Key Considerations in Using Real-World Evidence to Support Drug Development

(Draft for Public Review)

Center for Drug Evaluation, NMPA

May, 2019

Table of Contents

Key Considerations in Using Real-World Evidence to Support Drug Development

1. INTRODUCTION

1. Background and Purpose

Randomized Controlled Trials (RCTs) are considered the "gold standard" for evaluating drug efficacy and are widely used in clinical trials. With strictly controlled trial eligibility criteria and the utilization of randomization, RCTs minimize the impact of factors that potentially affect the causal inference, and hence result in more definitive conclusions and derive more reliable evidence. However, RCTs also have limitations: stringent entry criteria may reduce the representativeness of the trial population to the target population, the standard trial interventions used may not be completely consistent with real world clinical practice, the limited sample size and short follow-up time leads to insufficient evaluation of rare adverse events. These limitations bring challenges when extrapolating the RCT conclusions to real world clinical practice. In addition, for some rare and major life-threatening diseases that lack effective treatments, conventional RCTs may be difficult to implement, require substantial time costs, or raise ethical issues. Therefore, how to use real-world evidence (RWE) during drug R&D, especially as complementary evidence to RCTs in evaluating the efficacy and safety of drugs, has become a common and challenging question for global regulatory agencies, the pharmaceutical industry and academia.

First, we need to clarify the definition and scope of real-world evidence on a conceptual level.

Secondly, can and how will real-world data (RWD), as the fundamental basis of real-world evidence, provide sufficient support will

29 face many questions that need to be discussed, including data sources, data

30 standards, data quality, data sharing mechanism, data infrastructure and so

31 on.

32   Third, the lack of regulatory guidance. At present, there are no mature

33 and relevant regulations worldwide. Without sufficient experience, how to

34 formulate guidelines that fit the reality of China's pharmaceutical industry

35 requires active exploration and innovation.

36   Fourth, the methodologies for evaluating real-world evidence needs

37 to be streamlined. Real-world evidence stems from the correct and

38 adequate analysis of real-world data. The analysis methods used are mainly

39 for causal inference, which often requires more complex models and

40 assumptions, screening of corresponding covariates, identification of

41 confounding factors, definition of intermediate variables and instrumental

42 variables, etc., All these will put forward higher requirements for statistical

43 analysts as well as the urgent needs for regulatory guidelines.

44   Fifth, the scope of real-world evidence application remains to be

45 determined. The main role of real-world evidence is to complement,

46 instead of substitute, the evidence provided by conventional clinical trials,

47 and to form a complete and rigorous chain of evidence to further improve

48 the efficiency and scientific validity of drug development. Therefore, it is

49 necessary to clearly define the scope of application of real-world evidence

50 according to the stage of drug development, and in the meanwhile adopt

51 appropriate adjustment as the actual conditions evolve over time.

52   In light of the above, this guideline aims to provide clarity on the

53 definition of real-world research, outline the use and scope of real-world

54 evidence in drug R&D, explore the basic principles for the evaluation of

55 real-world evidence, and consequently provide scientific and practical

guidance for the industry to consider when utilizing real-world evidence to support drug development.

2. Progress in the development of related regulations or guidelines by domestic and foreign regulatory agencies

In February 2009, the American Recovery and Reinfection Act played a significant role in promoting Comparative Effectiveness Research (CER). Accordingly, the concept of real-world research (RWR, or real-world study RWS) was proposed given the context of the real world environment of CER.

In December 2016, the United States passed the 21st Century Cures Act (the Act), encouraging the Food and Drug Administration (FDA) to accelerate the development of pharmaceutical products by conducting research in the use of real-world evidence. Under the support of the Act, during 2017-2018 the FDA issued a series of guidelines, namely "Use of Real World Evidence to Support Medical Device Regulatory Decisions", "Guidelines for the Use of Electronic Health Record Data in Clinical Research" and "Framework for Real World Evidence Solutions".

In 2013, the European Medicines Agency (EMA) released the "Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease", discussing the technical details in using real-world observational data to establish disease progression models. In 2014, EMA also launched the Adaptive Licensing Pilot to assess the feasibility of using observational study data to assist decision-making. Later in 2016, the "Scientific Guidance on Post-authorisation Efficacy Studies" was released.

At the International Council for Harmonisation of Technical Requirements for Medicinal Products for Human Use (ICH), Japan's Pharmaceuticals and Medical Devices Agency (PMDA), proposed a

84 strategic approach for pharmacoepidemiology studies submitted to
85 regulatory agencies to advance more effective utilization of real-world data.

86   The systematic use of real-world evidence to support drug
87 development and regulatory decision-making in China is still under
88 development. However, the national drug regulatory agencies have already
89 begun to utilize real-world evidence in the review practices. For example,
90 the extended Bevacizumab treatment regimen in combination with
91 platinum-based chemotherapies was approved in 2018, using real-world
92 evidence from three retrospective studies. In another case, a drug was
93 further evaluated, after marketing, through a prospective, observational
94 real-world study to provide additional evidence on efficacy and safety.

95 **2. Relevant Definitions of Real-World Research**

96   Generally speaking, real-world research includes both research on
97 natural populations and on clinical populations; the latter yields real-world
98 evidence that can be used both to support medical product development
99 and regulatory decisions, as well as for other scientific purposes. For that
100 reason, this guidance focuses on real-world research that supports
101 healthcare product development and regulatory decisions (see figure
102 below).
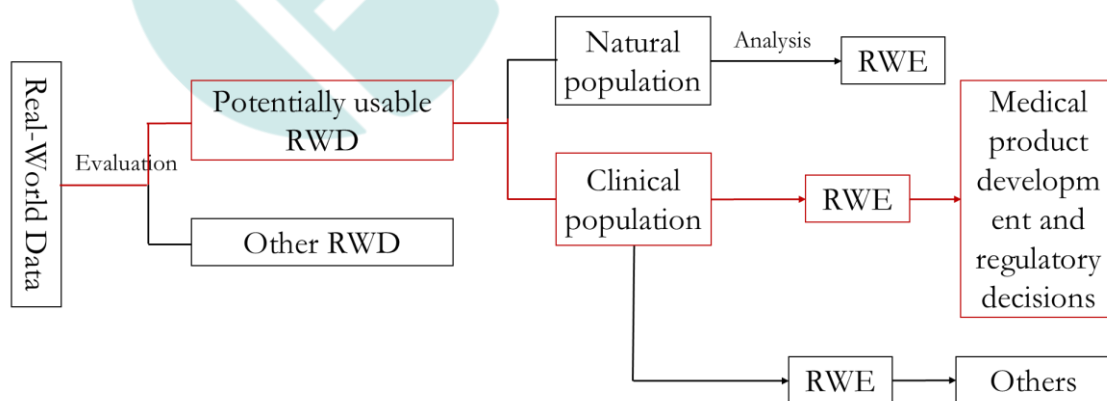
103

104   Figure 1 The path from RWD to RWE, which supports regulatory
105 decisions for medical products

We define real-world research as: collecting patient-related data in a real-world environment (real-world data), and obtaining clinical evidence (real-world evidence) of the value and potential benefits or risks of the medical products through analysis. The primary research type is observational, but it can also be pragmatic clinical trials.

1. Real-World Data

（1）Definition

Section 505F (b) of the Federal Food, Drug, and Cosmetics Act (FD&C Act) defines real-world data as "data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials". In "Framework for FDA's Real-World Evidence Program" and the "Use of Real World Evidence to Support Medical Device Regulatory Decisions.", the FDA defines real-world data as "data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources". For example, Electronic Health Record (EHR) data, Electronic Medical Record (EMR) data, medical insurance data, product and disease registry data, patient report data (including home environment), and other health tests (such as mobile devices) data.

We define real world data as: data collected from patients' medications and health status, and/or derived from various daily medical processes.

（2）Source of real-world data

Common sources of real-world data in China include:

1) Health Information System (HIS): similar to EMR/HER, digital patient records including structured and unstructured data fields, such as patient demographics, clinical characteristics, diagnosis, treatment, laboratory tests, safety and clinical outcomes.

2) Medicare system: structured data such as basic patient information, medical service utilization, prescriptions, billing, medical claims, and planned health care.

3) Disease Registry System: a database of patients with specific (usually chronic) diseases, often derived from a cohort registry of the disease population in the hospital.

