#266

Use of Real-World Data and Real-World Evidence to Support Effectiveness of New Animal Drugs

Guidance for Industry

Submit comments on this guidance at any time. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number FDA-2020-D-1400.

For further information regarding this document, contact AskCVM@fda.hhs.gov.

Additional copies of this guidance document may be requested from the Policy and Regulations Staff (HFV-6), Center for Veterinary Medicine, Food and Drug Administration, 7500 Standish Place, Rockville MD 20855, and may be viewed on the Internet at https://www.fda.gov/animal-veterinary, https://www.fda.gov/regulatory-information/search-fda-guidance-documents, or https://www.regulations.gov.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Veterinary Medicine (CVM)
October 2021

Table of Contents

Contains Nonbinding Recommendations

Use of Real-World Data and Real-World Evidence to Support Effectiveness of New Animal Drugs

Guidance for Industry

This guidance represents the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. Introduction

FDA is issuing this Guidance for Industry (GFI), as required under section 305 of the Animal Drug and Animal Generic Drug User Fee Amendments of 2018 (Pub. L. 115-234), to assist sponsors in incorporating real-world evidence (including ongoing surveillance activities, observational studies, and registry data) into proposed clinical investigation protocols[1] and applications for new animal drugs under the Federal Food, Drug, and Cosmetic Act (FD&C Act). Section 305 of Pub. L. 115-234, among other things, directed FDA to hold a public meeting for interested parties to discuss innovative animal drug investigation designs and to issue guidance addressing the incorporation of the use of such elements of investigations as complex adaptive and other novel investigation designs, data from foreign countries, real-world evidence (including ongoing surveillance activities, observational studies, and registry data), biomarkers, and surrogate endpoints into clinical investigation protocols and applications to support the effectiveness of new animal drugs.

The contents of this document do not have the force and effect of law and are not meant to bind the public in any way, unless specifically incorporated into a contract. This document is intended only to provide clarity to the public regarding existing requirements under the law. FDA guidance documents, including this guidance, should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. Background

In the *Federal Register* of July 9, 2019 (84 FR 32749), FDA's Center for Veterinary Medicine (CVM) published a notice of a public meeting entitled "Incorporating Alternative Approaches in Clinical Investigations for New Animal Drugs" giving interested persons until August 17, 2019, to comment on the topics discussed at the public meeting and the questions published in the

[1] Submission of protocols is not required; however, it is recommended for any study intended to support the approval or conditional approval of a new animal drug.

meeting notice (84 FR at 32750-32751).[2] On August 13, 2019, we published a notice announcing the extension of the comment period to September 16, 2019 (84 FR 40071). CVM received numerous comments on the topics discussed at the public meeting and the questions published in the meeting notice and those comments were considered as draft guidance was developed.

This document describes recommendations for designing, conducting, and reporting the results for investigations or studies for new animal drugs including real-world evidence (including ongoing surveillance activities, observational studies, and registry data), and also incorporating real-world data in protocols for these investigations or studies, to demonstrate substantial evidence of effectiveness or a reasonable expectation of effectiveness of drugs intended for use in animals and to support the approval of a new animal drug application (NADA) or an application for conditional approval of a new animal drug (CNADA). Other centers within FDA, including the Center for Drug Evaluation and Research (CDER), the Center for Biologics Evaluation and Research (CBER), and the Center for Devices and Radiological Health (CDRH), have released draft and final guidance documents on the topics of the use of electronic health records and healthcare data in studies, submitting documents containing real-world evidence to FDA, and/or the use of real-world evidence in regulatory decision making.[3]

CVM will consider all established and accepted methodologies associated with real-world data and real-world evidence in submissions to investigational new animal drug (INAD) files, NADAs, and CNADAs to demonstrate substantial evidence of effectiveness or a reasonable expectation of effectiveness.[4] This guidance document provides CVM's recommendations specific to investigations for animal drugs.

Some concepts and language in the recommendations for animal drugs are intended to be similar or the same as those in other guidance documents issued by FDA on the same or similar topics. Because these recommendations are specific to investigations for animal drugs, they have been tailored to the unique aspects of and considerations for animal drug development.

III. Scope

For the purposes of this guidance, CVM defines real-world data and real-world evidence as follows:

Real-World Data (RWD) are data routinely collected from a variety of sources relating to the health and productivity of animals, the delivery of veterinary care, or the management of livestock/animals for food. For the purposes of this guidance, the term animal(s) could refer to an individual animal or a herd/flock/tank/group, depending on the context in which data are collected.

---

[2] https://www.fda.gov/animal-veterinary/workshops-conferences-meetings/public-meeting-incorporating-alternative-approaches-clinical-investigations-new-animal-drugs

[3] https://www.fda.gov/regulatory-information/search-fda-guidance-documents

[4] 21 CFR 514.4; 21 U.S.C. 360ccc(a)(2)B)

Examples of RWD applicable to new animal drugs include:

- Data derived from health records of veterinary practices, farms, or livestock management companies (including handwritten paper records and **electronic veterinary medical records**[5] (EVMRs);
- Data from product and disease **registries**;
- Data from other sources that can inform on animal health status such as mobile and/or remote health sensing devices for animals;
- Data generated by animal owners;
- Data from diagnostic laboratory, slaughterhouse, and abattoir records;
- Companion animal and livestock insurance claims; and
- Data from surveillance programs.

Real-World Evidence (RWE) is the clinical evidence of the effectiveness of a new animal drug derived from analysis of RWD.

Because RWD are collected as part of the routine care and management of animals, including their health and/or productivity, these data may be useful to support effectiveness of a drug in a diverse population of animals and conditions of use.

This guidance discusses the following topics related to the potential use of various RWD sources to generate RWE to support regulatory decisions related to effectiveness or reasonable expectation of effectiveness including:

- RWD relevance: selection of one or more data sources that appropriately address the study question and sufficiently capture representative study populations, exposure (as used in this guidance the new animal drug), outcomes of interest, and key **covariates**;
- RWD reliability; and
- Design and analysis of studies utilizing RWD to generate RWE.

The purpose of this guidance is to describe how CVM intends to evaluate RWE to determine whether it is sufficient to provide substantial evidence of effectiveness for new animal drugs and reasonable expectation of effectiveness for conditionally approved new animal drugs. In addition, this guidance describes how sponsors may obtain feedback from CVM on technical issues related to the use of real-world data and real-world evidence before the submission of an application.

This guidance describes the circumstances under which real-world evidence may be used to support a variety of FDA decisions based on the existing evidentiary standards applicable to

[5] Words and phrases in bold throughout this guidance are defined in section VII. *Glossary*.

FDA's regulatory decision making. This guidance highlights some of the potential uses of real-world evidence and describes the factors that CVM may consider when evaluating whether a specific source of real-world data is relevant and of sufficient quality to inform or support a regulatory decision relative to the demonstration of substantial evidence of effectiveness or reasonable expectation of effectiveness for new animal drugs.

CVM encourages sponsors to explore a variety of design options in planning clinical effectiveness study designs and to proactively discuss their considerations with CVM. Some recommendations in this guidance may be technically involved and we recommend you consult with the appropriate experts to help facilitate use of these principles.

The use of RWD/RWE to support technical sections other than Effectiveness for NADAs or Reasonable Expectation of Effectiveness for CNADAs is outside the scope of this guidance. This guidance also does not address the use of nonclinical laboratory data or systematic reviews of published literature. Although published literature itself is not a common source of RWD, some approaches developed for assessing and analyzing information in the scientific literature may also be applicable to aggregating RWD across multiple sources.

