

# THE EFFICACY-EFFECTIVENESS DISTINCTION IN TRIALS OF ALCOHOL BRIEF INTERVENTIONS

**Nick Heather PhD,  
Department of Psychology,  
Northumbria University, UK  
Presentation at 10th INEBRIA Conference,  
Rome, 20<sup>th</sup> September, 2013**

## THREE 'DISAPPOINTING' RECENT FINDINGS (1)

- van Beurden, I., Anderson, P., Akkermans, R., Grol, R., Wensing, M., & Laurant, M. (2012). Involvement of general practitioners in managing alcohol problems: a randomized controlled trial of a tailored improvement programme. *Addiction, 107, 1601-1611.*
- Conclusions: “A tailored, multi-faceted programme aimed at improving general practitioner management of alcohol consumption in their patients failed to show an effect and proved difficult to implement. There remains little evidence to support the use of such an intensive implementation programme to improve the management of harmful and hazardous alcohol consumption in primary care.”

## THREE 'DISAPPOINTING' RECENT FINDINGS (2)

- SIPS
  - Kaner, E., Bland, M., Cassidy, P., & et al (2013). Effectiveness of screening and brief alcohol intervention in primary care (SIPS trial): pragmatic cluster randomised controlled trial. *British Medical Journal*, 346, e8501.
  - No superiority of brief counselling or brief advice over patient information leaflet + feedback
  - True for all 3 arms of trial: primary health care, accident & emergency departments, probation services

# THREE 'DISAPPOINTING' RECENT FINDINGS (3)

- PRE-EMPT
  - Butler, C.C., Simpson, S.A., Hood, K. *et al.* (2013). Training practitioners to deliver opportunistic multiple behaviour change counselling in primary care. *BMJ*, 346, f1191.
  - No differences in alcohol outcomes for hazardous and harmful drinkers over a 12-month period following interventions delivery by general practitioners and practice nurses.
  - Interpretation: “Enduring behaviour change and improvements on biochemical and biometric measures are unlikely after a single routine consultation with a clinician trained in behaviour change counselling, without additional intervention.”
- See also McCambridge, J. (2013). Brief intervention content matters (Editorial). *Drug & Alcohol Review*, 32, 339-341.

## Saitz letter to BMJ 17-1-13

- **“Particularly given the robust findings from systematic reviews that favor brief intervention (advice, counsel) when compared to no brief intervention in efficacy trials ..., the conclusion most consistent with these data is that even when great efforts are made to implement SBI in real world clinical care (eg, with less external researcher support), the effects seen in efficacy studies do not translate into effective interventions in practice.”**
- **“... the effect sizes in efficacy studies, while large from a public health perspective, are small enough (eg, 3 fewer drinks per week) that they could easily be erased when SBI is not implemented in practice exactly like it was in those studies.”**

## Saitz letter to BMJ 17-1-13 cont...

- **“Yet alcohol SBI can only reach its potential if the effects seen in efficacy studies can be achieved in real world practice.”**
- **“Kaner et al’s systematic review suggested that the practice was similarly effective in trials in which SBI implementation looked more like it would in clinical practice and less like research implementation, but none of those studies came close to being pragmatic trials like SIPS, so they couldn’t really inform that question.”**
- **“Policymakers should be leery of widespread implementation unless it is done well. And it will take a lot to do it well – saying we are doing it well without assuring high quality implementation ... will give us false reassurance that we have taken care of unhealthy alcohol use and will waste time and money.”**
- **“... researchers and educators should turn their attention to how to implement alcohol screening and brief intervention in clinical practice in a way that retains the efficacy seen in clinical trials.**

# SIPS authors' reply to Saitz

- **“In contrast to Professor Saitz, we feel that the brief intervention evidence-base to date has indicated ... a growing preponderance of effectiveness rather than efficacy trials.”**
- **(In the Cochrane systematic review) “... the majority of studies ... were judged to be clinically relevant effectiveness trials (with high external validity) rather than ideal world efficacy trials (with high internal validity). In a field that has evolved for over 25 years, it is to be expected that evaluations have increasingly reflected the variability and constraints of real world primary care.”**
- **They also pointed to the difference between the SIPS trial, in which the aim was to evaluate the impact of SBI on patients' drinking outcomes, and the van Beurden trial which was an implementation (service-delivery) trial to evaluate the impact of an intensive multi-faceted improvement programme on GPs' management of alcohol problems and where the primary outcome was the number of patients screened and intervened with.**
- **There therefore appears to be disagreement and some confusion over the meaning and applicability of the terms 'efficacy', 'effectiveness', 'implementation' and 'pragmatic' trials.**

# AIMS

- To clarify the meaning of the terms ‘efficacy’ and ‘effectiveness’ trials and other related concepts, and to try to dispel some of the confusion surrounding these terms.
- To review the method and findings on efficacy-effectiveness measurement in the Kaner et al. Cochrane review.
- To make suggestions for further research concerning the efficacy-effectiveness distinction.



