PharmaSUG China 2022 - Paper RW-119

Implementation of CDISC in Real World Data

Xiang Wang, Happy Life Tech

ABSTRACT

Recent advances in healthcare data access, technology and regulatory guidance have created an opportunitiy for the biopharmaceutical industry to use Real World Data (RWD) in clinical research, with profound impact.

In recent years, CDE released the guide '真实世界证据支持药物研发与审评的指导原则（试行）' '用于产生真实世界证据的真实世界数据指导原则（试行）' and '药物真实世界研究设计与方案框架指导原则（征求意见稿）', which clearly pointed out that Real World Evidence can be used to support drug regulatory decisions. In the guidance of "***Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry***" released last October, FDA encouraged the use of CDISC standards to process Real World Data.

What difficulties will CDISC standards face in the implementation of Real World Data Study?

From the perspective of programmers, combined with the experience of Happy Life Tech Real World Study, this paper will introduce from three parts: the data extraction process of Happy Life Tech Real World Study, the implementation path of CDISC in Real World Study data, and the difficulties and considerations of CDISC implementation.

INTRODUCTION

For many years, randomized controlled trials (RCTs) have been the "Gold Standard" for evaluating the safety and effectiveness of drugs. However, RCT is often time-consuming, expensive, and the clinical applicability of the output results is limited. At the meantime, it is subject to many restrictions due to ethical, feasibility and other reasons.

In recent years, the research methods of using Real World data (RWD) to carry out Real World Study (RWS) and then generate Real World Evidence (RWE) have developed rapidly. Compared with RCT, RWS has wide coverage, good representativeness, low cost, less time-consuming and strong operability. It can obtain the actual effect and long-term safety evidence of the application of the target population, and its application value has been widely recognized.

In 2020, CDE released the guide '真实世界证据支持药物研发与审评的指导原则（试行）', which pointed out that "for some rare diseases and major life-threatening diseases that lack effective treatment measures, single arm clinical trials based on Real World Evidence as external control are adopted". At the meantime, the number of approval decision cases based on external control studies of FDA and EMA in the United States is also increasing.

In the FDA's guidance of "***Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry***" released last October, which pointed out that '*Sponsors should refer to the specifications, recommendations, and general considerations provided in the Study Data Technical Conformance Guide when submitting study data in an applicable drug submission to FDA……. FDA recognizes that a range of approaches may be used to apply currently supported data standards (e.g., Clinical Data Interchange Standards Consortium's (CDISC's) Study Data Tabulation Model (SDTM)) to RWD sources such as EHR or claims data. With adequate documentation of the conformance methods used and their rationale, study data derived from RWD can be transformed to SDTM datasets and submitted to FDA in an applicable drug submission.*', encouraged the use of CDISC standards to process Real World Data.

Also, In guidance of '药物试验数据递交指导原则(试行)', CDE encouraged the use of CDISC standards to use for key clinical trials to support the registration and marketing of drugs.

This raises a question: What difficulties will CDISC standards face in the implementation of Real World Data studies?

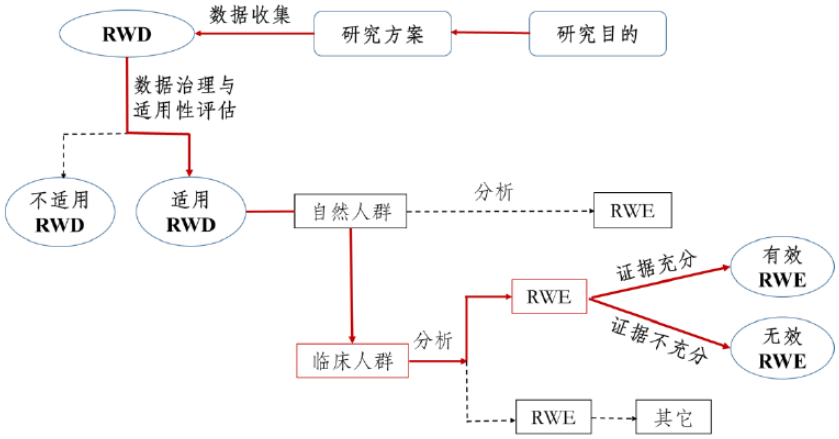
From the perspective of programmers, combined with the experience of Happy Life Tech Real World Study, this paper will introduce from three parts: the data extraction process of Happy Life Tech Real World Study, the implementation path of CDISC in Real World Study data, and the difficulties and considerations of CDISC implementation.

