Statistical approaches in implementation science

Donna Spiegelman, Sc.D.
Professor of Epidemiologic Methods
Departments of Epidemiology, Biostatistics, Nutrition and Global Health
Harvard T.H. Chan School of Public Health
stdls@hsph.harvard.edu
www.hsph.harvard.edu/donna-spiegelman/
First, a brief theoretical overview of translational research, implementation science and related disciplines

From https://www.iths.org/about/translational
Translational Science

Dissemination Science

Implementation Science

“Bench to bedside model” of NIH

M & E
Qualitative Research, Mixed methods

Operations Research

Impact Evaluation, Comparative effectiveness research

Economic Evaluation, cost-effectiveness research

Modeling to predict future health and economic impact
Definitions

• **Implementation science**
  – The scientific study of programs and interventions which promote the systematic uptake of clinical research findings and other evidence-based approaches into routine clinical practice and public health policy, hence improving the quality (effectiveness, reliability, safety, appropriateness, equity, efficiency) of health care
  – Implementation science is about determining what works, in real-life full-scale settings.

• **Efficacy (T2) → effectiveness and cost-effectiveness (T3, T4)**
Definitions

Operations research
– assess the effects of interventions on the processes which have been shown to influence health outcomes

Impact evaluation assesses due to a particular intervention, such as a project, program or policy, both the intended ones, as well as the unintended ones. Impact evaluation is structured to answer the question:
– how would outcomes such as participants’ well-being have changed if the intervention had not been undertaken?

This involves counterfactual or causal analysis, that is, “a comparison between what actually happened and what would have happened in the absence of the intervention.” That is, changes in outcomes that are directly attributable to a program

- Impact evaluations assess the effects of interventions on health outcomes
What does the NIH Director have to say about the importance of translational science?

“The United States urgently needs the evidence to design a system that offers health interventions that are both beneficial and cost effective.”
Comparative effectiveness research (CER)

- (CER) is the direct comparison of existing health care interventions to determine which work best for which patients and which pose the greatest benefits and harms. The core question of comparative effectiveness research is which treatment works best, for whom, and under what circumstances.

- The Institute of Medicine committee has defined CER as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels."

- An important component of CER is the concept of Pragmatic Trials. These clinical research trials measure effectiveness—the benefit the treatment produces in routine clinical practice, in contrast to most clinical trials, which measure efficacy, whether the treatment works or not under ideal circumstances.
Comparative Effectiveness Research

On February 17, 2009, President Barack Obama signed into law an initiative providing $1.1 billion to support research on the comparative effectiveness of drugs, medical devices, surgical procedures, and other treatments for various conditions.

“The opposition to CER found its voice in commentators who claimed that these studies will inevitably lead to government domination of the doctor–patient relationship, “cookbook medicine,” and “rationing”

The Partnership to Improve Patient Care, a coalition of 36 industry, patient-advocacy, and clinician organizations, raised concerns that CER will not take adequate account of individual patient differences and may impede the development and adoption of improvements in medical care and “stymie progress in personalized medicine”

“Far from impeding personalized medicine, CER offers a way to hasten the discovery of the best approaches to personalization, providing more and better information with which to craft a management strategy for each individual patient”
How will this new field be used?

1. To **develop** the evidence base for cost-effective **preventive** strategies to achieve domestic and global health, and for the elimination of both domestic and global health disparities

E.g., mitigate the global obesity epidemic, which is leading to skyrocketing rates of diabetes, CVD, cancer in the US and worldwide

- Diabetes accounted for a full 12% of health expenditures in 2010 (Zhang P Diabetes Res Clin Pract; 2010).
- 60% of diabetes can be prevented by eliminating obesity (Hu et al., *N Engl J Med*, 2001)
- **Lifestyle intervention trials have reduced diabetes incidence by 58% (US), 29% (India), 42% (China).**
How will this new field be used?

2. Ensure that our resources are effectively deployed to reduce health care costs. In 2011, the United States spent more on health care per capita ($8,608), and more on health care as percentage of its GDP (17.2%), than any other nation.

