Design, Implementation and Analysis of Observational Studies

Pragmatics vs Perfection

WHO Meeting

Global research and innovation forum: RESILIENCE against outbreaks and pandemics
October 23-24, 2023

Lindsey R. Baden, MD
Dana-Farber Cancer Institute
Brigham and Women’s Hospital
Harvard Medical School
Gold Standard – RCTs but...
Small Molecule Antiviral: Remdesivir

- **Gilead:** Compassionate use — NEJM 10Apr20
  - N=61, open label, hospitalized, hypoxemic. LD200/100mg for 9 days
- **NIAID-ACTT-1** — NEJM 22May (prelim) and 5Nov20
  - N=1062, RCT-pbo, LD200/100mg 9 days, hospitalized
  - **Time to recovery** —median 10 vs 15 days
- **Gilead:** 5 or 10 days, *Severe* Covid -- NEJM 27May2020
  - N=397, Randomized, open-label, hospitalized no IMV
  - Clinical status improvement d14 – 64% in 5D vs 54% in 10D
- **Gilead:** 5 vs 10 days vs pbo, *Moderate* Covid — JAMA 21Aug2020
  - N=596, RCT-pbo, hospitalized, O2>94%, LD200/100 for 5 or 10d (median 6d)
  - Clinical status d11 – 5d>pbo, 10d~pbo
- **WHO-Solidarity:** Inpatient — NEJM 11Feb2021, Lancet 02May2022
  - N= 2750 remdesivir (10d)+2708 SOC, RCT-SOC, hospitalized, moderate Covid
  - Mortality – 11.0% (14.5%) vs 11.1% (15.6%)
- **PineTree:** Outpatient — NEJM 27Jan2022
  - N=562, RCT, LD200/100 2 days. Outpatients
  - Hospitalization/death – 0.7% vs 5.3%
Forecasting Monoclonal Antibodies (mAb) Utility
Integrate Several Lines of Evidence

- Pathogen and Variant of Concern (VOC)
- *in vitro* activity of mAb
- PK/PD of mAb
- Clinical safety data
- Clinical efficacy data
  - In general vs against a specific VOC
  - Tempo of availability
In 1929 Finland was asked by Dr. Nye to join his laboratory at the Thorndike. Thus began one of the most remarkable careers in the field of infectious diseases. The first studies conducted by Max and his associates dealt with pneumonia. At that time the only treatment for pneumococcal pneumonia was administration of type-specific antiserum. The process of treating patients was cumbersome, to say the least. A naso-pharyngeal swab was taken and placed in a tube containing culture medium. After a few hours of incubation when enough bacteria had proliferated, material from the culture was exposed to type-specific antisera. If there was a match between the antiserum and the chemical composition of the polysaccharide on the surface of the bacterium, the capsule would swell and it could be seen with an ordinary light microscope (known as the Quellung reaction). If Quellung occurred, the corresponding antiserum (horse or rabbit) was administered to the patient. The patients usually survived the infection, but they invariably suffered from serum sickness, which could be most unpleasant. Finland and his fellows did a series of studies on the treatment of pneumococcal infection conducted with meticulous care, a hallmark of Finland’s research throughout. When sulfonamides became available
Classification of Dryvax Takes

Category 1

Category 2

Category 3
Observational Studies Needed to Fill in Gaps

• Safety
  • Event frequency
  • Large databases (healthcare systems), voluntary reporting (VAERS)
  • Sorting out from background rates
    • Reactivation zoster, Bell’s palsy
    • Serious
      • TTS, anaphylaxis, myocarditis

• Defining rates
  • Numerator
  • Population specific
Observational Studies Needed to Fill in Gaps

- Special populations
  - Immunocompromised patients, pregnancy, children

- Durability
  - *in vitro* (antibody titers) vs clinical activity (outcome of interest)
  - Boosting, interval

- Determine a Correlate of Protection (CoP)
  - ?Neutralizing antibody titer (nAb)

- Pathogen evolution
  - Variants
  - Implications for vaccines, monoclonal antibodies, small molecule antivirals
Observational Designs

• Case series
• Case control
• Cohort
  • Large databases (healthcare system, country wide)
Some Key Confounders/Biases

• Understanding of illness caused by pathogen (and VOCs)
  • Direct viral effects, immunopathogenesis, tempo

• Variable infrastructure across settings
  • Variable prevention/treatment over time

• End point of interest
  • Infection; hospitalization; mortality
  • Availability and reasons for SARS-CoV-2 testing; reasons for hospitalizations; reasons for discharge; availability of supportive care

• Data availability/completeness
  • Variable data collection in medical record
  • Accuracy: variable definitions used across settings

• Human behavior
  • Patient, practitioner, health system

• Speed
Design Challenges
14 September 2023 – WHO Meeting

1. Confounding
2. Healthy Vaccinee Bias
3. Misclassification
4. Selection Bias
5. Biases for Test Negative Design (TND)
6. Differential depletion of susceptibles
7. Waning immunity
Observational Designs