Finally, while this guidance describes the factors that sponsors should consider when evaluating the potential use of RWD or RWE, it does not provide a specific set of pass/fail criteria or other scoring tools for determining the suitability of RWD or RWE for a particular regulatory decision. This guidance does not elaborate on specific study designs or statistical analyses because of the rapidly advancing methodology for generating and interpreting RWD. When reviewing the use of RWE to support a regulatory decision, CVM may rely on scientifically robust methods and approaches to determine whether submitted RWE is of sufficient quality to support a particular regulatory decision. For all study designs, it is important to ensure the relevance and quality of the data used to support regulatory decisions pertaining to the effectiveness of new animal drugs.

IV. Real-World Data (RWD)

A. Sources of RWD in Veterinary Medicine

FDA does not endorse one RWD source over another or seek to limit the possible sources of data that may be relevant to answering regulatory questions pertaining to the effectiveness of new animal drugs. In veterinary medicine, both **unstructured RWD** and **structured RWD** are available and may be collected to generate RWE.

Although many animal health records exist as unstructured data in paper form (e.g., veterinarian's notes in patient records), veterinary practices increasingly are capturing both unstructured and structured animal health data in EVMRs. As the use of EVMRs expands, there may be more opportunities to use RWD collected on a specific drug or condition across many animals from multiple regions in addition to RWD on the health of individual animals that receive care from different veterinary practices or production facilities. Most large and medium scale food animal operations already use electronic record systems to maintain records of animal health and productivity. Smaller food animal operations and producers of minor/niche species are progressively transitioning

from paper to electronic records. In many cases, the records of the livestock producers are linked with records maintained by their veterinarians.

Diagnostic laboratory data, data from remote monitoring devices, and data from slaughterhouses provide additional RWD which may be used alone or linked to animal health and productivity records.

Diagnostic laboratory data include, but are not limited to, information from reference laboratories used by veterinary practices such as hematology, blood chemistry, and histopathology reports, and information from Federal, State, or privately-owned diagnostic laboratories. For example, clinical pathology and/or histopathology data may be linked to an individual animal's EVMR to provide critical diagnostic or treatment outcome information, or diagnostic laboratories may provide coordinated domestic and international disease surveillance information which could be useful to the design of an effectiveness study.

Remote monitoring devices are used to track both companion animal health and livestock health and productivity and may be a source of RWD. Devices to monitor blood glucose, activity and behavior, and other physiologic parameters are able to transmit information directly to veterinarians or researchers (Atanasov et al., 1996; Belda et al., 2018; Benjamin and Yik, 2019; Griffies et al., 2018; Lahdenoja et al., 2019; Theurer et al., 2013; Yamazaki et al., 2020). The information may be incorporated into the animals' health records or may be available for linkage with these records at a later time. Numerous other devices are now available that collect data, such as rumination rate in cattle, body temperature, feeding behaviors, measures associated with specific diseases (e.g., ketosis in cattle), and a variety of parameters associated with milking in dairy cows. For example, robotic milking systems are available which collect information from dairy cows on milk production, feed intake, and milk quality indicators including the electrical conductance of milk, which may identify cows with mastitis (King and DeVries, 2018).

Slaughterhouses are another potential source of RWD for livestock, as they typically collect data and provide reports to client livestock operations. These reports include information on animal/organ condemnation, live weight, carcass weight, dressing percentage, and other variables of interest to producers and managers.

Disease, product, and breed-specific health registries may also serve as a source of RWD for new animal drug development, although available data may be sparse in some registries. The specificity of some registries may also impact the generalizability of the information obtained. Nonetheless, some registries may serve as a source of cases or assist in establishing the natural course of a disease or typical standard of care. This information may be comprehensive enough to serve as a historical control in certain situations. Examples of existing registries in veterinary medicine include, but are not limited to, the Veterinary Committee on Trauma (VetCOT) registry (Hall, 2019), Orthopedic Foundation of America's (OFA) Canine Health Information Center (CHIC)[6]

[6] https://www.ofa.org/about/chic-program

and Canine Eye Registry Foundation (CERF),[7] the Cat Phenotype and Health Information Registry (PHIR),[8] and a lifetime health study known as the Golden Retriever Lifetime Study (Guy et al., 2015).[9] Some breed clubs (e.g., for dogs and cats) also maintain registries for health conditions of concern to the breed.

The comprehensiveness and quality of submitted and recorded information likely vary significantly among registries. We recommend referring to the Agency for Healthcare Research and Quality (AHRQ) publication, Registries for Evaluating Patient Outcomes: A User's Guide (Gliklich et al., 2020).[10] The User's Guide contains practical information to guide the design, operation, and analysis of patient registries and provides a framework for evaluating the quality of patient registries and evidence derived from patient registries.

Finally, insurance claim and billing data may also be a useful source of RWD. However, until health insurance becomes more common in veterinary medicine, this option may not be a good source of RWD.

B. General Considerations Regarding RWD Sources

Because RWD sources are generally not developed for the purposes of regulatory decision making, they often have several challenges, such as a lack of standardized terminology and/or completeness of records, concerns regarding data sharing/privacy, and difficulties with the **interoperability** of systems. Unlike the incentives provided by the Federal Government to promote the continued development of electronic health care systems used in human health care, at this time there is no Federal mandate or financial incentive for veterinary practices to invest in such systems. CVM encourages sponsors to consider these challenges and engage in early discussions with CVM to develop methods to overcome or work within the constraints that these challenges present in generating relevant and reliable RWE.

Within animal health records in particular, CVM recognizes that there is currently a lack of standardization in how data are entered, how conditions are characterized, and how outcomes are documented. Variability in animal health record documentation may exist across practices or even among individual veterinarians within a practice. In addition, CVM understands that animal health records can vary by veterinary practice, producer, State, and type/species of animal. With increasing corporate ownership of veterinary practices and multiple veterinary practices obtaining electronic records systems from a limited number of providers, efforts to standardize terminology in records systems, particularly EVMRs, are underway, but are not yet complete (Alpi et al., 2011; Santamaria and Zimmerman, 2011). Some examples of systems used or being evaluated for use in veterinary health records include Systematized Nomenclature of Medicine

[7] https://www.ofa.org/diseases/eye-certification

[8] https://www.vgl.ucdavis.edu/lyons/catphir.php

[9] https://www.morrisanimalfoundation.org/golden-retriever-lifetime-study

[10] https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/users-guide

Clinical Terms (SNOMED CT),[11] the American Animal Hospital Association (AAHA) Problem and Diagnostic Terms,[12] Health Level 7 (HL7),[13] Logical Observation Identifier Names and Codes (LOINC),[14] and Digital Imaging and Communications in Medicine (DICOM).[15] Completeness of case histories may also be a challenge, as owners may not report the outcome of treatment or may seek care at more than one veterinary clinic, and practices have limited time and resources for follow-up. In contrast, the Zoological Information Management System used by zoos and aquaria for husbandry and medical records allows for the sharing of records when animals move among institutions.

With regard to privacy, although veterinary practice records are not covered by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), State laws may prohibit the sharing of veterinary records, or veterinarians may be reluctant to share data without their clients' consent. In addition, herd health data from commercial animal production facilities such as farms, feed yards, and hatcheries and data from slaughterhouses are often proprietary and owners of these operations may also be reluctant to share data. CVM encourages sponsors to work with veterinarians and livestock producers to address their concerns regarding protecting the privacy of data.

