drugsatfda_docs/applletter/2010/022272s00oltr.pdf. Accessed 3/23/2017.

identify novel opportunities that will allow the Agency to take more forceful steps, in addition to the already ongoing work, to address this crisis.³

One important step toward the goal of creating safer opioids has been the development of opioid products with properties that are designed to deter abuse. In April 2015, the Agency issued a final guidance to assist industry in the development of opioid drug products with properties designed to deter abuse by different routes. The guidance for industry *Abuse-Deterrent Opioids — Evaluation and Labeling* explains the Agency's current thinking regarding studies that should be conducted to demonstrate that a given formulation can be expected to deter abuse by particular routes, makes recommendations about how those studies should be performed and evaluated, and discusses how to include data, together with an accurate characterization of what the data mean in product labeling.⁴

Evaluating the effects of opioid formulations designed to deter abuse in post-approval settings is important for understanding the impact of these products on abuse after approval, but this is a new area of inquiry, so standard approaches and methodologies are still being developed. The current approaches to these studies are challenging on many levels, and we seek to bring together scientific expertise to help illuminate ways to improve the data sources and the methodologies.

The purposes of this document are to:

- 1) Provide a brief overview of the currently available data resources that are used to evaluate the impact of opioid products with properties designed to deter abuse in post-approval settings,
- 2) Summarize some key data and methodological issues that have arisen in our review of studies in this area, and
- 3) Outline topics for discussion, including ways to improve currently available data sources and analytic methods and to develop new data resources and methods in this area.

3. OVERVIEW OF CURRENT DATA SOURCES USED IN POSTMARKET EVALUATION OF OPIOID FORMULATIONS WITH PROPERTIES DESIGNED TO DETER ABUSE

3.1 POISON CONTROL CENTER CALL DATA

The American Association of Poison Control Centers (AAPCC) maintains the National

³ <https://blogs.fda.gov/fdavoices/index.php/2017/05/fda-commissioner-asks-staff-for-more-forceful-steps-to-stem-the-opioid-crisis/>. Accessed 6/6/2017.

⁴ <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm334743.pdf>. Accessed 3/15/2017.

Poison Data System (NPDS), a data repository that includes reports of exposures to substances based on calls to Poison Control Centers (PCCs) in the United States. These calls come from individuals (or someone caring for them, including health professionals) exposed to pharmaceutical drugs or potentially hazardous substances. AAPCC states that all 57 regional PCCs upload case data to the system. Data fields include, but are not limited to, demographics, product or substance, call type, reason for exposure, and route of administration. Depending on the circumstances, some cases are followed-up after the initial call to determine the outcome (Mowry et al., 2016).

The Researched Abuse, Diversion, and Addiction-Related Surveillance (RADARS) System Poison Center Program obtains data on a subset of drugs, including opioids, from a large majority of regional U.S. poison centers, covering over 90 percent of the U.S. population. RADARS personnel then conduct additional quality checks on the call data, based on review of case narratives (Dart et al., 2015).

Strengths of the PCC data include their meaningful and clinically relevant abuse-related outcome measures, product specificity, and wide geographic coverage. In addition, they can capture information from individuals who might not participate in surveys or interact with the health care delivery system. However, these data have limitations that must be taken into consideration. First, an unknown, and likely small, fraction of abuse and overdose events result in a call to a PCC. It is unclear what factors might influence whether an opioid abuse-related event generates a call, or how these factors might vary over time or across drugs. The ability of these data to reliably distinguish specific product formulations and generic products is also unclear and is discussed further below. Although there is some evidence that trends in call rates are correlated with trends in rates of emergency department visits involving misuse and abuse of prescription opioids (Davis et al., 2014, Bau et al., 2016), there is also evidence suggesting that patterns of PCC use have been changing in recent years (Mowry et al., 2015), further complicating the interpretation of analyses using PCC data with regard to making inferences about abuse trends in the population. In addition, overdoses resulting in rapid, unattended death are unlikely to generate a call. Therefore, PCC may disproportionately fail to capture cases involving drugs with the highest risk of such fatal overdoses.

3.2 DATA COLLECTED FROM INDIVIDUALS ENTERING OR BEING ASSESSED FOR SUBSTANCE USE DISORDER TREATMENT

The National Addictions Vigilance Intervention and Prevention Program (NAVIPPRO) is a surveillance system that measures patterns of abuse for selected prescription and illicit drugs. The Addiction Severity Index—Multimedia Version (ASI-MV) is a computerized standard clinical intake assessment used by a dynamic network of treatment centers and other types of institutions, such as correctional facilities, to assess individuals for substance use disorders. Specifically, the ASI-MV assessment captures product-specific data related to past 30-day use and abuse of prescription opioid products using visual

images to aid in identification of products. Although it covers a majority of states, it is not nationally representative. The ASI-MV also includes questions on route of administration. ASI-MV questions are modified as changes in the prescription drug market occur (Cassidy et al., 2014).

The Comprehensive Health Assessment for Teens (CHAT) is a computerized behavioral health assessment targeted to adolescents aged 18 years and younger being assessed for substance use disorders also administered through NAVIPPRO. Similar to the ASI-MV, CHAT collects data on the use and abuse of opioids, as well as factors related to substance use disorder that are specific to this younger population. Also like the ASI-MV, data related to route(s) of administration are collected. The CHAT network of participating sites comprises treatment centers and other facilities, such as alternative schools and mental health programs. CHAT monitors the same prescription medications tracked by ASI-MV (Lord et al., 2011).