# Pragmatic trials

- Schwartz, D., & Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases*, 20, 637-648.
- Explanatory trials concerned primarily with *understanding*, whereas pragmatic trials concerned primarily with *decision*.
- In a pragmatic trial, treatments are compared “under the conditions in which they would be applied in practice.”
- Similar in many ways to the efficacy-effectiveness distinction but with special implications for the aims and design of the trial.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.

- *Efficacy trials* provide tests of whether a technology, treatment, procedure or program does more good than harm when delivered under optimum conditions
- *Effectiveness trials* provide tests of whether a technology, treatment, procedure or program does more good than harm when delivered under real-world conditions
- NB. Efficacy is necessary to but not sufficient for effectiveness (ie, if a treatment is effective, it must be efficacious but if it is efficacious, it need not necessarily be effective.)

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs.

*Preventive Medicine, 15, 451-474.*

- 4 LEVELS OF HEALTH PROMOTION

PROGRAM TESTING:

- a) Efficacy trials, under optimum conditions of program implementation and recipient participation;
- b) Treatment effectiveness trials, with expected variation in target audience acceptance;
- c) Implementation effectiveness trials, under varying conditions of implementation;
- d) Program evaluation of previously untested programs.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15, 451-474.*

- An efficacy trial provides a test of
  - a well-specified standardised treatment/program that
  - is made available in uniform fashion, within standardised contexts/setting, to a specified target audience which
  - completely accepts, participates in, complies with, or adheres to the treatment/program as delivered.
- Also usually involve randomised comparison or control groups and utilise, where possible, blinding procedures and placebos.

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15, 451-474.*

- An intervention will be *effective* only if an *efficacious* intervention is delivered/implemented in such a way as to be made *available* to an appropriate target audience in a manner *acceptable* to them (ie, that they will be receptive to, participate in, comply with or adhere to).
- Thus the observed effects, or lack thereof, of an intervention may be due to one or more of the following:
  - the efficacy level of the evaluated intervention;
  - the availability of the intervention to the target audience (which may be affected by the mode and extent of intervention delivery/implementation); or
  - the level of acceptance of (participation in, compliance with, or adherence to) the intervention by the target audience.
- Effectiveness trials can be experimental (RCT) or quasi-experimental and do not usually include double- or single-blind procedures and placebos.
- Some efficacious interventions will prove to be effective only for a subset of the target population or only under certain conditions of implementation.

TABLE 3

FOUR LEVELS OF EXPERIMENTAL ASSESSMENT DETERMINED BY VARIATION IN THREE FACTORS

Level of experimental assessment	Program implementation	Availability	Acceptance
Efficacy	Standardized	Optimized	Optimized
Treatment effectiveness	Efficacious	Optimized	Variable
Implementation effectiveness	Efficacious	Variable	Variable
Program evaluation	Unproven efficacy or nonstandardized	Variable	Variable

Flay, B. R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine, 15, 451-474.*

- 4 levels of testing, together with experience in one area of health promotion (smoking prevention), suggest 8 phases of research for the development of health promotion programs:
  - I. Basic research;
  - II. Hypothesis development;
  - III. Pilot applied research;
  - IV. Prototype evaluation studies;
  - V. Efficacy trials;
  - VI. Treatment effectiveness trials;
  - VII. Implementation effectiveness trials;
  - VIII. Demonstration evaluations.

Holder, H. D., Flay, B., Howard, J., Boyd, G., Voas, R., & Grossman, M. (1999). Phases of alcohol problem prevention research. *Alcoholism: Clinical & Experimental Research, 23, 183-194.*

- I. Foundational research: Basic studies to define and determine the prevalence of specific alcohol-involved problems, establish the causal factors that yield specific problems or increase the risk of a problem, and provide foundations for the development of effective preventive interventions;
- II. Developmental studies: Preliminary studies to develop and test new interventions or to assess the effectiveness of an existing intervention;
- III. Efficacy studies: Rigorous studies (of maximised internal validity) of the intervention under optimal conditions with maximal implementation (availability or enforcement) and acceptance (participation or compliance);
- IV. Effectiveness studies: Studies of real-world effectiveness of preventive interventions with purposeful or natural variations;
- V. Diffusion studies: Studies of the effects of different levels or types of implementation or acceptance on effectiveness.