DATA EXTRACTION PROCESS OF HAPPY LIFE TECH REAL WORLD STUDY

The types of Real World Study can be roughly divided into **Non Interventional Study**(Observational) and **Interventional Study**. The former includes retrospective and prospective observational research without any intervention, and the diagnosis and treatment of patients, disease management, information collection and so on are completely dependent on daily medical practice; The biggest difference between the latter and the former is the active implementation of some interventions, such as practical clinical trial (PCT).

Since **Interventional Study** is similar to RCT, we focus on **Non Interventional Study** here.

The following figure(Display 1) shows the Real World Study path to support drug regulatory decisions[1].

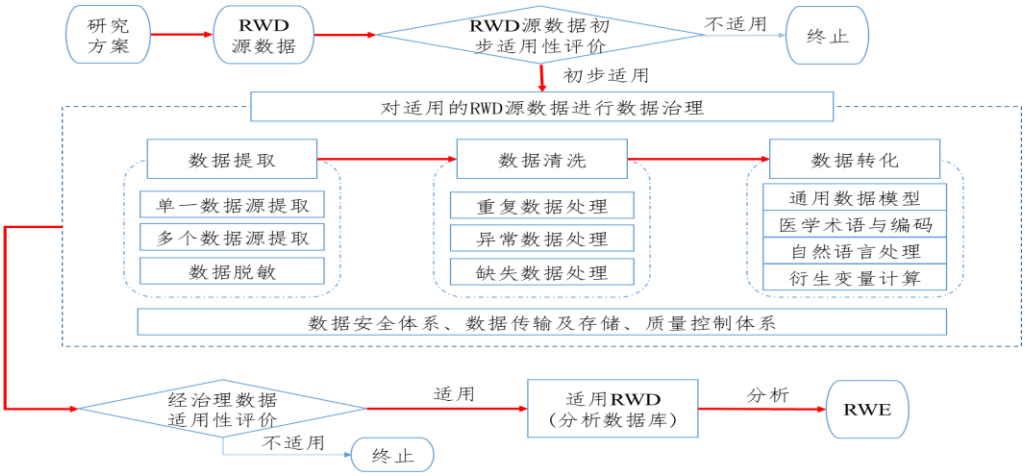


Display 1: Real World Study Path to Support Drug Regulatory Decision

You can simply divide these processes into protocol design, data curation or data management, statistical analysis and interpretation of the results after analysis etc.

Real World Study can be divided into prospective study, retrospective study and prospective + retrospective study according to the time of data acquisition. Whether it is prospective study, retrospective study or prospective + retrospective study, in Real World Study, our clinical data acquisition methods are basically similar: 1. Electronic extraction from EHR(electronic health record) 2. Manual transcription