The RADARS Opioid Treatment and Survey of Key Informants Programs (OTP and SKIP, respectively) record the specific prescription opioids reported by persons entering treatment for opioid use disorders. The OTP resource collects information primarily from public facilities that use medication-assisted treatment, while the SKIP includes primarily private facilities that do not use medication-assisted treatment. Each patient is offered the opportunity to complete an anonymous, standardized, self-administered questionnaire that solicits information on specific prescription drugs “used to get high” in the past 30 days (Dart et al., 2015).

The Treatment Episode Data Set (TEDS) is an admission-based census that includes data from facilities that receive public funds, are licensed or certified by states to provide treatment, or are tracked at the state level for other reasons. It provides information on demographic and substance use disorder characteristics for individuals aged 12 and older who are admitted for abuse of alcohol and/or drugs in facilities that report to state administrative data systems. Although TEDS collects information from all states, each state varies in the information that it submits; private facilities, doctor’s practices, and specific other types of facilities (e.g., programs within the criminal justice system, detox facilities) may not be required to report. An important limitation of TEDS is that the database only allows reporting of three substances of abuse for each admission, and this information is, in general, only available at the molecule level (i.e., brand and formulation data are not available). Because of these and other limitations, these data are of limited use in the postmarket evaluation of opioid formulations with properties designed to deter abuse.⁵

5

https://www.samhsa.gov/data/sites/default/files/2014_Treatment_Episode_Data_Set_National_Admissions_9_19_16.pdf. Accessed 3/8/2017.

An important limitation of data collected from people entering or being assessed for substance use disorder treatment, as well as those calling PCCs, is the potential for various types of misclassification, including the specific product(s) being abused. A certain amount of misclassification is inevitable in any self-reported data. However, if respondents are not able to reliably distinguish between original and reformulated products, extended- and immediate-release products, and/or branded and generic products—or if survey instruments change over time in such a way as to change the degree of product misclassification—comparisons over time and across products can be biased. FDA review of analyses of these data suggest that such misclassification may be substantial and may also be differential, influenced by factors such as the order in which products are presented to the respondent and the similarity in appearance between different opioid products.

Another important limitation of data collected from individuals entering or being assessed for substance use disorder treatment is that NAVIPPRO (including ASI-MV and CHAT) and RADARS (including SKIP and OTP) are convenience samples, with sites entering and leaving the surveillance network in a nonrandom fashion. Shifts in the geographic distribution of the sample, as well as in the distribution of the types of assessment sites contributing data to the system (e.g., public versus private treatment program, inpatient center versus probation office) have the potential to create bias when estimating trends in abuse rates over time. Furthermore, data from sources such as treatment centers are not based on a probability sample from a well-defined sampling frame or population. They are only captured when individuals interact with these surveillance systems. It is therefore difficult to characterize the underlying population about which statements regarding abuse and abuse-related outcomes are to be made. In addition, numerous factors—for example, judicial referral policies and availability and funding of substance use disorder treatment—can affect the probability that an individual who is abusing or addicted to prescription opioids is assessed for treatment and included in the sample. These factors may vary by time period, geographic region, and type of treatment center. The underlying population for these samples is sometimes described as consisting of individuals who are at high risk of prescription opioid abuse (Butler et al., 2013). This is reasonable, but some conceptual issues remain. For example, what characteristics does such a population exhibit and can we enumerate those characteristics? Can we quantify the selection process that drives individuals from this underlying high-risk population to interact with the surveillance systems? And finally, can we make valid inferences to some underlying population based on patterns we see in these samples?

3.3 ELECTRONIC HEALTH CARE DATA, INCLUDING ADMINISTRATIVE CLAIMS DATA

One type of electronic health care data used in the postmarket evaluation of opioid formulations with properties designed to deter abuse is patient-level claims for

reimbursement for services (consisting of billing codes, diagnoses, etc.) submitted by pharmacies and health care providers to insurance companies.

Although claims data are commonly used in many areas of pharmacoepidemiologic research, by themselves, they are of limited use when studying abuse and related outcomes. First, individuals who misuse or abuse prescription drugs often obtain them from sources other than their own prescription, so ascertaining exposure to a product using only prescription claims is not sufficient for fully evaluating the risks associated with exposure to a drug product.

Determining outcomes accurately can also be challenging. Misuse and abuse are not well captured in administrative claims data. Validated algorithms that can accurately identify opioid overdoses using claims data are still being developed, with results from these validation studies expected later in 2017. In addition, these ongoing studies are evaluating the ability of algorithms using coded claims data to reliably identify opioid use disorder diagnoses. Although opioid-related overdose can potentially be measured using claims data if linked to data on deaths, attribution of the overdose to a specific prescription opioid product is not possible (with the exception of methadone) based on claims data alone. FDA has also determined that claims data cannot be used alone to evaluate the impact of opioid formulations with properties designed to deter abuse. However, if claims linked with other data resources, including death registries, are analyzed using rigorous methods and are one component of a suite of studies evaluating opioid formulations with properties designed to deter abuse, they may in the future provide some valuable information on the impact of these products on clinical outcomes of interest.⁶

3.4 POPULATION-BASED SURVEYS

A number of large surveys collect information on the nonmedical use and/or abuse of opioids and other drugs. The National Survey on Drug Use and Health (NSDUH) is an annual, nationally representative survey of the civilian, noninstitutionalized, general U.S. population aged 12 years and older. The survey provides national estimates on respondents' self-reported drug-taking behaviors, including the prevalence of nonmedical use of prescription pain relievers as a class. NSDUH collects data through face-to-face interviews with a representative sample of residents of households and noninstitutional group quarters (e.g., shelters, rooming houses, dormitories) and civilians living on military bases. The survey excludes homeless persons who do not use shelters, military personnel on active duty, and residents of institutional group quarters, such as jails and

6

<https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm334743.pdf>. Accessed 3/17/2017.

hospitals. The NSDUH survey recently underwent a major redesign, which was implemented in 2015. These changes resulted in a trend break for annual estimates of nonmedical use of prescription opioids. The revised survey now collects information on past-year use and misuse of specific opioid molecules (e.g., hydrocodone, oxycodone, fentanyl). However, the lack of product- and formulation-specific data still limit the utility of these data in the evaluation of specific opioid products with properties designed to deter abuse. One exception is OxyContin, for which NSDUH has collected product-specific nonmedical use data since 2004 (Center for Behavioral Health Statistics and Quality, 2016).