4) China ADR Sentinel Surveillance Alliance (CASSA): the use of electronic data from medical institutions to establish an active monitoring and evaluation system for the safety of drugs and medical devices.

5) Natural population cohort database: the (to be) established natural population cohort and special disease cohort database.

6) Omics-related databases: databases that collect information on the physiology, biology, health, behavior, and possible environmental interactions of patients, such as pharmacogenomics, metabolomics, and proteomics.

7) Death registration database: a database formed by death registries jointly confirmed by hospitals, centers for disease control and prevention (CDC), and department of household registration.

8) Mobile devices: mobile devices such as wearable devices that measure relevant data.

9) Other special data sources: databases created for special purposes, such as national immunization program databases.

（3）Data Quality Evaluation

The quality of real-world data is mainly assessed by its relevance and reliability.

1) Relevance: Important relevant factors to assess the suitability of real-world data for regulatory use include, but are not limited to:

① the inclusion of important variables and information related to clinical outcomes, such as drug use, patient demographic and clinical characteristics, covariates, outcome variables, follow-up duration, sample size, etc.;

②whether the definition of clinical outcome is accurate and the corresponding clinical significance is meaningful;

③Accurate and representative definition of target population;

④ The study hypothesis can be evaluated through the study protocol and statistical analysis plan.

2) Reliability: The reliability of real-world data is mainly evaluated by data integrity, accuracy, quality assurance, and quality control.

①Integrity: missing data problems are inevitable in the real-world setting, but the amount of missing should have a certain limit. For different studies, the degree of missing data may vary. When the proportion of missing data within a specific study exceeds a certain limit, there is a great deal of uncertainty about its impact on the study conclusion. At this time, it will be necessary to carefully assess whether the data can be used as real-world data that produce real-world evidence.

②Accuracy: the accuracy of the data is critically important and needs to be identified and verified against authoritative sources of reference. For example, the measurement of blood pressure requires the use of a calibrated sphygmomanometer, for which and the measurement process is subject to the operating specifications; whether the endpoint event is determined by an independent endpoint event committee, etc.

③ Quality Assurance: quality assurance refers to the prevention, identification, and correction of data errors that occur during the course of

the research. Quality assurance is closely related to regulatory compliance and should run through every aspect of data management that needs to have a corresponding Standard Operating Procedures (SOPs).

④ Quality Control: data collection, modification, transmission, storage, and archiving, as well as data processing, analysis, and submission, are all subject to quality control to ensure that the real-world data are accurate and reliable. It is necessary to develop a complete, normative and reliable data management process or protocol.

（4）Data criteria

Data standards, in the form that information technology systems or scientific tools can use, help ensure that the submitted data are predictable and consistent. In order to manage real-world data from multiple sources, it is necessary to convert the data into a common format with a generic formulation (e.g., terminology, vocabulary, coding scheme, etc.).

In addition, whether the quality of real-world data can support drug development depend on key factors including (but not limited to): whether there is a clear process and qualified personnel for data collection; whether a common defining framework, i.e., the data dictionary, is used; whether the common time frame for key data points collection is followed; whether a study plan, protocol and/or analysis plan related to the collection of real-world data have been established; whether the technical approach used for data element capture, including integration of data from various sources, data records of drug use, links to claims data etc., is adequate; whether patient recruitment minimizes the bias and reflects the true target population; whether data entry and transfer are useable and timely; and whether adequate and necessary patient protection measures such as patient privacy protection and regulatory compliance with informed consent are in place.

214   2. Real-World Evidence

215   Real-world evidence is clinical evidence about the use and potential

216 benefits or risks of medical products, obtained through the analysis of real-

217 world data. This definition is not limited in concept to obtaining evidence

218 through retrospective observational studies, but also allows prospective

219 access to a wider range of data to form evidence, through particular study

220 designs including pragmatic clinical trials (PCTs).

221   **3. Scenarios where real-world evidence supports drug**

222   **development and regulatory decisions**

223 Real world evidence may support drug development through a variety of

224 ways, covering pre-marketing clinical development and post-marketing

225 evaluation. Any use of real-world evidence for the purpose of product

226 registration will require adequate communication in advance with

227 regulatory authorities to ensure alignment on the study objectives and

228 methodology.

229   1. Treatment for rare diseases

230   In addition to the challenges in subject recruitment, clinical trials for

231 rare disease also face difficulties in the choice of control arm, given the

232 few or lack of treatment options. Therefore, external controls established

233 based on real world data in natural disease cohorts can be considered.

234   External controls are primarily used for non-randomized single-arm

235 trials, as a historical or in-parallel control. Historical external controls are

236 based on real-world data obtained earlier; parallel external controls are

237 based on data from disease registries constructed simultaneously with the

238 single-arm trial. The use of external controls should take into account the

239 impact of the heterogeneity and comparability of the target population on

240 the corresponding real-world evidence.

241   2. Revision of indications or drug combination labeling

242   For drugs that are already marketed, long-term clinical practice may

243 find it necessary to expand the indication, and RCTs are often utilized to

244 support the indication expansion. When an RCT is not feasible or when

245 evidence it generates is not optimal, a PCT could be a reasonable choice.

246 For example, clinical practice may find that a new drug for diabetes can

247 potentially benefit patients with cardiovascular diseases (such as heart

248 failure). In that case the subject recruitment into an RCT will be difficult

249 with potential ethical issues and therefore the use of a PCT design may be

250 more feasible.

251   In terms of pediatrics medication, there are often cases of off-label

252 usage in clinical practice. For that reason, the use of RWE in supporting

253 the expansion of targeted population is also a viable strategy in drug

254 development.

255   A typical use of real-world evidence to support the development of

256 Bevacizumab, a humanized monoclonal antibody of the vascular

257 endothelial growth factor (VEGF). In 2015, Bevacizumab was approved in

258 China in combination with chemotherapy (carboplatin and paclitaxel) for

259 the first-line treatment of late stage unresectable advanced, metastatic or

260 recurrent squamous non-small cell lung cancer. However, the real-world

261 use of chemotherapy with Bevacizumab also includes Pemetrexed in

262 combination with platinum, Gemcitabine and Cisplatin. In October 2018,

263 Bevacizumab was approved to expand the treatment regimen with a

264 combination of platinum-based chemotherapy, based on the strong

265 supporting evidence from three real-world studies. These studies

266 retrospectively analyzed patient data from three hospitals and showed that

267 the combination of Bevacizumab with platinum-based chemotherapy

268 significantly prolonged PFS and OS compared with chemotherapy alone,

269 and no new safety issues were identified. This finding was consistent with

270 global population data. In addition, relevant real-world studies have also

271 provided data in different patient subgroups such as those with EGFR

272 mutations or brain metastases, confirming the efficacy and safety of

273 Bevacizumab combination therapy from multiple perspectives.

274   3. Post-marketing evaluation

275   Due to factors such as limited sample size, short study duration, strict

276 enrollment criteria, and standardization of intervention, drugs approved

277 based on RCTs usually have limited safety information, lack of

278 generalization of efficacy conclusions, less optimal drug regimen, and

279 insufficient health economic benefits. As a result, there is a need to use

280 real-world data for more comprehensive assessment of these aspects of the

281 approved drugs, and to refine the decision making based on the real-world

282 evidence from natural populations on a continuous basis.

283   For example, a drug for cardiovascular diseases has been approved in

284 more than 50 countries/regions worldwide. In the multi-regional clinical

285 trials that supported it approval, small number of Chinese subjects resulted

286 in limited number of cardiovascular events and short drug exposure in the

287 Chinese subgroup. This has led to greater variability in the efficacy results

288 in the Chinese population. As an overseas marketed drug with clinically

289 urgent needs in China, to further evaluate the efficacy of this compound in

290 Chinese patients, the applicant plans to conduct a prospective,

291 observational, post-marketing real-world study to evaluate the

292 combination of the compound with standard treatment versus standard

293 treatment alone, in the prevention of major adverse cardiovascular events

294 (MACE) in Chinese patients with cardiovascular disease.

