Data & Statistical Sciences

Sensitivity Analysis in Deriving RWE from the Analysis of RWD

Yixin Fang, AbbVie

MBSW 2025

abbvie

Disclaimer

The comments provided here are solely those of the author and are not necessarily reflective of the positions, policies or practices of the author’s employer.

Outline

- **Introduction**
- Taxonomy of assumptions
- Targeted learning framework
- Sensitivity analysis methods
- Sensitivity analysis of unmeasured confounding
- Discussion

Definition of sensitivity analysis

- ICH guidance entitled “E9(R1) Statistical principles for clinical trials: Addendum: Estimands and sensitivity analysis in clinical trials”
  - Two themes: **estimand** and **sensitivity analysis**
- ICH E9(R1) defines **sensitivity analysis** as:

  “*A series of analyses conducted with the intent to explore the robustness of inferences from the main estimator to deviations from its underlying modelling assumptions and limitations in the data*”

  <u>Sensitivity analysis is for either of the two issues:</u>

  1. Deviations from its underlying modelling assumptions ☑
  2. Limitations in the data

Two types of studies we commonly conduct

Randomized Controlled Trials (RCT)

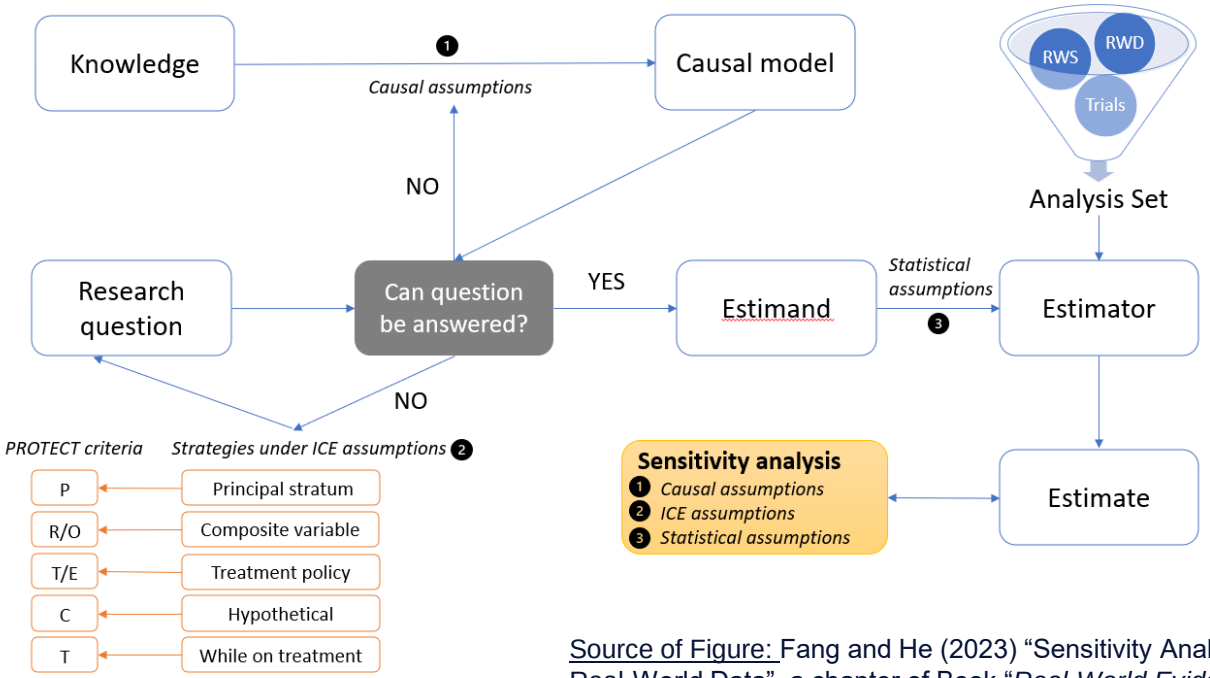
Non-interventional Studies (NIS)

Today's focus

Outline

- Introduction
- **Taxonomy of assumptions**
- Targeted learning framework
- Sensitivity analysis methods
- Sensitivity analysis of unmeasured confounding
- Discussion

