Levels of Evidence in Prevention Research

Brian R. Flay, D.Phil
Department of Public Health
OREGON STATE UNIVERSITY

Prepared for IOM/NRC Committee on Prevention of Mental Disorders
Washington DC, October 31 2007
Outline of Presentation

I. What is evidence?
   – Phases of research
     • Efficacy Trials, Effectiveness Trials, Dissemination Research

II. What kind of evidence exists about efficacy and effectiveness of preventive interventions?

III. Challenges and conclusions
I. What Do We Mean By Evidence?

It is necessary to distinguish efficacy, effectiveness and readiness for dissemination

– Flay, 1986; Greenwald & Cullen, 1985; NIH Roadmap; SPR Standards (Flay et al., 2005); What Works Clearinghouse

I will address these three “phases” in turn

“…studying what interventions work is an iterative and cyclical, rather than a linear, process.”

– Mercer et al., 2007
1. Efficacy Trials

- Primacy of RCTs (internal validity)
- Clearest approach to establishing that an intervention is the cause of an effect
- Efficacy trials, by their nature, have low external validity or generalizability
  - “Where did the field get the idea that evidence of an intervention’s efficacy from carefully controlled trials could be generalized as THE best practice for widely varied populations and settings?” (Green, 2001)
  - “The potential number of setting–population–problem combinations is infinite, so replication of efficacy trials on each intervention in all those combinatorial configurations will never be possible.” (Green, 2007)

- Generalizability is improved with
  - Multiple replications
  - Multiple effectiveness trials
  - Reviews/meta-analyses
    - Reviews/meta-analyses are the true unit of progress, not any individual study
When Efficacy Trials are not Possible (see Holder et al., 1995, 1999)

- Some kinds of interventions are not easily amenable to randomized efficacy trials
  - Large-scale community interventions
  - Policy changes at state or national level

- However, even some of these are amenable to efficacy trials – natural experiments
  - Interrupted time-series analyses
  - Multiple-baseline time-series, community by community or state by state

- Otherwise, get best estimate of efficacy from effectiveness trials – that often use the same methods as listed above
2. Effectiveness Trials

- Established efficacy is a necessary condition for effectiveness (SPR Standards – Flay et al 2005)
  - Even when an efficacy trial is not conducted before an effectiveness trial (e.g., a new policy)

- Other necessary conditions include:
  - Availability (intervention must be delivered to or be available to the target audience)
    - A combination of adoption (by some agency to deliver) and implementation in Glasgow et al’s (1999) RE-AIM framework
    - Efficacy x Implementation = Effectiveness
  - Acceptability (target audience must accept, participate in, comply with or adhere to the intervention).
    - = Reach in Glasgow’s terms
Generalizable to What?

- Patients/recipients/target audiences
- Setting/context conditions
- Clinicians/providers
- Outcomes – including economic
- Time
- Treatment nuances, including integrity of delivery
  - Cronbach, et al. (1972)
  - Shadish, et al. (2002)
# Comparison of Efficacy and Effectiveness Trials

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Efficacy Study</th>
<th>Effectiveness Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants are:</strong></td>
<td>Homogeneous, highly motivated sample; exclude those with complications, other comorbid problems</td>
<td>Broad, heterogeneous, representative sample; often using a defined population</td>
</tr>
<tr>
<td><strong>Interventions are:</strong></td>
<td>Intensive, specialized interventions that attempt to maximize effect size; standardized</td>
<td>Brief, feasible interventions, not requiring great expertise; adaptable to setting</td>
</tr>
<tr>
<td><strong>Settings are:</strong></td>
<td>Usually one setting to reduce variability; settings with many resources and expert staff</td>
<td>Appeals to and works in multiple settings; adaptability to fit setting</td>
</tr>
<tr>
<td><strong>Delivery is:</strong></td>
<td>By research staff closely following specific protocol</td>
<td>By variety of different staff with competing demands, using adapted protocol</td>
</tr>
<tr>
<td><strong>Cost and Sustainability are:</strong></td>
<td>Often not an issue</td>
<td>Major issues; setting level maintenance equally important as individual</td>
</tr>
</tbody>
</table>
3. Dissemination Research

- Established efficacy and effectiveness are necessary conditions for an intervention to be ready for dissemination (SPR Standards)

- Other necessary conditions
  - Evidence of practical importance
  - Intervention is standardized/manualized
  - Training and technical support is available
  - Monitoring and evaluation tools are available
  - Clear cost information
  - Provider has ability to “go-to-scale”

- Desirable conditions
  - Data on cost-effectiveness
All of the Above Argues For ...