295   4. Clinical development of traditional Chinese medicine hospital
296   preparations

297   Traditional Chinese medicine prepared and used in hospitals have
298 been widely used clinically for a long time without being approved for
299 marketing. This is a unique phenomenon in China. For the clinical research

and development of such drugs, if real-world research and randomized controlled clinical trials can be combined, scientific and feasible clinical R&D and regulatory decision-making pathways can be further explored.
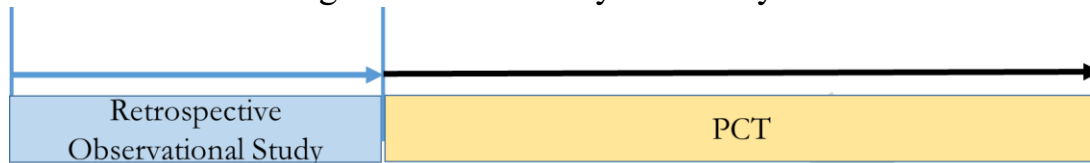
For the development of traditional Chinese medicine hospital preparation, there exist multiple R&D strategies that utilize real-world evidence. Figures 2 and 3 outline two potentially possible pathways. The pathway that combines observational studies and RCTs is illustrated in Figure 2. Specifically, stage 1 starts with retrospective observational studies. At this stage effort should be made to collect as much as possible existing real-world data related to the use of the product including all possible covariates, develop data cleaning rules, identify possible controls, assess data quality, and conduct comprehensive and detailed analyses using appropriate statistical methods. If the retrospective observational studies show that the drug has potential benefits for patients in clinical use, it may proceed to the next stage of the development, otherwise the process should be terminated. In stage 2, prospective observational studies can be conducted. Based on the stage 1 research, this second stage can be more carefully designed in terms of several aspects, including data acquisition and its system, data quality control, data cleaning rules, and clearer definition of controls. Once this stage 2 prospective observational research has progressed to certain phase, and if the data are consistent with the results of stage 1 retrospective observational studies by continuing to show clinically meaningful benefits, a third stage of RCT can be conducted in parallel. If needed, a pilot RCT may be conducted first to acquire sufficient information to support the design of the primary RCT. However, if existing evidence from previous observational studies is deemed sufficient, a confirmatory RCT may be designed and conducted directly. In terms of timing, the duration of the RCT may be covered by the stage 2 prospective observational studies, which can be completed at the same time as the RCT or extended for some time after the end of the RCT, depending on the maturity of the real-world evidence.
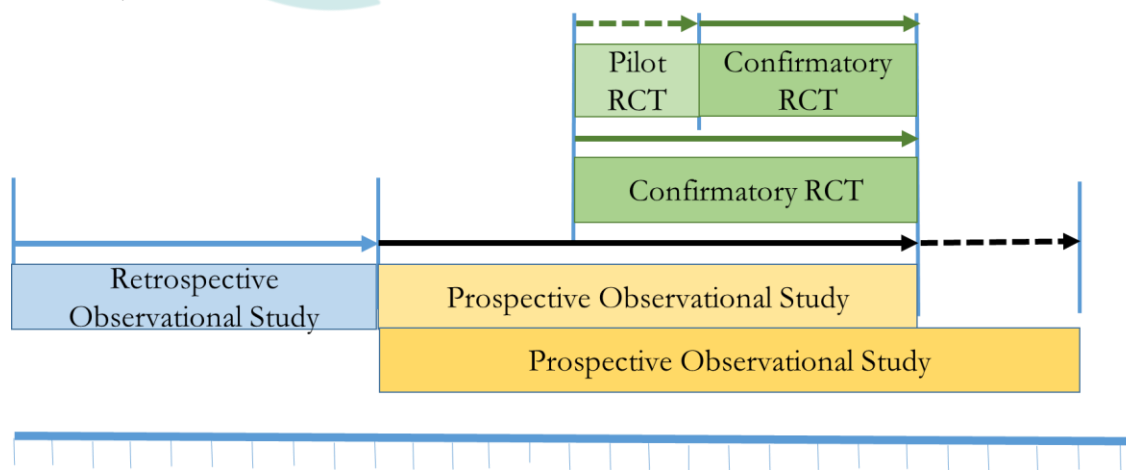
Figure 2 Potential development pathway for traditional Chinese medicine hospital preparations

Another potentially possible pathway, which combines observational studies with PCTs, is outlined in Figure 3. In the first stage, retrospective observational studies are conducted first. If it is concluded that the drug has potential benefits in clinical practice, it may proceed to the second stage, otherwise the process should be terminated. The second offstage consists of a PCT research, which provides evidence that can be used to support the evaluation of the drug's clinical efficacy and safety.

Figure 3 Potential development pathway for traditional Chinese medicine hospital preparations

5. Guiding clinical trial design

Compared with other potential applications, using real-world evidence to guide clinical trial design has more practical utilization. For example, the two potential pathways for the development of hospital-prepared traditional Chinese medicines described in the previous section have used the real-world evidence generated by retrospective observational studies, including for example the disease natural history, the disease prevalence in the target population, the effectiveness of standardized treatments, and the distribution and variation of key related covariates, to provide a basis for the next stage study design. More generally, real-world evidence can provide valid reference for inclusion and exclusion criteria, parameters for sample size estimation, and determination of non-inferiority margins, etc.

6. Identify the target population

Precision medicine aims to better predict the therapeutic benefits and risks of drugs to specific populations (subgroups), and real-world evidence based on real-world data provides the possibility for the development of precision medicine. For example, due to the limited sample size, regular clinical trials often ignore or have limited power to consider subgroup effects in the research plan. This prevents important information on potential treatment responders or high-risk populations with serious side

365 effects from being fully recognized. Through a thorough analysis of real-

366 world data, the treatment benefits and risks in different subgroups can be

367 more adequately assessed, and hence real-world evidence can be obtained

368 to support more precise identification of the target population.

369 The identification of biomarker is critical for preclinical and early
370 clinical studies of targeted therapies. Using real-world information such as
371 omics data, public gene bank information, and related clinical data in
372 population cohorts, real-world evidence can be generated through various
373 contemporary data mining techniques such as machine learning, which can
374 in consequence support the precise identification of population for targeted
375 therapies.

376 **4. The Basics of Real-World Research Design**

377 1. Pragmatic clinical trials

378 Pragmatic Clinical Trials (PCT), also known as practical clinical trials,

379 refer to clinical trials that are designed and conducted in an environment

380 close to the real-world clinical practice. They represent a type of study

381 between RCTs and observational studies. Unlike RCTs, PCT interventions

382 can be either standardized or non-standardized; subjects in the PCTs can

383 be randomized or allocated per pre-defined criteria; the inclusion criteria

384 for the subjects are often less restrict and considered more representative

385 of the target population, and the evaluation of intervention outcomes may

386 not be limited to clinical efficacy and safety. On the other hand, unlike

387 observational studies, PCTs are intervention studies, although the

388 interventions are often designed with additional flexibility.

389 Since a PCT needs to consider the impact of all potential factors,

390 including especially various biases and confounding factors, its study

391 design and statistical analysis are usually complicated, and the required

392 sample size can be much larger than a regular RCT design. PCTs, when

393 randomization is utilized, will reduce the impact and biases of the

394 confounders and thus provide a generally speaking robust causal inference.

395 In addition, PCTs do not adopt blinding in most cases, therefore sufficient

396 attention should be paid in estimating and adjusting the resulting detection

397 bias. Since PCTs are conducted in a setting close to real clinical practice,

398 the evidence obtained by PCTs is considered as the most reasonable and

399 practice real-world evidence compared to other research types.

400   2. Single-arm trial using real world data as control

401   The use of external controls has limitations, mainly including

402 different medical environments, changes in medical technology over time,

403 different diagnostic criteria, different outcome measures, different baseline

404 condition of patients, diverse interventions, data quality, etc. These

405 limitations result in additional challenges in the comparability of research

406 subjects, the accuracy of research results, the reliability and extrapolation

407 of research conclusions.

408   To address these limitations, it is first necessary to ensure that the

409 collected data meet the relevant quality requirements of real-world data.

410 Secondly, in terms of design, the use of parallel external controls is

411 generally superior to historical controls. Prospective parallel external

412 controls can use disease registration models to ensure that data records are

413 as complete and accurate as possible. Third, appropriate methods shall be

414 adopted for statistical analysis, such as the Propensity Scores (PS) method

415 and Virtual Matched Control method.

416   3. Observational studies

417   The data collected from observational studies are undoubtedly the

418 closest to the real world, but their most notable limitations are the existence

419 of various biases, data quality is difficult to guarantee, and observational

420 and unobserved confounding factors are difficult to identify. These

421 challenges leave the study conclusion with large uncertainty.