Flay, B., Biglan, A., Boruch, R., Castros, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175. (1)

- An efficacious intervention will have been tested in at least 2 rigorous trials that
  - (i) involved defined samples from defined populations;
  - (ii) used psychometrically sound measures and data collection procedures;
  - (iii) analyzed their data with rigorous statistical approaches;
  - (iv) showed consistent positive effects (without serious iatrogenic effects);
  - (v) reported at least one significant long-term follow-up.

Flay, B., Biglan, A., Boruch, R., Castros, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175. (2)

- An effective intervention will not only meet all standards for efficacious interventions but also will have:
  - (i) manuals, appropriate training and technical support available to allow third parties to adopt and implement the intervention;
  - (ii) been evaluated under real-world conditions in studies that include sound measurement at the level of implementation and engagement of the target audience (in both the intervention and control conditions);
  - (iii) indicated the practical importance of the intervention outcome effects; and
  - (iv) clearly demonstrated to whom the intervention findings can be generalized.

Flay, B., Biglan, A., Boruch, R., Castros, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151-175. (3)

- An intervention recognized as ready for broad dissemination will not only meet all standards for efficacious and effective interventions but will also provide:
  - (i) evidence of the ability to ‘go to scale’;
  - (ii) clear cost information;
  - (iii) monitoring and evaluation tools so that adopting agencies can monitor or evaluate how well the intervention works in their settings.

Kaner, E., Beyer, F., Dickinson, H., Pienaar, E., Campbell, F., Schlesinger, C., et al. (2007). Effectiveness of brief alcohol interventions in primary care populations . *Cochrane Database of Systematic Reviews Issue 2. Art. No.: CD004148. DOI:10.1002/14651858.CD004148.pub3.*

- **Subgroup analysis undertaken to assess the impact of brief interventions in efficacy (ideal world) and effectiveness (real world) trials using a coding scale developed from the work of Shadish and colleagues.**
- **Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological Bulletin*, 126(4), 512-529.**
- **These authors used 10 codes for clinical representativeness (qv) and applied these codes to 60 trials of psychological therapy.**
- **Conclusion: “... psychological therapies are robustly effective across conditions that range from research-oriented to clinically representative.”**
- **NB. Clinical representativeness dimension not the same as effectiveness vs efficacy.**

# ITEMS IN EFFICACY-EFFECTIVENESS SCALE (1)

From Kaner et al., Cochrane Review, 2007

## ***Patients and problems***

2 = clinically representative subjects initially present with a typically wide range of problems via self-referral or invitation for a health check

0 = research representative subjects may be paid patients, researcher-solicited volunteers (e.g. via advertisement) or referrals from specialist services.

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (2)

From Kaner et al., Cochrane Review, 2007

### ***Practice context***

2 = clinically representative is a community-based setting in which a range of clinical services are usually provided to patients;

0 = research representative is a setting in which the research function clearly dominates any clinical one (e.g. clinic at a university or hospital) .

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (3)

From Kaner et al., Cochrane Review, 2007

### ***Practitioners and therapists***

2 = clinically representative practitioners are practising doctors, nurses and qualified therapists who earn their main living by providing health services in primary care;

0 = research representative practitioners are non-clinicians, or clinicians in training, who are contracted to deliver interventions for the purposes of the study.

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (4)

From Kaner et al., Cochrane Review, 2007

### ***Intervention content***

2 = clinically representative intervention fits with current practice in terms of timing, content or style, e.g. 5-15 minutes for a GP; 20-30 minutes for a nurse or initial screening accompanied by a return visit for brief intervention;

0 = research representative treatment would not normally occur in routine practice, e.g. unusually long consultations.



## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (5)

From Kaner et al., Cochrane Review, 2007

### ***Therapeutic flexibility***

1 = clinically representativeness allows professional judgement in how an intervention is delivered, e.g. freedom to focus on particular issues according to patient need;

0 = research representativeness would be strict adherence to a prescribed protocol or script that does not allow for variability in practice.

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (6)

From Kaner et al., Cochrane Review, 2007

### ***Pre-therapy training***

1 = clinically representative training in intervention procedures occurs according to typical CPD/CME procedures, e.g. outreach visits, seminars, one-off training days;

0 = research representative training is unusually intensive or requiring of atypical levels of interest or motivation, e.g. prolonged or intensive courses, formal qualification.

