An Overview of Statistical Methods for Handling Nonadherence to Intervention Protocol in Randomized Control Trials: A methodological review

Journal of Clinical Epidemiology 108 (2019) 121e131, December 2019

Mohammod Mostazira,*, Rod S. Taylorb, William Henleyb, Ed Watkins

Passara Jongkhajornpong, M.D.
Clinical Epidemiology

11.09.2020

ELSEVIER

Journal of Clinical Epidemiology 108 (2019) 121–131

Journal of Clinical Epidemiology

REVIEW

An overview of statistical methods for handling nonadherence to intervention protocol in randomized control trials: a methodological review

Mohammod Mostazir[a,*], Rod S. Taylor[b], William Henley[b], Ed Watkins[c]

[a]College of life and Environmental Sciences (CLES), University of Exeter, EX4 4QG, Exeter, United Kingdom
[b]Institute of Health Research, University of Exeter Medical School, University of Exeter, EX1 2LU, Exeter, United Kingdom
[c]Mood Disorders Research, School of Psychology, College of Life & Environmental Sciences, University of Exeter, EX4 4QG, Exeter, United Kingdom

Accepted 4 December 2018; Published online 7 December 2018

Abstract

Objective: To undertake a methodological review of statistical methods used in randomized controlled trials (RCTs) for handling intervention nonadherence.

Study Design and Setting: Bibliographic databases were searched using predefined search terms.

Results: A substantive number of identified studies (56%) were excluded as they only used naive per protocol analysis for handling nonadherence. Our review included 58 articles published between 1991 and 2015. A total of 88 methodological applications were made by these studies. The two most used methods were complier average causal effect (56%) and instrumental variable (23%) predominantly with the use of maximum likelihood (ML) estimators. These alternative applications typically produced treatment effects greater than the intention-to-treat effect but as their standard errors were larger there was no statistical difference between the methods.

Conclusion: A substantive proportion of RCTs rely on naive per protocol for handling nonadherence. Recent years have seen an increasing number of applications of more appropriate statistical methods, in particular complier average causal effect and instrumental variable methods. However, these later methods rely on strong underlying assumptions that may be vulnerable to violation. More empirical studies are needed that directly compare the usability and performance of different statistical methods for nonadherence in RCTs.

Keywords: Nonadherence; Noncompliance; Randomized controlled trial; Methodological review; Causal effect modeling; Statistical methods

Introduction

- Randomized controlled trials (RCTs) and systematic reviews of RCTs are the highest level of evidence for assessing the effects of health care interventions.
- Randomization creates equal groups based on observed and unobserved variables.
- However, not everyone assigned to the treatment may take the treatment.
- Challenges: nonadherence/noncompliance and loss to follow-up
- Associated with poorer patient outcomes (higher mortality)
- MA across 569 trials estimated an average treatment nonadherence rate of 25%.

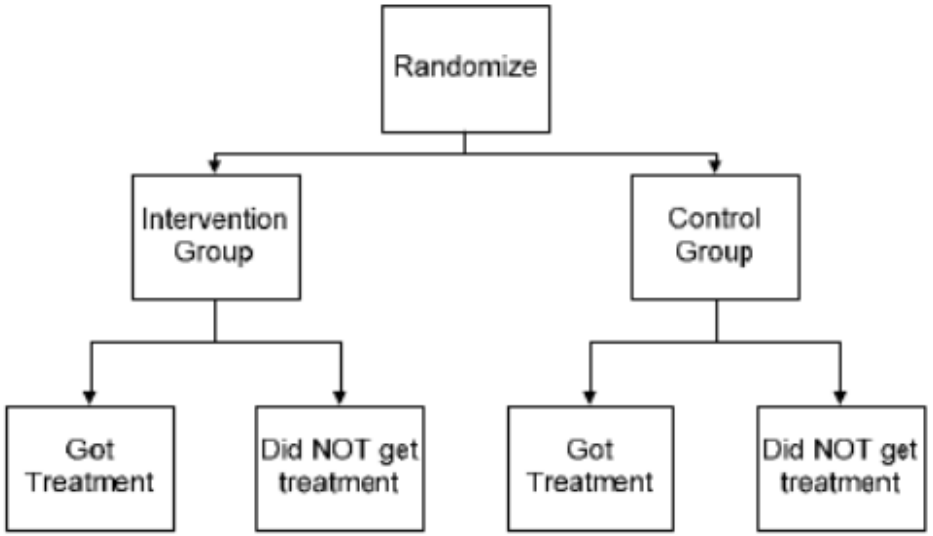
Evaluating effectiveness of an intervention in the nonadherence

- Consolidated standards of reporting trials [CONSORT] recommend the intention-to-treat (ITT).
- In case of non-compliance occurs; ITT provides an unbiased estimate of the effect of treatment allocation, rather than the effect of actual treatment use.
- If the treatment is effective, ITT analysis will underestimate the effect of treatment.