C. Determination of the Suitability of RWD for Regulatory Use

To determine the suitability of RWD for evaluating effectiveness of a new animal drug, it is recommended that sponsors assess the relevance and reliability of the RWD and its specific elements. CVM is adapting RWD relevance and reliability considerations developed for human data sources to animal data sources (Daniel et al., 2018; Girman et al., 2019; Miksad and Abernethy, 2018).[16,17,18,19] Not all considerations will be appropriate for a particular RWD source, and CVM encourages sponsors to discuss any unique issues or challenges encountered relating to the data relevance and reliability assessment with the relevant review division(s).

1. Relevance to the regulatory question

Relevance refers to whether the information captured by the RWD source(s) is adequate to address the regulatory question or requirement relative to the

[11] http://www.snomed.org/

[12] https://www.aaha.org/practice-resources/running-your-practice/diagnostic-terms/

[13] https://www.hl7.org/

[14] https://loinc.org/

[15] https://www.dicomstandard.org/

[16] See GFI, "Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data," (May 2013)

[17] See GFI, "Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices," (August 2017)

[18] Framework for FDA's Real-World Evidence Program

[19] See GFI, "Use of Electronic Health Record Data in Clinical Investigations," (July 2018)

effectiveness evaluation for a new animal drug. Depending on the proposed conditions of use for the new animal drug, there may be differences in the practice of veterinary medicine and animal production around the world and between veterinary practices or production facilities that may affect the relevance of the data source to the study question. Questions about the applicability of an RWD source within a specific drug development plan should be discussed with CVM.

Sponsors should consider the following factors to determine if the RWD are suitable for regulatory use (as appropriate):

a. Background information about the data source, including if there are standardized methods of disease diagnosis; procedures for prescribing or procuring veterinary drugs for therapeutic and/or production uses, including for approved indications, formulations, and doses; preferred treatments for the disease or indication of interest; standard methods to measure production variables, as appropriate; and the degree to which such information is collected in the proposed data source. Background information may also include prior documented (e.g., peer reviewed publications or practice guidelines) use of the RWD source.

b. Whether the animals in the RWD source are representative of the intended target animal/class.

   Because animals differ in a range of characteristics, such as age, breed, species/class, health conditions, risk factors, and other potential covariates, it is important to identify whether the data sources cover all populations relevant to the study if those sources are to be used to examine the study hypothesis.

   For foreign RWD sources, sponsors should consider any factors that could affect the generalizability of study results to the U.S. target animal population and conditions of use.[20]

c. Whether the RWD contain sufficient detail and completeness to capture the critical **data elements** related to exposures, key covariates, the outcomes of interest in the appropriate target animal population, and any other important factors (e.g., inclusion/exclusion criteria, timing of exposure, timing of outcome, etc.) that are relevant to the study question (hypothesis) and design.

   As part of this consideration, sponsors should evaluate whether the RWD source contains necessary elements to capture specific drug formulation information (e.g., proprietary name, manufacturer, lot and/or batch numbers, etc.).

[20] See CVM GFI #265, "Use of Data from Foreign Investigational Studies to Support Effectiveness of New Animal Drugs," (October 2021)

In addition, if animal health records are used as a source of RWD, sponsors should consider the continuity of care: whether an animal receives all or only a portion of care at a particular practice (i.e., use of multiple primary care practices, or referral or emergency practice). If animal health records from a veterinary referral or emergency practice are used without linkage to records from the animal's primary veterinarian, the amount of preventative care, comorbidity, and concomitant medication data available in the RWD source may be limited. The typical prescribing practices for prescription and over-the-counter (OTC) drugs should also be considered when using animal health records as a source of RWD. Many animal health records do not systematically capture the use of OTC drugs, supplements, online pharmacy prescription purchases, or some routine immunizations that may be administered by a separate vaccine clinic. If these exposures are relevant to the study question, the data source may not be suitable, or the protocol should describe how this information gap will be addressed. Other aspects to consider include whether the RWD source adequately captures animal health history and preexisting conditions, as well as follow-up information needed to evaluate the question being addressed.

The RWD source should also be evaluated to determine if sufficient data elements are collected to adjust for any confounding factors that may impact the exposure or outcomes of interest. These include factors that are well-captured in the proposed data source (measured confounders) and those that are not well captured (unmeasured or imperfectly measured confounders). Examples of confounders that are unmeasured or imperfectly measured in many animal health record data sources include management factors (e.g., diet or physical activity), certain physical measurements (e.g., body condition), diagnostic laboratory test results, concurrent drug administration by animal owners such as OTC drugs and supplements, and prescription drugs obtained from online pharmacies.

d. Whether the data source contains an adequate number of animals that represent the target animal/class with adequate length of follow-up to ascertain outcomes of interest based on the biologically plausible timeframe when the outcome, if associated with the exposure to a particular new animal drug, might be expected to occur.

   Information should be provided about the distribution of length of follow-up for animals within the data sources, because the length of follow-up may dictate whether the selected data sources are appropriate or whether additional supportive data are needed to evaluate outcomes that require long follow-up periods.

e. Whether supplemental data sources are available if needed.

   When critical information in a data source is absent, the data source may not be sufficient to achieve the study objectives. In this case, using alternative

data sources, linking multiple data sources, or prospectively collecting additional information may be necessary. The linking of RWD sources should be scientifically appropriate and account for differences in coding and reporting across sources.

2. RWD reliability

The reliability of RWD is determined by the quality and integrity of the data and the source from which it is derived. In general, high quality data, regardless of whether it is collected manually or with electronic systems, are attributable, legible, contemporaneous, original, and accurate. FDA's acceptance of the final study-specific dataset for regulatory decision-making purposes depends on FDA's ability to verify the quality and integrity of the data. The threshold for sufficient quality will depend on the specific regulatory use of the evidence. For example, a specific RWD source might be leveraged for reasonable expectation of effectiveness, but not be adequate to support a determination of substantial evidence of effectiveness.

In the context of RWD, both the quality of the data source and the final dataset are considered. However, CVM understands that in most cases the RWD source itself (e.g., animal health records including EVMRs) belongs to veterinary clinics, producers, or livestock management organizations and is not under the control of the sponsor or investigator. The majority of the data quality considerations discussed are therefore focused on the collection of the RWD by the sponsor using paper or electronic systems including electronic data capture (EDC) systems.

For RWD, data integrity characteristics are encompassed by the following attributes: **accuracy**, **completeness**, **provenance**, and **traceability**/transparency of data processing.

The accuracy of RWD is established by assessing (1) the validity of the data elements in the RWD source and any algorithms used to transform the data (e.g., data elements are appropriately defined using accepted clinical criteria); (2) the code-based algorithms used to extract the data from the RWD source; (3) the plausibility of the data (e.g., a laboratory result is biologically possible); (4) data consistency for each animal (e.g., correlation of laboratory results and unstructured data from EVMR and trends of laboratory data over time); and (5) **conformance** of the data to any applicable standards.

A standardized process of data extraction from the RWD source, using qualified and trained personnel, should be used to ensure the accuracy of the data. A standard and reproducible process is critical for minimizing intra- and inter-observer variability, especially when RWD are extracted from multiple sites or multiple systems.

FDA is aware that advances in the evolving field of **machine learning** and **artificial intelligence** may permit the processing of unstructured health care data (Shah et al., 2019), and systems are being explored for veterinary medicine (Nie et al., 2018). If AI or other derivation methods are used to extract unstructured text from an RWD

source, the protocol should specify the assumptions and parameters of the business rules and algorithms used, the data source from which the information was used to build the algorithm, whether the algorithm was supervised (i.e., using input and review by experts) or unsupervised, and any validation and metrics associated with it.