Monitoring the Future (MTF) is a nationally representative, annual survey of secondary school students, college students, and young adults intended to monitor emerging substance use disorder problems and understand the effectiveness of policy and intervention efforts designed to address them. The survey has been conducted continuously since 1975. Although the survey asks about a wide variety of substances, in general it is not possible to distinguish products by brand and/or formulation.⁷

The RADARS College Survey Program assesses the nonmedical use of specific prescription opioids and stimulants in undergraduates. It began data collection in 2008, and individuals who are enrolled as undergraduates in 2- or 4-year college, online, or technical schools at least part time are eligible. It is administered online, three times annually. Data include drugs used, reasons for use, sources and routes, chronic pain assessment, and the Drug Abuse Screening Test. Data are self-reported, and the population is self-selected. The underlying population for this type of volunteer *opt-in* internet survey sample remains unclear, and response rates are unknown (Dart et al., 2015).

With the exception of the RADARS College Survey Program, these national surveys are not sufficiently detailed to examine specific opioid products and formulations (other than OxyContin), and information on route of administration is very limited. In addition, individuals with advanced substance use disorders may be underrepresented, particularly if they are homeless, incarcerated, or in a residential treatment facility.

3.5 OTHER POTENTIAL SOURCES OF INFORMATION

Although they may generate abuse-related safety signals or provide anecdotal or supportive information on drug abuse levels and trends, the data resources discussed below have limitations that generally preclude conducting formal pharmacoepidemiologic studies. The base population typically cannot be well defined,

⁷ <http://www.monitoringthefuture.org/pubs/monographs/mtf-overview2016.pdf>. Accessed 3/8/2017.

nor can information be objectively verified. In some cases, individuals may report events that happened to others or historical events.

3.5.1 Spontaneous Adverse Events

The FDA Adverse Event Reporting System (FAERS) is a database that contains information on adverse event and medication error reports submitted to FDA. The database is designed to support the FDA's postmarket safety surveillance program for drug and therapeutic biologic products.

Although FAERS data are quite useful in identifying drug safety signals, they have a number of important limitations that preclude their use as formal metrics, even for calculating the incidence of an adverse event or medication error. No proof of a causal relationship between a product and event is needed in order for the FDA to require the event be reported, and reports do not always contain enough detail to properly evaluate an event. Further, FDA does not receive reports for every adverse event or medication error that occurs with a product. Many factors can influence whether or not an event will be reported, such as the length of time a product has been marketed and publicity about a safety issue. Therefore, FAERS data cannot be used to calculate the incidence of an adverse event or medication error in the U.S. population (Dal Pan et al., 2012).

3.5.2 Drug Diversion Data

The RADARS Drug Diversion Program (DDP) gathers surveillance data on prescription drug diversion. Approximately 300 drug diversion investigators across 49 states and Puerto Rico submit data quarterly on the number of documented drug diversion cases within their jurisdiction for specific prescription drugs of interest. A case results in a written complaint or report containing information about specific drugs found outside of controlled distribution channels. Drug diversion investigators represent municipal police departments, multijurisdictional drug task forces, county sheriffs' departments, regulatory agencies such as state medical and pharmacy boards, state police agencies, prosecutors' offices, and departments of health. In addition to the number of diversion cases, the DDP provides information on the cost of diverted products on the street, based on reports by diversion investigators (Dart et al., 2015).

FDA considers drug diversion to be supportive information, as drug diversion rates are not direct measures of abuse or related clinical outcomes, but rather a measure of law enforcement activity. It is unclear how funding or local law enforcement priorities may influence the number of drug diversion cases, the drugs on which investigators focus their efforts, and how these may vary over time.⁸

⁸ <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm334743.pdf>. Accessed 3/15/2017.

3.5.3 Web Monitoring Programs

Web monitoring programs collect and analyze internet data from internet-based recreational drug use message boards and discussion forums. Programs provide quantitative evaluation designed to examine the level of discussion of the target products on the monitored forums, as well as coding of qualitative sentiment and themes related to abuse of particular drugs. The evaluation and postmarket surveillance of opioid formulations with properties designed to deter abuse also typically includes examination of *recipes* discussed online for defeating or circumventing these properties.

Although this is a rich arena for understanding what might be happening in the community of individuals who misuse or abuse drugs, including prescription drugs, it has substantial limitations for quantifying risks. It is difficult to define the base population, because the geographic location of respondents is not always apparent, and individuals who participate in online forums might be substantially different from those who do not. Another unknown is how the characteristics and use of specific forums affect the amount and type of information provided. None of the information provided can be verified, and finally, it is not clear how online endorsements of a particular product translate into real-world abuse (McNaughton et al., 2012).

3.5.4 Street Price

StreetRx is a RADARS program that uses a crowdsourcing website to gather information on prices paid for specific drug products. The site is anonymous and open to all, but individuals are asked for the product, formulation, dose, city, and state for the transaction.