Alternative analysis of an intervention in the nonadherence

- 47% adopt per protocol (PP) analysis; outcomes are compared according to initial random allocation, excluding participants who do not adhere to the intervention protocol.
- As treated (AT) analysis; classifies patients according to the recieved intervention regardless to their random allocation.
- Both PP and AT are subjected to selection bias.

ITT vs PP vs AT



Interntion-to-Treat	YES	YES	NO	NO
Per protocol	YES	DROP	DROP	NO
As Treated	YES	NO	YES	NO

$$\hat{\delta}_{ITT} = \bar{y}_1 - \bar{y}_0$$

$$\hat{\theta}_{AT} = \bar{y}_{treated} - \bar{y}_{untreated}.$$

$$\hat{\theta}_{PP} = \bar{y}_{\text{observed compilers in the intervention}} - \bar{y}_{\text{observed compilers in the control}}.$$

Break randomization (serious selection bias)

Statistical framework for causal inference in RCTs

- Developed by Rubin, referred to as Rubin's causal
- Each participant is assumed to have a set of counterfactual outcomes
- Preserve randomization with the accounting for potential confounding
- Several methods developed for handling *nonadherence*
  - Instrumental variable (IV)
  - Complier average causal effect (CACE)
  - Structural mean models (SMMs)
- To date, there has been no comprehensive review of the use of these statistical methods and their pros and cons.

Objectives of the review

- 1) To assess the range of statistical methods reviewed and applied in RCTs to handle nonadherence
- 2) Review the relative pros and cons of these methods
- 3) Make pooled comparison of the treatment effects estimated by ITT and proposed statistical methods for handling nonadherence.

Methods

- Literature search
  - Database: EMBASE (OvidSP), PsycInfo (OvidSP), MEDLINE (OvidSP), CINAHL (EBSCOHOST), and Cochrane Library for methodological studies (Wiley Online Cochrane Library) from inception to June 2015.
  - Key terms: ‘‘intention to treat’’, ‘‘as-treated’’, ‘‘per protocol’’, ‘‘non-adherence’’, ‘‘complier average causal effect’’, ‘‘CACE’’ (and synonyms)
- **Include:** RCTs that reviewed statistical methods for handling nonadherence and applied to actual/simulated data
- **Exclude:** 1) no full publication, 2) no information on the statistical basis of the methods, 3) unrelated to nonadherence to intervention protocol

Data analysis & Presentation

- Descriptive approach for data presentation.
- For comparison of treatment effect between ITT and the proposed methods: compared whether the treatment effect by ITT was larger/smaller compared to the effect estimated by the proposed method (coded ‘‘yes/no’’).
- Absolute z-statistic was calculated for each method and the pooled mean z-statistic was compared between ITT and proposed methods.

Data analysis & Presentation

- This pooled comparison accounted for within-study variance by subtracting each proposed method z-statistic from ITT z-statistic before calculating pooled mean z-statistic, that is,

$$\frac{\sum_{i=1}^{n}\left(z_{pi}-z_{ITTi}\right)}{n},$$

Z is the z-statistic for proposed (p) or ITT
n is the total number of method applications

- The pooled z-statistic was also used to compare treatment effect between IV vs CACE method by metaregression accounting for nonadherence rate.
- Analyses were undertaken using statistical software Stata, version 15

Flow of studies through inclusion and exclusion process

Initial title/abstracts screened (n = 4,664)

Manual screening (n = 27)

Excluded (n = 4508)
- Title/abstract unavailable: 143
- Conference/symposium abstracts: 20
- Study protocol/design: 39
- Un-related to handling non-adherence/Not RCTs: 1605
- Per-protocol (PP) analysis: 2591 (56%)
- Modified intention-to-treat analysis (mITT): 84 (1.8%)
- As-treated (AT) analysis: 26 (0.56%)

Full-text assessed for eligibility (n = 183)

Excluded
- Methods applied for other purposes: 63
- Lacks methodological details/Pure application: 23
- Commentary/review/discussion: 36
- Methods not applied to any dataset:3