- US compared to the other 16 highest income countries (National Research Council and Institute of Medicine. *U.S. Health in International Perspective: Shorter Lives, Poorer Health;* 2013)
  - U.S. men had the shortest average life expectancy
  - U.S. women had the second-shortest life expectancy
  - Americans face the lowest probability of living up to age 50

Photo Credit: Sandra Cohen-Rose and Colin Rose, flickr @colros
Design strategies

- Adaptive intervention designs
- Adaptive research designs
- Clinical trial design
- Cohort-sequential (accelerated longitudinal) designs
- Cross-over designs
- Cross-sectional designs
- Effectiveness and implementation studies
- Factorial designs
- Fractional factorial designs
- Group- or cluster-randomized trial
- Mixed or multi-method designs for qualitative and quantitative studies
- Multiple baseline designs
- Natural experiments
- Quasi-experimental designs
- Regression discontinuity designs
- Sequential multiple-assignment randomized trials (SMARTs)
- Stepped wedge design
- Time series designs
- Two-stage designs
Example: the case of maternal mortality

A woman receives antibiotics from the pharmacy at a health facility in Lusaka, Zambia. © 2014 Gareth Bentley/SCMS, Courtesy of Photoshare
Maternal Mortality – the world’s biggest health disparity (WHO)
Maternal mortality is nearly entirely preventable

Causes of maternal death

- Severe bleeding (haemorrhage) 25%
- Infections 15%
- Eclampsia 12%
- Obstructed labour 8%
- Unsafe abortion 13%
- Other direct causes 8%
- Indirect causes 20%

Traditional Birth Attendants in Pakistan

- Cluster-randomized 20,000 pregnant women in 7 sub-districts
- Result: -25% RR (95% CI -65% to +23%), p=0.24
- Inconclusive? Null?
Sample size calculation to detect effect of TBA training

• Assumes power=0.9, 5% loss to follow-up

• Assumes maternal mortality rates from 2007 WHO statistics
  – Pakistan 113,378
  – Nigeria 70,943
  – Uganda 64,204
  – Tanzania 37,460
  (assumes RR=0.74 and an ICC of 0 since 0.001 was observed in Jokhio et al., 2005)

• Jokhio et al. under-powered to detect effect (n=20,000, 1/5 the size needed)
Proposed: Multi-factorial CRT (cluster-randomized trial)
to prevent maternal mortality

- 2x2x2 factorial design will be sufficiently powered for each intervention
- Unit of randomization could be village, district, or individual

**Interventions:**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBA training &amp; supplies</td>
<td>0.74</td>
</tr>
<tr>
<td>Calcium supplement</td>
<td>0.17</td>
</tr>
<tr>
<td>Oral misoprostol</td>
<td>0.33</td>
</tr>
<tr>
<td><strong>Combined effect</strong></td>
<td><strong>0.04</strong></td>
</tr>
</tbody>
</table>
Quantitative analysis methods

- Agent based modeling
- Analysis of high-dimensional data
- Analysis of small sample data
- Bayesian methods
- Complier Average Causal Effect (CACE) analysis
- Causal inference
- Cost-effectiveness methods
- Data mining
- Decision analysis
- Econometric methods
- General linear modeling (including regression, multivariate analysis)
- Gene-environment interactions
- Generalized linear modeling (logistic, Poisson, Gamma, etc.)
- Genome-wide statistical analysis
- Geospatial analysis
- Growth modeling
- Individual person-level meta-analysis
- Integrative data analysis
- Item response theory
- Latent class and latent variable modeling
- Measurement theory and methods (EFA, CFA, etc.)
- Mediation analysis
- Meta-analysis (of summary statistics)
- Methods for analysis of intensive or long longitudinal data
- Microsimulation methods
- Missing data methods (imputation, full information maximum likelihood, etc.)
- Mixture models (including growth and regression mixture models)
- Moderation/Effect Modification/Interaction analysis
- Multi-level/hierarchical/mixed-model regression
- N of 1 experiments
- Network analysis
- Propensity score methods (matching, weighting, etc.)
- Psychometric methods
- Simulation methods
- Statistical power analysis
- Structural equation models
- Subgroup analysis
- Survey data analysis
- Survival analysis
- System dynamics
- Systems engineering methods
Three widely held myths limit widespread implementation and dissemination of translational research:

(1) Evaluations must be randomized,

(2) Valid results require individual level data

(3) Evaluation of population-level health interventions is time- and cost-prohibitive.
Myth 1: Evaluations must be randomized

- New developments in causal inference for observational research provide a wealth of methods for reducing or eliminating confounding, selection bias and information bias
  - Computationally feasible causal methods are needed for the evaluation of program impacts in the presence of time-varying confounding, using marginal structural models and g-causal models as the point of departure (Robins, 1986+)
  - Causal methods need to be extended to handle correlated data for estimating total effects (i.e. to obviate the no interference assumption)

