Routinely collected data in clinical trials: methods, opportunities and challenges

Dorcas Kareithi – Research Associate (Biostatistics)
Newcastle university

Dorcas.Kareithi@newcastle.ac.uk

IASC Africa Member Group Webinar – March 2024
Background – Dee

Dorcas Kareithi, Msc
Research Associate (Biostatistics), Newcastle university

• Over a decade of experience as a statistician, researcher and analyst. Worked as an analyst, head of research and consulting epidemiological analyst in several organizations and companies over the years.

• My research interests are in cancer research and use of routinely collected data in clinical research.

• Mentor, current president of Young African Statisticians Association (YASA) and member of several African and international statistical associations.

• https://www.linkedin.com/in/dorcaskareithi/
The PHSI-BRG consists of 33 statisticians (and growing) with a variety of backgrounds, experience, and expertise.

The PHSI-BRG provides statistical expertise to a large portfolio of clinical trials and observational studies, working closely with Newcastle Clinical Trials Unit and the NIHR Research Design Service North East and North Cumbria in United Kingdom.

As well as supporting real clinical trials, the PHSI-BRG has a growing methodology research group and a newly established Artificial Intelligence in Health Research group.

Twitter: @nclBiostats

Website: www.newcastle-biostatistics.com
Clinical trials are studies that evaluate interventions and their effects on human health outcomes.

These interventions could be focused on prevention, diagnosis or treatment.

Clinical trials involving medicine mostly have 3 or 4 stages, before licensing and are subject to regulatory approval.

Other clinical trials may have 1 or 2 stages, depending on intervention and objectives.

Clinical trials allow researchers to assess the impact/effect of an intervention with minimum bias.

Source: www.lifetrans.org
Background – Randomized Clinical Trials

- The analysis of difference in outcomes gives the average effect of experimental arm across the trial population (treatment effect)

\[ Average\ Treatment\ Effect\ (ATE) = E[Y^1 - Y^0] \]

- \( Y^1 \) is the outcome for a participant if given intervention
- \( Y^0 \) is the outcome if they were given the control

- RCTs have been very successful and has resulted in a lot of changes in how conditions are diagnosed, treated or prevented

- However, one may
  - not have enough resources
  - need to evaluate the effectiveness or safety of interventions in routine clinical practice (external validity)
  - want to investigate long-term outcomes beyond the trial period.

- Sometimes RCTs may not be feasible or ethical eg for rare diseases or very small subsets of participants.
**INTERNAL VALIDITY**

The extent to which the differences observed between groups can be correctly attributed to the intervention under investigation.

**EXTERNAL VALIDITY**

The extent to which the results of an RCT can be generalized into clinical practice and the general population.

- Average treatment effect in RCTs is reliably estimated for participants randomised and who don’t drop out. In practice, this is not the case.

- RCTs have **high internal validity** but **low external validity**.

- Statistical considerations to improve external validity include conducting network meta-analysis, using observational or routinely collected data to perform **target trial emulation** or **create synthetic controls**.
RCTs and Long-Term Effectiveness

- Costs and other factors make RCTs relatively short.
- Sometimes in clinical trials, it’s possible to extend the duration through “Open Label Extension (OLE) studies”.
- Sometimes it is a regulatory requirement by regulatory bodies such as the U.S. Food and Drug Administration- FDA, Medicines and Healthcare products Regulatory Agency-MHRA (UK) European Medicines Agency- EMA (EU) or national bodies.
- The OLE becomes a “single arm trial” (participants are given the option to switch treatment and be all on the intervention).
- In such cases, routinely collected data can become synthetic controls.

Long-Term Safety and Efficacy of Bempedoic Acid in Patients With Atheroaclerotic Cardiovascular Disease and/or Heterozygous Familial Hypercholesterolemia (from the CLEAR Harmony Open-Label Extension Study): [https://doi.org/10.1016/j.ajjcard.2022.08.020](https://doi.org/10.1016/j.ajjcard.2022.08.020)

Synthetic Controls

- Synthetic control methods are used in various fields, not only health research.
- The synthetic control method builds a counterfactual using a weighted combination of potential control arms.
- It is a control group that is matched to the intervention arm from other sources of data, such as routinely collected data.
- Requires good quality data to exactly match the design of the trial in mind (e.g., participant characteristics, inclusion and exclusion criteria, frequency of collecting outcomes).
RCD as Synthetic Controls in RCTs