Total articles included (n = 58)

Characteristics of included RCTs

Table 1. Summary of included study characteristics

Characteristics	Number	Percent (%)
Number of articles	58	100
Year of publication (n = 58)	–	–
1991-1999	12	21
2000-2007	21	36
2008-2015	25	43
Journals (n = 58)	–	–
Statistics in medicine	20	34
Biometrics	8	14
Biostatistics	5	9
Journal of the American Statistical Association	4	7
Controlled Clinical Trials	2	3
Journal of the Royal Statistical Society	2	3
Psychological Methods	2	3

Higher at the later periods compared to the 1991-1999

IRR for 2000-2007: **1.56** (1.00 to 2.45), P < 0.05

IRR for 2008-2015: **1.72** (1.11 to 2.65), P < 0.01

IRR = incidence rate ratio

Most included studies were published in statistical/methodological journals.

Wide range of patients, intervention types, study sizes and outcome types (continuous/binary/count/time-to-event)

Characteristics of included RCTs

Table 2. Statistical methods (a) and their estimators (b) as stated by the authors, applied for handling nonadherence

I #	a) Methods	Method elaboration	Number (%)	b) Estimators	Estimator elaboration	Number (%)
1	CACE	Complier average causal effect model	49 (56)	ML	Maximum likelihood	29 (33)
2	IV	Instrumental variable model	20 (23)	MOM	Method of moments	19 (22)
3	SNMM	Structural nested mean model	7 (8)	IV	Instrumental variable estimator	17 (19)
4	ATR	Adjusted treatment received model	4 (5)	BI	Bayesian inference	9 (10)
5	RPSFTM	Rank preserving structural failure time model	3 (3)	G-estimator	G-estimator	5 (6)
6	C-PROPHET	Rank preserving structural failure time model	2 (2)	Cox-PH	Cox-proportional hazard estimator	4 (5)
7	CALM	Compliers proportional hazards effect of treatment with proportional Hazards model	1 (1)	WLS	Weighted least square	2 (2)
8	Cox-Reg1	Causal accelerated life model	1 (1)	GSMM	Generalized structural mean model estimator	1 (1)
9	Cox-Reg2	Regression adjustment with Cox-model	1 (1)	ISM	Intensity score method	1 (1)
				WGSNM	Weighted generalized structural mean model	1 (1)
	Total		88 (100)			88 (100)

Complier average causal effect (CACE)

- Introduced by Angrist et al. 1996 for estimating causal effects in the presence of nonadherence based on counterfactual outcome
- Potential adherence are stratified into **4 principal strata**
  - **"Compliers"** receive treatment when they are assigned to it
  - **"Never-takers"** do not receive treatment when they are assigned to it
  - **"Always-Takers"** always receive the treatment regardless of randomization
  - **"Defiers"** always do the opposite of what is assigned and *assumed to be nonexistent*

$$C_i = \begin{cases} n \text{ (never-taker)} & \text{if } M_i(1) = 0, \text{ and } M_i(0) = 0, \\ d \text{ (defier)} & \text{if } M_i(1) = 0, \text{ and } M_i(0) = 1, \\ c \text{ (complier)} & \text{if } M_i(1) = 1, \text{ and } M_i(0) = 0, \\ a \text{ (always-taker)} & \text{if } M_i(1) = 1, \text{ and } M_i(0) = 1. \end{cases}$$

Complier average causal effect (CACE)

- Apart from randomization and *stable unit treatment value assumption* (SUTVA)
- *SUTVA = the observation [potential outcome] on one unit should be unaffected by the particular assignment of treatments to the other units
- **Two key assumptions** need to be fulfilled CACE model

(1) Exclusion restriction (ER) = the effect of treatment assignment on outcomes *entirely operates through treatment compliance*

(2) Monotonicity assumption = no "defiers" who do the opposite of what they have been told (Angrist & Pischke, 2014 → rare or nonexistent)

Complier average causal effect (CACE)