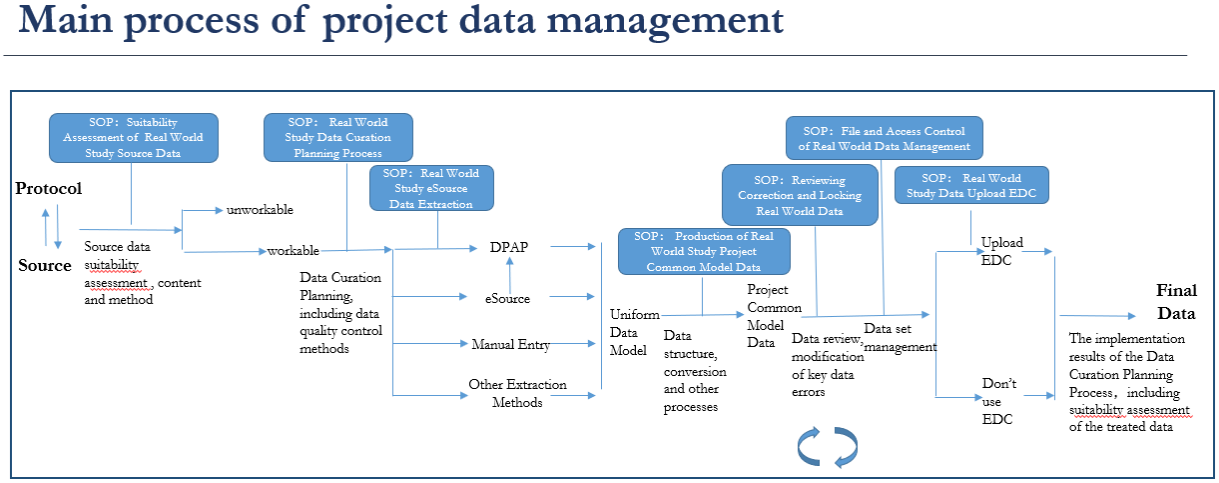
You can refer to the following figure(Display 2) for the specific data curation process[3].



Display 2: Schematic Diagram of Suitability Assessment and Data Curation Process of Real World Data

1. HAPPY LIFE TECH DATA EXPORT PROCESS

Combined with various guidelines, laws and regulations for Real World Study and the reality of Happy Life Tech, we summarized Happy Life Tech data export process.



- The source data will form the final data set through the processes of **Data Suitability Assessment, Data Curation Planning Formulation, Data Extraction, General Model Data Production, Data Reviewing, Correction and Locking.**

Display 3: Schematic Diagram of Data Management Process of Real World Data in Happy Life Tech

(1) Find the appropriate source data according to the protocol, and then improve the protocol according to the source data. It is necessary to assess the suitability of the source data. Once the evaluation is completed, the protocol and source data cannot be changed at will. The suitability of source data is mainly evaluated through data correlation and reliability. For specific considerations of correlation and reliability, please refer to "《真实世界证据支持药物研发与审批的指导原则》".

(2) If the suitability assessment of source data is feasible, the protocol will be finalized. After that, we will prepare data curation plan, including data quality control methods.

(3) Then do the data curation and import the data which after curation into the DPAP* platform. At the meantime, if there is eSource data, it also can be imported into the DPAP platform. The imported data of DPAP platform, as well as the data of manual entry and data which using other extraction methods build a Uniform Data Model together.

(4) From the Uniform Data Model, through the conversion of data results and some other processes, we get the Project Common Model data at the project level.

(5) Then medical expert will have a medical review of the project level data and DM will make the correction of errors in some key data with the consent of PI. It is a repetitive process to correct the errors in key data. Try to ensure that the data can objectively and accurately reflect the objective facts and meet the needs of analysis.

(6) Decide whether to use Happy Life Tech EDC or other EDC system according to the requirements of the study.

(7) Here we use Happy Life Tech EDC by default, and refer to the standard database of CDASH. After a series of data curation and data quality control, we can get the data that can be used for subsequent analysis.

*DPAP: The DPAP medical data intelligent platform integrates the system data of medical record home page, HIS, EMR, inspection, examination, ICU, mobile nursing, hand anesthesia, etc., and uses AI technology to process and analysis huge medical data to form a uniform medical model and shield the difficulties of clinical data application caused by multi systems, multi standards and low structure. On this basis, it provides clinical and scientific research doctors with data visualization capabilities based on the full amount of governance data and the whole life cycle of patients, and provides reliable, fast, flexible and easy-to-use data retrieval capabilities based on big data search engines, so as to meet the functional needs of doctors in clinical, scientific research, management and other scenarios, such as medical record search, medical record details browsing, patient panoramic view, data export and analysis.

