A practical framework for external control trials

EMA Workshop on the use of external control analyses in regulatory decision-making

Hans Ulrich Burger
Medical University Graz
Amsterdam, November 3, 2025

Overview

- Introduction
- 10 key potential differences
- Some recommendations
- Summary

Introduction

- External control analyses are discussed as an alternative to randomized clinical trials (RCTs)
- One hope is that with more and better quality real world data, randomized trials often no longer needed
- In addition, methods are available which apparently make external control “similar” to randomized control
  - Multivariable modeling
  - Inverse probability weighting based on propensity scores
  - …

Introduction

- Without doubts when these things all materialize and tools would work, external control could be the golden standard of the future
- But as often, things are much more complex
- When we perform hypothesis tests, we compare groups
  - Patients in one group are treated with an experimental treatment and patients in another one with a standard treatment or placebo
  - A p-value then indicates if differences between groups are by chance
  - A RCT makes everything in both groups the same apart form the treatment given, at least at baseline. Hence, differences between groups should be caused by differences in treatment. The p-value then indicates if differences between treatments are by chance
  - In non-randomized settings, we cannot easily make this last step!

Introduction

- The key question then is a causal one:

  Are observed differences or the absence of such due to the different treatments applied or due to other underlying differences?

- We are not interested in differences between groups when differences cannot be associated to treatment differences
- When we are interested in the treatment effect, then other differences, which change the outcome, introduce bias for the treatment comparison
- This makes non-randomized settings or external control analyses challenging
  - Conclusions are much more complex as we need to factor in other differences than treatment, including their impact on outcome which we would usually then call bias
  - A p-value then is no longer directly associated with treatment differences
  - Even when we control for most of potential differences there is always increased uncertainty compared to a RCT

Introduction

- Hence, the first question for using external control is

  How sure are we that there are no differences between study group and external control group (at least which could influence the treatment effect) other than treatment?

- Two principles are key at **planning and analysis stage**

  - **Transparency** on ALL potential differences between study group and external control group
  - **Evaluation of the impact (bias)** on outcome of each potential difference noted

10 key potential differences

Source	Content
Population selection	Differences between patient populations in terms of demographics and disease characteristics
Calendar time	Different standard of care at different calendar times
Region	Different standard of care
Assessment	Differences in the knowledge about treatment group
Endpoint definition	Differences in how the assessment of an endpoint was done
Start of study	Differences in what we call day 0
Retrospective design	Differences due to the knowledge of outcome
Type of data	Study or registry data as patients may be treated differently
Intercurrent events	Differences in intercurrent events and their frequency
Data quality	Differences in data quality and completeness

10 key potential differences

- This is a long list and could be even longer
- But at the end we need to go through it step by step
  - To identify areas of potential differences
  - To identify areas where differences matter and differences may substantially impact the outcome
- We need to do this at planning and analysis stage
  - At planning it may help select the most appropriate external controls
  - And it may prevent us from doing a useless analysis when we could upfront realize that there are too many differences
  - At analysis stage we may need to do additional sensitivity analyses to evaluate impact. Differences need to be factored into conclusions
- Each of these differences can introduce bias, and bias can go often in any direction, but there are also often mitigation steps to reduce or minimize bias

10 key potential differences

- Population selection
  - Patients enrolled are different in demographics and baseline characteristics
  - Potential mitigation:
    Prospectively set clinical trial inclusion/exclusion criteria (I/E) based on measures used for external control (e.g. collected in routine clinical practice) or apply I/E to external control database. Adjust for remaining population differences e.g. by using propensity scores methods
- Calendar time
  - Patients treated in the past do differently than those treated today because of differences in standard of care
  - Potential mitigation:
    Taking external control patients from the same calendar time as clinical trial patients or having evidence that standard of care in this setting did not change.
    Otherwise: Provide data indicating that differences do not matter

10 key potential differences

- Region
  - Patient outcome may vary between regions due to different standard of care
  - Potential mitigation:
    Taking external control patients from the same region or having evidence that standard of care and outcomes do not vary between chosen regions
- Type of data (study versus observational data)
  - Patients in clinical trials have often different outcomes than in clinical practice. (e.g., due to placebo effect, different level of care, e.g. more regular assessments)
  - Potential mitigation:
    Use same data type as external control.
    Otherwise: Provide data indicating that differences do not matter in this case

10 key potential differences

- Intercurrent events
  - Differences in intercurrent events after study entry may be correlated with outcome and jeopardize comparability with external control
  - Potential mitigation:
    Same bias as for RCTs. Same methods (see ICH E9 addendum) should be applied.
    Use of study data as external control with similar profile of intercurrent events
    Otherwise: Provide data indicating that differences do not matter (which may be difficult)
- Data quality
  - Differences in data quality caused for example by differences in data cleaning procedures (e.g. differences in missing data or length of follow up) can introduce significant bias
  - Potential mitigation:
    Clean external control patient data to same extent as study patient data, or choose external control data with similar level of quality
    Otherwise: Provide data indicating that differences do not matter