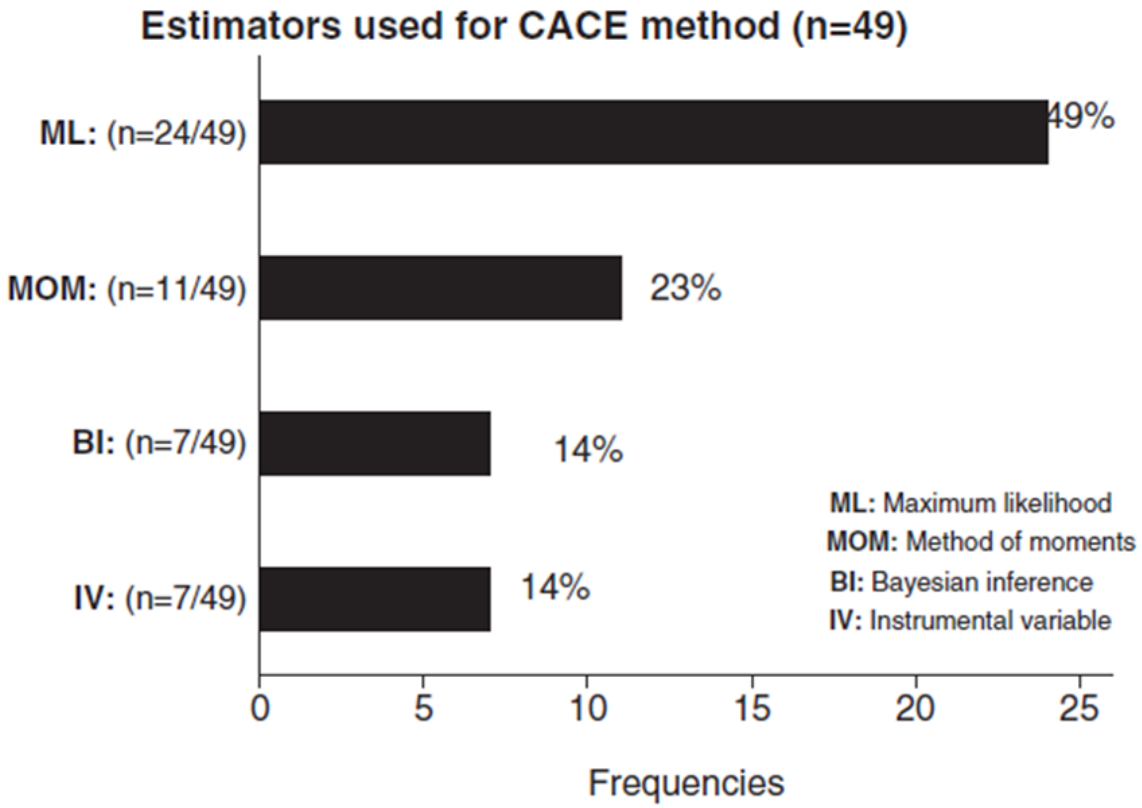
The CACE is then the difference between the observed outcomes among compliers in the intervention arm and the expected outcomes among the anticipated compliers in the control arm.

Compliance	Intervention (N = 200)	Control (N = 200)	Treatment Effect Estimates
Compliers (120/200 = 60%)	12/120 (10%)	*40/200-20/80 = 20/120 (16.7%)*	*CACE:* 10% - 16.7% = –6.7%
Non-compliers (80/200 = 40%)	20/80 (25%)	*20/80 (25%)*	*Per-Protocol :* 10% - 20% = –10%
Overall	32/200 (16%)	40/200 (20%)	*ITT:* 16% - 20% = –4%

Random assignment ensures that there is an equal proportion in both groups

CACE = ITT estimate/compliance rate in the intervention arm= -4/(0.6) = -6.7%

Complier average causal effect (CACE)



- Although the initial CACE estimator proposed by Angrist et al. was an IV estimator, other estimators can be applied in different settings.
- ML-base estimation was applied more often.
- ML estimates may be more efficient than 2SLS-based IV estimators.

Complier average causal effect (CACE)

CONS:

- The underlying assumptions (ER & monotonicity) are not easily testable.
- If compliance rate is very low, violation of the ER can cause a substantial bias of the results.
- In case of multiple arms, CACE may suffer from non identifiability issues/require complex modeling assumptions & Bayesian methods.
- Missing data add another level of complexity in presence of treatment nonadherence (authors should provide guidelines for handling missing data).

Instrumental variable (IV)

- IV is an exogenous variable that influences the outcome solely through a binary post treatment variable that identifies whether participants adhered to treatment or not.
- Typically in RCTs;
  - IV is the randomizing variable.
  - Participants' adherence status is the endogenous variable through which outcome is affected.
- The assumption that outcome solely depends on adherence status is equivalent to the ER assumption in the CACE.
- Therefore, in a two-arm trial design where post-randomization switching between arms is restricted, an IV estimates alternate CACE estimates, given same estimator applied.