422   Whether the data collected from observational studies are appropriate

423 to generate real-world evidence to support regulatory decisions depend on

424   a few areas of focus: ①What are the data characteristics? (e.g., collection

425   of relevant endpoints, consistency of records, description of missing data,

426   etc.) ②What are the characteristics of the research design and analysis?

427   (e.g., is there an appropriate positive control? Is the non-inferiority design

428   applicable considering potential untested confounders as well as potential

429   measurement variability?) ③What sensitivity analyses and statistical

430   diagnostic methods are pre-determined to analyze real-world data?

431   The key technique for analyzing real-world data from observational

432   studies is causal inference. The statistical analysis methods commonly used

433   in real-world studies are summarized in Appendix 2.

434   **5. Evaluation of Real-World Evidence**

435   The evaluation of real-world evidence should follow two main

436   principles: whether the real-world evidence can support the scientific

437   questions that need to be answered; and whether the existing real-world

438   data can be scientifically analyzed to obtain the required real-world

439   evidence.

440   1. Real world evidence and the scientific questions it supports

441   Prior to the decision to use any evidence including real world

442   evidence, the scientific questions under evaluation should first be clearly

443   defined. For example, the safety considerations for the use of drugs in

444   combination with other drugs after marketing; the expanded indications

445   for approved products; and the establishment of robust and reliable

446   historical controls for a single arm clinical trial. The original intention of

447   using real-world evidence should be considered: is it because the

448   corresponding scientific question is facing real world, or it is because

449   traditional clinical trials cannot be effectively implemented. If the latter,

450   whether or not the real-world evidence can replace traditional clinical

451   trials, answer the same questions and arrive at robust conclusions, should

452   be used as important guidelines for measuring real-world evidence

453   applications.

454   2. How to transform real-world data to real-world evidence

455   To answer this question, a few key factors need to be considered:①

456   The research environment and data acquisition need to be close to the real

457   world, such as a more representative target population, diversity of

458   interventions compatible with clinical practice, or natural selection of

459   interventions; ②Use of appropriate controls; ③More comprehensive

460   evaluation of drug effectiveness; ④Effective bias control, such as the use

461   of randomization, harmonization of measurement and evaluation

462   methods, etc.; ⑤Appropriate statistical analyses, such as the correct use

463   of causal inference methods, reasonable handling of missing data,

464   adequate and sufficient sensitivity analyses, etc.; ⑥Reasonable

465   interpretation of results; ⑦Consensus among the key stakeholders.

466   Finally, it should be emphasized again that all study designs,
467   assumptions, and specific definitions and methodologies relevant to the
468   generation of real-world evidence should be clearly defined in advance in
469   the study protocol. In the meanwhile, any use of real-world data and
470   evidence with the ultimate expectation of drug registration would require
471   sufficient communication with regulatory authorities in advance, in order
472   to ensure mutual agreement on study objectives and methods. Post-hoc
473   remedial data citation, definition, analysis, and interpretation are generally
474   not acceptable for regulatory decisions.

475

References

[1] Sun YX, Wei FF, Yang Y. Opportunities and Challenges in Using Real-World Evidence to Support Regulatory Decision Making for Drug and Medical Devices [J]. Chinese Journal of Pharmacovigilance, 2017(06):353-358.

[2] Yang W, Xie YM. An Interpretation of AHRQ "Registries for Evaluating Patient Outcomes: A User's Guide (Second Edition)" [J]. China Journal of Chinese Materia Medica, 2013,38(18):2958-2962.

[3] CFDA. Opinions on Deepening the Reform of the Evaluation and Approval System and Inspiring Innovation of Drugs and Medical Devices[S]. 2017.

[4] Beckie TM, Mendonca MA, Fletcher GF, et al. Examining the Challenges of Recruiting Women Into a Cardiac Rehabilitation Clinical Trial[J]. Journal of Cardiopulmonary Rehabilitation & Prevention, 2009,29(1):13-21, 22-23.

[5] Berger M, Daniel G, Frank K, et al. A FRAMEWORK FOR REGULATORY USE OF REAL-WORLD EVIDENCE[M]. 2017.

[6] CenterWatch. The value of real world evidence in rare disease research[EB/OL]. https://www.iqvia.com/-/media/library/experts/the-value-of-real-world-evidence-in-rare-disease-research.pdf?vs=1&hash=8E3A6A07E4ABA0E05F72B64D4789281134EF1A72.

[7] Chalkidou K, Tunis S, Lopert R, et al. Comparative effectiveness research and evidence-based health policy: experience from four countries[J]. Milbank Q, 2009,87(2):339-367.

[8] Daniels M. Streptomycin treatment of pulmonary tuberculosis[J]. Med Times, 1949,77(6):251.

[9] Dreyer NA. Advancing a Framework for Regulatory Use of Real-World Evidence[J]. Therapeutic Innovation & Regulatory Science, 2018,52(3):362-368.

[10] EMA. Draft scientific guidance on post-authorization efficacy studies[EB/OL]. (2015-11-30)https://www.ema.europa.eu/en/news/supporting-better-use-medicines.

[11] EMA. New cell-based therapy to support stem cell transplantation in patients with high-risk blood cancer[Press Release][EB/OL]. https://www.ema.europa.eu/en/news/new-cell-based-therapy-support-stem-cell-transplantation-patients-high-risk-blood-cancer.

[12] FDA. Use of Real World Evidence to Support Regulatory Decision Making for Medical Devices. [S]. 2017.

[13] FDA. Framework For FDA's Real-world Evidence Program[EB/OL]. https://www.fda.gov/downloads/ScienceResearch/SpecialTopics/RealWorldEvidence/UCM627769.pdf.

[14] Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process[J]. Am J Prev Med, 2001,20(3 Suppl):21-35.

[15] Institute OM. Initial National Priorities for Comparative Effectiveness Research[M]. Washington, DC: The National Academies Press, 2009.

[16] IMI GetReal. http://www.imi-getreal.eu/About-GetReal/Overall-objectives.

[17] James S. Importance of post-approval real-word evidence[J]. European Heart Journal Cardiovascular Pharmacotherapy, 2018,4(1):10.

[18] Jelínek T, Maisnar V, Pour L, et al. Adjusted comparison of daratumumab monotherapy versus real-world historical control data from the Czech Republic in heavily pretreated and highly refractory multiple myeloma patients[J]. Current Medical Research and Opinion, 2018,34(5):775-783.

518 [19] Makady A, de Boer A, Hillege H, et al. What Is Real-World Data? A Review of Definitions Based
519 on Literature and Stakeholder Interviews[J]. Value in Health, 2017,20(7):858-865.

520 [20] Olariu E, Papageorgakopoulou C, Bovens SM, et al. Real World Evidence in Europe: A Snapshot
521 of its Current Status[J]. Value in Health, 2016,19(7):A498.

522 [21] Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence - What Is It and What Can It
523 Tell Us?[J]. The New England journal of medicine, 2016,375(23):2293-2297.

524 [22] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research
525 for decision making in clinical and health policy. JAMA 2003; 290 (12): 1624-1632

526 [23] USA. National Institutes of Health Health Care Systems Research Collaboratory[EB/OL]. https:/ /
527 www .nihcollaboratory .org/about-us/ Pages/ default .aspx.

528 [24] USA CONGRESS. H.R.34—21st Century Cures Act[EB/OL]. https://www.congress.gov/bill/
529 114th-congress/house-bill/34/text?q=%EF%BC%857B%EF%BC%8522search%EF%BC%8522%
530 EF%BC%853A%EF%BC%855B%EF%BC%852221st+Century+Cures+Act%EF%BC%8522%E
531 F%BC%855D%EF%BC%857D&r=3.

532 [25] Velentgas P, Dreyer NA, Nourjah P, et al. Developing a Protocol for Observational Comparative
533 Effectiveness Research: A User's Guide.[J]. Methods, 2013.

534 [26] Victor L Serebruany, Moo Hyun Kim, Thomas A Marciniak, Worldwide reporting of fatal
535 outcomes after ticagrelor to the US Food and Drug Administration, European Heart Journal -
536 Cardiovascular Pharmacotherapy, Volume 4, Issue 1, January 2018, Pages 6–
537 9, https://doi.org/10.1093/ehjcvp/pvx024

538 [27]von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies
539 in Epidemiology (STROBE) statement: guidelines for reporting observational studies[J]. Journal of
540 Clinical Epidemiology, 2008,61(4):344-349.