Like the internet survey, this effort uses innovative methods to gather information on drugs of abuse in the community, which is quite informative. However, because of both the subject and the collection method, these data are unverifiable with regard to both the specific product involved and the purchase price. It is also not clear how changes in street price relate to the outcomes that are of specific interest to FDA: misuse, abuse, addiction, overdose, and death.⁹

3.5.5 Prescription Dispensing Data: Doctor and/or Pharmacy Shopping

A growing body of literature explores the use of prescription dispensing, or drug utilization, data to help understand the risk of abuse, diversion, and related adverse clinical outcomes, such as addiction and overdose. Various algorithms for so-called *doctor shopping* or *doctor-pharmacy shopping* have been proposed as indicators of increased risk of abuse-related outcomes, based on an individual receiving prescriptions

⁹ <http://www.radars.org/home2/programs/streetrx>. Accessed 3/17/17.

for opioid products from multiple physicians and filling those prescriptions at multiple pharmacies (Cepeda et al., 2012). These types of analyses have been used to try to identify individuals at risk for adverse outcomes as well as a possible indicator of relative levels of misuse and abuse for different opioids (Cepeda et al., 2013). If rates of doctor shopping involving a particular drug decline following its reformulation, it has been proposed that this change may be the result of the abuse-deterrent properties of the drug making it less desirable for manipulation (Coplan et al., 2016).

Although doctor shopping and pharmacy shopping may be associated with an elevated probability of aberrant drug-seeking behavior, they do not currently constitute a meaningful indicator of abuse or abuse-related clinical outcomes. Rather their use as metrics relies on assumptions that link these metrics to levels of abuse in the community. A small body of qualitative research—as well as the experience of prescribers, pharmacists, and law enforcement—suggests that some individuals seek controlled substance prescriptions from multiple prescribers and pharmacies for abuse or diversion (Inciardi et al., 2007, 2009). However, the exact relationship between the number of doctors and/or pharmacies associated with overlapping prescriptions for a product and the probability of that product being abused has not been established. Further research is needed to characterize this association and determine how well doctor-shopping metrics distinguish between therapeutic and nontherapeutic use of controlled substances and to what degree they can be considered proxy measures of abuse. Studies to better define and validate doctor-shopping metrics are currently underway as part of the extended release/long acting opioid analgesic class postmarketing study requirements.¹⁰

4. CURRENT METHODOLOGICAL APPROACHES AND CHALLENGES

4.1 OUTCOMES

The 2015 guidance for industry *Abuse-Deterrent Opioids — Evaluation and Labeling* states that “[t]he goal of postmarket studies . . . is to determine whether the marketing of a product with abuse-deterrent properties results in meaningful reductions in abuse, misuse, and related adverse clinical outcomes, including addiction, overdose, and death in the post-approval setting.”¹¹ In this guidance, FDA defines *abuse* as “the intentional, non-therapeutic use of a drug product or substance, even once, to achieve a desirable psychological or physiological effect.” Abuse is distinguished from *misuse*, which refers to “the intentional therapeutic use of a drug product in an inappropriate way and specifically excludes the definition of abuse.” However, every data source does not define

¹⁰ <https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf>. Accessed 3/17/17. <https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf>. Accessed 3/17/17.

¹¹ <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm334743.pdf>. Accessed 3/15/2017.

these outcomes the same way, and therefore, the definitions of abuse, misuse, and related outcomes depend in part on the source of the outcome data being used for a study.

Opioid formulations with properties designed to deter abuse have, to date, focused primarily on deterring non-oral abuse, most commonly via the nasal and injection routes. Therefore, in addition to evaluating the impact on overall abuse, postmarket evaluation of abuse-deterrence must also include measurement of *route-specific abuse outcomes*. An important consideration in evaluating the effect of an abuse-deterrent opioid formulation is whether the data suggest that it could be shifting abuse to a more dangerous route, for example, from nasal to intravenous abuse.

The 2015 guidance refers to showing a “meaningful reduction” in abuse to support labeling describing properties designed to deter abuse, recognizing that whether or not an effect is meaningful may, from a clinical or public health standpoint, vary depending on a variety of factors. In the postapproval setting, these factors include the baseline levels and patterns of use, abuse, and related outcomes for the drug of interest and other available opioids that do and do not have properties designed to deter abuse. From an epidemiologic standpoint, reduction is a relative term that raises the fundamental question: “Compared to what?” In making causal inferences about the effect of an opioid formulations with properties designed to deter abuse, the concept of the counterfactual is useful. The counterfactual refers to the hypothetical scenario in which the exposure or intervention being evaluated had not occurred yet everything else remains the same. The counterfactual question to be answered here, then, is: “Is the risk of abuse (overall and route-specific) and related adverse outcomes lower than it would have been without the introduction of the product with abuse-deterrent properties, and, if so, to what extent?” The most appropriate time periods, comparators, and analytic approaches for answering this question depend on the specific drug being evaluated and are rarely straightforward.

As part of the workshop, FDA would find it useful to discuss the outcomes currently being studied, how they are measured, and whether different or additional outcomes would be helpful in evaluating the impact of opioid formulations with properties designed to deter abuse. For example, the postmarket evaluation of opioid formulations with properties designed to deter abuse focuses primarily on the effects on abuse, misuse, addiction, overdose, and death associated with the product itself. We are, however, interested in scientific discussion about the value and feasibility of assessing the public health effects of such products more broadly. For example, such effects could include changes in prescribing patterns, increased use of illicit drugs (e.g., heroin), or specific changes in how a product is manipulated or used that could alter the risk of blood-borne pathogen transmission or other adverse health effects.¹²

¹² <https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AnestheticAndAnalgesicDrugProductsAdvisoryCommittee/UCM545760.pdf>. Accessed 3/17/17.