Instrumental variable (IV)

- Typically, IV estimators are implemented with two stage least square (2SLS) estimators.

CONS:

- IV with 2SLS is valid only when missing data is ignorable.
- If the compliance rate is low, 2SLS-base IV estimator produces large effects and large variances compared to ITT, which makes it a less attractive estimator. *ML is a more efficient estimator of IV.*
- A variation of IV: adjusted treatment received, using the error terms from first stage endogenous regression is added to the model as a covariate to allow adjustment for any unmeasured confounding.

Other statistical methods

- SMM/structural-nested mean model was introduced by Robins; a framework provides causal treatment effect for observed adherence comparing with conditional reference level of adherence.
  - Linear additive framework for continuous outcomes
  - Multiplicative framework for binary outcomes
  - Rank preservative structural failure time model (RPSFTM) for time to event survival outcome
- G-estimators (GE) have not been widely adopted because of their high complexities.
- For handling nonadherence in survival data includes Cox-reg, complier proportional hazard effect of treatment (C-PROPHET) model, and causal accelerated life model (CALM).

Comparison of estimated treatment effects

- 68 studies were able to compare.
  - 71% (n = 48) produced treatment effects *GREATER* than estimated by ITT
  - 16% (n = 11) produced *SIMILAR* treatment effects estimated by ITT
  - *All methods produced overlapped 95% CIs* with the CIs of ITT either at lower or upper bound region, *except Bayesian applications*.
- 64 studies contributed to s.e. and z-statistics analysis.
  - 83% (n = 53) produced *LARGER* s.e. than the s.e. of ITT estimates
  - After accounting for within study variation, average z-statistic from proposed methods were *GREATER* by +0.13 SD (0.99 to 1.71) compared o ITT
  - 12% (n/N = 7/58) achieved *significant treatment effect* by applying an alternative method, which was not achieved by the ITT method.

Comparison of estimated treatment effects

- In metaregression, when accounted for percent nonadherence rate:

  Z-statistic*

  - IV was not different compared to ITT (0.01, 95% CI: 0.27-0.26)
  - CACE was *GREATER* by +0.18 SD compared to ITT (0.18, 95% CI: 0.01-0.35)
  - CACE was *GREATER* by the same amount when compared to IV.

*A z-score describes the position of a raw score in terms of its distance from the mean, when measured in standard deviation units. The p-value is the probability that have falsely rejected the null hypothesis. If Z score is between -1.96 and +1.96, p-value will be larger than 0.05.

Discussion

- The median intervention nonadherence 38% (2% to 78%).
- Two most commonly used methods: *CACE & IV*
- Overall, *no significant difference* between the pooled z-statistics from ITT vs the alternative methods.
- In general, the most of the proposed applications (83%) produced *LARGER* error variance compared to the error variance produced by ITT.
- CACE resulted in *LARGER* z-statistics compared to IV when accounting for nonadherence rate.
- One of the benefits of CACE is that cell-specific treatment effect can be obtained, whereas limited for the IV.

Discussion

Limitations:

- The comparison of pooled z-statistic may not be an ideal approach, but this provides an indication of location of treatment effect estimated by different methods around the region of significance.
- The authors excluded studies using statistical methods for handling general confounding e.g. propensity score (PS), inverse probability weighting (IPW), which are wider applied in observational studies for adjusting general confounding because they directly do not contribute to the formation of causal frameworks for handling nonadherence.
- Details of statistical programs used for each methods have not been clearly described >> need further information to apply in real practice.

