N-of-1 Trials

Christopher H Schmid
Brown University School of Public Health
Department of Biostatistics
Center for Evidence Based Medicine

Christopher_schmid@brown.edu

University of Massachusetts
4 February 2014
Worcester

Joint work with Naihua Duan, Richard Kravitz, Deborah Zucker, Ida Sim and others

Funded through grants R01 NR01393B from the National Institute of Nursing Research
Contract No. 290-05-0016 from Agency for Healthcare Research and Quality
Outline

- Motivation for N-of-1 trials
- Examples
- Design features
- Analysis
- Presentation of Results
- Mobile Application
- Combining N of 1 Trials in Meta-Analysis
Problem: Heterogeneity of Treatment Effects

- RCTs generate average effects for people in trials
- Average effects *may* not (and in some cases, demonstrably *do* not) apply to the individual patient
- Patients want to know “what treatment is best for me?”
Research vs. Practice

- Clinical practice often anecdotal, local, based on trial and error
  - Need to apply treatments to individuals

- Research knowledge usually produced centrally
  - Relevance to local practice might not be compelling
  - Clinical trials address specialized populations and conditions
  - Community setting underrepresented in clinical research
    - Can reach wider spectrum of patients

- Missed opportunities to learn systematically from local practice

- Clinical practice or human subjects research?
N of 1 Trials

- Single patient multiple period crossover trials to estimate individual treatment effects
- Personalized protocol
  - Clinician and patient can design own study
  - Can select own outcomes
  - Patients have more control over study design and may be more motivated/connected
- Pragmatic design for decision-making
- Contrast to usual practice
Indications

- Substantial therapeutic uncertainty about treatment
- Treatment effect heterogeneous across patients
- On-going treatment for chronic conditions
- Clinical or patient reported outcomes can be easily collected
- Symptoms wax and wane but are relatively stable
  - Outcome expected to return to baseline after each period
- Negligible persistence of treatment effect (no carryover)
- Patient and clinician willing to put in time to learn “what treatment works better for me?”
Contraindications

- No symptoms, signs or lab tests to track/follow
- Condition is rapidly progressive
- Effects of treatment persist a long time
Structured single patient (N-of-1) trial

Treatment with Drug A (Amitriptyline)
Treatment with Drug C (Combo:AM+F)

Randomized treatment periods

Disease status measurements

Measurements on Amitriptyline

Measurements on Placebo

Multiple crossover design for single patient
Patients block randomized to two (or more treatments)
Examples

- Fibromyalgia
  - Amitryptoline vs. Amytryptoline + Fluoxetine
- Sleep for ADHD Patients
  - Melatonin vs. no Melatonin
- Chronic Pain
  - Mobile Health Application
  - Any treatments set up by patient and clinician
Key Design Elements

- Pairing within patient
- Randomization or systematic counterbalanced design (AB/BA)
  - Usually each treatment once in each block
- Blinding
- Replication
  - Length and number of study periods
  - Number of measurements per period
- Washout period to control for carryover effects
- Adaptive/play the winner
Randomization/Counterbalance

- Randomize interventions within blocks of size 2 (or greater)
- Systematically counterbalanced design (AB/BA)
- Blocking/balanced assignment controls temporal effects and minimizes consequences of early termination
- Randomization may not achieve balance for each design
- Counterbalance poor if unbalanced with respect to unknown confounder
Blinding

- Patients very involved with study
- May try to guess treatment especially as receiving both
- Subjective outcomes influenced
- But may try to be objective in order to gain most benefit
- Without pure blinding, preferences of individual patients may generalize to themselves in future, but not to others as in parallel trial (e.g. treatment preferred for convenience)
Blinding

- Can use different sized blocks if worried that patient will figure out equal block assignments
- Importance depends on:
  - Whether trial is established solely for benefit of individual patient (and not mandated/encouraged by someone else)
  - Whether one takes a more pragmatic or explanatory stance
Replication

- Need sufficient number of crossovers and measurements within period to have enough information about within and between-period variability

- Optimal allocation depends on:
  - Expected size of variance components
  - Measure validity on different time scales
  - Likelihood of dropout
  - Tendency to become less careful about measurement and following protocol over time
Carryover and Washout