4.2 CURRENT STUDY DESIGNS

The ecologic time series is the most common study design that has been used in postmarket studies evaluating opioid formulations with properties designed to deter abuse (Coplan et al., 2016). Ecological studies—in this case that compare aggregate measures of abuse across time periods—are commonly used in public health and policy arenas to assess the impact of community-level interventions. This type of study is quite different from a clinical trial or cohort study, in which a group is followed over time to assess the individual-level association between some intervention or exposure and the outcome of interest. Ecologic designs have some particular limitations compared to studies that link an exposure/intervention and an outcome at the individual level. Because associations and patterns seen at the aggregate or group level may not reflect associations at the individual level, caution is warranted in drawing inferences from an observed reduction in aggregate abuse prevalence or rates about the *risk* of an individual abusing a product, of transitioning from one route to another, or of progressing to more severe opioid use disorder. We are particularly interested in discussing study designs that might be capable of measuring changes in the *risk* of abuse and related adverse outcomes at an individual level that can be attributed to a product’s abuse-deterrent properties.

4.3 METRICS AND DENOMINATORS

In addition to the previously described issues with ecologic time series studies, there are unanswered questions about which metrics and methods are most appropriate and useful for assessing the effect of opioid formulations with properties designed to deter abuse over time. In postmarket abuse-deterrence studies, three broad categories of metrics are typically used. First, what is sometimes referred to as a drug’s *route of abuse profile* examines the proportion of those indicating abuse of the drug who report using it via specific routes (e.g., oral, nasal, smoking, injecting). Second, the prevalence of abuse represents the proportion of the total study population—for example, the number of individuals surveyed or the population covered by a set of PCCs—that report or are identified as abusing a particular drug during a specified time period. Finally, utilization-adjusted abuse rates measure the number of abuse reports or other related outcomes relative to the amount of product dispensed from pharmacies to patients within the study coverage area during the time period of interest.

A drug’s utilization volume is most commonly measured as the number of individuals prescribed the drug, the number of prescriptions, or the number of dosage units (e.g., tablets, capsules) dispensed. Utilization data generally come from large databases that collect data on drugs dispensed from retail pharmacies—for example, IMS Health, National Prescription Audit—to estimate the number of prescriptions or dosage units dispensed in a particular geographic area. Utilization-adjusted analyses are critical to the assessment of abuse-deterrence because (1) adjustment for varying levels of drug

availability is important in comparing abuse levels across products, and (2) many factors other than reformulation can affect prescribing trends for a product. Some of these factors might include drug marketing practices, the availability of generic products, insurance reimbursement policies, use of prescription drug monitoring programs, and drug shortages. Reformulation of a product to include abuse-deterrent properties itself might also affect utilization, for example, if the reformulation results in a decline in demand by those intending to abuse the product. Adjusting for utilization controls for all of these factors.

Methodological uncertainties remain about the best way to measure community-level availability and to model the relationship between utilization and abuse rates at varying levels of market share (Secora et al., 2014, 2017). In evaluating the impact of opioid formulations with properties designed to deter abuse, FDA generally considers the number of dosage units dispensed to be superior to the number of prescriptions dispensed or the number of individuals receiving a prescription, because every dosage unit presents an opportunity for abuse, and the average number of dosage units per prescription may vary across opioids. However, different utilization denominators may be appropriate for different types of products or study questions. Because prescription drugs may be used in areas remote from where they are dispensed, the best catchment area for utilization data can be unclear in studies that do not have nationally representative samples. Despite these limitations and uncertainties, accounting for differences in the volume of drug prescribed in the community is important to enable valid comparisons across products and time periods. However, we are interested in public and scientific discussions regarding the best methods for accounting for prescribed drug availability, and how best to interpret observed changes in population- and utilization-adjusted abuse rates in assessing the impact of opioid formulations with properties designed to deter abuse on abuse and related outcomes in the post-approval setting.

4.4 MISCLASSIFICATION/ASCERTAINMENT OF PRODUCTS

In evaluating changes in abuse and abuse-related metrics for a product over time, an important consideration is the potential for misclassification of products. A certain amount of misclassification is inevitable in any self-reported data; however, it is important to consider how false positives and negatives might affect study results. For example, when a product is reformulated with abuse-deterrent properties, some misclassification would be expected to occur in both pre- and post-reformulation periods. Nondifferential misclassification is expected to attenuate differences seen across products and time periods, resulting in a bias toward the null. If misclassification differs across time periods or products, however, bias may occur in a direction that is difficult to predict. In some cases, if misclassification is quantifiable, then it may be possible to apply bias correction procedures (Greenland, 2008; Lyles et al., 2010, 2011; Lash et al., 2009). We are interested in discussing whether and how misclassification bias correction methods might apply to data sources currently used to evaluate opioid formulations with

properties designed to deter abuse in the postmarket setting. In addition, we are interested in discussing strategies for minimizing misclassification, for example, the use of pill photos, design of tablets with distinctive shapes and colors, and use of indices to make specific products and formulations more easily distinguishable. Given that some misclassification is unavoidable, we are also interested in discussing design and analysis approaches to help ensure that misclassification is likely to be nondifferential, for example, by randomizing the order in which products are presented to respondents. And finally, we would like to discuss the possibility of sensitivity analyses that might generate a range of possible effect sizes that take possible misclassification bias into consideration.

