

Supporting Clinical Trial Recruitment using Real-World Data

Yi-Chien Lee

RWD Programming Lead

Pfizer Inc.



24-Mar-2026

Agenda

01 RCT recruitment challenges

02 How RWD can help

03 Examples

04 Limitation

04 Conclusion

RCT recruitment challenges

Estimate RCT population

- Traditionally, feasibility questionnaires are distributed to investigators to estimate
- Often lack alignment with actual conditions
- Keeping clinical sites active with low recruitment is highly inefficient

Regulator-requested RCT

- Regulatory authorities may require to conduct RCTs within specific populations
- Important to present regulators with appropriate supporting evidence if recruitment is not practical

How RWD can help

- Estimate the attainable sample size using empirical data, rather than relying solely on investigators' projections
- Explore criteria to balance feasibility with clinical relevance
- Can be presented to regulatory bodies to demonstrate when conducting an RCT in a specified population is impractical

Examples

Despite these significant advantages, translating RCT eligibility criteria into operational business rules for use with RWD involves considerable complexity.

We will discuss some examples:

- Age
- Diagnosis of interest
- Lab values

AGE

RCT

- Age as of screening / randomization

RWD

- Only the year of birth available
- Need to specify the exact point at which patient age should be assessed

Example

RCT criteria

- Patients at least 18 years old

RWD proxy

- Patients at least 18 years old as of their last T2DM diagnosis on/before 31-Dec-2024

DISEASE / TREATMENT OF INTEREST

RCT

- Clinical term
- Physician discretion

RWD

- Billing codes
- ex. ICD-9, ICD-10 diagnosis codes; NDC, HCPCS, and CPT-4 codes
- Require 1IP/2OP to find true case

Example

RCT criteria

- Patients with T2DM

RWD proxy

- Patients with at least two outpatient diagnoses (at least 30 days apart) or one inpatient diagnosis of T2DM between 2020-2024.
- T2DM diagnosis codes:
 - ICD-9: 250.*
 - ICD-10: E11.*

LAB VALUES

RCT

- Typically from central lab or local lab
- Or based on what's available on the medical chart

RWD

- Claims data linkable to lab database; or EHR
- LOINC codes, or keyword search
- Require cleaning of unwanted records
 - Irrelevant test
 - Invalid value
 - Invalid unit
- Standardization of units are also need before applying the cutoff
- Require substantial exploration and iterative communication between scientists and programming teams.

Example

RCT criteria

- Patients with Total IgE ≥ 30 KU/L

RWD proxy

- Identify the IgE test
 - Use LOINC code 19113-0 or 2462-0
 - Also check using keyword search, including keywords such as: 'IMMUNOGLOBULIN E TOTAL(EXT)', 'BKR IGE TOTAL', 'TOTAL IMMUNOGLOBULIN IGE', etc
- Remove invalid units, ex. %?%
- Convert 'U/L' to 'KU/L' by multiplying by 1/1000
- Remove outliers (ex. values below 0.001 or above 100,000)

Limitations

- Information not available in RWD
 - Fertility
 - Severity scores
 - ex. Patient Global Activity (PGA), Child Health Assessment Questionnaire
- Generalizability

Take away

- Leveraging RWD enhances clinical trial recruitment by assessing feasibility, refining criteria, and providing regulatory evidence.
- RWD offers data driven estimates of eligible populations and aids site selection and recruitment challenges.
- Limitations include missing clinical details and representativeness concerns.
- Collaboration among scientists, clinicians, and programmers is vital to manage data complexities, ensure quality, and improve RWD integration for better trial outcomes.

Thank You