In a word, the role of DPAP is to transform data from different sources and heterogeneous into uniform data with the same structure, and to structure the text and standardize the data in RWD processing.

After the actual operation summary of several studies, some experience has been summed up:

(1) Try to build the database follow the CDASH standard.

(2) It is necessary to ensure the integrity and accuracy of data

(3) The data is divided into subjective data and objective data. Subjective data will be collected as much as possible in the data collection stage. For objective data, when key data is missing, it is necessary to communicate with PI about the reasons for the missing data and try to improve the integrity of the data. If there are errors in the data, the data will also be corrected in the data management stage.

IMPLEMENTATION PATH OF CDISC IN REAL WORLD DATA

When doing statistical analysis, we often use SDTM and ADaM standards to process data.

Three strategies implemented after data is imported from the database:

1. aCRF->SDTM->ADaM->Table, Listing, Figure

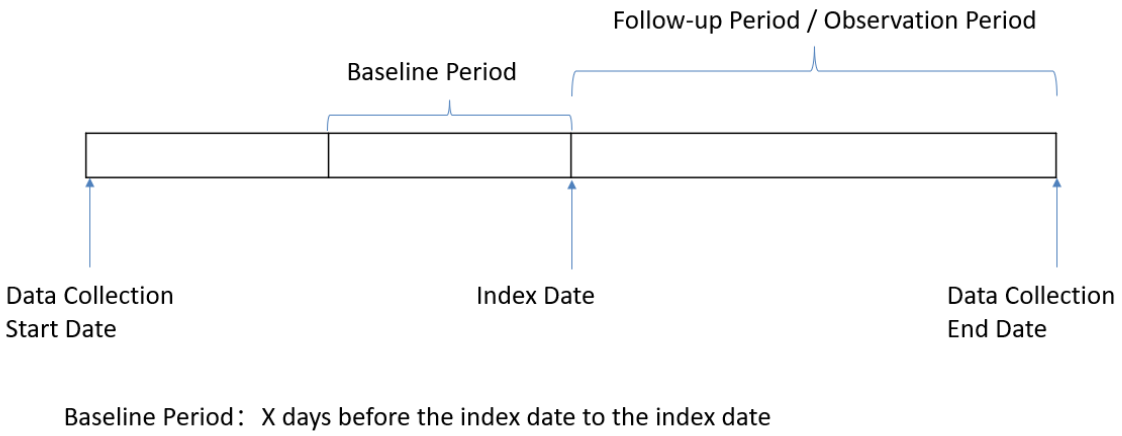
Prospective Study:

Because prospective study can have unified diagnostic criteria, detection criteria and evaluation criteria, so when you follow the above process, you won't have too much trouble, it's not much different from doing an ordinary pre-market clinical trial.

Retrospective Study:

Due to the selection of similar clinical data in a certain period in the past for sorting and analysis. Therefore, the screening time is after all events, and the data quality is not controllable.

The following figure is the schematic diagram of common retrospective Real World Study trail design.



Display 4: Schematic Diagram of Common Retrospective Real World Study Trail Design

From this schematic diagram, we can get those information:

(1) There are several key time points: Data Collection Start Date, Index Date, Data Collection End Date.

(2) Baseline Period is a period of time before index date.

(3) After Index Date, it has Follow-up Period/Observation Period.

Due to the characteristics of retrospective research, we need to pay attention to the following points when doing statistical analysis:

(1) aCRF

We might need to Custom some new Domains.

(2) SDTM

Time Level

Index Date: It will be the first date that the enrolled patients meet the requirements of this study protocol during the data collection period.

Baseline Period: X days before the index date to 1 day before the index date, X will define in Protocol.

Follow-up Period/Observation Period: From index date to data collection end date(or event occur date).

VISIT: A patient has multiple visit records on the same day. For example, a subject has both outpatient and inpatient visits on the same day. Outpatient and inpatient examinations were conducted respectively. Some visits are specified in the plan (for example, in the tumor study, the first, second, and third visits of tumor evaluation), and some may not be specified in the plan (for example, visits within a period of time).