- Carryover always threatens validity in crossover trials
- Bias toward null if carry-over present and not controlled for
  - Hard to know if it exists
  - Tests for it are not very powerful
Carryover and Washout

- Can try to design away by washout
  - Deviates from pragmatic design principle
  - Might not be appealing to users, esp. with active control
  - Washout increases study length and may compromise study completion
  - Removal from treatment may be ethical issue and may also reduce patient’s willingness to participate
  - But otherwise analysis depends on model validity
Analyzing Carry-over

- Can take multiple measurements within each period, then deal with carry-over analytically
- Weight measurements in follow-up periods
- More weight given to measures further from previous period
- Or, all weight given to last measurement
Analyzing Carry-over

- Flexible number of crossovers
- Early stopping
- Skew randomization ratio according to interim results
- Incorporate results from other patients; equipoise
Multiple Outcomes

- Patients often interested in several outcomes
- Standardization vs. customization
- Composite outcomes hard to interpret and may not be patient-centered
- Could weight different outcomes
- Multiple testing not an issue because decision made on basis of all evidence
Forms of Statistical Analysis

- Structured time series with treatment factor
- Descriptive analysis with graphics (visual inspection)
- Paired t or Wilcoxon test
- Time trends and time-varying treatment effects
- Carryover and correlation
- Time series analysis
  - Serial correlation
  - Trends over time
Rationale for Using Bayesian Models

- Personalized nature of decision
- Need to incorporate external information (patient, clinician)
- Interpretation of probability that one treatment better than other
- Lack of sufficient data for frequentist methods to return “significant” result
- Joint posterior distribution for composite statements about multiple outcomes
- Can also combine multiple N-of-1 studies together to get both average treatment effect and better individual treatment effects through borrowing of strength
Results to Present

- Graphs
- Probabilities/odds of A vs. B or A > B+k
- Not hypothesis testing, but decision analysis
  - Bayesian models
  - Effect sizes
  - Uncertainties
- Flexibility to accommodate user preferences (customized/menu-ized statistical feedback?)
Linear Models

Treatment effect only \( y_t = \alpha + \beta X_t + \varepsilon_t; \varepsilon_t \sim N(0, \sigma^2) \)

Treatment and linear time effects

Without interaction: \( y_t = \alpha + \beta X_t + \gamma t + \varepsilon_t; \varepsilon_t \sim N(0, \sigma^2) \)

With interaction: \( y_t = \alpha + \beta X_t + \gamma t + \delta X_t t + \varepsilon_t; \varepsilon_t \sim N(0, \sigma^2) \)
Linear Models

Treatment and block/period effects

\[ y_t = \alpha + \beta X_t + \sum \gamma_j B_j + \sum \sum \delta_{k(j)} P_{k(j)} + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma^2) \]

\[ \gamma_j \sim N(0, s^2) \text{ and } \delta_{k(j)} \sim N(0, \sigma_{k(j)}^2) \]

- Can add interaction terms for treatment*block and treatment*period
- Can add term for treatment sequence
Serial Correlation Models

\( y_t = \alpha + \beta X_t + \varepsilon_t; \varepsilon_t = \varphi \varepsilon_{t-1} + \nu_t; \nu_t \sim N(0, \sigma^2) \)

\( y_1 = \alpha + \beta X_1 + \varepsilon_1; \varepsilon_1 = \varphi \varepsilon_0 + \nu_1; \nu_1 \sim N(0, \sigma^2); M = \varphi \varepsilon_0 \sim N(0, 1000) \)

-1 ≤ \( \varphi \) ≤ 1 is correlation between consecutive errors

- M is latent (unobservable data) imputed by treating as parameter

- Can add serial correlation to all models
Lagged Dependent Variable Models

\[ y_t = \alpha + \beta X_t + \phi y_{t-1} + \varepsilon_t; \varepsilon_t \sim N(0, \sigma^2) \]
\[ y_1 = \alpha + \beta X_1 + \phi y_0 + \varepsilon_1; \varepsilon_1 \sim N(0, \sigma^2); M = \phi y_0 \sim N(0,1000) \]