Taxonomy of assumption



Source of Figure: Fang and He (2023) "Sensitivity Analysis in the Analysis of Real-World Data", a chapter of Book "*Real-World Evidence in Medical Product Development*" edited by He, Fang, and Wang (2023)

Three sets of assumptions

- (A1) Identifiability assumptions
  - For example:
    1. Consistency assumption
    2. No-unmeasured-confounding assumption
    3. Positivity assumption
- (A2) Assumptions behind missing data and intercurrent event handling
  - For example: Missing-at-random assumption
- (A3) Statistical model assumptions
  - For example: Generalized linear model; Mixed-effect model; Cox PH model

Outline

- Introduction
- Taxonomy of assumptions
- **Targeted learning framework**
- Sensitivity analysis methods
- Discussion

A cohort study

- Cohort $A = 1$ vs. cohort $A = 0$
- Pre-treatment covariates: W
- Outcome: Y
- Missing status: Δ (1: missing; 0: observed)

Potential outcomes: $Y^a, a = 1,0$

Causal estimand (e.g., average treatment effect---ATE):

$$\theta^* = E(Y^1) - E(Y^0)$$

Identifiability assumptions

- Consistency assumption: $Y^a = Y$, if $A = a$ and $\Delta = 0$
- No-unmeasured-confounding (NUC) assumption: $Y^a \perp\!\!\!\perp A|W$
- Positivity assumption: $P(A = a|W) > 0$
- Assumption on missing status: $Y^a \perp\!\!\!\perp \Delta|W, A$

Under these assumptions,

Causal estimand

$$\theta^* = E(Y^1) - E(Y^0) = E[E(Y^1|W)] - E[E(Y^0|W)]$$
$$= E[E(Y^1|A = 1, W)] - E[E(Y^0|A = 0, W)]$$
$$= E[E(Y^1|A = 1, \Delta = 0, W)] - E[E(Y^0|A = 0, \Delta = 0, W)]$$
$$= E[E(Y|A = 1, \Delta = 0, W)] - E[E(Y|A = 0, \Delta = 0, W)] = \theta$$

Statistical estimand

Targeted learning

- Our goal estimate statistical estimand

$$\theta = E[E(Y|A=1,\Delta=0,W)] - E[E(Y|A=0,\Delta=0,W)]$$

- We propose to consider the targeted minimum loss estimator (TMLE): $\hat{\theta}_{TMLE}$
- TMLE has nice properties:
  - ✓ We make no extra model assumption besides those identifiability assumptions
  - ✓ Asymptotic consistency: $\hat{\theta}_{TMLE} \to \theta$, as $n \to \infty$
  - ✓ Asymptotic efficiency: the asymptotic variance of $\hat{\theta}_{TMLE}$ is the smallest among all the regular and asymptotic linear estimators

Outline

- Introduction
- Taxonomy of assumptions
- Targeted learning framework
- **Sensitivity analysis methods**
- Sensitivity analysis of unmeasured confounding
- Discussion

Causal gap

- Our goal is to estimate causal estimand θ^*
- Based on observed data and TMLE, we estimate θ
- To explore how well we estimate θ^* via estimating θ, we conduct **sensitivity analysis** to evaluate those identifiability assumptions

Sensitivity analysis

- Honestly write down all the underlying assumptions
  - The consistency assumption
  - The no-unmeasured-confounding (NUC) assumption
  - The positivity assumption
  - Missing at random (MAR) assumptions
- Conduct sensitivity analysis for all these assumptions or some of them
- For example,
  - *E-value method* for the NUC assumption
  - *Tipping-point method* for the MAR assumption

Reference: Fang and He (2023) “Sensitivity Analysis in the Analysis of Real-World Data”, a chapter of Book “Real-World Evidence in Medical Product Development” edited by He, Fang, and Wang (2023)