Element: Element can be defined in a general way, it does not need to be as specific as arm. For example, a RWS trial evaluated the efficacy and safety of second-line and above standard treatment for lung cancer, a variety of therapies will be collected in the data, you can't define elements as those treatments. You can define Element as 'second-line and above standard treatment'.

If the index date is the start date of the treatment plan during the treatment period

RFXSTDTC: Date/Time of First Study Treatment, you can assign it equal to index date.

RFXENDTC: Date/Time of Last Study Treatment, you can assign it equal to treatment end date.

RFPENDTC: Date/Time of End of Participation, equal to the last non missing date(including data collection end date).

Domain Level

Drug Information with Coding: When encountering drugs with coding, whether therapeutic drugs or non-therapeutic drugs specified in the protocol, we'd better put the drug related data set into the CM dataset.

(3) ADaM

Using SDTM data as the source data to further generate ADaM data.

The baseline definition should be written within X days before the index date, X will define in Protocol.

In tumor trials, we usually reconfirm CR PR, but we cannot do the same in RWS.

(4) Table, Listing, Figure:

Using SDTM data and ADaM data as the source data to further generate Table, Listing, Figure.

2. SDTM Like ->ADaM-> Table, Listing, Figure

(1) SDTM Like

Write a program to read the database file. You can generate a program for each domain through the method of program generation program, and simply convert the original data set in batch processing to meet certain requirements: for example, the length of the original data set is 2, and the length of the label is no more than 40. For specific requirements, refer to '药物临床试验数据递交指导原则（试行）'[4].

Of course, you need to perform QC on the data processed by SDTM like.

(2) ADaM

Using SDTM like data as the source data to further generate ADaM data.

(3) Table, Listing, Figure

Using SDTM like data and ADaM data as the source data to further generate Table, Listing, Figure.

3. ADaM-> Table, Listing, Figure

(1) ADaM

This method can be used to write Adam directly from the data exported from EDC if it is a simple study(the data structure is relatively simple, and there are relatively few data points). If it is a complex study(there are many data points, and there are many and complex endpoints), it will be difficult to use this method.

(2) Table, Listing, Figure

Using ADaM data as the source data to further generate Table, Listing, Figure.

In general, the prospective RWS data is more ideal and more suitable for directly making SDTM format. We recommend to choose path 1 or path 3 to meet different trail purposes. In the retrospective study of data quality is not controllable, we can choose paths 1,2 or 3 to meet different trail purposes. Of course, it is a simpler method for RWS to directly select path 3 if it is not for the purpose of registration, but when it comes to complex derive, it will be more difficult to query the original data.

THE DIFFICULTIES AND CONSIDERATIONS OF CDISC IMPLEMENTATION

The following experience is a summary of several projects and cannot cover all situations.

1. MAIN PROTOCOL AND SAP

In order to avoid result driven bias and ensure the transparency of the research process, Real World Study emphasizes that at least the main analysis plan should be determined synchronously with the protocol, which is very different from the regulation in RCT that the statistical analysis plan can be completed before the database is locked[4].

So the protocol and SAP should be written as detailed as possible, considered as comprehensively as possible, and it is best to work closely with data manager to consider the missing value and whether it needs to be filled.

Clinical trials for submission purposes, such as single arm clinical trials based on Real World Evidence as external controls, mockup shell must be finalized before locking the database.

2. COMPLEX ENDPOINT

Due to the variety of RWD sources and their inconsistent formats (e.g., EHR, registry) and the complex endpoint, some results are too complex to be derived by programming or need medical expertise to judge.

You can consider exporting the required data, judging it by medical expert, integrating into an external file and then merging it back to the data set for further analysis.

Of course, the steps of medical manual interpretation, data export and import need to be supported by quality control, convincing processes and documents etc.

3. UNIT CONVERSION

Including the diversity of data sources such as laboratory examination and concomitant medication. Descriptive analysis often involves unit conversion and unified unit.