- Matching participants after inclusion and exclusion criteria has been applied can be done through methods such as:
  - Direct (Hard) matching
  - Multivariable Outcome Regression
  - Frequentist Propensity score matching (PSM)
  - Propensity score stratification
  - Doubly robust propensity score
  - Bayesian propensity score matching
  - Nearest-neighbor propensity score matching
  - Inverse Probability of Treatment Weighting (IPTW) using Propensity Scores

Propensity score matching methods:

- Matching treated and untreated participants on the estimated probability of being in the treatment group (propensity score)
  - \( p(X) = \Pr(Z = 1 | X) \)
  - \( Z \) is the indicator representing treatment.
  - \( Z = 1 \) if the group is intervention and \( Z = 0 \) if control
  - \( X \) is a vector of covariates

- Aims to reduce/remove selection bias

- Instead of matching each participant based on the same value of \( X \), we match them on the probability of being treated.

- Key assumption: participation \( (y_0) \) is independent of outcomes conditional on \( X \), i.e. adjusting for \( X \) (all confounders are included in the covariates) is sufficient to eliminate all confounding

\[
E[y_0 | X, Z = 1] = E[y_0 | X, Z = 0]
\]
RCD as Synthetic Controls in RCTs

Propensity score matching methods:

- $p(X)$ is estimated using predicted values of regression models such as the logistic regression models or Classification and Regression Tree Analysis (CART).

- Logistic regression is most commonly fitted on the combined data (treated and untreated):

$$\log\left(\frac{p(X_i)}{1-p(X_i)}\right) = \log\left(\frac{\Pr(Z_i=1|X_i)}{1-\Pr(Z_i=1|X_i)}\right) = \alpha + \beta X_i$$

- In this logistic regression, the dependent variable is binary, $Z_i=1$ is the value for the treatment and the value for the control is $Z_i=0$, $X_i$ is the vector of covariates.

- The CART is a non-parametric decision tree method that partitions populations into homogenous subgroups. It is not commonly used as it is computationally and statistically complex.

- After obtaining the predicted values, $\hat{p}(X)$ of the logistic regression model, produce density plots to assess the distribution for each group, and check that the assumption $0 < p(X) > 1$ has not been violated.

- After that we can adjust the propensity scores through **Stratification, Matching, Covariate/Regression adjustment**, or **Inverse Probability of Treatment Weighting (IPTW)**
RCD as Synthetic Controls in RCTs

Inverse Probability of Treatment Weighting (IPTW) using Propensity Scores:

- In IPTW, the treated and control observations are re-weighted to make them more representative of the population.

- The weight of a treated participant \((Z = 1)\) when computing Average Treatment Effect (ATE) is the inverse of its propensity score:
  \[
  w_i^1 = \frac{1}{\hat{p}(X_i)}
  \]

- The weight of a control participant \((Z = 0)\) when computing ATE is the inverse of one minus its propensity score:
  \[
  w_i^0 = \frac{1}{1 - \hat{p}(X_i)}
  \]

- Sometimes IPTW results in very large weights, this can be “moderated” by stabilising the weights by using the \(n\), trimming the weights or both.
  \[
  \text{stabilised } w_i^1 = \frac{n_1}{n_1 + n_0} \times w_i^1, \text{ and stabilised } w_i^0 = \frac{n_0}{n_1 + n_0} \times w_i^0
  \]

- Stabilising and trimming the weights helps to further reduce variance and possible bias.

- After this, you proceed to estimate ATE using methods such as Cox-proportional hazards models, and the trimmed and stabilised IPTW as your weights.
Primary biliary cholangitis (PBC) is a rare autoimmune cholestatic liver disease predominantly affecting women over the age of 40 years (approximately 6 in 1000).

Without treatment, patients can progress to end-stage cirrhosis, resulting in hepatic decompensation and, without transplantation, death.

First-line therapy for PBC is ursodeoxycholic acid (UDCA), a bile acid found to improve clinical outcomes and improve transplant-free survival.