Sensitivity analysis within the targeted learning framework

Only need to conduct sensitivity analysis for those causal assumptions

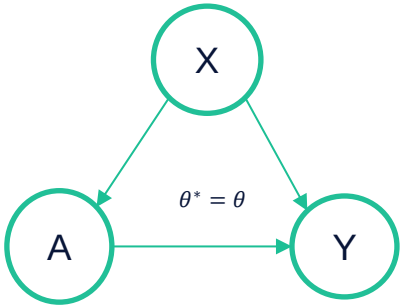
There is no extra model assumptions behind the estimator process

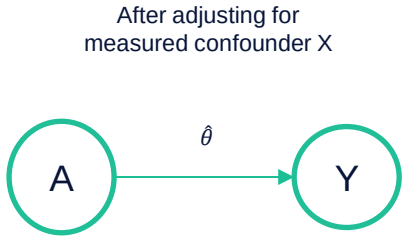
Outline

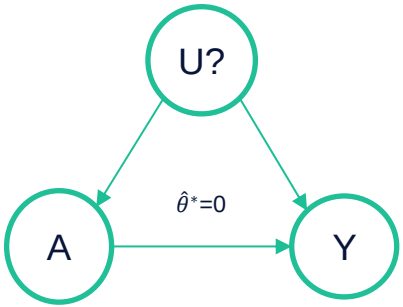
- Introduction
- Taxonomy of assumptions
- Targeted learning framework
- Sensitivity analysis methods
- **Sensitivity analysis of unmeasured confounding**
- Discussion

Unmeasured confounding

- X – measured confounders; U – unmeasured confounder
- A – treatment (or exposure); Y – outcome (a.k.a. endpoint)







Three categories

Category One

- Based on internal data only
- Parametric methods

Remark: Assume models that parametrize the relationships between unmeasured confounder(s) and one or more of treatments, outcome, and measured confounders

Category Two

- Based on internal data only
- Non-parametric methods

Remark: Without any underlying model for relationships among outcome, treatment, and unmeasured confounders. For example, **E-value** is a method belonging here

Category Three

- Based on internal data and some level of external data

Remark: Some methods can rely on external data only. These 3 categories are not exhaustive. There are methods which don't belong to these categories and need new assumptions instead of NUC assumption

Definition:
Internal data are the data that used in the main estimator for the estimand
External data are the extra data that are used or collected for the purpose of sensitivity analysis

E-value

Ann Intern Med. 2017 Aug 15;167(4):268-274. doi: 10.7326/M16-2607. Epub 2017 Jul 11.

Sensitivity Analysis in Observational Research: Introducing the E-Value

Tyler J VanderWeele [1], Peng Ding [1]

Abstract

Sensitivity analysis is useful in assessing how robust an association is to potential unmeasured or uncontrolled confounding. This article introduces a new measure called the "E-value," which is related to the evidence for causality in observational studies that are potentially subject to confounding. The E-value is defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and the outcome to fully explain away a specific treatment-outcome association, conditional on the measured covariates. A large E-value implies that considerable unmeasured confounding would be needed to explain away an effect estimate. A small E-value implies little unmeasured confounding would be needed to explain away an effect estimate. The authors propose that in all observational studies intended to produce evidence for causality, the E-value be reported or some other sensitivity analysis be used. They suggest calculating the E-value for both the observed association estimate (after adjustments for measured confounders) and the limit of the confidence interval closest to the null. If this were to become standard practice, the ability of the scientific community to assess evidence from observational studies would improve considerably, and ultimately, science would be strengthened.