541 [28] White R. Building trust in real-world evidence and comparative effectiveness research: the need for
542 transparency[J]. Journal of Comparative Effectiveness Research, 2017,6(1):5-7.

543

544

545

Appendix 1: Glossary

Patient Registry: A system of collecting standard clinical and other data, using an observational research approach, to evaluate specific disease, condition, or specific outcome in the exposed population, for one or more predefined scientific, clinical, or policy objectives.

Single-arm (One-arm) Clinical Trial: A non-randomized clinical trial where only the experimental group is set up. A single-arm trial, usually uses external controls based on historical data or in a parallel manner.

Observational Study: A study that explores the outcomes in natural or clinical populations without active intervention, based on specific research objectives.

Clinical Trial: An interventional clinical research in which one or more interventions, possibly including placebo or other controls, are prospectively assigned to human subjects to assess the impact of these interventions on health-related biomedical or behavioral outcomes.

Retrospective Observational Study: An observational study based on historical data (generated before the start of the study).

Prospective Observational Study: An observational study based on data to be collected prospectively based on a preset research plan.

Comparative Effectiveness Research: A research method, by considering both individuals and the population in an environment as close as possible to the real world, that evaluates the clinical effectiveness and safety, social effects, and economic benefits of a particular intervention. Such evaluation helps key stakeholders such as patients, doctors, policy makers, and service consumers to improve healthcare services so that the most appropriate interventions or strategies can achieve the optimal outcomes in the most appropriate target population and timing.

The comparative effectiveness research is based on the real world, with a wide range of applications focusing on the decision-making for the natural population. Therefore, it is necessary to consider the impact of many factors on the outcome as comprehensively as possible. The designs are often more complex with usually a large sample size. In the meanwhile, there are clear requirements for valid statistical analysis in terms of causal inference.

Pragmatic Clinical Trial (PCT, a.k.a. Practical Clinical Trial): A clinical trial that is designed and conducted in an environment as close as possible to the clinical real world. It is a type of research between RCTs and observational studies. Unlike RCTs, PCT interventions can be either standardized or non-standardized; subjects in the PCTs can be randomized or allocated per pre-defined criteria; the inclusion criteria for the subjects

are often less restrict and considered more representative of the target

population, and the evaluation of intervention outcomes may not be limited

to clinical efficacy and safety. On the other hand, unlike observational

studies, PCTs are intervention studies, although the interventions are often

designed with additional flexibility

Data Standard：A set of rules on how to construct, define, format, or

exchange specific types of data between computer systems. Data standards
allow the submission of information to be predictable and consistent, and
in forms that information technology systems or scientific tools can use.
Randomized Controlled Trial (RCT): A clinical trial that utilizes a
randomization method in subject assignment to experimental and
appropriate control groups.
External Control: The control in clinical trials established based on data
outside the scope of the study, such as real-world data, to evaluate the
effects of the interventions under investigation. External controls can be
historical data or data obtained during the same period of time in a parallel
manner.
Medical Claims Data: A compilation of information on medical claims
submitted to insurance companies for access to claims for treatments and
other interventions.
Causal Inference: An inferential action, often based on real-world data,
that characterizes the causal relationship between interventions or
exposures to clinical or health outcomes, taking into account the effects of
various covariates and measured or unmeasured confounders and
controlling possible biases. Appropriate statistical models and analytical
methods should be used to establish the conclusions and causal relationship.
Real World Data (RWD): Data collected for a patient's health status
and/or derived from various routine medical processes that can be analyzed
to potentially form real-world evidence.

Real World Research/Study (RWR/RWS): As part of the CER, an

RWR/RWS refers to the collection of patient-related data in a real-world

environment to, through analysis, acquire clinical evidence (real-world

evidence) of the value and potential benefits or risks of medical products.

The main research type is observational, but it can also be pragmatic

clinical trials.

Real-World Evidence (RWE): Clinical evidence on the use and potential

622 benefits or risks of medical products obtained through the analysis of real-
623 world data.
624

Appendix 2: Common Statistical Methods for Real-World Research

As compared with RCTs, causal inference in real-world studies requires special attention to adjustment for confounding effects. Therefore, there is often a need for relatively complex statistical models and analytical methods. These methods include both classical statistical methods, such as conventional multivariate regression, and also some relatively more cutting-edge and sophisticated ones, such as propensity score matching and instrumental variables. This guidance only provides a general description of these statistical methods. More specific methods and application details can be found in the references provided and do not preclude the appropriate use of methods that are not described here.

1. Descriptive and Unadjusted Analyses

For descriptive analysis, appropriate descriptive statistics and statistical plots can be selected according to different data types, including: the range for continuous/numerical variables, dispersion and central tendency, counts and percentages for categorical variables, and graphs that describe the distribution of data. For real-world research, correct and effective descriptive statistical analyses can play an important role. For example, in disease registry cohort studies, stratified descriptive statistics of relevant covariates by levels of exposure factors can help to examine their distribution balance; in propensity score matched datasets, summary statistics by group of relevant covariates by exposure factors can help to identify imbalances in residuals after the matching, etc.

Univariate or unadjusted hypothesis testing, such as two-sample *t* test, can be used to assist in the identification of covariates related to exposure factors and/or study outcomes. For real-world studies, where possible confounding effects often need to be identified and considered from within a large number of covariates, extensive and comprehensive exploratory

653 analyses of relevant subject characteristics using descriptive statistics are
654 generally necessary.

655   2. Adjusted Analyses

656   （1） Selection of Covariates

657   - When using causal inference methods that adjust for covariates, the
658 selection of covariates selection is often a frontend question. Generally,
659 methods for covariate selection belong to one of the two categories. One
660 is, based on a causal network based on the exposure-to-outcome
661 relationship, to identify risk factors, confounders, intermediate variables,
662 time-varying confounders, collider variables, and instrumental variables.
663 Risk factors and confounders should be included as covariates in the model,
664 while the inclusion of intermediate variables, collider variables, and
665 instrumental variables should be avoided:

666   - **Risk Factor:** Baseline covariates that are predictive of the outcome
667 variable but have no effect on the level of the treatment/exposure factor. In
668 the causal relationship as shown in Figure 1, where R denotes the risk factor,
669 A indicates treatment or exposure factors, Y denotes the outcome
670 variable. Any adjustment to R does not affect the estimation of the effect
671 from $A \rightarrow Y$, i.e., such adjustment does not introduce or reduce bias, but
672 instead can improve the estimation precision and model efficiency.

673
674

$$A \longrightarrow Y$$
$$R \nearrow$$
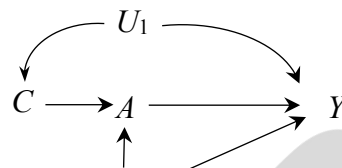
675   Figure 1. Causal relationship between risk factors (R) and outcome variables (Y)

676

677   - **Confounder**: Factors that affect both the level of
678 treatment/exposure factors and are predictive for outcome variables.
679 Certain confounders are measured, but there are also those that have not
680 been measured. In the causal relationship as shown in Figure 2, where A

681 indicates treatment or exposure factors, Y denotes the outcome variable, U_1

682 and U_2 are two unmeasured confounders, C represents a measured

683 confounder. In such case, (C can be a proxy variable for U_1 such that an

684 adjustment on C can eliminate the confounding impact of U_1 on the

685 outcome Y.

686

687

688

689 Figure 2. Causal relationship between measured confounding variables (C), and

690 unmeasured confounding variables (U_1, U_2)

691 **- Intermediate Variable:** Variables that may or may not be on the

692 treatment-outcome causal pathway after treatment or exposure. As shown

693 in Figures 3a and 3b, respectively, where A indicates treatment or exposure

694 factors, Y represents the outcome variable at the moment of measurement,

695 M denotes the intermediate variable, U indicates an unmeasured

696 confounder between M and Y. To estimate the total effect of $A \to Y$, in

697 case of Figure 3a, Fig. $A \to Y$ The total effects are divided into direct effects

698 and indirect effects. an adjustment on M may eliminate the indirect effect,

699 resulting in a biased estimation of the total effect; and in case of Figure 3b,

700 an adjustment on M A will introduce correlation between A and U,

701 which are originally independent, and consequently U into a confounding

702 factor in the causal relationship from $A \to Y$, and result in a biased

703 estimation of the total effect if no appropriate adjustment to U is made.