- Phased research
- Rigorous research designs at all phases
- Quantitative research
- Qualitative research to
  - Supplement/complement quantitative research
  - Help understand what works
  - Help understand conditions under which interventions work
  - Help understand why interventions work
II. Evidence of Efficacy and Effectiveness of Adopted Preventive Interventions

Many systematic reviews and meta-analyses of substance use, violence and mental health problem prevention have now been conducted. They generally show moderate positive effects (.1 - .5). Some are really “treatment effects” – Gillham et al. 2000.

Some major differences:

- Larger ESs for cognitive-behavioral approaches
- Larger ESs for interactive programs
- ESs for high-risk groups (indicated/selective) about 3 times ESs for the general population (universal)
  - Indicated selective/interventions include only people already at highest risk for or already doing what is being prevented
- Larger ESs from RCTs (usually research-driven programs)

Adoption of effective interventions is a slow process.
## Selected Results from Tobler et al., 2000 MA of Drug Prevention

<table>
<thead>
<tr>
<th></th>
<th>N ESs</th>
<th>ES*</th>
<th>99% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Non-interactive&quot; programs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge only</td>
<td>18</td>
<td>0.11</td>
<td>-0.04, 0.26</td>
</tr>
<tr>
<td>Affective only</td>
<td>7</td>
<td>-0.04</td>
<td>-0.40, 0.32</td>
</tr>
<tr>
<td>Decisions/Values/Attitudes</td>
<td>13</td>
<td>-0.02</td>
<td>-0.16, 0.12</td>
</tr>
<tr>
<td>Knowledge + Affective</td>
<td>14</td>
<td>0.07</td>
<td>-0.12, 0.26</td>
</tr>
<tr>
<td>DARE-type</td>
<td>17</td>
<td>0.03</td>
<td>-0.04, 0.10</td>
</tr>
<tr>
<td>&quot;Interactive&quot; Programs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Influences</td>
<td>82</td>
<td>0.14</td>
<td>0.10, 0.18</td>
</tr>
<tr>
<td>Comprehensive Life Skills</td>
<td>47</td>
<td>0.17</td>
<td>0.12, 0.22</td>
</tr>
<tr>
<td>System-wide change</td>
<td>9</td>
<td>0.22</td>
<td>0.11, 0.32</td>
</tr>
</tbody>
</table>

* All ESs shown are from the higher quality studies
Durlak & Wells (1998)

Selected effect sizes from different types of indicated prevention for children and adolescents at immediate posttest and follow-up

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>Post</th>
<th>Follow-up</th>
<th>Months FU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonbehavioral</td>
<td>11</td>
<td>0.09</td>
<td>0.11</td>
<td>24</td>
</tr>
<tr>
<td>Behavioral</td>
<td>12</td>
<td>0.51</td>
<td>0.44</td>
<td>5</td>
</tr>
<tr>
<td>Cognitive-behav</td>
<td>12</td>
<td>0.8</td>
<td>0.83</td>
<td>3</td>
</tr>
</tbody>
</table>
Skowron & Reineman (2005)

Effect sizes by type of outcome measure for child maltreatment prevention

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>N</th>
<th>N ES</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child cognitive process</td>
<td>7</td>
<td>22</td>
<td>0.28</td>
</tr>
<tr>
<td>Child personality report</td>
<td>8</td>
<td>43</td>
<td>0.44</td>
</tr>
<tr>
<td>Parent ratings child behav</td>
<td>9</td>
<td>34</td>
<td>0.42</td>
</tr>
<tr>
<td>Behav observations child</td>
<td>4</td>
<td>13</td>
<td>0.3</td>
</tr>
<tr>
<td>Parent self-report</td>
<td>7</td>
<td>42</td>
<td>0.53</td>
</tr>
<tr>
<td>Behav observations family</td>
<td>2</td>
<td>14</td>
<td>0.21</td>
</tr>
</tbody>
</table>

The table above shows the effect sizes (ES) for different types of outcome measures in child maltreatment prevention studies. The effect sizes are calculated based on the number of studies (N) included in each category.
Lists of Evidence-Based Programs

- CSAPP/NREPP – recently revised – multiple behaviors
- US Dept Educ – Safe and Drug Free Schools
- Blueprints – U of Colorado - violence prevention
- NIDA-funded research-based guide
- U of Maryland What Works in crime prevention
- What Works Clearinghouse – US Dept Educ
  - Includes two behavioral areas to date – character education and dropout prevention
- OJJDP Model Programs Guide – multiple behaviors
- And dozens of others!