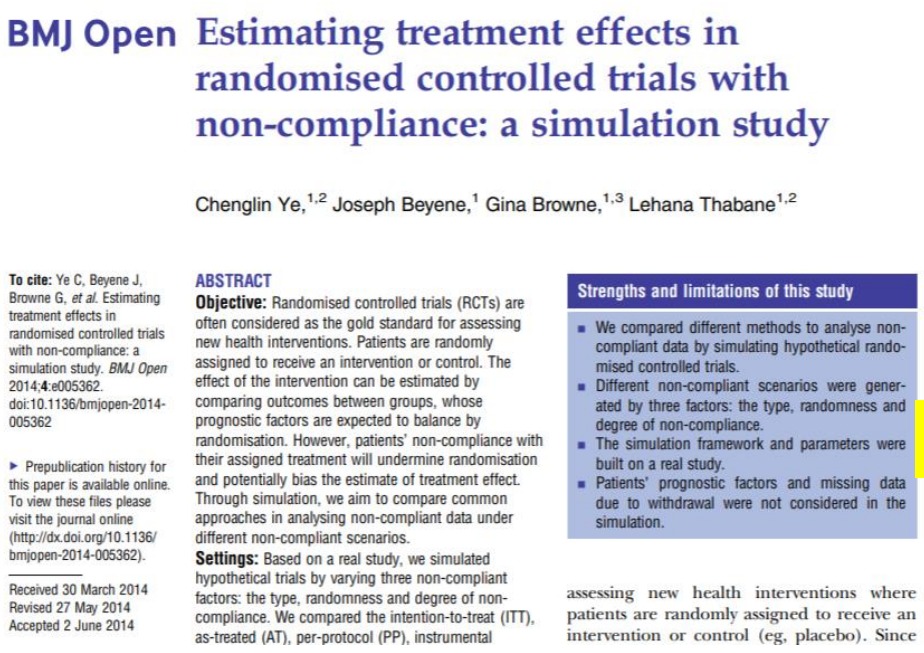
Implications for practice and policy

- Usually the ITT estimate of a treatment effect will be *SMALLER than the "true" effect* because if the treatment works, noncompliance to treatment means *suboptimal effects*.
- CACE and IV methods are two important unbiased alternatives to ITT when adherence to treatment is suboptimal.
- Both suffer from strong underlying assumptions.
- Always reported in addition to ITT analysis and regarded as *a sensitivity analysis.*

Conclusions

- A large proportion of RCTs continue relying on PP method for handling intervention nonadherence.
- Statistical applications based on causal framework are *more appropriately adjust treatment effect* for nonadherence in RCTs esp. *Maximum likelihood-based CACE and IV.*
- Their strong assumptions should not be violated.
- More empirical studies that directly compare the usability and performance of different statistical methods for nonadherence in RCTs are needed to standardize the optimal analysis approaches.

Optimal analyses for different scenarios



BMJ Open

Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study

Chenglin Ye,[1,2] Joseph Beyene,[1] Gina Browne,[1,3] Lehana Thabane[1,2]

To cite: Ye C, Beyene J, Browne G, *et al*. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open* 2014;**4**:e005362. doi:10.1136/bmjopen-2014-005362

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2014-005362).

Received 30 March 2014
Revised 27 May 2014
Accepted 2 June 2014

ABSTRACT

Objective: Randomised controlled trials (RCTs) are often considered as the gold standard for assessing new health interventions. Patients are randomly assigned to receive an intervention or control. The effect of the intervention can be estimated by comparing outcomes between groups, whose prognostic factors are expected to balance by randomisation. However, patients' non-compliance with their assigned treatment will undermine randomisation and potentially bias the estimate of treatment effect. Through simulation, we aim to compare common approaches in analysing non-compliant data under different non-compliant scenarios.

Settings: Based on a real study, we simulated hypothetical trials by varying three non-compliant factors: the type, randomness and degree of non-compliance. We compared the intention-to-treat (ITT), as-treated (AT), per-protocol (PP), instrumental

Strengths and limitations of this study

- We compared different methods to analyse non-compliant data by simulating hypothetical randomised controlled trials.
- Different non-compliant scenarios were generated by three factors: the type, randomness and degree of non-compliance.
- The simulation framework and parameters were built on a real study.
- Patients' prognostic factors and missing data due to withdrawal were not considered in the simulation.

assessing new health interventions where patients are randomly assigned to receive an intervention or control (eg, placebo). Since

Ye C, Beyene J, Browne G, et al. BMJ Open 2014;4:e005362.

1 Expected effect of the intervention relative to the control

- A moderate or larger effect: 25% or more improvement to the control
  - 2 Random non-compliance
    - The AT, PP, IV, and CACE analyses
    - Dependent non-compliance
      - One sided non-compliance
        - 3 Only never-takers
        - The PP analysis
      - Two sided non-compliance
        - Always- and never-takers
        - 4 How does the patients condition affects the compliance?
          - Scenarios C and E
            - The PP and CACE analyses
          - Scenarios A and D
            - The IV analysis
- A null to small effect: 0% to less than 25% improvement to the control
  - The ITT analysis

A. Patients with good conditions would always get the intervention while patients with poor conditions would always reject it.
C. Patients with poor conditions would always reject the intervention.
D. Patients with good conditions would always reject the intervention while patients with poor conditions would always get it.
E. Patients with good conditions would always reject the intervention.

ITT, PP & IV could be universally applied as sensitivity analysis.

Thank *YOU*