Completeness of RWD refers to the amount and type of missing data. There are two broad situations that result in information being missing from an RWD source. The first case is when the information was intended to be collected but is absent from the data sources. The second case is when the information was not intended to be collected in the RWD source (e.g., EVMR) and is therefore absent. When critical information of a data source is absent, the data source may not be sufficient to achieve the study objectives.

Although FDA does not endorse any particular set of guidelines or checklists, researchers should evaluate the completeness and accuracy of the data, including verifying data against their original source (e.g., veterinary clinician notes, pathology reports, registry records, etc.) and consensus-based data standards (e.g., Veterinary Extension of SNOMED CT), where applicable. Researchers should provide scientific justifications for the selected standards and should articulate how the selected standards are sufficient to ensure the completeness and accuracy of the relevant data source. The acceptability of various degrees of accuracy and completeness depends on the specific research question and regulatory purpose.

The accurate and complete collection of RWD can be enhanced when the RWD source has an operational manual or other documentation, such as standard operating procedures, that pre-specifies the data elements to be collected, data element definitions (i.e., data dictionary to provide a common definitional framework), methods for data aggregation and documentation (e.g., common case report form, abstraction from verifiable sources, etc.), and the relevant timeframes for data element collection (i.e., common temporal framework). In many cases, RWD sources available in veterinary medicine will not have operational manuals or standard operating procedures; however, these sources may still have sufficient quality and integrity to support regulatory decision making.

Provenance and transparency of data processing are established through the use of an appropriate **quality control** (QC) and **quality assurance** (QA) plan throughout the data life cycle, including but not limited to **data accrual** (source data), **curation** of the RWD into an electronic data set such as a **clinical data repository**, **transformation** of data (including the processing of data into a **common data model** (CDM) and creation of a **data warehouse**), creation of a study specific dataset, archiving, and disposition.

In human medicine, RWD may be accrued, curated, and transformed by outside organizations to create "research-ready" RWD (Daniel et al., 2018). At this time, such datasets are in the early stages of development in veterinary medicine (Kwong et al., 2019). For example, the Veterinary Companion Animal Surveillance System

(VetCompass™)[21] is collecting real-world data routinely recorded by United Kingdom (UK) veterinary practices in a de-identified format and merging the data into a single, searchable dataset. A similar approach is being used for horse health data in the UK (Equine VetCompass)[22] and companion animal health data in Australia (VetCompass Australia).[23] At this time, CVM anticipates that sponsors will handle the RWD throughout the data life cycle.

Regardless of whether single or multiple sources of data are used to create the study specific dataset, sponsors should ensure that the data management procedures used for curation and transformation do not alter the meaning of data or lose important contextual information. The procedures should include safeguards or checks to ensure that animal data are not duplicated or overrepresented.

At a minimum, the data curation and transformation QC/QA plan should include:

- A description of the processes used for data curation and transformation including procedures for cleaning, linking, and transforming, as appropriate to the dataset;
- General procedures used by study personnel to ensure completeness, consistency, and accuracy of data accrual, curation, transformation, and management;
- Any changes allowed in key data elements and their potential effect on the study;
- The frequency and type of any data error corrections or changes in data processes;
- Any updates and changes (version control) in electronic data repositories or data warehouse systems that are relevant to the outcomes of interest;
- The checks in place to assure the steps taken to verify accuracy and handle errors identified during curation and transformation; and
- Procedures on how to assess and handle missing and uninterpretable data.

V. Real-World Evidence (RWE)

A. General Considerations

CVM may consider the use of RWE to support regulatory decision making for a new animal drug when the RWD are relevant and of sufficient quality, and valid study designs are used to generate the RWE as appropriate to the particular regulatory decision.

Some purposes for which RWE may potentially be used include the following:

[21] https://www.vetcompass.org/

[22] https://www.rvc.ac.uk/vetcompass/projects/vetcompass-equine

[23] https://www.vetcompass.com.au/

- To characterize a dosage regimen and define conditions of use;
- To develop historical, external, or concurrent control;
- To justify sample sizes, define appropriate enrollment criteria, and provide the basis for the evaluation of endpoints in a clinical study;
- To provide reasonable expectation of effectiveness for a CNADA; or
- To provide substantial evidence of effectiveness for an original or supplemental NADA.

Studies using RWD to generate RWE for effectiveness may provide adequate evidence alone or may be combined with other studies, including traditional clinical studies, to provide substantial evidence of effectiveness for a particular indication for use.

Traditional clinical studies conducted outside of routine animal care or management (typically using separate or otherwise controlled study sites) are designed to control variability through detailed eligibility criteria and carefully designed clinical protocols implemented by specially trained and qualified research personnel. By contrast, studies leveraging RWD to generate RWE utilize routine animal care and management environments, typically have wider eligibility criteria and a larger number of animals, and can potentially provide information on a broader animal population and conditions of use than a traditional clinical study. In addition, when using RWD, the effectiveness of a drug may more likely be evaluated under conditions of use closer to or identical to the conditions of use expected following approval.

B. Types of Studies to Generate RWE for Regulatory Decision Making

A variety of study designs can be used to generate RWE, including **interventional studies** (clinical studies) and **observational studies** (observational studies**)**. The study question of interest (the study objective) should be established first, then the data source and study design most appropriate to address this question should be determined. The study objective should not be tailored to fit a specific data source because limitations of an RWD source may restrict the options for study design and may limit the inferences that can be drawn.

Interventional study designs used to generate RWE potentially include clinical studies with pragmatic design features (pragmatic clinical studies), clinical studies using a hybrid design, or single-arm studies using an external or historical control. Pragmatic clinical studies allow for prospective data collection within the real-world environment (e.g., veterinary clinic, registries, etc.) and may have wider inclusion/exclusion criteria than a traditional clinical study, but still utilize randomization and concurrent controls such as an active control. Hybrid clinical studies include elements of both traditional randomized controlled studies and pragmatic clinical studies and may allow for some data to be collected from RWD sources.

There are many types of observational study designs and only certain designs produce measures of association that approach establishment of a causal relationship between the exposure of interest, the investigational drug, to the outcome of interest, and the observation used as a measure of effectiveness. An assessment of causality is necessary to establish that the new animal drug has the intended effect as proposed in the labeling. Observational study designs most likely to be useful to generate RWE for new animal drugs include the cohort and case-control study designs. Cohort and case-control studies could be particularly suitable to demonstrate drug effects such as a decrease in disease episodes over several weeks or months; improvements in production characteristics; or improvement in clinical signs for a condition that requires months to properly adjust medication levels. These study designs may also be useful to evaluate the effectiveness of drugs that are widely used by veterinarians, and not yet approved for animals, including conditionally approved drugs. The observational study designs relying on RWD may also be better suited to studies including active comparators, particularly for conditions where veterinarians or producers will select some type of intervention rather than no intervention, euthanasia, or slaughter. Situations where veterinary clinicians or producers have no good basis for choosing a particular treatment (also referred to as clinical equipoise) may allow for a more efficient identification of an adequate number of cases within a database of RWD.

Observational studies can be conducted prospectively or retrospectively, both of which come with their own strengths, limitations, and ability to control treatment exposure and important covariates. Whether a study is prospective or retrospective depends on when the exposure and outcomes occur in relation to the time at which the study protocol is initiated. Sponsors should consider the use of analytical techniques that seek to control for differences in baseline characteristics (i.e., covariates) between treatment groups. In a prospective study, the exposure and the outcome of interest has not yet occurred when the study begins. Studies begin with groups identified as having received the exposure of interest (the investigational drug) or an appropriate control (comparator group). Those groups are then followed for the outcome of interest. With a prospective study, there may be some ability to direct the collection of a measure of interest, which may be used to compare the study groups to each other or the target population or to allow a better assessment of drug effect. In a retrospective study, both the exposure and outcome have occurred when the study begins, and the study relies on previously recorded information. A study could be a combination design where the exposure is evaluated in existing records and follow up and measurement of the outcome occurs in the future for some or all patients. The information recorded and the quality of the information recorded impact what can be measured from these studies. While there may be a desire to add to the available information, any information obtained via follow up contacts would be prone to significant bias (e.g., recall bias).

