

Evaluation of Engagement

No engagement -> no outcome

No outcome -> might still have engagement

What to measure (all over time): (*what have we missed?*)

Some inspiration from web stats ... what else is similar?

Behavioral:

- Touch
- Length of time
- Traffic (different people)
- Traffic individual (day/month)
- Response time (some applications)
- How they respond (length of time in specific app or specific measurable activity)
- Last state prior to disengagement at that moment
- % of feature space used (versus expected % used based on level of expertise)
- (If you know why they are doing they are doing) then measure # of things they don't have to do (voluntary)
- Psychophysiology measurement (e.g. HR when using, eye movements)
- Number of people they engage (challenges if need special equipment; IRB issues)

Subjective self report

- Desire to continue work with the system (motivation)
- Gather information about barriers?
- Keep going...

Outcomes (if engagement leads to outcome)

- E.g., Weight
- "Useful" data generated

Some design options:

Compare multiple designs: compare against...

- Straw man with very different behavior (easy to criticize)
- Same app with one feature different
- Best known "gold standard" (if exists)
- "Standard of care" (if an option)

Wait list control

(get baseline then...)

- Condition A
- Condition B
- "We'll call you"

Only seems to work for measurement of outcome, not engagement

Evaluating engagement based on voluntary use at end of study

- May be costly to do if special equipment needed
- IRB challenge?

Related "free market" approach

(BEWARE: only try if your system is REALLY GOOD):

Setup experiment to assess how much person will pay to keep the access to the system

(Reality check: May be more likely to pay not to use the system)

Overcoming and/or Measuring Novelty

Characterizing long term vs. novelty

Want to tease apart experimental interest vs. novelty (how?)

What is the minimum length of time to have any hope of getting beyond novelty?

- 3 weeks?
- Does every application have a novelty peak? Should you keep going until you see it? Reasonable?

Release early - release often

- Hard (because has to be good and flashy ... need to get good PR)
- May feel very out of control
- Complicates analysis (what about "lost" people)
- Major subject selection bias
- Separate social effects of peer pressure from actual effects

Would like to get data on usage/novelty effects of products like Nike/iPod, Steelcase desk, etc.

Talked about viral/grass roots strategy (if special equipment not needed). UV agent ... sneak up by starting with sports, People magazine, etc. Measure engagement and uptake.

- Still may be expensive/difficult to run study