Trials and Tribulations: The Good, The Bad, and The Patently Absurd

A Provocative Look at RCTs, Retrospective Studies, and Everything In Between

BASS Biostatistical Conference 2025
Savannah, GA

Stephan Ogenstad, PhD

Stephan.Ogenstad@gmail.com

Introduction

- **Goal of the Talk:** Provide an insightful, provocative, and humorous look at different study designs.

- **What We'll Cover:**
  - RCTs: Gold standard or just expensive?
  - Retrospective Studies: Useful or riddled with bias?
  - Single-Arm Studies: A necessary evil?
  - Registries: Rich data sources or garbage in, garbage out?
  - Propensity Score Matching: Magic bullet or statistical wishful thinking?
  - Historical Controls: Clever shortcut or a recipe for misleading conclusions?
  - Patents: Protecting innovation or stifling progress?

Randomized Controlled Trials (RCTs)

- Randomization reduces bias, Control Groups establish causality, blinding increases objectivity, multi-center trials enhance generalizability, prespecified statistical hypotheses and rigor (reduces cherry-picking), and regulatory acceptance (FDA, EMA, etc.).
- **Pros:**
  - Gold standard for causal inference
  - Reduces confounding bias
  - Regulatory agencies love them
- **Cons:**
  - Expensive and time-consuming
  - May not reflect real-world patients
  - Ethical concerns in some cases (e.g., placebo use in life-threatening conditions)
- **Commentary:**
  - Are RCTs the only way to get "truth," or are we fooling ourselves?
  - Is external validity (generalizability) sacrificed in the pursuit of internal validity (causal inference)?

Real-World Patients

Why RCTs May Not Reflect Real-World Patients

- Strict inclusion/exclusion → younger, healthier, fewer comorbidities
- Close monitoring → better adherence than in practice
- Limited diversity → underrepresents minorities, elderly, special populations
- Artificial treatment conditions → less physician flexibility
- Shorter duration → misses long-term effects
- High **internal validity**, but limited **generalizability**

Retrospective Studies

- **Pros:**
  - Faster and cheaper than RCTs
  - Large datasets available (e.g., EHR, claims data)
  - Can explore rare events or long-term outcomes
- **Cons:**
  - Confounding and bias are major threats
  - Data quality issues (missing data, misclassification, etc.)
  - No control over how data was collected
- **Commentary:**
  - Just because it is real-world data doesn't mean it is real good data.
  - Propensity score matching can ameliorate the results.

- Can clever statistical adjustments compensate for a flawed study design? No, statistical adjustments can reduce bias or improve precision, but they cannot fully rescue conclusions from a fundamentally flawed study design

Single-Arm Studies

- **Pros:**
  - Feasible for rare diseases or breakthrough therapies
  - Faster approval pathways

- **Cons:**
  - No comparator group means no direct estimate of treatment effect
  - Highly dependent on historical data or external controls

- **Commentary:**
  - Are we lowering the evidence bar too much?
  - When does "uncontrolled" become "out of control"?

Statistical Reasoning vs. Common Sense

- **Not all knowledge requires a RCT**.

- The **counterfactual is obvious**: wisdom tooth extraction without anesthesia → severe pain, always. (e.g. lidocaine, Astra Pain Control)

- The **mechanism is known**: lidocaine blocks sodium channels, preventing nerve conduction → no pain signals are transmitted.

- The **effect size is massive**: going from extreme pain to no pain doesn't require subtle statistical testing to detect.

Statistical Reasoning vs. Common Sense (2)

- **RCTs are essential when effects are modest, confounded, or uncertain.**

- **The Sledgehammer Effect**
  "If you want to test a cholesterol drug, run a RCT. If you want to test whether lidocaine blocks pain, just pull a tooth. Some effects are so big they don't need statistics—they need common sense."

- The **Gleevec trial** was single-arm because CML's **(chronic myeloid leukemia)** prognosis was dismal, standard therapies barely worked, and imatinib's effect was so dramatic that randomization was both unethical and unnecessary.

Registries

- **Pros:**
  - Large real-world datasets
  - Longitudinal follow-up
  - Can complement RCT evidence
- **Cons:**
  - Data inconsistency across sites
  - Selection bias
  - Difficult to control for confounding
- **Commentary:**
  - Are registries a data goldmine or a statistical landfill?
  - If the data quality is poor, can we trust the conclusions?