The increasing use of EVMRs in both companion and food animal medicine to capture patient information, diagnostic test results (including those from outside facilities), and prescribing and treatment information may provide greater opportunities to design studies to generate RWE using RWD in prospective or retrospective observational study designs. This also includes studies where information from remote monitoring devices may be of interest.

Although observational studies (both prospective and retrospective) are prone to bias and confounding, new study designs and analytic techniques that include matching, stratification, weighting, and/or multivariate regression models have made progress in managing such limitations. These methods may be able to address some of the potential biases inherent in non-randomized studies and potentially increase the applicability of these studies in the regulatory context. Understanding the impact that certain data limitations may have on the ability to make inferences will also be integral to selecting appropriate analytical techniques.

C. Study Design Considerations for Substantial Evidence of Effectiveness

Studies generating RWE from RWD which are intended to demonstrate substantial evidence of effectiveness of a new animal drug for an original or supplemental NADA should be adequate and well-controlled (21 CFR 514.117).[24] A study protocol and analysis plan should be created before accessing, retrieving, and analyzing RWD, regardless of whether the RWD are already collected or if they are to be collected in the future. Protocols and analysis plans for studies using RWD should address the same elements as those used for a traditional randomized controlled clinical study. The protocol for a study generating RWE should provide the following details relative to the study design, conduct, and analysis in accordance with the characteristics of an adequate and well-controlled study[25] in order to meet regulatory requirements for substantial evidence of effectiveness:

- A clear statement of the study objective or question of interest;
- A description of an acceptable standard of conduct[26] (e.g., Good Clinical Practice[27]);
- Descriptions of training for study personnel relative to data accrual, curation, and transformation;
- Standardized definitions (i.e., data dictionary if possible) for data elements (e.g., inclusion/exclusion criteria for study population, exposure, outcomes, covariates, and potential confounding factors), and results of validation studies, as appropriate; and
- Description(s) of RWD source(s) that will be used and a justification for the relevance and quality of the data source(s) in accordance with the study objectives.

[24] 21 CFR 514.117 Adequate and well-controlled studies

[25] 21 CFR 514.117(b) Characteristics.

[26] In this case, the standard of conduct applies to the protocol for collecting and analyzing the RWD to generate RWE, not necessarily to the RWD itself.

[27] See CVM GFI #85 (VICH GL9), "Good Clinical Practice," (May 2001)

The RWD source description should include historical experience with and use of the selected data source for research purposes, including how well the selected data source has been shown to capture key study elements (e.g., inclusion and exclusion criteria, exposures, outcomes, key covariates, etc.), and how the accuracy of the RWD will be assessed (see section IV.C.2. *RWD reliability*).

If RWD sources will be linked, the protocol should describe how the data, including those with heterogenous data structures, will be obtained and integrated with acceptable quality. This description should include the procedures used to assess and address linkage quality including quantification of errors (e.g., false-matches and/or missed-matches) that may lead to biased study findings (Harron et al., 2017) and plans to handle linkage discrepancies. The protocol should also specify the details of the linkage algorithm and appropriate metrics (e.g., linkage error rates, match rates, and comparison of characteristics of linked and unlinked data).

If data are from multiple RWD sources, the protocol should describe how data will be harmonized across the sources or systems.

- Time periods (the RWD source reporting schedule, including time interval between database close and release, and length of reporting periods);

- Descriptions of the procedures for unstructured and structured data processing including the operational definitions and techniques employed (e.g., manual review or automated techniques) to abstract unstructured data (e.g., clinician notes) and structured data (e.g., diagnostic laboratory data). These procedures should include data standardization procedures (e.g., data types, sizes, and formats) for consistency of data elements and semantics, including semantics of local codes to a target terminology (e.g., for laboratory data);

- A description/justification of the sample size necessary to measure the effect with sufficient statistical power, including the methods used to calculate the sample size;

- Identification of possible sources of bias and confounding and plans to address them;

- A description of data management and data QA plans, including plans for site and data monitoring, availability of standard operating procedures, algorithms used to transform and clean the data, and procedures for source data verification. The protocol should specify the curation and transformation procedures used throughout the life cycle of the data and describe how these procedures could affect data integrity and the overall validity of the study. Procedures for adequate documentation (e.g., **audit trail**) of the creation of the final dataset should be described to verify proper handling of data;

- Information about the intended use of any computerized systems used to curate or transform real-world source data, a description of the security measures employed to protect the data, and a description or diagram of the electronic data flow. In

addition, a description of how the data integrity is maintained when information is transferred to a format different from that originally collected;

- Processes used for managing and preparing the final analytic dataset. This includes any study-specific data transformations performed on a real-world dataset that are applied to a subset of animal health records within a larger database (e.g., manual extraction of data from unstructured clinical pathology reports for study-specific patients). For all programs that will be used in the study (e.g., written by the analysts), the protocol should describe the intent or purpose of each data management and analysis step written in the program (e.g., annotate each data step in a statistical analysis program);

- A description of data checks implemented on the final analytic dataset for implausible values for data elements (e.g., body weight or age), completeness of data for key analytic variables, and/or the extent, percentage, and pattern of missingness and implausible data. Assumptions regarding the information content underlying the statistical analysis for study endpoints and important missing data (e.g., missing at random or missing not at random) should be supported and the implications of missing data considered. Procedures to mitigate missing data should be described if needed; and

- The statistical analysis plan including the criteria to determine the effectiveness of the drug (basis of study conclusion). The protocol should describe the methodology used to analyze RWD and assess clinically relevant differences as well as statistical significance.

The following provides additional detail for certain concepts specific to adequate and well-controlled effectiveness studies based on RWD/RWE.

Study objective or question of interest:

The study question (study objective) should be structured to include the following elements: population (target animal species/class); intervention (new animal drug and conditions of use as appropriate); comparator (control); and outcome (endpoint specific to intended use) (Girman et al., 2019).

Appropriate standard of conduct

For studies designed to use RWD to generate RWE that are intended to provide substantial evidence of effectiveness, CVM recommends that studies be conducted in accordance with GFI #85 (VICH GL9), "Good Clinical Practice," (May 2001).[28] GFI #85 describes the responsibilities of the investigator, sponsor, and monitor; the contents of a study protocol; the contents of the final study report; and the handling of study documentation. The concepts presented in GFI #85 can be adapted and applied to study designs such as observational studies. For example, even if RWD are evaluated

[28] https://www.fda.gov/media/70333/download

retrospectively, the investigator should ensure that the study to generate RWE is executed in accordance with the protocol. In these situations, while the investigator may not have oversight of the original data collection by the veterinarian or producer, they are responsible for ensuring the process of RWD curation, transformation, and analysis is implemented in accordance with the protocol and documented in accordance with established data quality and integrity principles.

Any gaps in the quality standards followed should be identified and explained in the final study report. This explanation should include a description of the steps taken to mitigate the gaps and the impact on the outcome of the study results, if any.

New animal drug formulation

For the purposes of this guidance, the terms exposure or intervention refer to the product and regimen of interest being evaluated in the proposed study. Sponsors must be able to demonstrate an ability to identify the specific products of interest in the proposed data source.