- Loss to follow-up can restrict the validity of results from longitudinal studies
  - Contextually-specific instrumental variables for participant selection need to be identified and used to adjust for bias due to loss to follow-up, using Heckman’s selection model

- Non-adherence is a form of exposure measurement error and may be differential
Myth 2: Valid results require individual-level data

Two stage designs are well suited for translational science settings, where a mixture of individual-level and group-level data can be integrated to maximize learning at minimum cost and speed. Routinely available group-level health data can be augmented by smaller sub-studies with individual-level data, to produce valid and efficient estimates of intervention.

• Application: Familia Salama is a cluster-randomized study of maternal health and prevention of mother-to-child transmission of HIV among more than 150,000 pregnancies in DSM, Tanzania (2012-2014). Lifelong or shorter-term maternal ART may have indirect effects resulting from reduced HIV transmission to partners.
  – To assess these indirect effects, repeated administrative cross-sectional surveys conducted by the Tanzanian government can be linked to the study database in a two-stage design and analysis, to assess the indirect and total effects on lowering the community’s HIV infection rates; alternatively, population surveys can be used
Myth 3: It is too time-consuming, difficult, expensive, even unethical to collect and process the data needed for evidence-based health care decisions in real time

The stepped wedge design, in which an intervention is phased in over randomly selected administrative units over time, is another useful approach in IS

- Methods for binary, and survival endpoints need to be developed
- Software is needed for these methods
Table 2: Power for the SWD versus CRD with a baseline risk of 0.05. Assume both designs have the same total number of clusters and total sample size. RD=risk difference, ICC=intracluster correlation coefficient, I=number of clusters, J=number of steps, N=total sample size per cluster over all steps

<table>
<thead>
<tr>
<th>RD</th>
<th>ICC</th>
<th>I</th>
<th>J</th>
<th>N</th>
<th>SWD</th>
<th>CRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>8</td>
<td>3</td>
<td>15</td>
<td>0.086</td>
<td>0.165</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>80</td>
<td>3</td>
<td>15</td>
<td>0.471</td>
<td>0.872</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>8</td>
<td>3</td>
<td>45</td>
<td><strong>0.235</strong></td>
<td><strong>0.326</strong></td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>80</td>
<td>3</td>
<td>45</td>
<td><strong>0.975</strong></td>
<td>0.998</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>8</td>
<td>5</td>
<td>45</td>
<td><strong>0.362</strong></td>
<td><strong>0.326</strong></td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>80</td>
<td>5</td>
<td>45</td>
<td>0.999</td>
<td>0.998</td>
</tr>
<tr>
<td>0.05</td>
<td>0.001</td>
<td>8</td>
<td>5</td>
<td>90</td>
<td>0.869</td>
<td>0.689</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>8</td>
<td>5</td>
<td>90</td>
<td>0.869</td>
<td>0.461</td>
</tr>
<tr>
<td>0.05</td>
<td>0.01</td>
<td>8</td>
<td>3</td>
<td>90</td>
<td>0.997</td>
<td>0.910</td>
</tr>
<tr>
<td>0.025</td>
<td>0.01</td>
<td>8</td>
<td>3</td>
<td>90</td>
<td>0.645</td>
<td>0.461</td>
</tr>
<tr>
<td>0.0125</td>
<td>0.01</td>
<td>8</td>
<td>3</td>
<td>90</td>
<td>0.084</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Kunz and Spiegelman, In preparation, 2015
The cost-effectiveness ratio \((CER)\)

- The CER is an essential quantity in implementation science, integrating estimates of the effectiveness of two alternative approaches to health care delivery with their costs.

- The \(CER\) is a function of per capita costs and life expectancy under two alternative interventions, perhaps adjusted for quality of life, under each intervention.

\[
CER = \frac{Cost_B \times E_B(T) - Cost_A \times E_A(T)}{E_B(T) - E_A(T)}
\]

- Methods for estimating life expectancy are under-developed.

- Typically, health economists make the very strong assumption of an exponential survival distribution, while biostatisticians have long abandoned such parametric assumptions in the analysis of time to event data, for semi-parametric methods.