704 Also, especially in real-world studies, bias can be introduced due to over-

705 adjustment if the covariates being adjusted for are not those measured at

706 baseline.

707

708
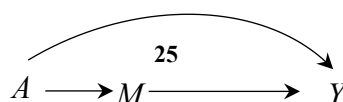
Figure 3a. Causal relationship between treatment (A) and outcomes (Y), with an

intermediate variable (M)

711

712

713

714

715

716
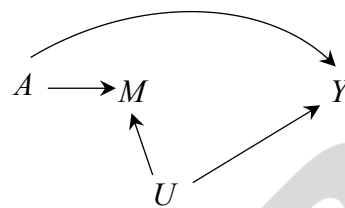
Figure 3b. Causal relationship between treatment (A) and outcomes (Y), with an

intermediate variable (M) and an unmeasured confounder (U)

719

- **Collider Variable**: In a causal relationship, if a variable has two independent parental nodes, then such variable is considered a collider. An adjustment to the collider may introduce correlation between the parental nodes, which are originally independent, and may bring additional confounding effect between the exposure and outcome, leading to a biased estimation of causal relationship. In a causal relationship as shown in Figure 4, where U_1 denotes an unmeasured confounder between variable L and outcome Y, U_2 denotes an unmeasured confounder between variable L and exposure factor A. In such case the variable L becomes a collider, with U_1 and U_2 being two independent parental nodes. An adjustment to L will introduce correlation between U_1 and U_2, which are originally independent, and may bring additional confounding effect between the exposure and outcome, leading to a biased estimation of the causal relationship between

26

733 $A \rightarrow Y$. It might be noted that the intermediate variable *M* in Figure 3b is

734 also a collider variable.

735

736

737

738

739
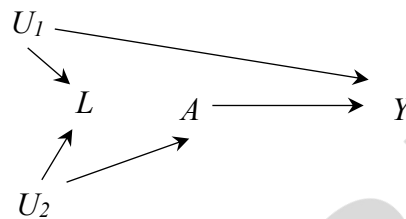
741 Figure 4.Causal relationship between treatment (*A*) and outcomes (*Y*), with a collider
742 variable (*L*)

743

744 **- Instrumental Variable:** A pre-treatment variable that has a causal

745 effect on the level of a treatment or exposure factor, but has no causal

746 association with the outcome variable other than indirectly affecting the

747 outcome variable through the effect of the exposure factor. The

748 instrumental variable is independent of confounders of exposure and

749 outcome. In a causal relationship as shown in Figure 5, where U

750 indicates the confounding factors between exposure factors, A and

751 outcome Y. In this case, Z is an instrumental variable. If the instrumental

752 variables are adjusted in a statistical analysis by being directly

753 incorporated into the model, the confounding impact of U might be

754 enlarged. On the other hand, certain analysis methods for

755 instrumental variables may be used to eliminate confounding effects

756 (see Estimation of instrumental variables).

757

758

759

760

761

$$U$$

$$Z \longrightarrow A \longrightarrow Y$$
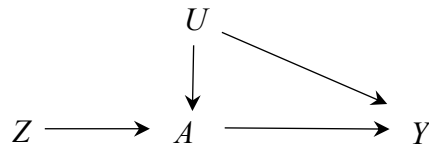
762

763   Figure 5.Causal relationship between treatment (A) and outcomes (Y), with an
764   instrumental variable (Z)

765

766   In reality, the true complete network structure is unknown. During

767   practical applications, when part of the causal structure is known, existing

768   covariate selection methods can be used, based on relevant professional

769   background knowledge, to adjust all observed baseline variables that may

770   be associated with the outcome, known outcome-related risk factors, and

771   all direct dependent variables for treatment or outcome. Another type of

772   covariate selection method is based on high-dimensional variable selection.

773   The principle is to use the degree of association between response variables

774   to empirically learn the correlation between variables from the data, and

775   select the variables related to the treatment factors and/or outcome

776   variables. Typical methods include forward selection, backward selection,

777   machine learning (such as Boosting, random forest, LASSO method, etc.)

778   and methods for automatic high-dimensional proxy adjustment. These two

779   types of methods can also be used in combination, i.e., first use

780   professional experience to identify a set of variables, and then use

781   appropriate empirical learning methods to further select the covariates to

782   be included in the final analysis model. This has the advantage of limiting

783   reliance on empirical learning, reducing the risk of over-adjustment while

784   also reducing confounding effect.

785   （2） Conventional Multivariate Regression

786   Regression analysis is a common strategy for adjusting the influence

787   of potential confounding variables and estimating treatment effects.

788   Generally, the variables to be adjusted are variables that are simultaneously

789   related to the study's treatment factors and outcome measures, and are

790   located before the treatment factors on the causal pathway. If an

791   intermediate variable is located on the treatment-to-outcome pathway, an

792   adjustment to it may eliminate some of the treatment effects, resulting in a

793   bias due to over-adjustment. There are extensive applications in

794   observational studies where traditional multivariate regression methods are

795   used to directly adjust for potential confounding and effect modifying

796   factors. These methods are also applicable in real-world studies. The use

797   of regression analysis methods requires attention to whether the

798   corresponding model assumptions are valid. For example, the linear

799   regression model assumes that the mean of the outcome variable is a linear

800   function with respect to the covariates. Therefore, this assumption needs to

801   be verified before choosing a linear regression approach. In addition,

802   whether to choose a regression model or other methods also depends on

803   the characteristics of the data. For example, if the number of events in a

804   study is sufficiently large (e.g., 8-fold or more than 10-fold the number of

805   covariates) relative to the number of covariates included in the model and

806   the treatment factor is not uncommon, the traditional logistic regression

807   approach is a reasonable option and may be considered as the primary

808   analysis method. Otherwise, alternative methods that are more appropriate

809   should be considered. In addition, all regression analysis methods have

810   potentially the risk in extrapolation, that is, the support of the fitted model

811   is actually outside the range of the sample data. To assess the risk of

812   extrapolation, statistical methods such as propensity scores can be used.

813   In the situation where the number of covariates is large, methods like

814 the stepwise approach may help in establishing a more efficient model.

815 However, it should always be noted that there may be certain level of

816 subjectivity, depending on the actual variable selection method and criteria

817 (e.g., p-value ≤ 0.1 for the corresponding parameter of interest). Also, for

818 covariates with a meaningful but relatively modest effect on disease risk,

819 the final model identified using independent variable selection methods

820 may miss these important covariates. Furthermore, the use of a stepwise

821 regression approach may lead to an underestimation of the standard error

822 in the estimation of the model parameters. Another strategy is to use

823 composite covariates such as Propensity Score (PS) or Disease Risk Score

824 (DRS) in the regression. In cases where the outcome event is relatively rare

825 (eg, less than 8-fold of the number of covariates), the propensity score

826 method is often superior to the traditional logistic regression method;

827 however, in cases of rare treatment/exposure (ie, only a small number of

828 subjects in a particular treatment group) but the number of outcome events

829 is large, the traditional logistic regression method is generally superior to

830 the PS method.

831   （3）Propensity Score

832   The propensity score method, proposed by Rosenbaum and Rubin, is

833 a method that adjusts the effect of confounders in the situation where a

834 large number of covariates exist. Let X denote all observed covariates, T

835 indicates the treatment or exposure factors of interest ($T = 1$ indicates

836 exposure), then the propensity score is defined as the probability that an

837 observed subject receives a certain treatment (or exposure) under the

838 observed covariate condition $\mathrm{PS} = \mathrm{Pr}[T = 1 \mid X]$. The propensity score

839 provides a composite summary of the effects of characteristic variables and

840 reflects the level of balance of all observed covariates between the two

groups. Rosenbaum and Rubin have demonstrated that, if the adjustment for raw covariates effectively controls the confounding effects, adjusting only the propensity scores based on these covariates is also sufficient to control for confounding effects. Propensity scores can often be estimated by regression models, such as commonly used logistic regression models with observed covariates as independent variables and treatment as dependent variables:

$$\text{logit}[P(T = 1 \mid X)] = \alpha_0 + \alpha_1 x_1 + ... + \alpha_p x_p + e$$

Propensity score methods are particularly appropriate in cases where treatment (or exposure) factors are common but outcome events are rare, or where multiple outcomes may exist. Propensity-Score Matching, Stratification/Subclassification, Inverse Probability of Treatment Weighting (IPTW), and the method of including Propensity Score as the sole covariate in the statistical model for adjustment analysis are all commonly used.