Some recommendations

- Transparency is key!
  - We need to go through all such differences for each external control analysis
  - We need to evaluate impact/bias on outcome for all differences
  - Comparability in some sources of differences can increase confidence in result
- It is not expected that all the sources of bias can be controlled
  - There will always be sources difficult to address
  - Interpretation of results must be made in the light of these limitations
  - Limitations can go to the extent that we cannot really conclude anything
  - P-values could be misleading when we do not sufficiently control other differences/bias
- Generally, even when all possible measures taken, there is still uncertainty around the assumptions of no differences/bias
  - There is still no strict type 1 error control
  - The additional uncertainty requires specific justification of the design versus other possible designs without external controls

Some recommendations

- However:
  It is also true that in the absence of data from randomized clinical trials, an external control analysis is better than doing nothing
  - Comparing just single arm study data with literature results is worse (no control of any population differences)
- There are areas where the use of external control is more adequate. This includes
  - Rare diseases where we hardly can afford a randomized trial
  - Parachute situation: When there is no effective treatment available and disease is live threatening
  - Longitudinal studies in chronic diseases when we cannot maintain randomization long enough
- There are specific examples (and probably more than expected) in which external control really make sense
- In addition, it is always helpful when we expect an exceptionally strong effect

Summary

- External control should not be thought of as a replacement for randomization, rather as a complement
- When using external controls, transparency about potential confounders, i.e. any differences between both populations, is key, in the planning as in the analysis phase
- Justification of external control analyses should always include RCT as a benchmark

References

1. Burger HU, Gerlinger C, Harbron C, Koch A, Posch M, Rochon J, Schiel A. ‘The use of external controls: To what extent can it currently be recommended?’ 2021. Pharmaceutical Statistics. 2021;1–15. https://doi.org/10.1002/pst.2120.
2. Zhou X, Pang H, Drake Ch, Burger HU, Zhu J. ‘Estimating treatment effect in randomized trial after control to treatment crossover using external controls” 2024, J Biopharm Stat. 2024 Oct;34(6):893-921. doi: 10.1080/10543406.2024.2330209

Back up

10 key potential differences

- Population selection
  - Patients enrolled are different in demographics and baseline characteristics
  - Potential mitigation:
    Prospectively set clinical trial inclusion/exclusion criteria (I/E) based on measures used for external control (e.g. collected in routine clinical practice) or apply I/E to external control database. Adjust for remaining population differences e.g. by using propensity scores methods
- Calendar time
  - Patients treated in the past do differently than those treated today because of differences in standard of care
  - Potential mitigation:
    Taking external control patients from the same calendar time as clinical trial patients or having evidence that standard of care in this setting did not change.
    Otherwise: Provide data indicating that differences do not matter

10 key potential differences

- Region
  - Patient outcome may vary between regions due to different standards of care
  - Potential mitigation:
    Taking external control patients from the same region or having evidence that standard of care and outcomes do not vary between chosen regions
    Otherwise: Provide data indicating that differences do not matter
- Assessment
  - Knowledge of therapy can influence the outcome assessment.
  - Potential mitigation:
    Choose objective endpoints, for example, OS, independent review (similar to strategies for open-label randomized studies).
    Conduct sensitivity analyses.

10 key potential differences

- Different endpoint
  - Endpoint definition (e.g. Response and progression) in clinical trials are measured differently than in routine clinical practice. (This bias not applicable to OS)
  - Potential mitigation:
    Use the same definition of procedure (e.g. obtain consent for tumor images to allow RECIST-like review).
    Conduct sensitivity analyses (for example, frequency of tumor assessments)
- Start of study (Immortal bias)
  - Study start is often difficult to determine as an anchor for patient specific time scale in non-study data but patient's outcome may vary given where they are in the course of the disease
  - Potential mitigation:
    Need to clearly define time origin.
    Evaluation of potential bias introduced

10 key potential differences

- Retrospective selection
  - Selection of sources of external data and type of study end points and analysis are done retrospectively
  - Potential mitigation:
    Generate external control data in parallel. If not possible pre-specify endpoints and analyses before searching literature to identify external control sources
- Type of data (study versus observational data)
  - Patients in clinical trials have often different outcomes than in clinical practice. (e.g., due to placebo effect, different level of care, e.g. more regular assessments)
  - Potential mitigation:
    Use same data type as external control.
    Otherwise: Provide data indicating that differences do not matter in this case

10 key potential differences

- Intercurrent events
  - Differences in intercurrent events after study entry may be correlated with outcome and jeopardize comparability with external control
  - Potential mitigation:
    Same bias as for RCTs. Same methods (see ICH E9 addendum) should be applied.
    Use of study data as external control with similar profile of intercurrent events
    Otherwise: Provide data indicating that differences do not matter (which may be difficult)
- Data quality
  - Differences in data quality caused for example by differences in data cleaning procedures (e.g. differences in missing data or length of follow up) can introduce significant bias
  - Potential mitigation:
    Clean external control patient data to same extent as study patient data.
    Use of study data as external control with similar profile of intercurrent events
    Otherwise: Provide data indicating that differences do not matter