Approximately of 40% of patients prescribed UDCA experience an inadequate response and require second-line therapy.
Objective: Evaluate the long-term efficacy of OCA in OCA-treated patients in the POISE trial and comparable non-OCA-treated external controls, comparing time to first occurrence of liver transplant or death among those inadequately responding to UDCA

HEPATOBILIARY
Greater Transplant-Free Survival in Patients Receiving Obeticholic Acid for Primary Biliary Cholangitis in a Clinical Trial Setting Compared to Real-World External Controls

C. Fiorella Murillo Perez,1,2 Holly Fisher,1,2 Shaun Hui,2 Dorcas Kareth,1 Femi Adekunle,1 Tracy Mayne,1 Elizabeth Malecha,1 Erik Nesi,3 Achmad J. van der Meij,4 Willem J. Lammers,4 Palak J. Trivedi,2 Pier Maria Battezzati,5 Frederik Nevens,6 Kris V. Kowdley,7 Tony Brun,9 Nara Cazzagon,6 Annarosa Floreani,6 Andrew L. Mason,1 Albert Parisi,1 Maria-Carola Londoño,1,15 Pietro Invernizzi,5,15 Marco Carbone,1 Ana Lioe,17,18 Marilyn J. Mayo,9 George N. Dalekos,10,15 Nikolaos K. Gatselis,13,15 Douglas Thorburn,21 Xavier Verhelst,22 Alysa Gulamhussein,22 Harry L. A. Janssen,22 Rachel Smith,24 Steve Flack,24 Victoria Mulcahy,22 Michael Trauner,22 Christopher L. Bowls,26 Keith D. Lindor,27 Christophe Corpechot,28 David Jones,9 George Mills,29 Gideon M. Hirschfield,1 James Watson,16 and Bettina E. Hansen,16 on behalf of the GLOBAL PBC Study Group and the members of the UK-PBC Consortium

1Toronto Centre for Liver Disease, Toronto General Hospital, University of Toronto, Toronto, Ontario, Canada; 2Newcastle University, Newcastle upon Tyne, United Kingdom; 3Ipsen Pharmaceuticals, Montevideo, New Jersey; 4Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; 5University of Birmingham, Birmingham, United Kingdom; 6University of Padua, Padua, Italy; 7University Hospital Nijmegen, Nijmegen, The Netherlands; 8University of Trieste, Trieste, Italy; 9University of Glasgow, Glasgow, United Kingdom; 1010 Hospital of Athens, Athens, Greece; 11Department of Gastenterology, University of Texas Health Science Center at Houston, Houston, Texas, USA; 12Department of Gastroenterology, Catholic University of Malaga, Malaga, Spain; 13University of Padua, Padua, Italy; 14University of Genoa, Genoa, Italy; 15Department of Medicine and Surgery, University of Genoa, Genoa, Italy; 16Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 17Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 18Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 19Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 20Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 21Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 22Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 23Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 24Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 25Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 26Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 27Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 28Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 29Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 30Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands; 31Department of Transplantation, Gastroenterology, and Medicine, University of Amsterdam, Amsterdam, Netherlands.

Example: POISE Study

- The only approved (accelerated approval) second-line therapy for PBC is obeticholic acid (OCA).
- Full approval depended on a post-approval requirement to confirm benefit by assessing the effect of OCA on clinical outcomes, such as hepatic decompensation, liver transplantation, and death.
- Recruitment challenges: As a rare disease, recruitment is difficult and retention, especially when there are multiple trials competing for a small pool of qualifying patients.
- Participant retention challenges: patients must remain in the trial for years in order to accumulate a sufficient number of clinical events.


IASC Africa Member Group Webinar – March 2024
Example: POISE Study

1,381 Global PBC patients and 2,135 UK PBC patients met POISE inclusion criteria, while 209 POISE DB trial patients were in the final sample. These data sources were used to create synthetic control groups to get a longer-term evaluation of OCA.
Example: POISE Study

- Each qualifying patient visit (diamond) was examined to determine whether the patient met eligibility criteria at that visit (yellow diamond) or did not meet criteria (white diamond).

- An eligibility period was established between the first visit that each control patient met POISE inclusion/exclusion criteria, and the last visit the patient met criteria.

- A random visit was selected between those dates (inclusive) which served as the index date. Sensitivity analyses conducted for different eligible visits (first, last and random).
Cox-proportional hazards models (with Firth correction) were run to establish the hazard ratio. A Wald test of cohort effect performed using the stabilized IPTWs.

Weighted Kaplan-Meier estimates (using the stabilized IPTWs) of the distribution of the time-to-event by treatment group.

Logistic regression was performed. The outcome was treatment group. Resulting propensity scores estimated IPTW. Extreme values (>10) were trimmed.