- The \(CER\) can be estimated with novel semi-parametric methods, avoiding bias due to incorrect parametric assumptions about the distributions of age at death or other relevant health outcomes.
The cost-effectiveness ratio (CER)

- Let $T$ denote the random variable, age at death, and $t_k, k = 1, ..., m$, are the observed ages at death, where we first assume that the survival probability at $t_m$ is close to zero.

- Then, the life expectancy for those who die on or after $t_1$ can be estimated by

$$\hat{E}(T|X, \mathbf{W}, T \geq t_1) = \sum_{s=t_1}^{t_m} s \left[ \hat{p}(T = s|T \geq t_1, X, \mathbf{W}) \right],$$

where $X$ is the binary indicator for the two interventions to be compared, and $\mathbf{W}$ is the vector of other risk factors included in the semi-parametric Cox regression model used to estimate the survival distribution, and

$$\hat{p}(T = t_k) = \hat{S}(t_{k-1}|X, \mathbf{W}) - \hat{S}(t_k|X, \mathbf{W}),$$

where $S$ is the survival function, and the baseline survival functions, $S_0(t_k)$, will be estimated by the product-limit method and the survival function will be estimated by

$$\hat{S}(t_k|X, \mathbf{W}) = \hat{S}_0(t_k)^{\exp\{(X, \mathbf{W})^T\hat{\beta}\}},$$

where $\hat{\beta}$ is the vector of Cox model regression coefficients.
The cost-effectiveness ratio (CER)

- Causal inference methods -- marginal structural models and g-causal models -- will be applied to control for bias due to confounding by time-varying covariates and ensure that $\hat{\beta}$ can be causally interpreted.

- The denominator of the CER will be obtained as

$$\hat{YLL}(W) = \hat{E}(T|X = 1, W, T \geq t_1) - \hat{E}(T|X = 0, W, T \geq t_1).$$

- $A\text{Var}[\hat{E}(T|X, W)]$ will be derived using asymptotic estimating equation (EE) theory, stacking up the EEs for $\hat{\beta}$ and $\hat{S}_0[t_1; \hat{\beta}], \ldots, \hat{S}_0[t_m; \hat{\beta}]$, to obtain the sandwich variance estimator. Then, the multivariate delta method will be used to derive $A\text{Var}[\text{CER}]$.

- We will develop methods for relaxing the assumption that the survival probability at $t_m$ is zero by exploring a new conditional CER, $E(T | t_1 \leq T \leq t^*)$, where $t^*$ is a pre-specified age $\leq t_m$, and compare this to the restricted CER, $\sum_{s = t_1}^{t^*} s [\hat{p}(T = s|T \geq t_1, X, W)]$ (Zucker, 1998)
The cost-effectiveness ratio \((CER)\)

At the Harvard PEPFAR site in Dar es Salaam, we have developed an extensive database of clinical, laboratory and pharmacy data from nearly 7,000 children and nearly 110,000 adults seen between 2004 – present.

Bottom line:
- Standard method; 2\(^{nd}\) line ArVs are not cost-effective in this setting (CER=$1723 relative to Tanzanian GDP of $1400)
- New method: 2\(^{nd}\) line ARVs are cost-effective in this setting (CER=$924)
- Methods matter! 2\(^{nd}\) line ARVs are cost-effective even in this severely resource-limited setting
Worksite Lifestyle Program for Reducing Diabetes and Cardiovascular Risk in India (with Drs. Narayan, Weber, Ali and others here at Emory)

- Goals: Evaluate success, scalability, effectiveness, and cost-effectiveness of a worksite intervention program
- Design: Pre-post design at 8 diverse worksites, 250 at each site, 4 in New Delhi and 4 in Chennai, India, with an external time-concurrent control
- Intervention includes individual-level interventions (healthy lifestyle education, diabetes screening) and worksite interventions (e.g., healthy foods in cafeteria, worksite lunchtime exercise classes)

Aim 1 Measure the **success** of implementation and assess the **scalability** of this intervention program by evaluating:

(a) program adoption  
(b) fidelity to the program; and  
(c) acceptability of the program through

Aim 2 To measure the **effectiveness** of the program among participants by evaluating the change in number of individuals reaching two or more of cardiometabolic risk goals

Aim 3 To measure the **value** and **return on investment** of the intervention for employers
New proposals: Substantive