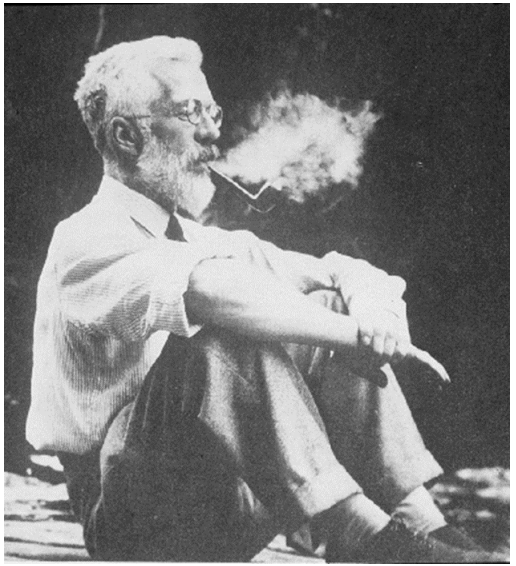
Famous example: Smoking and lung cancer

- Hammond and Horn (1958) “Smoking and Death Rates—Report on 44 Months of Follow-up of 187,783 Men: 2. Death rates by cause”; JAMA; 166(11):1294–1308
- Variables
  - Y: lung cancer status (1: Yes; 0: No)
  - A: smoking status (1: Yes; 0: No)
  - X: a list of measured confounders such as age, gender
- Hammond and Horn (1958) obtained the estimated relative risk of smoking A on lung cancer Y after adjusting for measured confounders X

$$\hat{\theta}=10.73 \text{ (95\% } CI\text{: 8.02, 14.36)}$$

Fisher's argument (or his excuse for being a smoker)

- Hammond and Horn (1958) obtained the estimated relative risk of smoking A on lung cancer Y after adjusting for measured confounders X: $\hat{\theta}$=10.73 (95% CI: 8.02, 14.36)
- Fisher (1958) "Cancer and Smoking"; *Nature*;182-596. Smoker Fisher thought the smoking-lung cancer relationship could be explained by **a genetic variant U**
- If we were to agree with Fisher, we should ask: how large should the association between the genetic variable with both smoking and lung cancer be to explain away the above estimate $\hat{\theta}$=10.73 (95% CI: 8.02, 14.36)?



E-value to resolve the above debate

❖ With an observed relative risk $\hat{\theta}$ between A and Y, we have that if RR_{UY} and RR_{AU} are greater than

$$E - value = \hat{\theta} + \sqrt{\hat{\theta} \times (\hat{\theta} - 1)}$$

then this confounding would explain away the observed exposure-outcome association, but weaker confounding could not

❖ Hammond and Horn (1958) obtained the observed smoking-lung cancer relative risk

$$\hat{\theta} = 10.73\ (95\%\ CI: 8.02, 14.36)$$

❖ E-value for the estimate is 20.9

$$10.73 + \sqrt{10.73 \times (10.73 - 1)} = 20.9$$

❖ E-value for the CI estimate is 15.5

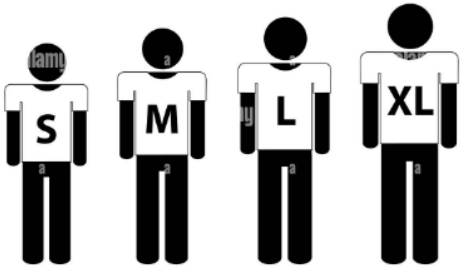
$$8.02 + \sqrt{8.02 \times (8.02 - 1)} = 15.5$$

“With an observed relative risk 10.73, an unmeasured confounder that was associated with the outcome and exposure by a relative risk of 20.9 each, above and beyond the measured confounders, could explain away the estimate, but weaker confounding could not”

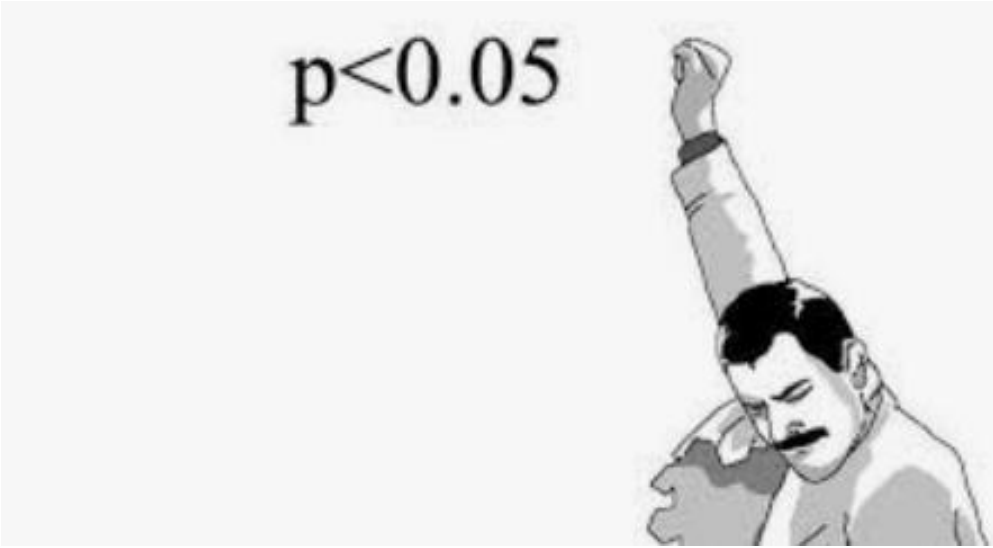
What is missing?