When utilizing the propensity score for causal estimation, it is important to first judge whether the covariate distribution is balanced between treatment groups for patients with a propensity score close to each other. The methods of judgment include, but are not limited to, visual inspection of the distribution of propensity scores across treatment groups after PS adjustment, or a statistical test of subject covariates across treatment groups. If the coincidence of the propensity score distribution between different groups is not high, the effect estimate obtained from the adjusted analysis using the propensity score remains at the risk of bias. Remediation schemes such as restricting the range of study subjects to overlapping regions of the distribution of propensity scores across groups may be considered in case of poor coincidence.

868   When possible, matching is a good application method for propensity

869   scores. If it can be coupled with the previously mentioned methods that

870   limit the range of the study subject, the overlap of propensity score

871   distributions among groups may be further improved. In addition, if the

872   summary results of the between-group equalization of all study covariates

873   after matching are provided, such as plotting the statistics or calculating

874   the standardized differences for each covariate before and after adjustment

875   (after-adjustment standardized difference is usually expected to be lower

876   than 20%), and comparing them with the results of the covariate balance

877   of randomized clinical trials, it will be helpful to evaluate the matched

878   effect. However, propensity score matching methods can only control the

879   known and observed covariates. Their impact on unknown or unobserved

880   confounders, the effect of the balancing, and the robustness of the analysis

881   results will need to be evaluated using other approaches. Note that the

882   standard error of the causal effect estimate based on the matched design

883   will be different from the unmatched case.

884   Covariates included in the propensity score model should be the

885   confounding variables or those associated with the outcome variables.

886   Otherwise, the variance of the estimator will increase if only the variables

887   that are related to the exposure factor are included. Traditional regression

888   adjustment method and propensity score matching method each has

889   advantages and disadvantages. The former does not guarantee that the

890   study covariates are balanced, and the latter may lead to a decrease in

891   sample size. Therefore, further sensitivity analysis is very necessary.

892   （4）Disease Risk Score

893   Disease risk scores are similar to propensity scores and are a

894   composite measure based on all covariates. Let X denote all observed

895   covariates, T denote the treatment or exposure factors of interest ($T = 1$

denote exposure), then the disease risk score is defined as the probability of an outcome event under the assumption of no treatment/exposure or specific covariate conditions $\mathrm{DRS} = \Pr[Y = 1 \mid X, T = 0]$.

Generally speaking, the methods for estimating DRS can also fall into one of the two categories. The first type of method uses all observations of the study sample in fitting a regression model, taking treatment and covariates as independent variables, study outcomes as the dependent variable. For example, for a logistic regression model

$$\mathrm{logit}[P(Y = 1 \mid X, T)] = \alpha_0 + \alpha_1 x_1 + ... + \alpha_p x_p + \beta T + e,$$

Once fitted, the DRS values for each study subject can be calculated by substituting the covariate values into the model and setting the treatment to be the control group. With that, the treatment-to-outcome causal effect can be estimated by analyzing the data stratification by DRS. The second type of method uses only the study data of the control (non-exposed) group, historical data before the treatment factor occurs, or sample data without (or low incidence of) treatment factor to fit the DRS model. For example, for a logistic regression model as follows

$$\mathrm{logit}[P(Y = 1 \mid X)] = \alpha_0 + \alpha_1 x_1 + ... + \alpha_p x_p + e$$

Once fitted by using only the control group data, the DRS values for each study subject can be calculated by substituting the covariate values into the model.

Different from the PS method, for studies where outcome events are common but treatment (exposure) factors are rare or there may be multiple levels of treatment, the DRS approach is a good option to balance baseline disease risk across groups. In particular, in case of multiple levels of treatment (exposure) factors, where some of them are sparse, it is often recommended that the DRS method be selected instead of the PS method.

923　　　　　（5）Instrumental Variables

924　　　　One common limitation of the previously mentioned methods

925　(conventional regression, PS, DRS) is that only measured confounding

926　factors can potentially be controlled. On the other hand, the causal

927　inference based on instrumental variables does not require the specification

928　of what confounders/covariates to be adjusted, and so the impact of

929　unmeasured confounders can also be potentially controlled during the

930　analysis. A variable is considered an instrumental variable if it is related to

931　the treatment factor, and the effect on the outcome variable can only be

932　achieved by influencing the treatment factor without being correlated with

933　the potential confounders. After the instrumental variables are identified,

934　even with the existence of unmeasured confounders, the treatment-to-

935　outcome causal effect can be estimated by separately estimating the effect

936　of instrumental variables on the treatment and that on the outcome, and

937　then contrasting the two estimated effects.

938　　　　The biggest challenge in using instrumental variables to estimate

939　causal effects lies in the identification of suitable instrumental variables.

940　　　　First, instrumental variables cannot be associated with any observed

941　or unobserved confounders of treatment and outcome, otherwise . Second,

942　instrumental variables cannot have a direct effect on the outcome but only

943　an indirect impact through the treatment-to-outcome pathway, otherwise

944　the estimated causal effect may again be biased. Finally, instrumental

945　variables need to be highly correlated with the treatment factor. If the

946　correlation is too weak, in which case the variable is referred to as a weak

947　instrumental variable, the corresponding estimator of the causal effect may

948　perform poorly especially with small sample size, with large estimation

949　variation and potentially enlarged bias. Variables that satisfy the above

950　three conditions can be used as instrumental variables to estimate the

951 treatment-to-outcome causal effects. In practice, however, it might be

952 difficult to find variables that meet the above conditions, and there is no

953 particularly appropriate statistical method to evaluate whether these

954 conditions are completely satisfied.

955 Once instrumental variables are identified, the estimation of causal

956 effects usually utilizes a two-stage least-squares approach:

957 Stage 1: Fit a regression that links the treatment factors (A) and

958 instrumental variables (Z) $E[A|Z] = \alpha_0 + \alpha_1 Z$ and obtains the predicted

959 value of the treatment factor $\hat{E}[A|Z]$;

960 Stage 2: Build a regression that links the outcome variables Y with the

961 predicted value of treatment factors based on the instrumental variable, i.e.,

962 $E[Y|Z] = \beta_0 + \beta_1 \hat{E}[A|Z]$. Wit that, the regression coefficient $\hat{\beta}_1$ is an unbiased

963 estimate of the treatment-to-outcome causal effect.

964 The selection of instrumental variables is particularly important to the

965 estimation of causal effects. The impact of instrumental variables to the

966 treatment factors is expected to be homogeneous and consistent across the

967 entire study population. Otherwise, the estimated causal effect may not

968 represent the average causal effect in the overall population, but only the

969 effect within a certain subpopulation in which the impact of instrumental

970 variables is meaningful, i.e., the Local Average Treatment Effect (LATE).

971 It should also be noted that when the treatment factor is a non-continuous

972 variable, the estimated causal effect and the estimated error obtained by the

973 two-stage least squares method may have potential statistical bias.

974 3. Missing data consideration

975 The missing data problem is often inevitable in real-world studies.

976 Not only the outcome variables, but covariates may also be missing. This

977 makes it difficult to assess the comparability of treatment groups, which in

turn may lead to biased estimation of treatment effect. Investigators and the Sponsor should optimize the trial design to minimize the missingness rate.

Before conducting the primary analysis, an attempt should be made to determine whether the data are truly missing and, if yes, the reason for the missing. First of all, no data does not mean that the data are missing. For example, a patient did not have a certain examination, or a doctor did not perform certain examination at all. These data should not exist, nor should they be considered as missing data. This is common in real-world data. If there indeed exist missing data, an analysis of the missingness mechanism should be performed. Generally, there are three types of missing mechanism: Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR). Missing completely at random means that the missing data are independent of the measured or unmeasured covariates and outcome variables. Let Y denote the outcome variable (Y_{mis} for missing data and Y_{obs} for the observed data) and X the treatment and associated baseline covariates. Let R be an indicator for missingness ($R = 0$ for missing and $R = 1$ for non-missing), then the missing completely at random can be expressed as: $\Pr[R \mid X, Y_{obs}, Y_{mis}] = \Pr[R]$. Missing at random refers to the case that the missing data are independent of the potential outcome conditional on the measured covariates and outcome variable, i.e. $\Pr[R \mid X, Y_{obs}, Y_{mis}] = \Pr[R \mid X, Y_{obs}]$. Finally, if the data are missing not at random, the missing data may depend on the value of the missing data themselves, and may also be related to the measured covariates and outcome data.