The samples were closely aligned on baseline characteristics before propensity scoring.
The samples were closely aligned on baseline characteristics before propensity scoring.
Example: POISE SMDs

- Demographic and disease-related variables were predictors and study arm (POISE OCA-treated vs external control) as the outcome in the logistic regression.

- Post-weighting standardized mean difference (SDM) of ±0.25 was pre-specified as indicative of adequate comparability between treatment arms.

- Any variables outside this range were included in the final Cox regression model as covariates.

- After weighting, all variables were within the pre-specified threshold.

- Although PBC duration was slightly outside the prespecified range (SMD=0.259), it was retained in the propensity score and weighting and was not added as a separate covariate in the outcomes model.
Kaplan-Meier curves for transplant-free survival comparing POISE to Global PBC (A) and UK PBC (B)

The graphs begin separating at 12 to 18 months.

Example: POISE Study Results
Example: POISE Study Results

- Patients treated with OCA in a trial setting had **significantly greater transplant-free survival** than patients in external control groups.

- All analyses for the random index date produced similar point estimates.

- Three sensitivity analyses were performed:
  1. **The impact of index date on outcomes**
  2. A **subgroup analysis in patients with and without cirrhosis in the Global PBC dataset (cirrhosis was not consistently quantified in the UK PBC registry)**
  3. **Tested the hypothesis that there was residual, unmeasured selection bias**, in which POISE study investigators selected healthier patients for the trial, and that any effect observed was due to the external controls representing patients with more progressed disease
Conclusion: Limitations

**Incomplete or missing data:** RCD sources may have missing or incomplete data, which can introduce bias or limit the ability to analyse certain variables or outcomes. Efforts to address data quality and completeness are crucial when using RCD. Gives opportunity to develop stats methods for different scenarios.

**Confounding factors and biases:** RCD may contain confounding factors or biases due to the non-random nature of data collection in routine clinical practice. Proper statistical methods, such as propensity score matching or regression modelling, are needed to address these potential confounders.

**Limited control over data collection:** Researchers have less control over how data are collected, leading to potential variability in data quality, standardization, and coding practices. This can affect the reliability and accuracy of the data used in the trial.

**Lack of specific trial-related variables:** RCD sources may not capture all variables of interest for clinical trials or use cases. This can limit the ability to analyse certain outcomes or evaluate specific objectives.

**Data Accessibility and Availability:** Access to RCD may be limited, particularly if data are controlled by different institutions or require complex data sharing agreements. This can restrict the feasibility and timeliness of conducting research using RCD.
Conclusion: Opportunities

RCD allows for the evaluation of interventions in routine clinical practice and increases the generalizability of trial findings and inform the applicability of trial results to broader populations.

This can help identify potential safety concerns or adverse events that may not have been observed in clinical trials, since RCD offer long-term follow-up data, allowing researchers to assess outcomes beyond the trial period.

Large sample sizes: RCD often covers a large and diverse population, providing access to a larger sample size than traditional clinical trials. This allows for increased statistical power and generalizability of findings.

Cost-effectiveness: Utilizing existing RCD sources can reduce the cost and time required for data collection, as data are already being routinely captured as part of clinical care.

Ethical considerations: RCD can help minimize additional burden on patients, as it relies on data collected during routine clinical encounters, eliminating the need for additional data collection procedures.
Resources and Acknowledgement

• **POISE paper**: [https://www.gastrojournal.org/action/showPdf?pii=S0016-5085%2822%2901060-5](https://www.gastrojournal.org/action/showPdf?pii=S0016-5085%2822%2901060-5)

• **Acknowledgement:**
  - Colleagues at BRG: Prof James Wason, Shaun Hiu, Dr Holy Fisher (POISE study stats team, UK)
  - Colleagues and partners at Global PBC, UK PBC, Intercept Pharmaceuticals (POISE study research and Toronto stats team).

• **Course on Leveraging external information using frequentist and Bayesian methods that BRG runs yearly (28th-29th May 2024):** [https://www.newcastle-biostatistics.com/courses/external_information/](https://www.newcastle-biostatistics.com/courses/external_information/)
Scan the QR code or register on our website to become a member
https://youngstatisticians.org/membership-form/

info@yasastatisticians.org
Young African Statisticians Association
@AfricanYas
www.youngstatisticians.org
THANK YOU

Dorcas.Kareithi@newcastle.ac.uk

www.newcastle-biostatistics.com

Population Health Sciences Institute
Biostatistics Research Group
Newcastle University

African Members Group (IASC AMG)