Data about exposure should include information about the dosage. Depending on the exposure and the question of interest in the study, it may be useful to describe the dose of each administration or a daily dose and duration of exposure.

The most useful information to support effectiveness of a new animal drug employs the use of the final intended formulation of that drug (i.e., the formulation the sponsor intends to market). Where possible, RWD which are intended to be used to support effectiveness of a new animal drug should be derived from the use of the final drug formulation. If data using the final formulation are not available, or if the sponsor wishes CVM to consider data generated using other formulations, additional information may be needed to determine whether data using other formulations are applicable to support the effectiveness of the final formulation.

Study population/animals

The protocol should include a detailed description of how inclusion and exclusion criteria will be implemented to identify appropriate patients meeting these criteria from the data source. The protocol should address the completeness and accuracy of the information collected in the proposed data source to fulfill the inclusion and exclusion criteria.

Sponsors should also address whether animal selection and enrollment criteria minimize bias, such as selection bias, and ensure a representative real-world population (e.g., all-comer's design which includes wide inclusion criteria and few exclusion criteria or consecutive patient enrollment).

Finally, the potential for differential loss to follow-up should be evaluated, and the impact of loss to follow-up on the conclusions drawn by the study should be considered. Sufficient longitudinal data should be available to capture the main outcomes of interest.

Minimization of bias and confounding

In traditional randomized clinical trials, bias and confounding can be reduced by utilizing masking, randomization, and strict patient inclusion and exclusion criteria. However, these methods may not be available or feasible in real-world studies.

With a proper design, procedures, and measures, bias and/or confounding can be avoided or identified and considered in assessing outcomes and reaching conclusions based on the study results. The final study report should discuss any identified or potential sources of bias and confounding and their potential impact on study conclusions.

Bias is frequently cited as the greatest concern with observational study designs (Grimes and Schulz, 2002; Weigler, 2001). In general, **selection bias** and **information bias** are the most likely to occur and cause the most concern to regulatory reviewers. Selection bias could cause the study animals to differ in important ways from the target population, impacting the generalizability of the study. Selection bias may also result in the exposed and unexposed groups differing in more than just the exposure which ultimately influences the outcome of the study. Controlling for selection bias at analysis is possible, but only when factors of interest are measured and recorded. When using RWD, researchers will be relying on what is measured and recorded in the source data.

Information bias can alter the magnitude or direction of estimates of association in ways that are not intuitive. Information bias occurs when factors are measured with error or have some inherent variability. These differences in measurements may impact the classification of study animals, which may impact study conclusions. Well-designed evaluation of endpoints and outcomes or the use of objective outcome assessments which leverage information typically measured or recorded in the study setting may help avoid significant information bias. Certain types of bias might be managed or identified in the statistical analyses because values about the bias may be measured during the study. In an observational study, particularly a retrospective study, protocols should ensure similar information is collected for each animal enrolled in the study, particularly if an animal's data will come from multiple facilities (e.g., veterinary practices or livestock operations).

Confounding occurs when the study groups being compared differ in the frequency of the outcome for reasons other than the exposure of interest. Confounders are associated with the exposure to the drug and the observed outcome but are not in the causal pathway. The reason for the difference is frequently unknown or undetected. When designing an observational study, consideration should be given to identification of potential confounders in study subjects or populations so that they can be considered in the analysis and interpretation of the study results. Generally, the distribution of confounders across exposure or treatment groups cannot be determined. One of the largest concerns with observational study designs of therapeutics is confounding by indication where patients with a certain presentation or disease severity receive a particular treatment exclusively or more frequently, or no treatment at all (Etminan and Samii, 2004; Strom and Kimmel eds, 2006). This may or may not be an issue with RWE as the study animal population may much more closely match the population that may receive the new animal drug once approved. Propensity scores provide a way to deal with confounding by indication. With propensity scores, the populations that would

always get the drug and never get the drug are removed from the dataset, potentially limiting the effect of confounding by indication.

Randomization has long been accepted as a method to allocate known and unknown confounding factors between groups by chance so that groups receiving a different treatment vary only by that treatment. In a prospective real-world clinical trial such as pragmatic clinical trial, randomization is an effective method that can be used to ensure the distribution of confounding variables at baseline to be similar for each of the groups being compared.

For observational studies of drug effects, random allocation to a treatment group does not occur, so the use of another method such as random sampling or analyzing the data using propensity scores may be appropriate (D'Agostino, 1998; Dohoo et al., 2009; Klungel et al., 2004; Schneeweiss, 2007).

VI. Obtaining Feedback on Uses of RWD and RWE for NADAs and CNADAs

There are various approaches that sponsors may take to open a discussion with CVM on the use of RWD and RWE as part of their development program to demonstrate effectiveness or a reasonable expectation of effectiveness. The sponsor's decision regarding which approach to select may be affected by where the project is in the development process. Communication about RWD and RWE may occur at any point in the development process.

The Office of New Animal Drug Evaluation (ONADE) project managers (PMs) serve as a central point of contact for drug sponsors and can provide information about the new animal drug review process and ONADE's regulatory procedures. If you have questions about the approval process and do not have an ONADE PM assigned to your company, you can contact the PMs through the CVM mailbox AskCVM@fda.hhs.gov.

A. When to submit information regarding the use of RWD and RWE

There are a variety of points in the development process and a variety of submission types that can be used to obtain feedback. CVM encourages sponsors interested in using RWD and RWE as part of their development program for a new animal drug to inform CVM as early in the product development process as possible.

Sponsors planning to incorporate RWD and RWE to demonstrate effectiveness or reasonable expectation of effectiveness are encouraged to inform CVM of their intent either as part of their initial request to open a General Correspondence (GC) file or an INAD file (A-0000), or as part of their initial presubmission conference[29] with CVM to discuss the drug product development plan (Z-submission product development meeting). If one or more studies incorporating RWD and RWE are already complete, CVM recommends sponsors submit the information described in section VI.B. *How to submit information regarding the use of RWD and RWE* prior to the initial presubmission conference. While CVM cannot make a determination if existing data would satisfy

[29] See 21 CFR 514.5 Presubmission conferences

technical section requirements outside of a data submission, if the sponsor submits sufficient information about the existing data early, we can provide feedback to help inform the development plan.[30] Sponsors are also encouraged to contact their assigned PM for assistance in determining the most appropriate method for obtaining feedback from CVM.

B. How to submit information regarding the use of RWD and RWE

There are several ways that sponsors may submit detailed information about plans for incorporating RWD and RWE into their development program to demonstrate substantial evidence of effectiveness or reasonable expectation of effectiveness. The regulatory pathway selected (CNADA versus NADA), the stage of development, the information available, and the feedback being sought from CVM, among other factors, may influence the submission type selected.

Sponsors may seek general guidance on the use of RWD and RWE in a GC file prior to opening an INAD file. Sponsors may submit information to support use of RWD and RWE as part of their initial request to open an INAD file; as part of a meeting request for a presubmission conference (Z-submission) to discuss the Effectiveness technical section requirements; or as part of an information submission (H-submission) or meeting request (Z-submission) to discuss study protocol design.

Sponsors considering incorporating RWD and RWE into future studies to demonstrate effectiveness or reasonable expectation of effectiveness should, prior to conducting a study, submit a study protocol for review (E-submission). Obtaining CVM input regarding study design will make reaching protocol concurrence more efficient.

Sponsors may also open a Veterinary Master File (VMF) to hold detailed information regarding a specific study design, including those regarding pre-investigational discussions about the use of RWD to generate RWE, or if the information will be used in the development of multiple applications.[31] The VMF is confidential and is typically used when a holder wishes the material in the VMF to remain proprietary, although the material may be referenced by multiple third-party products or files (INAD, NADA, or CNADA). Alternatively, if multiple sponsors are cooperating on product development, sponsors may establish a Public Master File (PMF) to allow all cooperators to reference the information. As suggested by the name, the information in a PMF is publicly available.