Registries (2)

- Registries can be a data goldmine — rich, real-world, and long-term. However, without curation and careful methods, they quickly become a statistical landfill.

Propensity Score Matching (PSM)

- **Pros:**
  - Can help mimic randomization in observational studies
  - Balances measured confounders
- **Cons:**
  - Doesn't balance unmeasured confounders
  - Requires large datasets
- **Commentary:**
  - Are we just sprinkling statistical fairy dust on observational studies?
- If PSM is so great, why do regulators still prefer RCTs?

If PSM is so great, why do regulators still prefer RCTs?

- RCTs eliminate both **known** *and* **unknown** confounding by randomization
- PSM only balances *observed* covariates → hidden bias remains
- Quality depends on data completeness & model assumptions
- Easy to "game" by choices of variables, calipers (distance), and matching ratios (# matches per treated patients).
- Regulators see RCTs as the most objective, transparent, and reproducible
- PSM is applicable when RCTs aren't feasible — but it's not a substitute. It can be used as a sensitivity analysis.

Historical Controls

- **Pros:**
  - Useful when RCTs are impractical (e.g., rare diseases)
  - Reduces patient burden (no placebo group needed)
  - Often used for devices using Meta-analysis.
- **Cons:**
  - Treatment standards change over time
  - Hard to ensure comparability with past data
- **Commentary:**
  - When does a historical control become a historical artifact?
  - Are we cherry-picking past data to fit our narrative?

When does a historical control become a historical artifact?

- When *standard of care* has changed

- When *diagnostics* or *endpoints* have evolved

- When *patient populations* no longer match today's standard

- When *supportive care* improvements change outcomes

- When data are *too old, sparse, or selective*

- A historical control is only helpful if it's still relevant — otherwise it belongs in a museum. They are commonly used for the design of device studies.

Patents & Drug Development

- **Pros:**
  - Encourage innovation
  - Provide financial incentives for R&D
- **Cons:**
  - Can be used to delay generics (evergreening, secondary patents)
  - Legal battles create uncertainty. E.g., UTC vs. Liquidia, PAH, treprostinil dry powder.
- **Commentary:**
  - Are patents protecting patients or profits?
  - When do exclusivity strategies cross ethical lines?

Patents: Protecting Patients or Profits?

- **Patients**: Incentivize innovation → new drugs & devices reach the clinic

- **Profits**: Extend exclusivity → delay generics, drive up prices

- Tension between ***rewarding innovation*** and ***ensuring access***

- Reality: both forces are always at play

- The answer depends on your perspective — inventor, regulator, payer, or patient.

- **Patents: lifesavers for companies, life-delayers for patients.**

When Do Exclusivity Strategies Cross Ethical Lines?

- Evergreening: minor tweaks → new patents, no real innovation
- Pay-for-delay: companies pay generics to stay out
- Patent thickets: overlapping claims → block competition
- Orphan drug "gaming": slicing populations to extend rights
- Pricing strategies that limit patient access
- Protecting innovation is ethical. Protecting profits *at the expense of patients* is not.
- **It's innovation when patients benefit, it's exploitation when only the balance sheet does.**

Should Regulators Be More Flexible?

- **Regulatory Flexibility: Pros & Cons**
- **Faster access** for serious/rare diseases
- **Encourages innovation** & new therapies
- **Patient-centered outcomes**
- **Risks:** safety/efficacy uncertainty, trust issues
- **Key:** Flexible methods + conditional approvals = smarter, safer regulation

Statistics: Illuminate or Obfuscate?

- **To Illuminate:**
- Clarifies complex data & uncertainty
- Supports evidence-based decisions
- Reveals hidden patterns
- **To Obfuscate:**
- Cherry-picked endpoints & selective reporting
- Overcomplicated models or jargon
- Manipulates perception of risk/effect
- **Bottom Line:**
  A **tool for insight** when transparent, **misleading** when misused.
- **If the data quality is poor, statistics don't clarify — they obfuscate. The model may look elegant, the p-value may sparkle, but all you've done is put lipstick on a dataset that can't be trusted.**

Final Thoughts

- There is no one-size-fits-all study design.
- The best approach depends on the question, context, and trade-offs.
- **Let's discuss:**
  - Are we too fixated on RCTs?
  - Should regulators be more flexible?
  - Are we using statistics to illuminate or obfuscate?

Thank you!   Qs & As.