- It is easy to calculate $E-value = \hat{\theta} + \sqrt{\hat{\theta} \times (\hat{\theta} - 1)}$
- It is also easy to interpret E-value: **the larger E-value the more robust** the main inference to the violation of the NUC assumption
- But it is hard to reach conclusion on robustness

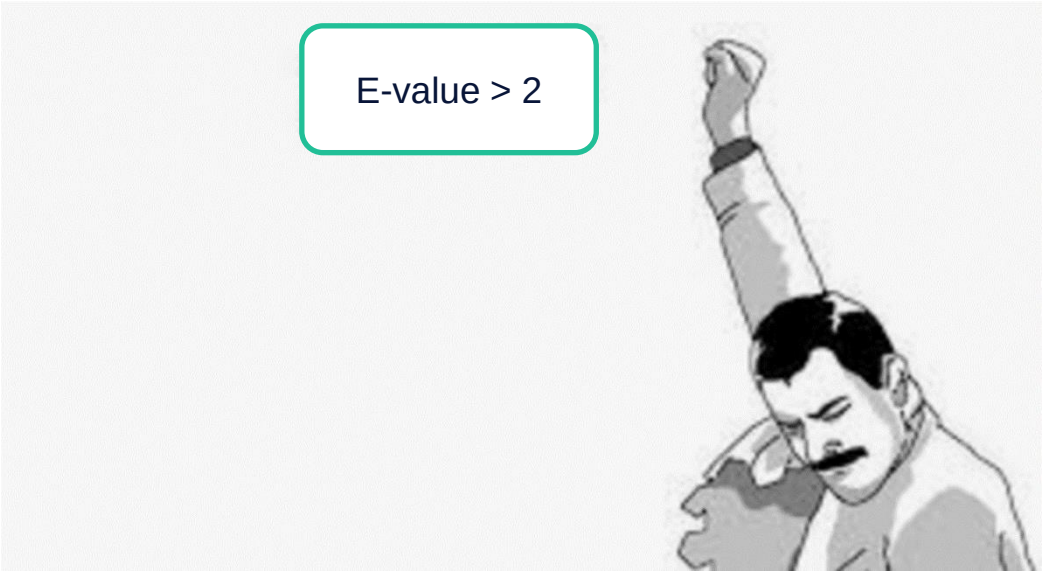
How large is large?



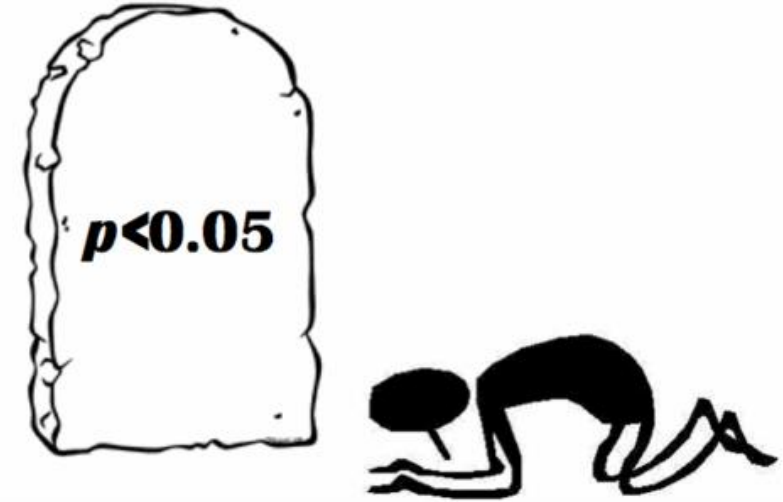
Think about p-value and 0.05



Do we have a benchmark for E-value to claim robustness?