4.5 SAMPLING AND SELECTION BIAS

For an opioid product reformulated with properties designed to deter abuse, studies typically evaluate how the level of abuse and related outcomes changed for this product, comparing a period before the reformulation (pre) to a period after the reformulation (post). Similarly, it is important to assess levels of abuse via specific routes, and how these patterns changed from the pre- to the post-periods. However, without a probability sample from a well-defined population, changes seen in abuse and abuse-related outcomes may not reflect what transpired in the underlying population from which those data arose, unless key assumptions are made. For example, comparison of the pre- and post-reformulation abuse rates for a product based on treatment center data may not be valid unless it can be assumed that the likelihood of an individual abusing this product interacting with treatment centers in the surveillance network (or in the case of PCC data an abuse or overdose event generating a call) would be the same in both time periods. Similarly, comparing abuse rates across comparators requires the assumption that the likelihood of being assessed for treatment (or in PCC data the likelihood of an abuse event generating a call) does not depend on the product being abused.

Geographic heterogeneity in abuse patterns can complicate the analysis and interpretation of the postmarket data, particularly in data sources that use nonprobability samples. When an estimate based on a coarse unit of analysis (e.g. national survey) shows a change in one direction while estimates based on a more granular unit of analysis (e.g., an analysis stratified by geographic region) show a change in a different direction, a paradox can result. A similar issue can arise in the context of multicenter clinical trials when patient treatment assignment varies from center to center (Rosenbaum, 2002). With treatment center data, for example, the number of individuals contributing data at the state level can decrease or increase simply by new treatment centers joining or existing treatment centers dropping out of the surveillance network over the course of the study period. This can lead to conflicting results between the direction of change between state-specific estimates and estimates aggregated over all states. Restricting the analysis to sites that contribute information throughout the study period may alleviate this issue to some extent, but not eliminate it. Note, for multicenter clinical trials, Rosenbaum discussed a standardization approach using direct adjustment that may be potentially useful. Additionally, it may not always be feasible or efficient to seek national

estimates. There may be value in approaches for valid estimation in some well-defined cohorts or subpopulations. We are interested in discussing these and other approaches, either in the design or analysis phase, to address these types of sampling issues that arise in postmarket abuse-deterrence studies.

4.6 CONFOUNDING AND SECULAR TRENDS

Many factors can affect trends in opioid prescribing and abuse, for example, changes in formularies and insurance coverage policies; provider education initiatives and clinical practice guidelines; increasing use of state-level Prescription Drug Monitoring Programs (PDMPs); large-scale initiatives to reduce rogue opioid prescribing and dispensing (e.g., *pill mill* crackdowns); and the availability of alternative drugs, including heroin. These and other larger community forces can be described as secular trends. Isolating the effect of a specific product's reformulation from the effects of secular trends is a challenging endeavor.

One approach is to compare the change in abuse metrics for the reformulated product from the pre- to the post-period with that of a comparator opioid product without properties designed to deter abuse. The change observed in the comparator product is hypothesized to reflect the effects of secular trends while that of the reformulated product reflects both the effects of reformulation and secular trends so that a comparison between two such products provides insights about the effect due to reformulation.

Unlike clinical trials or well-designed observational studies in which the comparator is similar to the treatment cohort by virtue of randomization or by some matching procedure, the most appropriate comparator product in postmarket observational studies of opioid abuse can be difficult to determine. Some of these difficulties stem from the fact that:

- No fixed -exposure cohorts (one exposed to the product in question and one exposed to the comparator product) are followed over time.
- Characteristics may differ between those who abuse the product in question and those who abuse the comparator product.
- Market characteristics of the product in question may be very different than those of the comparator product (e.g., how long it has been available, prescription volume, regional differences).
- Rate of abuse of the product in question may be trending in one direction before the reformulation, while that of the comparator may be trending in a different direction in the same period.

In addition, the effect of various components of secular trends may act in differential ways across opioid products. If that is the case, even if it were possible to identify an

appropriate comparator at baseline, the notion of fair comparison may still fail because of the differential influence of secular trends on different opioid products (e.g., local crackdowns or institutional or payer restrictions on specific prescription opioid products).

Because it is not possible to control for these factors, selection of an appropriate comparator is very challenging. Often, multiple comparators are used because no single comparator is sufficient. The use of multiple comparators can provide valuable information about the overall landscape of opioid prescribing and abuse trends; however, it greatly complicates hypothesis testing analyses and overall interpretation of findings. The use of composite comparators is another approach to this dilemma; however, these can be problematic for a variety of reasons. When multiple drugs are combined, the resulting pattern may be a reasonable representation of broad trends in a sector of the opioid market. However, composite comparator groups may include drugs that vary widely with respect to market share, length of time on the market, and trends in utilization and abuse. Therefore, the actual composition of the comparator group can change over time with respect to the relative contributions to abuse estimates. Adjusted abuse rates for composite categories may also be dominated by just one or two products with a large market share, masking large relative changes in other comparators.

We are interested in discussing approaches to control for confounding by secular trends, including the appropriate selection and use of comparators as well as other strategies.

4.7 STATISTICAL MODELS

Adding to the uncertainties about the most useful metrics for abuse and related outcomes is the question of how best to evaluate changes over time. One approach is to compare the prevalence or rate of abuse between a specified pre-reformulation period and a post-reformulation period, and then to compare the magnitude of this change to that for a comparator. However, when abuse metrics are trending up or down before the reformulation, these comparisons can be highly influenced by the duration and timing of the pre-period selected. Such analyses do not take into consideration such preexisting trends, either for the drug of interest or for comparators.