Regardless of how information is submitted to CVM, sponsors should submit an organized and focused information package. This will allow CVM the best opportunity to provide appropriate recommendations in response. Although full information may not be available in the early stages of the development process, the amount of information

[30] CVM Program Policy and Procedures (P&P) Manual 1243.2200 Submission and Review of Early Information (EI) Prior to Presubmission Conferences and Protocol Review (June 2020) and CVM P&P Manual 1243.3050 Determining Technical Section Requirements for New Animal Drug Product Approval (May 2019)

[31] See CVM P&P Manual 1243.2400 Veterinary Master Files with Manufacturing Information (August 2019)

provided and the level of detail of the information provided should be commensurate with the submission type. The information should address some or all of the following elements, as appropriate for the submission type:

1. The proposed regulatory use of the RWD or RWE, including the study question and study type if a study using RWD to generate RWE of effectiveness is proposed;
2. A description of the source or sources of RWD, including the sponsor's reason for selecting the source and an assessment of relevance and quality of the source for the evaluation of the effectiveness of the new animal drug in the identified regulatory context (e.g., as a sole source or partial source of evidence); and
3. A description of the format (e.g., paper or electronic) of the data source and the methods used for data accrual, curation/transformation, analysis, QC, and QA, as applicable to the proposed use.
4. If applicable, an outline of the design of the study proposed to generate RWE, including, but no limited to, the methods to minimize bias and confounding and the statistical analysis plan.

VII. Glossary

The following definitions are supplied to provide the reader with an understanding of the specific terms used in this guidance as applicable to new animal drugs. These definitions should not be construed to be new interpretations or clarification of the use of similar words or phrases in the FD&C Act, related code or regulation, other Federal, State, or local laws, or other guidance documents.

Accuracy: Closeness of agreement between the measured value and a true value of what is intended to be measured.[32]

Artificial Intelligence: The science and engineering of making intelligent machines, especially intelligent computer programs.

Audit Trail: Secure, computer generated, time-stamped electronic record that allows reconstruction of the course of events, including any changes made to the original data recorded.

Bias: Bias is any systematic error in the design, conduct, analysis, interpretation, publication, or review of a study and its data that results in a mistaken estimate of a treatment's effect on disease. This systematic error results from flaws in the method of selecting study participants, in the procedures for gathering data, and in the decision of how and whether to publish the results. These flaws can lead to observed study results that tend to be different from the "true" results. Some biases can be minimized by ensuring that the study design is appropriate for addressing the study hypotheses and establishing and carefully monitoring procedures of data collection that are valid and reliable. (Szklo and Nieto, 2000)

[32] Adapted from the Joint Committee for Guides in Metrology guidance International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 3rd edition, 2012.

Clinical Data Repository: A database that consolidates data from disparate clinical sources, such as those within an Electronic Health Record (EHR) system, to provide a broader picture of the care a patient has received. (Shortliffe and Cimino, 2014)

Common Data Model (CDM): Comprehensive framework that includes definitions, specifications, and operational rules for data to be presented and used in a common manner which ensures interoperability. (Daniel et al., 2018)

Completeness: "Presence of the necessary data." (NIH, 2014)

Conformance: Data congruence with standardized types, sizes, and formats. (Daniel et al., 2018)

Confounding: "A situation in which a non-causal association between a given drug exposure or treatment and an outcome is observed as a result of the influence of a third variable designated as a confounder. The confounding variable needs to be related to both the treatment and the outcome under study." (Gordis, 1996)

Covariate: Data used to characterize animal populations, balance groups, and/or control for confounding. (Daniel et al., 2018)

Curation: Processing (unstructured and structured data processing) of source data such as from EMVRs into an electronic dataset. The curation process involves the application of standards for the exchange, integration, sharing, and retrieval of source data, often from various sources. For example, the application of standard medical diagnostic codes to adverse events, disease staging, the progression of disease, and other medical and clinical concepts in an EVMR.

Data Accrual: The process by which the data was collected.

Data Element: A piece of data corresponding to one animal within a data field. (Daniel et al., 2018) Examples include species, animal class, breed, age, reproductive status, identification of drug product and dosage of exposure received, outcomes such as a pain score, and other observations made and documented during a study.

Data Integrity: The completeness, consistency, and accuracy of data.[33]

Data Warehouse: Data from the clinical data repository that has undergone data transformation and de-identification, if necessary.

Electronic Data Capture (EDC) systems: Electronic systems designed to collect and manage clinical trial data in an electronic format.

Electronic Veterinary Medical Record (EVMR): A veterinary patient record contained within an electronic veterinary medical record system which is designed and used for documentation of the animal's medical information as well as owner contact and billing information. EVMRs may

[33] See GFI, "Data Integrity and Compliance With Drug CGMP Questions and Answers," (December 2018)

include an animal's medical history, diagnoses, treatment plans, immunization dates, allergies, radiology images, pharmacy records, and laboratory and test results. (Krone et al., 2014)

Electronic Source Data: Data initially recorded in an electronic format.

Information bias: Systematic distortions in the data arising from measurement error. (Daniel et al., 2018) For discrete variables (variables with only a countable number of possible values, such as indicators for sex), measurement error is usually called *classification error* or *misclassification*. (Rothman et al., 2008)

Interoperability: The ability to communicate and exchange data accurately, effectively, securely, and consistently with different information technology systems, software applications, and networks in various settings, and exchange data such that clinical or operational purpose and meaning of the data are preserved and unaltered.[34]

Interventional Study: An experimental study or clinical trial in which the researcher intercedes as part of the study design and in which participants are assigned to groups that receive one or more treatments (interventions), or no treatment, so that researchers can evaluate the effects of the treatments.

Machine learning: The ability of a program to learn from experience, that is, to modify its execution on the basis of newly acquired information. In epidemiology and bioinformatics, examples include artificial neutral network, support vector machines, Bayesian networks, and other methods that update their procedures as new data are provided. (Porta, 2014)

Misclassification: The erroneous classification of an individual, value, or attribute into a category other than that to which it should be assigned. (Porta, 2014)

Missing Data: Data that would be used in the study but were not observed, collected, or accessible. This refers to information that is intended to be collected but is absent, and information that is not intended to be collected and is therefore absent.

Observational study: An epidemiologic study in which the investigator does not act upon study participants, but instead observes natural relationships between factors and outcomes. A type of study in which participants are identified as belonging to study groups and assessed for outcomes. Participants may receive treatment (intervention) but the investigator does not assign participants to a specific treatment.

Pragmatic clinical trial: A prospective clinical study that compares the clinical intervention and a relevant comparator in animals that are similar to those affected by the condition(s) under study and in settings that are similar to those in which the condition is typically treated (i.e., in routine clinical practice settings) (adapted from Califf and Sugarman, 2015; NIH, 2014).

[34] See GFI, "Providing Regulatory Submissions in Electronic Format - Standardized Study Data, Study Data Technical Conformance Guide," (May 2019)

Provenance: "A record trail that accounts for the origin of a piece of data (in a database, document or repository) together with an explanation of how and why it got to the present place." (Gupta, 2009)

Quality Control (QC): The steps taken during study conduct to ensure that it meets the prespecified standards and that the data is reproducible.