Hardly any listed programs have replicated results
  - LST has the most (>11 studies) (Gorman, 2005a)
Adoption of effective interventions is a slow process

- It takes an average of 17 years for results of RCTs to move into practice
  - (Balas & Boren, Yearbook of Medical Informatics, 2000)
- This is the 2\textsuperscript{nd} translational gap
- the 1\textsuperscript{st} is the gap between basic science and testing new ideas in clinical research
- The “gap between research and practice is the result of several interacting factors, including
  - limited time and resources of practitioners,
  - insufficient training,
  - insufficient incentives for use of evidence-based practices,
  - inadequate infrastructure and systems organization to support translation.”
  - Glasgow et al., 2003, pg 1261
Improving the evidence base by design – beyond efficacy

- Representative groups and participants
  - Represent diversity of population by race, ethnicity, age, gender, literacy
  - Reflects health status of population
  - Includes those who might benefit most
- Multiple and diverse settings
  - Include multiple and representative community settings
  - Typical, no-research staff involved
  - Study variations in process and outcomes across settings
- Implementation monitoring
  - Assess extent program or policy was delivered or implemented
  - Variability in enforcement
  - Document time and expense data
- Comparison groups, conditions realistic
  - If comparison intervention, select realistic alternative
  - Assess what was delivered/implemented in comparison condition
  - No such thing as a pure control group
- Broad range of outcome measures relevant to decision makers
  - Meaningful evaluation outcomes for participants, community members, and decision makers
  - Monitor intended and unintended consequences, include quality of life
  - Economic analysis
- Replication, replication, replication
Examples of Independent Successful Replication Trials

- Studies by investigators independent of the initial investigator or the intervention developer with positive effects provide greater confidence.
- Spoth’s replication of LST and Strengthening Families.
- Multisystemic therapy [Cochrane collaboration review].
- Ted Taylor’s trial of The Incredible Years.
Why Replications Might Fail

- Ineffective approach applied due to insufficient evidence
- Implementation of the approach is incomplete so full effect not gained
- Approach was effective only for specific measures of selected outcomes
- Tested approach is not feasible/practical in real world settings
  - Hallfors et al 2006a, Gottfredson et al 2006
- Approach is not generalizable to local community so effects don’t replicate
- Evaluation fails to capture effect or does not assess meaningful outcomes
Independent Non-Replications

Example of failed replications in independent evaluations
- Chicago Project Northland (Komro et al at SPR 2006)
- Life-skills training in PA (Smith et al 2004)
- Project Alert in PA (St. Pierre et al, 2004)
- Strengthening Families Program (Gottfredson et al 2006)
- Nurse Home Visitation (Alper 2002)
- Reconnecting Youth (Halfors et al 2006)

What does non-replication of results mean?
- Depends on how well conducted and how many
  - It was never Popper’s intent to suggest that a theory (or program) be rejected on the basis of a single discordant observation (Cook & Campbell, 1979)
  - It is unwise to give up on a theory (or program) too easily as one might never discover the possibilities inherent in it (Popper, 1985)
  - We need to retain the ability to “self-correct” (Bunge 1998, Gorman 2005)

Important methodological and substantive observations
- Document level and integrity of implementation
- Track adaptations (changes, deletions, additions)
III. Issues, Challenges and Conclusions

- We need comprehensive prevention
  - Problem behaviors and mental health problems have many common causes
- Challenges of program development, evaluation and implementation
- Conclusions
We Need Comprehensive Prevention: Problem Behaviors and Mental Problems Have Common Risk and Protective Factors

- Catalano et al., 2004
- Durlak, 1998
- Elliott et al., 1985
- Flay, 2002
- Jessor & Jessor, 1977
- Integrated or comprehensive theories
- Theory of Triadic Influence
  – Flay & Petraitis, 1994
The Theory of Triadic Influence (TTI): Integration of Theories (Flay & Petraitis 1994)

See whole talk at [http://www7.nationalacademies.org/bocyf/090805.html](http://www7.nationalacademies.org/bocyf/090805.html)
Challenges

- “Fidelity” vs. “Customization to Local Settings”
- Adaptive and Tailored vs. Standardized, Generic, Static Evidence-Based Programs
- Importance and Evaluation of Program Implementation Level and Integrity
- Importance of replication and reviews
  - Accumulation of knowledge
  - Critical rational examination of it
Conclusions

The future is multiple (interventions, populations, settings, conditions, behaviors, interactive modalities)

Important question is less. “What works?” and more “What combination of approaches works best, for whom, and under what conditions of delivery?”

The future is complex (we ignore complexity at our peril)
- “All models (and designs) are only partial”
  - Sterman, 2002
- Greater tolerance, respect, and creativity are needed

“The significant problems we face cannot be solved by the same level of thinking that created them.”
- A. Einstein
My question is: Are we making an impact?
References - 1


References - 2


Gorman DN (2005a) Does measurement dependence explain the effects of the Life Skills Training program on smoking outcomes? Preventive Medicine, 40, 479-487.


References - 3


References - 4