Another approach for assessing trends is by approximating the behavior of the abuse metrics over time using interrupted time series (ITS), which allows different intercepts and slopes for each period. Interpretation of results based on ITS models may not be straightforward, however. For example, it is not always clear which of the following quantities can be interpreted as the effect of reformulation:

- The difference in slopes between a trend line in the pre-period and a trend line in the post-period,

- The change in level at the time of the reformulation, or
- The difference between the product of interest and comparator(s) for each of these quantities.

A further complication is the question of how to perform these pre-post comparisons or ITS modeling. For pre-post comparisons as well as comparisons across products, two general approaches based on generalized linear model concepts have been suggested. One approach models prevalence in the log or logit scale as a function of a pre-post time period indicator, a drug product indicator, and their interaction; some function of utilization is included in the linear predictor as a way of adjusting for drug availability. A second approach directly models utilization-adjusted rates using Poisson regression, where the linear predictor is a function of a pre-post time period indicator, a drug product indicator, and their interaction; the log of the utilization enters the linear predictor as an offset. Similar approaches have been used for ITS modeling where in addition to describing time as that pertaining to a pre- and post-reformulation period of the product under consideration, it is also expressed on a finer scale, often quarterly.

In general, the underlying sampling process giving rise to the data enters the linear predictor as an offset of the sampling probabilities. In practice, these nuisance parameters are ignored because the selection probabilities are often unknown. Although this simplifies analyses, it is equivalent to incorrectly assuming that selection is random.

We are interested in discussing the best analytic methods to enable valid estimation without requiring knowledge of the selection probabilities and allow a reasonable expectation of causality in interpreting changes in abuse levels and trends over time that can be directly attributable to the abuse-deterrent properties of the drug product.

5. ORGANIZATION OF MEETING AND DISCUSSION TOPICS

As outlined in the previous sections, FDA is interested in discussing improving the quality of data and the methods currently being used in the evaluation of the real-world effect of formulations of prescription opioids properties designed to deter abuse to actually deter abuse, as well as on the appropriate interpretation of the complex findings these studies are likely to generate.

Day One

The first day of the meeting will focus on how existing data sources may be better used or enhanced to overcome the challenges and limitations that have been described here. The day will include brief presentations on the scientific issues that have emerged around postmarket evaluation of opioid formulations with properties designed to deter abuse and the current data sources, study designs, and methods used to evaluate these products in postapproval settings. Following discussion of the strengths and limitations of the current approaches, FDA is interested in discussing strategies to overcome some of the

identified challenges. In particular, FDA is interested in the following in the context of opioid formulations with properties designed to deter abuse:

- *Appropriate Outcomes*: The scientific value of the outcomes currently studied, how they are measured, and whether different or additional outcomes may be helpful in evaluating the impact of opioid formulations with properties designed to deter abuse.
- *Data Quality*: Strategies to improve the measurement of exposure and outcomes in these data sources, including instrument design and other approaches to reduce and/or adjust for misclassification.
- *Sampling*: Strategies to better understand, and possibly improve, the representativeness of the data sources currently used in evaluating opioid formulations with properties designed to deter abuse to produce valid national estimates or valid estimates in well-defined subpopulations, including understanding the catchment areas or leveraging of other data sources, such as national surveys, to create sample weights or in some other way reduce biases that can arise in dynamic convenience samples.
- *Metrics and Denominators*: Strategies to account for variation in prescription volume across products and over time—including discussion of various utilization denominators and coverage areas used for utilization adjustment—and how to synthesize and interpret results from prevalence (population-adjusted) and utilization-adjusted analyses.
- *Causality and Control for Confounding*: Strategies to account for secular trends and other interventions in the prescription opioid abuse area, including the use and interpretation of means and trend analyses (e.g., ITS), the use of comparators, and appropriate modeling approaches.

Day Two

The second day of the meeting will focus on what new data sources and study designs could be developed to enhance existing approaches. The day will include discussion of potential changes to existing national surveys, development of new surveys, the use of cohort studies, data linkages, and novel data sources and methods.

Postmarket studies designed to evaluate the impact of an opioid formulation with properties designed to deter abuse should enable an evaluation of whether the product results in meaningful reductions in abuse, misuse, and related adverse clinical outcomes, including addiction, overdose, and death in the postapproval setting. Although no single data source is expected to be sufficient, the most useful data systems for such studies would feature a number of characteristics:

- Identification of product-specific information, including brands, formulations, and routes of abuse.

- Ability to be rapidly modified, in response to changes in the opioid market, and to expand to evaluate potential product formulations with abuse-deterrent properties that are not opioids.
- Rigorous instrument and algorithm development and validation, with innovative methods to minimize misclassification and ensure that any misclassification occurring is nondifferential over time and across products.
- Use of probability sampling methods, enabling the generation of national and regional estimates for abuse of specific products that can be reliably trended over time.
- Ability to estimate risk of abuse and related outcomes in different populations (e.g., geographic regions, demographic groups, severity of substance use disorders).

FDA is interested in discussing the following in the context of opioid formulations with properties that are designed to deter abuse:

- Utility, strengths, and limitations of the existing national surveys in studying abuse deterrence.
- Approaches and feasibility to modifying existing national surveys versus developing new surveys, including internet panel surveys and other designs, to assess the impact of opioid formulations with properties designed to deter abuse on abuse and misuse, with attention to route of abuse.
- Novel study designs, including longitudinal cohort studies, and leveraging/linking multiple data sources such as PDMPs, electronic health care databases, and death registries, to evaluate the impact of opioid formulations with properties designed to deter abuse on clinical outcomes.
- Smaller local or regional studies and other types of information, for example, data from street ethnography or other qualitative studies, outbreak investigations, or small cohort studies.