In “p<0.05” we (don’t) trust?



When and how to claim whether the main result is robust?

Our proposals for you to write the sensitivity analysis section in your SAP (caution: not FDA's recommendation)

1. Specifying a predetermined threshold
2. Specifying a fixed threshold supported by historical results
3. Specifying a method for identifying a threshold based on internal data
4. Specifying a method for identifying a threshold based on external data

Reference: Fang et al. (2024) "Sensitivity analysis for unmeasured confounding in medical product development and evaluation using real-world evidence." Under journal review. (One manuscript contributed the ASA BIOP Section Real-World Evidence Scientific Working Group.)

A toy example

- Y is the response outcome (1 response; 0 otherwise)
- A is treatment variable (1 treatment; 0 control)
- X is measured confouder
- Estimand $\theta = P(Y^1 = 1)/P(Y^0 = 1)$
- Unadjusted estimate $\hat{\theta}_{un} = 1.87$, with 95% CI (1.64, 2.13)
- Adjusted estimate $\hat{\theta}_{adj} = 1.68$, with 95% CI (1.55, 1.83)
- E-value associated with point estimate 1.68 is 2.75
- E-value associated with the lower level 1.55 is 2.47

Proposal one: pre-specify a fixed threshold in the SAP

- In the SAP, we pre-specify a fixed threshold after discussion between the sponsor and the regulatory agency
- For example, after the discussion (say, between sponsor and regulatory), we pre-specify $\Delta = 2$ in the SAP, just like we would pre-specify $\alpha = 0.05$ as the threshold to claim statistical significance
- If the resulting E-value is larger than Δ (in the above toy example, the E-value associated with CI's lower level is 2.47>2), we claim that the result of the main estimator is robust to the deviation of the NUC assumption

Proposal two: pre-specify a fixed threshold supported by historical results in the SAP

- In the SAP, we pre-specify a fixed threshold supported by the preliminary results or the meta-analysis of the results of some similar studies provided in the literature, just like the way by which we pre-specify the non-inferiority margin
- For example, using two cohort studies reviewed in the literature with similar outcome variables and similar treatments, we calculate the ratio of the unadjusted estimate ($\tilde{\theta}_{un} = 1.75$) and the adjusted estimate ($\tilde{\theta}_{adj} = 1.40$)

$$B = \max\{\frac{\tilde{\theta}_{un}}{\tilde{\theta}_{adj}}, \frac{\tilde{\theta}_{adj}}{\tilde{\theta}_{un}}\} = 1.25$$

- Then use the above ratio to calculate a threshold, $\Delta = B + \sqrt{B(B-1)} = 1.81$
- Hence, we can pre-specify this $\Delta = 1.81$ supported by the preliminary results in the SAP. If the resulting E-value is larger than Δ (in the example, the E-value associated with CI's lower level is 2.47>1,81), we claim that the result of the main estimator is robust

Proposal three: pre-specify a rule for determining a threshold based on the internal data

- In the SAP, we pre-specify a fixed rule for determining a threshold using the results to be obtained from the current study
- For example, we pre-specify the following rule: based on the results to be obtained from the current study, the ratio of the unadjusted estimate and the adjusted estimate is to be calculated as $B = \max\{\frac{\hat{\theta}_{un}}{\hat{\theta}_{adj}}, \frac{\hat{\theta}_{adj}}{\hat{\theta}_{un}}\}$ and the threshold is to be determined as $\Delta = B + \sqrt{B(B-1)}$
- In the toy example, the current study provides $\hat{\theta}_{un} = 1.87$ and $\hat{\theta}_{adj} = 1.68$, we calculate the ratio as $B = 1.11$ and then determine the threshold as $\Delta = 1.46$. Since the E-value associated with CI's lower level is 2.47>1,46, we claim that the result of the main estimator is robust