- Causal inference from observational data
- Bradford Hill (9) Criteria
  - Strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, analogy
- Design innovations
  - Comparator group selection (access, EMR)
  - Propensity weighting
  - Emulated randomized (targeted) trials

The International Journal of Biostatistics

Volume 4, Issue 1 2008

Causal Inference from Longitudinal Studies with Baseline Randomization

Miguel A. Hernandez* and James M. Robins

*Correspondence to: Dr. Miguel A. Hernandez, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115
e-mail: miguel_hernandez@post.harvard.edu

Practice of Epidemiology

Using Big Data to Emulate a Targeted Trial When a Randomized Trial Is Not Available

Miguel A. Hernandez, MD, DrPH, Wei Wang, PhD, David E. Leaf, MD, MMSc

The Environment and Disease: Association or Causation?

by Sir Austin Bradford Hill (1965)

President’s Address

I wonder whether the pendulum has not swung too far — not only with the attentive pupils but even with the statisticians themselves. To decline to draw conclusions without standard errors can surely be just as silly? Fortunately I believe we have not yet gone so far as our friends in the USA where, I am told, some editors of journals will return an article because tests of significance have not been applied. Yet there are innumerable situations in which they are totally unnecessary — because the difference is grotesquely obvious, because it is negligible, or because, whether it be formally significant or not, it is too small to be of any practical importance. What is worse the glitter of the t table diverts attention from the inadequacies of the fare.
Inhaled Fluticasone Furoate for Outpatient Treatment of Covid-19

David R. Boulware, M.D., M.P.H., Christopher J. Lindell, Ph.D.,
Thomas G. Stewart, Ph.D., Adrian F. Hernandez, M.D., M.H.S., Sean Collins, M.D.,
Matthew William McCarthy, M.D., Dushyantha Jayaweera, M.D.,
Nina Gentile, M.D., Mario Castro, M.D., M.P.H., Massimo Sulliowski, M.D.,
Kathleen McGigue, M.D., M.P.H., G. Michael Feller, M.D., M.H.S.
Adit A. Ginde, M.D., M.P.H., Sarah E. Durumore, Ph.D., Stacey J. Adam, Ph.D.,
Allison DeLong, B.S., George Silver, M.D., M.H.S.
Rhonda Wilder, M.S., J.
Susanna Naggs, M.D., M.H.S.

Figure S4. Kaplan-Meier plot of time to recovery with matched and unmatched placebos

- Fluticasone
- Matched placebo
- Unmatched placebo

![Kaplan-Meier plot](image)

Figure 2. Times to Sustained Recovery with Inhaled Fluticasone Furoate or Placebo.

<table>
<thead>
<tr>
<th>No. at Risk</th>
<th>Fluticasone</th>
<th>Matched placebo</th>
<th>Unmatched placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>621</td>
<td>503</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>656</td>
<td>546</td>
<td>284</td>
</tr>
<tr>
<td>Fluticasone furoate</td>
<td>656</td>
<td>546</td>
<td>284</td>
</tr>
<tr>
<td>Placebo</td>
<td>621</td>
<td>503</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>656</td>
<td>546</td>
<td>284</td>
</tr>
</tbody>
</table>

Kaplan-Meier curve for time-to-recovery primary endpoint stratified by treatment and type of placebo. This exploratory analysis used a 3-level treatment variable (active, matched placebo, unmatched placebo) in place of the pre-specified 2-level treatment variable (active versus placebo with matched and unmatched placebos combined). The unadjusted log-rank test comparing the 3 groups resulted in a p-value of 0.1. Excluding the active group and comparing just the 2 placebo groups resulted in a p-value of 0.04. The covariate-adjusted Cox model with the 3-level treatment variable resulted in a 2 degrees of freedom chi-squared test p-value of 0.01, suggesting possible heterogeneity between the placebo groups. The covariate-adjusted Cox model was consistent with the Kaplan-Meier curves in that the time to recovery for the active treatment group fell in between the time to recovery profiles of the two placebo groups. Specifically, the treatment effect hazard ratio when compared with matched placebo was 1.12 (95% CI: 0.97, 1.30). When compared with the unmatched placebo, the hazard ratio was 0.95 (95% CI: 0.72, 1.29). On the absolute scale, the unadjusted estimate of median time to recovery was 12 days (95% CI: 12, 13) for the active arm, 14 days (95% CI: 13, 16) for the matched placebo arm, and 12 days (95% CI: 10, 13) for the unmatched placebo arm.
Conclusions

• For **Safety** observational data essential
  • Ability to detect rare events (i.e., <1:1,000,000)

• For **Efficacy** need a mix (totality) of data sources
  • Utilize multiple lines of evidence
    • Mechanistic, pre-clinical, observational, clinical, and RCT
  • Ethical
  • Efficacy end point
  • Define critical settings
  • Big data does not equal unbiased/unconfounded data
  • Develop a meaningful CoP
  • Utilize randomization when possible
  • Novel designs raise novel threats to validity