- Can add lagged dependent variable to all models
- M is latent (unobservable data) imputed by treating as parameter
- Or could put prior on \( y_0 \) directly, e.g. use \( U(0, 30) \).
Fibromyalgia Study

- Amitriptyline vs. amitriptyline + fluoxetine (Zucker, 2006)

- Earlier crossover trial with 19 patients showed combination treatment better (Goldenberg, 1996)
  - Mean FIQ change better on both AM and FL compared to placebo
    \[
    \begin{align*}
    AM: 6.1 \\
    FL: 10.9
    \end{align*}
    \]
  - Combination (AM+FL) better (additive) compared to either AM or FL alone
    \[
    AM+FL: 20.5
    \]
Fibromyalgia Study

- Not all patients responded
  - Improvement of >25% compared to baseline in:
    - 5% Placebo
    - 24% AM alone
    - 32% FL Alone
    - 62% AM+FL

- Want to design individualized treatment
- Verify earlier result and get patient-specific efficacy
Fibromyalgia Study Design

- One referral practice (34 pts), 7 community practices (24 pts)
- Six intervention periods of 6 weeks each
- Two week washout periods
- Randomized in blocks of two periods
- Information collected during patient visit at end of each period
- Main outcome: Fibromyalgia Impact Questionnaire (FIQ)
## Bayesian Analysis

<table>
<thead>
<tr>
<th>Probability that:</th>
<th>No prior assumptions (using only your data):</th>
<th>Using historical info from previous studies:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug B is more effective than Drug A:</td>
<td>95%</td>
<td>97%</td>
</tr>
<tr>
<td>Drug A is more effective than baseline:</td>
<td>96%</td>
<td>98%</td>
</tr>
<tr>
<td>Drug B is more effective than baseline:</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
</tbody>
</table>
Clinical Feedback in Fibromyalgia

- Of those completing their trials:
  - 24% Chose AM alone
  - 64% Chose Combo
  - 12% Chose Other

- 88% of completers would undertake another N-of-1 and would recommend it to a friend
MYNAP  Melatonin in Youth

- Randomized, double-blind, placebo-controlled, multicenter trial of aggregated N-of-1 Trials compared to parallel group RCT
- Effects of melatonin on sleep onset latency in children and adolescents with ADHD receiving stimulant medication
- 6 weeks (3 pairs over 6 week period) for each individual
- Hypotheses
  - Melatonin alleviates initial insomnia
  - N-of-1 trials provide similar treatment effect with less uncertainty than parallel group RCT.
- 300 participants:
  - 100 < 12 yr; 50 >=12 yrs each in Australia and Canada
PREEMPT Study

- Problem: Chronic Non-Malignant Pain
- Highly prevalent (>100 million Americans)
- Costly (> $160 billion/year)
- Clinically vexing
  - Associated with disability and lost productivity
  - Difficult to predict chronicity
  - Interaction with mental health conditions
  - Can be hard to treat
  - Fraught clinician-patient interactions
  - Risks: opiate misuse, addiction, diversion, accidental harm
PREEMPT Study

N-of-1 Trials Using mHealth in Chronic Pain

Aims

- Develop mobile application to conduct N-of-1 trials among adults with chronic musculoskeletal pain

- In RCT, assess effects of using app on
  - Pain,
  - Quality of life
  - Participatory decision making
  - Satisfaction
  - Trust
  - Adherence
PREEMPT N-of-1 Study Protocol

- Compare 2 interventions
- 1-2 week treatment periods
- Cycle of 2 periods (2 to 4 weeks long, AB or BA)
- Study of 2-4 cycles (4-16 weeks)
- All choices including primary outcome made by patient/clinician
Examples of Comparisons for Testing

- **Comparison 1**
  A: Tylenol (acetaminophen) 650mg 4 times daily vs.
  B: Motrin (ibuprofen) 600 mg 3 times daily

- **Comparison 2**
  A: Vicodin (hydrocodone/acetaminophen) 5/325 (to 8 tabs daily) vs.
  B: Tylenol, acetaminophen (up to FDA recommended limit of 2600 mg daily)

- **Comparison 3**
  A: Vicodin (hydrocodone/acetaminophen) 5/325 (to 8 tabs daily) vs.
  B: Percocet (oxycodone/acetaminophen) 5/325 to 8 tabs daily