Example 1. Unit conversion and unified unit

The unit conversion of laboratory test value is common, and we usually make a descriptive analysis of the laboratory test value after unit conversion. When we make descriptive statistics of a person's daily dose of a drug, we first have to convert the unit, assuming that we all convert to milligrams(mg).

Here is an example of converting the units of drugs.

EXDECOD (Standardized Name)	**EXDOSE (Single Dose)**	**EXDOSU (Dose Unit)**	**EXDOSUO (Other Dose Unit)**	**EXDOSFRQ (Dosing Frequency per Interval)**	**EXDSPEC (Drug Specifications)**	**CONVFACT (Convert Factor)**	**EXSTRESU (Standard Unit)**	**EXSTRESN (Standard Unit)**
Budesonide	1	ml		QD	3ml*5	1	ml	1
Budesonide	1	Other	spray	QD	64ug*120 spray	0.064	ml	0.064
Budesonide;Formotero	1	Other	absorb	QD	(160ug:4.5ug)*60 absorb	0.16	ml	0.16
Budesonide	1	ml		QD	3ml*5	1	ml	1
Budesonide	2	ml		QD	3ml*5	2	ml	2
Budesonide	3	ml		QD	3ml*5	3	ml	3

Example 2. Medical expert judge

Because the unit conversion is the Real World Data and is not the preset value in clinical trials, it may also need medical manual judgment. For example, EXDSPEC=' (160ug:4.5ug)*60 absorb', after medical expert judge, we get the Convert Factor = 0.16.

4. HANDLING OF MISSING DATA

Lack of data is common in Real World Study.

Time Filling

Time filling is also very common in RCT. You need to communicate with medical experts to find a clinically meaningful filling method.

General method: According to the logic of the time before and after the visit, if the "day" is missing, it can be taken as 15. If the "month" is missing, fill in the latest visit date.

TLF Missing Data Display

In CRF design, data manager often designs "Unknown" to collect unknown data. When display missing data in TLF, you better display 'Unknown' and 'Missing' to differ that.

5. QUALITY CONTROL

In real world study, quality control needs to be ensured through the whole study period, not only in data curation, statistical analysis.

CONCLUSION

This Paper introduction, in Real World Study, Happy Life Tech data export process, summarizes the process of statistical analysis, and summarizes the difficulties and considerations of statistical analysis. Randomized controlled trials (RCTs) is still "Gold Standard", but RWS has its unique advantages. RWS trials also have the risk of failure. What I want to say here is: put aside the shackles of so-called Real World Study and return to the original intention of clinical trials, based on real diagnosis and treatment needs. As long as we collect samples as comprehensively as possible, and improve data quality as much as possible, use as perfect statistical methods to correct confounding and bias, and use as objective data as possible to improve research efficiency, further reduce the cost of clinical trials, and speed up the early listing of drugs, so that more and more can get benefits.

REFERENCES

[1]"真实世界证据支持药物研发与审评的指导原则（试行）" Accessed Jan 7, 2020. https://www.cde.org.cn/zdyz/domesticinfopage?zdyzIdCODE=db4376287cb678882a3f6c8906069582.

[2,5] "药物真实世界研究设计与方案框架指导原则（征求意见稿）" Accessed Jul 7, 2022. https://www.cde.org.cn/main/news/viewInfoCommon/ea778658adc3d1ae3ffe3f1cc0522e5e.

[3]"用于产生真实世界证据的真实世界数据指导原则（试行）" Accessed Apr 13, 2021. https://www.cde.org.cn/main/news/viewInfoCommon/2a1c437ed54e7b838a7e86f4ac21c539.

[4]"药物临床试验数据递交指导原则（试行）" Accessed Jul 20, 2020. https://www.cde.org.cn/main/news/viewInfoCommon/f649995d3a9ade8dcd67f6a2ced36f0b.

ACKNOWLEDGMENTS

I will thank my colleague Lingjia Zeng and Xiaohui Rong for the review and feedback of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Xiang Wang
HappyLifeTech
xiang.wang@hlifetech.com