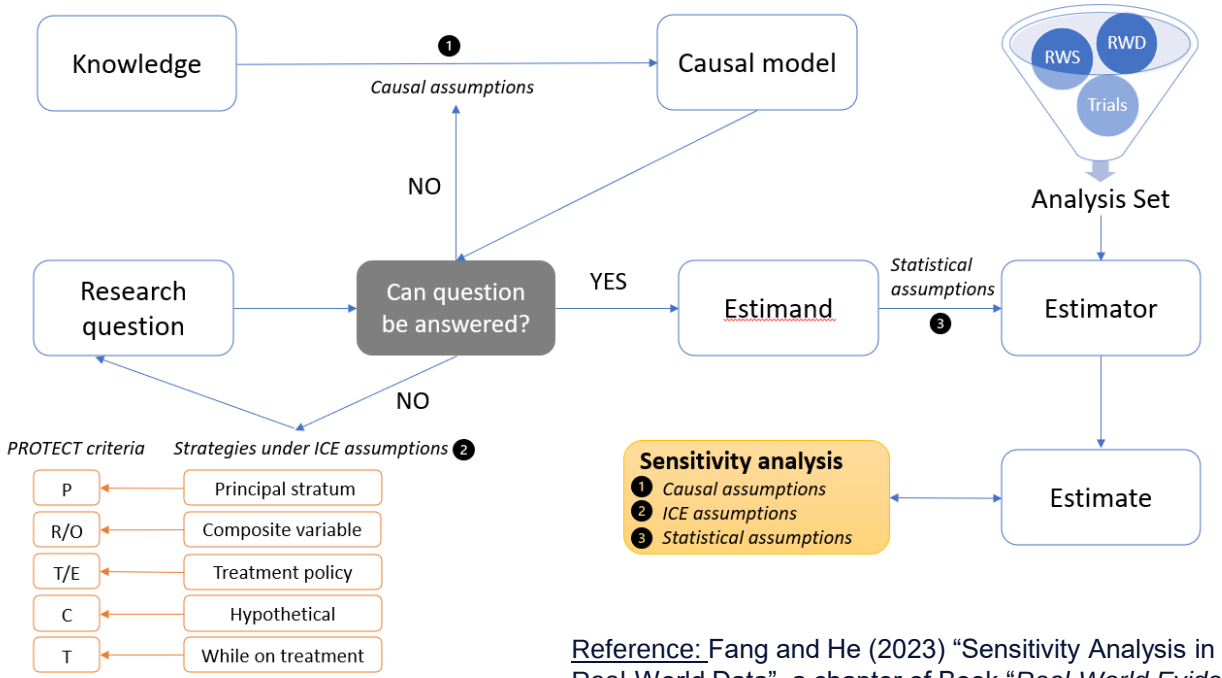
Proposal four: pre-specify a rule for determining a threshold based on external data

- In the SAP, we pre-specify a fixed rule for determining a threshold using some external data
- For example, we pre-specify the following rule: based on some external data of some unmeasured confounders (which are not collected in the current study but are collected from another data source for the purpose of sensitivity analysis), the new adjusted estimate is denoted as $\hat{\theta}_{new}$, the ratio of the new estimate and the adjusted estimate is to be calculated as $B = \max\{\frac{\hat{\theta}_{new}}{\hat{\theta}_{adj}}, \frac{\hat{\theta}_{adj}}{\hat{\theta}_{new}}\}$ and the threshold is to be determined as $\Delta = B + \sqrt{B(B-1)}$
- In the toy example, if combining the internal and external data provides $\hat{\theta}_{new} = 1.43$ (and $\hat{\theta}_{adj} = 1.68$ based on only the internal data), we calculate the ratio as $B = 1.17$ and then determine the threshold as $\Delta = 1.62$. Since the E-value associated with CI's lower level is 2.47>1.62, we claim that the result of the main estimator is robust

Outline

- Introduction
- Taxonomy of assumptions
- Targeted learning framework
- Sensitivity analysis methods
- Sensitivity analysis of unmeasured confounding
- **Discussion**

Why, where, and when do we need sensitivity analysis?



Reference: Fang and He (2023) "Sensitivity Analysis in the Analysis of Real-World Data", a chapter of Book "*Real-World Evidence in Medical Product Development*" edited by He, Fang, Wang (2023)


abbvie