1004   For missing data problems, selecting the appropriate methods for

1005   imputation and analysis is an effective way to avoid bias and information

1006   loss. If no imputation is performed and only observations with no missing

1007   data are analyzed, then regardless of the missing mechanism, the study

1008   efficiency will be reduced due to reduced sample size. When the

1009   characteristics of subjects with missing data differ from those with

1010   complete data, excluding missing data also results in biased treatment

1011   effect estimates. Imputation methods should be established based on

1012   appropriate assumptions on missing mechanisms and clinical problems. In

1013   general, for missing completely at random cases, imputation with sample

1014   means or predicted values of generalized estimating equations will suffice.

1015   Or, the analysis can be based on the complete data only. For missing at

1016   random cases, a statistical model can be constructed to predict the value of

1017 $E[Y \mid X, R = 1]$ with covariates. Multiple Imputation (MI) methods are

1018   generally recommended, such as traditional regression model methods,

1019   Markov Chain Monte Carlo (MCMC) methods, and Fully Conditional

1020   Specifications (FCS). In addition, for the missing at random case in a

1021   longitudinal study, the Mixed Model for Repeated Measures (MMRM) can

1022   be used. It should be noted that although the MMRM method is

1023   recommended for handling missing data, it does not impute the missing

1024   data. For the case of missing not at random, the Pattern Mixture Models

1025   (PMM) method can be applied to construct different statistical models for

1026   missing and non-missing data.

1027   In addition, there is a single value imputation method, which utilizes

1028   simple principles and is easy to implement. However, even under the

1029   assumption of missing at random, the single value imputation cannot

1030   guarantee a valid result, and the variability of missing data is not

1031   considered, either. Therefore, it is generally not recommended for the

1032   primary analysis.

1033   In observational studies with missing covariates, according to the

1034   specific pattern of missingness, a number of existing statistical methods

1035   may be considered, including complete data analysis, multiple imputation

1036   (MI) and propensity score (PS).

1037   The complete data analysis method performs statistical analysis by

1038   excluding patients with missing covariates (or patients with missing

1039   follow-up in cohort studies). This will reduce the power of the statistical

1040   test. Note that this method can provide unbiased estimates of treatment

1041   effect only when the missing data are not correlated with the study design

1042   nor the treatment factors.

1043   Multiple imputation method (MI) takes into account the uncertainty

1044   of the missing values and impute the missing data multiple times with

1045   possible values. As previously stated, the MI is typically performed under

1046   the assumption of missing at random, implying that the missing data may

1047   potentially associate with observed covariates but not with unobserved

1048   variables. Since MI produces multiple datasets, two methods can be used

1049   for estimating propensity scores, i.e., estimating based on each imputed

1050   data, or estimating based on all imputed data. Rubin's method may be used

1051   to combine multiple treatment effects that simultaneously account for

1052   variability within and between imputed data.

1053   It needs to be clarified that the assumption on any of the three types

1054   of missing mechanism (MCAR, MAR, and MNAR) are generally not

1055   verifiable and can only be justified through a correct description and

1056   understanding of the data collection process.

1057   It should be noted that there is no optimal way to deal with missing

1058   data, and no method can yield the same robust and unbiased estimates as

1059 the one based on the complete data. The best strategy to deal with missing

1060 data is not to plan how to analyze the data, but rather to control the chance

1061 of missing data by optimizing the study design and implementing it with

1062 good practice.

1063   4. Sensitivity Analysis

1064   The various causal inference methods mentioned previously all have

1065 their own applicable conditions and model assumptions. For example, the

1066 propensity score matching method does not need to satisfy the model

1067 assumptions of the instrumental variable method, while the instrumental

1068 variable method is able to handle situations where the propensity score

1069 method is not applicable (eg, with the existence of unmeasured

1070 confounders). Therefore, for the choice of statistical methods for causal

1071 inference, sensitivity analyses can be performed to evaluate the robustness

1072 of the analysis by using different statistical models, thereby prioritizing

1073 statistical models with good robustness. A more comprehensive sensitivity

1074 analysis can be found in the Guidelines for the Development of an

1075 Observational Effectiveness Comparative Study Plan.

1076   Finally, like other confirmatory studies, the interpretation of analysis

1077 results for real-world studies should be as comprehensive, objective,

1078 accurate, and adequate as possible, not only emphasize statistical

1079 significance (such as P-values and confidence intervals), but also focus on

1080 Clinical practical significance; not only   depend on the final conclusion,

1081 but also on the logic and integrity of the entire evidence chain that supports

1082 the conclusion; not only depend on the overall effect, but also on the

1083 subgroup effect. In addition, a detailed elaboration on the control and

1084 impact of various possible biases and confounding should be provided as

1085 well.

1086

Appendix 3: Chinese-English Vocabulary

English	中文
21st Century Cures Act	21 世纪治愈法案
FDA Adverse Event Reporting System, FAERS	FDA 不良事件报告系统
Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease	阿尔茨海默病疾病进展和临床试验评估的数据驱动模型新方法的意见书
Standard Operation Procedure, SOP	标准操作规程
Standardized Differences	标准化差
Patient Registry	病例登记
Single-arm/One-arm Trial	单臂临床试验
Electronic Medical Record, EMR	电子病历
Electronic Health Record, EHR	电子健康档案
Multiple Imputation, MI	多重填补
Missing Not At Random, MNAR	非随机缺失
Stratification/Subclassification	分层法
Risk Factor	风险因子
Instrumental Variable	工具变量
Observational Study	观察性研究
Center for Drug Evaluation, CDE	国家药监局药品审评中心
CASSA	国家药品不良反应监测哨点联盟
Patient Reported Outcome, PRO	患者报告结局
Retrospective Observational Study	回顾性观察性研究
Confounder	混杂因素
Baseline Observation Carried Forward, BOCF	基线观测值结转
Disease Risk Score, DRS	疾病风险评分
Regulatory Compliance	监管合规性
Local Average Treatment Effect, LATE	局部平均处理效应

Clinical Trial	临床试验
Markov Chain Monte Carlo, MCMC	马尔科夫链蒙特卡洛模拟
The American Recovery and Reinvestment Act	美国经济复苏刺激法案
Federal Food, Drug, and Cosmetic Act, FD&C	美国联邦食品，药品和化妆品法
Food and Drug Administration, FDA	美国食品药品监督管理局
Pattern Mixture Models, PMM	模式混合模型
Last Observation Carried Forward, LOCF	末次观测值结转
Inverse Probability of Treatment Weighting, IPTW	逆概率加权方法
European Medicines Agency, EMA	欧盟药物管理局
Collider Variable	碰撞节点变量
Prospective Observational Study	前瞻性观察性研究
Propensity Scores, PS	倾向性评分
Propensity-Score Matching	倾向性评分匹配法
Hot-Deck Imputation	热卡填补
International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, ICH	人用药品注册技术要求国际协调会
Pharmaceutical and Medical Devices Agency, PMDA	日本医药品医疗器械综合机构
Time-varying Confounder	时变型混杂因素
Comparative Effectiveness Research, CER	实效比较研究
Pragmatic Clinical Trial, PCT	实用/实操临床试验
Adaptive Licensing Pilot	适应性许可试点项目
Data Standard	数据标准
Randomized Controlled Trials, RCT	随机对照临床试验
Missing At Random, MAR	随机缺失
Conditional Mean Imputation	条件均值插补
External Control	外部对照

Extrapolation	外推
Missing Completely At Random, MCAR	完全随机缺失
Completeness	完整性
Health Information System, HIS	卫生信息系统
Vascular Endothelial Growth Factor, VEGF	血管内皮生长因子
Medical Claims Data	医保数据
Causal Inference	因果推断
Real World Data, RWD	真实世界数据
Real World Research/Study, RWR/RWS	真实世界研究
Real World Evidence, RWE	真实世界证据
Quality Assurance	质量保证
Quality Control	质量控制
Intermediate Variable	中介变量
Mixed Model for Repeated Measures, MMRM	重复测量混合效应模型
Accuracy	准确性
Worst Observation Carried Forward, WOCF	最差观测值结转

1088