Quality Assurance (QA): Activities conducted to evaluate the quality control and the level of adherence to the study protocol, standard operating procedures and to the specified standard of conduct. It should be noted, that QA is not required for effectiveness studies. However, CVM supports the use of QA to enhance the level of data quality and study integrity.

Registries: Organized systems that use observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serve one or more predetermined scientific, clinical or policy purposes.

Selection bias: Bias due to systematic differences between baseline characteristics of the groups that are compared, within the study or between the study population and target population.

Source Data: All information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents (original records or certified copies).

Structured RWD: Data that resides in a pre-defined data model or is organized in a pre-defined manner which makes it easily searchable.

Traceability: Characteristics that permit an understanding of the relationships between the analysis results (tables, listings and figures in the study report), analysis datasets, tabulation datasets, and source data.

Transformation: Process of data extraction, cleaning, verifying, linking (if necessary), and standardizing RWD within a dataset. The transformation process includes conversation of multiple data sources into a CDM if necessary.

Unstructured RWD: Data that does not reside in a pre-defined data model and is not organized in a pre-defined manner. Examples include narrative text in a patient record, diagnostic reports, images, videos, or email communication.

Validation: The process of establishing that a method is sound or that data are correctly measured. (Porta, 2014)

VIII. References

Alpi, KM, HA Burnett, SJ Bryant, and KM Anderson. 2011. Connecting knowledge resources to the veterinary electronic health record: opportunities for learning at point of care. J Vet Med Educ 38(2):110-122.

Atanasov, P, S Yang, C Salehi, AL Ghindilis, and E Wilkins. 1996. Short-term canine implantation of a glucose monitoring-telemetry device. Med Eng Phys 18(8):632-640.

Belda, B, M Enomoto, B C Case, and BDX Lascelles. 2018. Initial evaluation of PetPace activity monitor. Vet J 237:63-68.

Benjamin, M and S Yik. 2019. Precision Livestock Farming in Swine Welfare: A Review for Swine Practitioners. Animals (Basel) 9(4).

Califf, R M and J Sugarman. 2015. Exploring the ethical and regulatory issues in pragmatic clinical trials. Clin Trials 12(5):436-441.

D'Agostino, RB. 1998. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 17(19):2265-2281.

Daniel, G, C Silcox, J Bryan, M McClellan, M Romine, and K Frank. 2018. Characterizing RWD Quality and Relevancy for Regulatory Purposes. DMC. f. H., ed.

Dohoo, I, W Martin, and H Stryhn. 2009. Veterinary Epidemiologic Research, 2nd edition. Charlottetown, Canada: VER.

Etminan, M and A Samii. 2004. Pharmacoepidemiology I: a review of pharmacoepidemiologic study designs. Pharmacotherapy 24(8):964-969.

Girman, CJ, ME Ritchey, W Zhou, and NA Dreyer. 2019. Considerations in characterizing real-world data relevance and quality for regulatory purposes: A commentary. Pharmacoepidemiol Drug Saf 28(4):439-442.

Gliklich, R, M Leavy, and N Dreyer. 2020. Registries for Evaluating Patient Outcomes: A User's Guide. 4th edition. Rockville (MD): Agency for Healthcare Research and Quality (US). Available from: https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/users-guide

Gordis, L. 1996. Epidemiology. Philadelphia: Elsevier/Saunders.

Griffies, JD, J Zutty, M Sarzen, and S Soorholtz. 2018. Wearable sensor shown to specifically quantify pruritic behaviors in dogs. BMC Vet Res 14(1):124.

Grimes, DA and KF Schulz. 2002. Bias and causal associations in observational research. The Lancet 359(9302):248-252.

Gupta, A. 2009. Data Provenance. In L. Liu and MT ÖZsu (Eds.) Encyclopedia of Database Systems. Boston, MA: Springer.

Guy, MK, RL Page, WA Jensen, PN Olson, JD Haworth, EE Searfoss, and D. E. Brown 2015. The Golden Retriever Lifetime Study: establishing an observational cohort study with translational relevance for human health. Philos Trans R Soc Lond B Biol Sci 370(1673).

Hall, K. 2019. VetCOT: The veterinary trauma registry. Topics in Companion Animal Medicine.

Harron, KL, JC Doidge, HE Knight, RE Gilbert, H Goldstein, DA Cromwell, and JH van der Meulen. 2017. A guide to evaluating linkage quality for the analysis of linked data. Int J Epidemiol 46(5):1699-1710.

King, MTM and TJ DeVries. 2018. Graduate Student Literature Review: Detecting health disorders using data from automatic milking systems and associated technologies. J Dairy Sci 101(9):8605-8614.

Klungel, OH, EP Martens, BM Psaty, DE Grobbee, SD Sullivan, BH Stricker, HG Leufkens, and A de Boer. 2004. Methods to assess intended effects of drug treatment in observational studies are reviewed. J Clin Epidemiol 57(12):1223-1231.

Krone, LM, CM Brown, and J M Lindenmayer. 2014. Survey of electronic veterinary medical record adoption and use by independent small animal veterinary medical practices in Massachusetts. J Am Vet Med Assoc 245(3):324-332.

Kwong, M, HL Gardner, N Dieterle and V Rentko. 2019. Optimization of Electronic Medical Records for Data Mining using a Common Data Model. Topics in Companion Animal Medicine.

Lahdenoja, O, T Hurnanen, M Kaisti, J Koskinen, J Tuominen, M Vaha-Heikkila, L Parikka, M Wiberg, T Koivisto, and M Pankaala. 2019. Cardiac monitoring of dogs via smartphone mechanocardiography: a feasibility study. Biomed Eng Online 18(1):47.

Miksad, RA and AP Abernethy. 2018. Harnessing the Power of Real-World Evidence (RWE): A Checklist to Ensure Regulatory-Grade Data Quality. Clinical Pharmacology & Therapeutics 103(2):202-205.

Nie, A, A Zehnder, RL Page, Y Zhang, AL Pineda, MA Rivas, CD Bustamante, and J Zou. 2018. DeepTag: inferring diagnoses from veterinary clinical notes. NPJ Digit Med 1:60.

NIH. 2014. National Institutes of Health Collaboratory. Assessing Data Quality for Healthcare Systems Data Used in Clinical Research. Available from https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Assessing-data-quality_V1%200.pdf#search=Assessing%20data%20quality. (Accessed December 7, 2020)

Porta, M, Ed. 2014. A Dictionary of Epidemiology. New York, NY: Oxford University Press.

Rothman, K, S Greenland, and T Lash, 2008. Modern Epidemiology, 3rd edition, Philadelphia: Lippincott Williams & Wilkins.

Santamaria, SL and KL Zimmerman. 2011. Uses of informatics to solve real world problems in veterinary medicine. J Vet Med Educ 38(2):103-109.

Schneeweiss, S. 2007. Developments in post-marketing comparative effectiveness research. Clin Pharmacol Ther 82(2):143-156.

Shah, P, F Kendall, S Khozin, R Goosen, J Hu, J Laramie, M Ringel and N Schork. 2019. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ Digit Med 2:69.

Shortliffe, E and J Cimino. 2014. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 4th edition. New York, NY: Springer.

Strom, B and S Kimmel, eds. 2006. Textbook of Pharmacoepidemiology. West Sussex England: John Wiley & Sons, Ltd.

Szklo M, and F Nieto. 2000. Epidemiology: Beyond the basics. Gaithersburg, MD: Aspen Publisher, Inc.

Yamazaki, AK, K Edamura, K Tanegashima, Y Tomo, M Yamamoto, H Hirao, M Seki, and K Asano. 2020. Utility of a novel activity monitor assessing physical activities and sleep quality in cats. PLos ONE 15(7): e0236795.