- **Comparison 4**
  A: Low dose Vicodin hydrocodone/acetaminophen 5/325 (to 8 tabs daily) vs
  B: High dose Vicodin hydrocodone/acetaminophen 10/325 (to 8 tabs daily)
Main N-of-1 Outcomes Measures

- Daily pain (3 questions scored 0-10)
- Daily sleep disturbance (5 item Likert scale)
- Daily fatigue (5 item Likert scale)
- Daily drowsiness (6 items)
- Daily constipation (5 items)
- Daily cognitive functioning (4 items)
- Neuropathic pain (3 questions scored 0-10)
- Self-reported Adherence (4 items)
mHealth

- Make participation easy and fun
- Create new opportunities for patient engagement in care
- Awaken “inner scientist” in both patients and clinicians
## Overcoming Barriers to Participation

<table>
<thead>
<tr>
<th>Challenge</th>
<th>mHealth Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confusion about purpose</td>
<td>Online education (YouTube?)</td>
</tr>
<tr>
<td>Need for significant staff support</td>
<td>Self-contained mHealth modules</td>
</tr>
<tr>
<td>Cumbersome data collection procedures</td>
<td>Data seamlessly uploaded to server</td>
</tr>
<tr>
<td>Missing data</td>
<td>Automated reminders, “ecological momentary assessment,” passive measures</td>
</tr>
<tr>
<td>Outcomes not personally relevant</td>
<td>Patients/clinician involvement in selecting outcomes</td>
</tr>
<tr>
<td>Results not interpretable</td>
<td>Customized results reporting</td>
</tr>
<tr>
<td>Sparse data from single trial</td>
<td>Potential for Bayesian “borrowing from strength”</td>
</tr>
</tbody>
</table>
Interim Results Display
Final Display

Red = A better
Purple = A trivially better
Green = B trivially better
Blue = B better

Pain
Dizzy
Constipated
Sleep

0 50 100

Outcomes with Confidence Intervals

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Better Treatment</th>
<th>Likelihood Better</th>
<th>Likelihood Better by 20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain</td>
<td>A</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Fatigue</td>
<td>B</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Depression</td>
<td>B</td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>Sleep</td>
<td>B</td>
<td>54%</td>
<td>54%</td>
</tr>
<tr>
<td>Thinking</td>
<td>B</td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>Constipation</td>
<td>B</td>
<td>60%</td>
<td>60%</td>
</tr>
</tbody>
</table>
Automated Statistical Analyses

- Not enough time for statistician to separately analyze each trial in real time
- Need to automate analyses
- Try several models and see which fits best
- Check for significance of new effects, do model diagnostics
- Need to impute missing data for some models
- Program will store data and analyses for recall at decision
- Leave a bit of lag time so statistician can check if error occurs
Combining N of 1 Studies

population

- center 1
  - Doctor 1
    - patient 7*
  - Doctor 2
    - patient 1*
    - patient 2
- center 2
  - Doctor 3
    - patient 3*
  - Doctor 4
    - patient 4
    - patient 5
    - patient 6*
Population Treatment Effects

- N of 1 can be considered a type of repeated measures design or as a type of meta-analysis design
  - Like repeated measures, unit of analysis is patient
  - Like meta-analysis, inference made to individual unit

- Repeated measures models assume common covariance structure across subjects
  - Correlated observations within subjects
  - Equal weights across subjects

- Meta-analysis models assume different variances across studies
  - Each study has univariate outcome so single variance
  - Unequal weights across studies
Multilevel Model Combining N-of-1 Studies

- Population estimate of treatment efficacy
- Improved estimates for individuals by borrowing strength
- Compromise between population estimate (complete pooling) and individual's observed results (no pooling)
  - Weighted to observed if low variation or many crossovers
  - Weighted to pooled if little information for individual
- Helps make treatment decision if individual outcomes equivocal
- May want to use common variance within-patient
- Similarity of approach to way clinicians treat their patients
Social Benefits

- Engage patients in own care
- Enhance scientific literacy in population
- Personalized medicine
- Improve clinical practice through learning environment
- Enhanced communication
- Generate randomized data
- Borrowing from strength
Thank You!