**Optimizing Integrated Cervical Cancer Prevention at Tanzanian HIV Clinics**

- Goals: Evaluate effectiveness and cost-effectiveness of “see and treat” versus “screen and refer” for screening for and preventing cervical cancer through early stage treatment
- Design: Cluster randomized controlled trial to evaluate “see and treat” versus “screen and refer”, include 20 HIV treatment and care facilities across Dar es Salaam, Tanzania, with 670 patients/facility

- **Aim 1:** To evaluate clinic-based screening combined with referral to higher-level cryotherapy clinics facilitated by CHWs to maximize retention/adherence, vs. “see and treat” site-based screening with immediate cryotherapy, when indicated. We will compare the effectiveness of the two strategies with respect to screening coverage and appropriate treatment rates.
- **Aim 2:** To conduct cost-effectiveness analyses (CEA) to evaluate clinic site-based screening combined with referral to cryotherapy clinics facilitated by CHWs vs. “see and treat” site-based screening with immediate cryotherapy.
New proposals: methodologic

New Methods for the Design and Evaluation of Large HIV Prevention Interventions

Goals: Develop novel network methods to improve evaluation of the impact of HIV prevention interventions, to evaluate effectiveness and cost-effectiveness of interventions

• **Aim 1** To improve the evaluation of the causal effects of specific interventions that comprise a ‘package’.

• **Aim 2** To augment prevention trial designs to adjust for bias due to loss to follow-up. Refusal to test and selective loss to follow-up can bias effect estimates.

• **Aim 3** To develop advanced statistical methods for cost-effectiveness estimation and stepped wedge designs, Semi-parametric survival data analysis methods will provide improved estimates of comparative effectiveness ratios (CER) and their uncertainty.

• **Aim 4** To develop and validate novel network methods to improve the evaluation of large combination prevention clinical trials. Based on social linkages in cell phone data, we will investigate the attenuating effect of cross-community mixing on CRHPT effect estimates, and use this information in analyses of study results to help better evaluate the effectiveness of HIV prevention interventions.
Many Funding Opportunities
(NIH and other opportunities
5/14/2014 from HSPH ORSD)

**CDC: Evaluation of Health Promotion and Prevention Programs for Blood Disorders Grant**
Closing Date: May 21, 2014

*New* - **DHHS: Minority AIDS Initiative Continuum of Care Pilot - Integration of HIV Prevention and Medical Care into Mental Health and Substance Abuse Treatment Programs for Racial/Ethnic Minority Populations at High Risk for Behavioral Health Disorders and HIV**
Closing Date: June 4, 2014

**DHHS: Agency for Health Care Research and Quality Exploratory and Developmental Grant to Improve Health Care Quality through Health Information Technology (IT) (R21) Grant**
Closing Date: November 16, 2016

**NIH: NIJ FY 14 Developing Knowledge About What Works to Make Schools Safe**
Closing Date: July 10, 2014

**NIH: mHealth Tools for Underserved Populations with Chronic Conditions to Promote Effective Patient-Provider Communication, Adherence to Treatment and Self-Management (R01) Grant**
Closing Date: May 7, 2017

**NIH: Healthy Habits: Timing for Developing Sustainable Healthy Behaviors in Children and Adolescents (R21) Grant**
Closing Date: May 7, 2017

**NIH: Improving Delivery of HIV Prevention and Treatment through Implementation Science and Translational Research (R01)**
Closing Date: January 7, 2017
Many funding opportunities

*New - USDA: Competitive Grant to Establish a USDA Center for Behavioral Economics and Healthy Food Choice Research
Closing Date: June 30, 2014

NIH: Strengthening Adherence to Antiretroviral-Based HIV Treatment and Prevention (R01)
Closing Date: January 7, 2017

NIH - Planning Grants for Pragmatic Research in Healthcare Settings to Improve Diabetes Prevention and Care (R34) Grant
Closing Date: November 1, 2016

NIH: Translational Research to Improve Diabetes and Obesity Outcomes (R01)
Closing Date: January 7, 2017

NIH: Dissemination and Implementation Research in Health
Closing Date: January 7, 2016

NIH: Obesity Policy Evaluation Research (R01)
Closing Date: May 7, 2016
Conclusion

- **Implementation science** and its related disciplines are new and growing areas of research, ideally situated in schools of public health. Epidemiology and biostatistics have much to offer, both in terms of the methods we’ve developed and methods that grow out of them, and in terms of translational research projects that stem directly from our existing more etiologic research.
- There is much to learn.
- There are many new opportunities for finding support for translational public health research, both methodologic and substantive.