6. REFERENCES

Bau G, Bucher-Bartelson B, Severtson SG, Green JL, Dart RC. Comparison of population rates between the Drug Abuse Warning Network (DAWN) and the RADARS System Poison Center Program. RADARS System Technical Report, 2016-Q4.

Butler SF, Cassidy TA, Chilcoat H, Black RA, et al. Abuse rates and routes of administration of reformulated extended-release oxycodone: Initial findings from a sentinel surveillance sample of individuals assessed for substance abuse treatment. *J Pain*, 2013;14(4):351-358.

Califf RM, Woodcock J, Ostroff S. A proactive response to opioid abuse. *NEJM*, 2016;374(15):1480-1485.

Cassidy TA, DasMahapatra P, Black RA, Wieman MS, Butler SF. Changes in prevalence of prescription opioid abuse after introduction of an abuse-deterrent opioid formulation. *Pain Med*, 2014;15(3):440-451.

Center for Behavioral Health Statistics and Quality. 2015 National Survey on Drug Use and Health: Methodological summary and definitions. Rockville (MD): Substance Abuse and Mental Health Services Administration, 2016.

Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: A large cohort study from a medication dispensing database in the US. *Drug Saf*, 2012;35(4):325-334.

Cepeda MS, Fife D, Vo L, Mastrogiovanni G, Yuan Y. Comparison of opioid doctor shopping for tapentadol and oxycodone: A cohort study. *J Pain*, 2013;14(2):158-164.

Coplan PM, Chilcoat HD, Butler SF, Sellers EM, et al. The effect of an abuse-deterrent opioid formulation (OxyContin) on opioid abuse-related outcomes in the postmarketing setting. *Clin Pharmacol Ther*, 2016;100(3):275-286.

Dal Pan GJ, Lindquist M, Gelperin K. Postmarketing spontaneous pharmacovigilance reporting systems. In: BL Strom, SE Kimmel, S Hennessy, eds. *Pharmacoepidemiology*, 5th ed. Oxford (UK): Wiley-Blackwell, 2012.

Dart RC, Surratt HL, Cicero TJ, Parrino MW, et al. Trends in opioid analgesic abuse and mortality in the United States. *NEJM*, 2015;372(3):241-248.

Dasgupta N, Kramer ED, Zalman MA, Carino S Jr, et al. Association between non-medical and prescriptive usage of opioids. *Drug Alcohol Depend*, 2006;82(2):135-142.

Davis JM, Severtson SG, Bucher-Bartelson B, Dart RC. Using poison center exposure calls to predict prescription opioid abuse and misuse-related emergency department visits. *Pharmacoepidemiol Drug Saf*, 2014;23(1):18–25.

Greenland S. Maximum-likelihood and closed-form estimators of epidemiologic measures under misclassification. *J Stat Plan Inference*, 2008;138(2):528–538.

Inciardi JA, Surratt HL, Kurtz SP, Cicero TJ. Mechanisms of prescription drug diversion among drug-involved club- and street-based populations. *Pain Med*, 2007; 8(2):171–183.

Inciardi JA, Surratt HL, Cicero TJ, Kurtz SP, et al. The “black box” of prescription drug diversion. *J Addict Dis*, 2009;28(4):332–347.

Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York (NY): Springer-Verlag, Springer Science & Business Media, 2009.

Lord SE, Trudeau KJ, Black RA, Lorin L, et al. CHAT: development and validation of a computer-delivered, self-report, substance use assessment for adolescents. *Subst Use Misuse*, 2011;46(6):781–794.

Lyles RH, Lin J. Sensitivity analysis for misclassification in logistic regression via likelihood methods and predictive value weighting. *Stat Med*, 2010;29(22):2297–2309.

Lyles RH, Tang L, Superak HM, King CC, et al. Validation data-based adjustments for outcome misclassification in logistic regression: An illustration. *Epidemiology* 2011;22(4):589–597.

McNaughton EC, Black RA, Zulueta SH, Budman SG, Butler SF. Measuring online endorsement of prescription opioids abuse: an integrative methodology. *Pharmacoepidemiol Drug Saf* 2012;21(10):1081–92.

Mowry JB, Spyker DA, Brooks DE, McMillan N, Schauben JL. 2014 Annual Report of the American Association of Poison Control Centers’ National Poison Data System (NPDS): 31st Annual Report. *Clin Toxicol (Phila)*, 2015;53(10):962–1147.

Mowry JB, Spyker DA, Brooks DE, Zimmerman A, Schauben JL. 2015 Annual Report of the American Association of Poison Control Centers’ National Poison Data System (NPDS):33rd Annual Report. *Clin Toxicol (Phila)*, 2016;54(10):924–1109.

Pearl J, Glymour M, Jewell NP. *Causal Inference in Statistics: A Primer*. West Sussex (UK): John Wiley & Sons, Ltd, 2016.

Rosenbaum PR. *Observational Studies*, 2nd ed. New York (NY): Springer Verlag, 2002.

Secora AM, Dormitzer CM, Staffa JA, Dal Pan GJ. Measures to quantify the abuse of prescription opioids: A review of data sources and metrics. *Pharmacoepidemiol Drug Saf*, 2014;23(12):1227-1237.

Secora A, Trinidad JP, Zhang R, Gill R, Dal Pan G. Drug availability adjustments in population-based studies of prescription opioid abuse. *Pharmacoepidemiol Drug Saf*, 2017;26(2):180-191.