Design Options: Studying Welcome Baby Impacts in the Best Start LA Communities

Todd Franke, Christina Christie, Lourdes Brown, UCLA
Embry Howell, Urban Institute

23 April 2014
Goal of an Impact Study

• To establish Welcome Baby as an evidence-based model meeting HomVEE Standards
  – Impact on key maternal outcomes
  – Impact on key child outcomes
  – Identify variation in impact by demographic variables
  – Impact by varying dosage levels
Research Questions

1. What is the impact of the Welcome Baby program:
   - On key maternal outcomes?
   - On key child outcomes?
   - On key family outcomes?

2. To what extent does the impact of Welcome Baby:
   - Vary by demographic subgroups?
   - Vary by other subgroups?
   - Vary by dosage? (timing, duration, intensity)
Overview of Presentation

• Designs that meet “high” HomVEE ratings:
  1. Randomized Control Trial
  2. Regression Discontinuity Design

• Designs that meet “moderate” ratings:
  1. Propensity Score Matching
  2. Matched Comparison Groups

• Designs that may not meet HomVEE standards

• Additional considerations for an impact study
Highly Rated HomVEE Designs

Designs having highest internal validity:

1. Randomized Control Trial (RCT)
2. Regression Discontinuity Design (RDD)
Randomized Control Trial (RCT)

- Welcome Baby participants are enrolled and then randomly assigned to either intervention group or one or more control groups ("arms")
- Assignment occurs after eligibility assessment, before services
- Size of treatment and control groups could vary
Randomized Control Trial (RCT)

Feasibility

PROS

• Highest rigor for causal inference
• Arms could be used to examine dosage

CONS

• May be difficult to get community buy-in, especially if some are denied services
• Possibly difficult to implement rigorously (requires program changes)
Regression Discontinuity Design (RDD)

• Uses risk screener that is related to program outcomes (the “forcing variable”)
• Women assigned to treatment or comparison group depending on cut-off of forcing variable, (e.g. Modified Bridges for Newborns Screening Tool)
Regression Discontinuity Design (RDD)

Alternative approaches:

• Use only women close to cut-off score, with those above in the treatment group and those below in comparison group
• If forcing variable is continuous, use all women and assign to one or more arms based on alternative cutoff scores (eg. to Select Home Visiting, Welcome Baby, Welcome Baby Lite, and Referral Only)
Regression Discontinuity Design (RDD)

Could maintain current Welcome Baby model and examine dosage:

• Compare Select Home Visiting to Welcome Baby for those who live in the Welcome Baby community:
  >=60 get Select Home Visiting and <60 get Welcome Baby

• Compare Welcome Baby Lite to Referrals Only for those that Live Outside the Welcome Baby community:
  >=60 get Welcome Baby Lite and <60 Get Referrals Only
Regression Discontinuity Design (RDD)

Feasibility

PROS

• High rigor for causal inference if implemented well
• May help overcome resistance to random assignment
• Could be used to examine dosage

CON

• Hard to implement strictly
Moderate HomVEE Designs Using Matched Comparison Groups

• Two Alternatives:
  1. Propensity Score Matching
  2. Hospital-Based Comparison Group

• Data collection methods are standardized across treatment and comparison groups
Propensity Score Matching

- Estimate the effect of Welcome Baby by accounting for the observed factors that predict receiving treatment
- Comparison group identified from another data base such as WIC records, hospital electronic health records, vital statistics, or Medi-Cal (for example)
- Welcome Baby mothers enrolled and their characteristics are used for matching to other files
- Matched comparison group then recruited into study before infant’s first birthday
Propensity Score Matching

Feasibility

PROS

• Does not require program modifications

CONS

• Will only meet “moderate” standards if groups are matched on baseline outcomes; identifying appropriate outcomes is difficult
• Requires cooperation/intensive effort of organizations providing data
• Organizations or IRBs may not agree to contact with comparison group mothers without consent
• Susceptible to unreliability in administrative data used to assign mothers to comparison group
• Does not control for unobservable factors associated with outcomes
Hospital-Based Comparison Group

- One or more Welcome Baby hospitals recruited for study
- Baby’s birth date used to assign mother to one or more arms, for example:
  - May births offered Welcome Baby
  - June births offered Welcome Baby Lite
Hospital-Based Comparison Group

Feasibility

PROS
• Does not require major program modifications
• Could be used to study dosage

CONS
• Will only meet “moderate” standards if groups are matched on baseline outcomes; identifying appropriate outcomes is difficult
• Comparison group offered different program; could lead to selection
• Does not control for unobservable factors associated with outcomes
Other Quasi-Experimental Designs that May Not Meet HomVEE Standards

- Community-based comparison group
- Concurrent hospital comparison group
- WIC-based comparison group
Additional Considerations for an Impact Study

- Selection of Best Start communities
- Sample size
- Attrition
- Contamination
- Primary and secondary data collection
- Analysis
- Timeline
- Cost implications
Selection of Best Start Communities - Multiple BSLA Sites

PROS

• Multiple sites provide broader scope and likely recruit a more diverse sample
• Diverse sample enhances generalizability of findings and lends itself to conducting subgroup analyses

CONS

• May be more challenging to identify and control for community-level variance
• More costly due to increased data collection demands (travel)
• Management of day-to-day operations more challenging with multiple sites
Sample Size

• Power analyses will determine final sample size needed for chosen design

• Sample size determined by the number of communities, comparison groups, sub-groups analyzed, and budget constraints
Attrition and Contamination

• Attrition should be minimal and similar in treatment and comparison groups (regardless of design)

• Contamination: the comparison group may get alternative services similar to Welcome Baby; must track that in data collected
Primary Data Collection

Baseline Survey:

• Administer baseline survey at recruitment to measure differences in treatment and comparison groups

• Could administer by telephone
Primary Data Collection

Follow-up Surveys:

• For example, administer at 12-, 24-, 36-, and/or 48-month time points.

• Could conduct brief phone surveys at 6-month intervals to improve retention
  – 12-month: immediate impact on key outcomes (breastfeeding, mother’s knowledge of child development)
  – 24-month: allows additional outcomes (child language, child nutrition) and minimizes attrition due to engagement
  – 36-month: allows for additional outcomes on parenting practices, developmental progress over time, and school readiness
  – 48-month: comprehensive school readiness evaluation
Secondary Data Collection

• Acquire and match administrative datasets
• Examples:
  – Stronger Families LA Database
  – Medi-Cal
  – Dept. of Child and Family Services (DCFS)
  – WIC
  – Dept. of Mental Health (DMH)
Analysis

- Examine impact of Welcome Baby on child and family outcomes, controlling for other factors affecting outcomes
- Examine effect of different levels (dosage) of Welcome Baby
- Examine differential impacts by subgroups, for example risk levels, hospital/Best Start community
- Annual reports provided at end of each data collection time point
Timeline or Phasing

- A proposed schedule could be as follows but different design option may require different timelines
  - 4 months: study design and BS community selection
  - 4 months: design survey instrument, program CAPI
  - 2 months: enroll moms for pilot, pilot instrument
  - 3 months: revise data collection tools, train staff, begin sample recruitment
  - 1 year: baseline data collection
  - 1 year: 12-month data collection
  - 1 year: 24-month data collection
  - 1 year: 36-month data collection
  - Final report: 6 months after completion of data collection at each time point (12-, 24-, 36-month)
Cost Implications

- Most expensive designs are the those with multiple communities (for generalizability) and those with multiple “arms” (for studying dosage)
- Cost of data collection is comparable across designs
- Primary data collection costs vary based on size of study sample, length of survey, and number of data collection points
Questions?