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (7)

From Kaner et al., Cochrane Review, 2007

### ***Intervention support***

1 = clinically representative support occurs within standard practice resources, e.g. colleague assistance with screening, IT flagging;

0 = research representative support would not typically be available, e.g. researcher help to flag notes, extra staff for period of the trial.

## ITEMS IN EFFICACY-EFFECTIVENESS SCALE (8)

From Kaner et al., Cochrane Review, 2007

### ***Intervention monitoring***

1 = clinically representative monitoring of intervention delivery does not interfere with practitioners' behaviour or their relationship with patients;

0 = research representative monitoring would be direct observation of therapist behaviour or ongoing/immediate feedback to practitioners after each session.

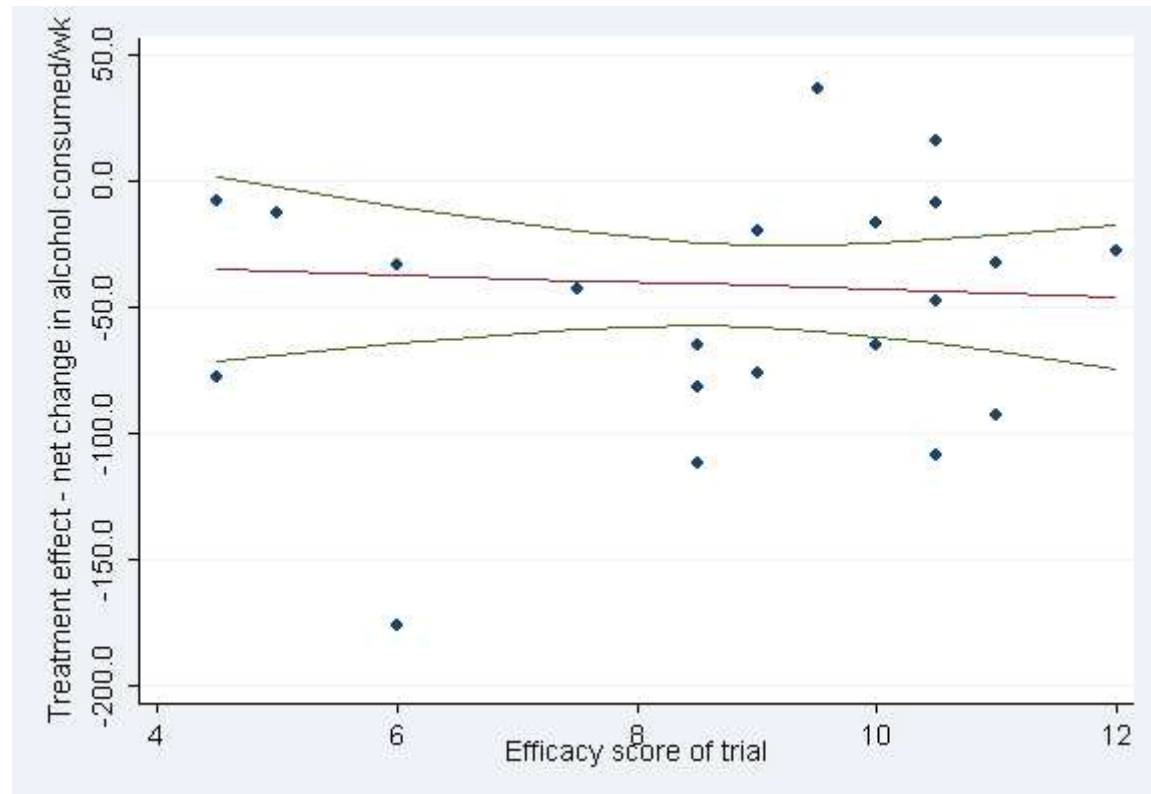
## DERIVATION OF EFFICACY-EFFECTIVENESS SCALE

- 4 items with greater apparent relevance to effectiveness scored 2/0; 4 items with less apparent relevance scored 1/0. Range of total score = 0 - 12.
- If an item appeared to be partially clinically representative on any item, then a midpoint score was given (either 1 or 0.5 as applicable). Similarly if authors did not report data relating to a particular item.
- Each trial was independently classified by two authors. If there was disagreement concerning classification, this was resolved through discussion in order to gain consensus. (NB, no measure of agreement reported in Cochrane review.)
- No pilot work to establish reliability.

## EFFICACY-EFFECTIVENESS ANALYSIS

- Efficacy/effectiveness scores ranged from 4.5 (Fleming 2004; Romelsjo 1989) to 12 (Lock 2006); the median was 9 and the inter-quartile range 8-10.5
- Scores were reported visually in the form of a graph of effect sizes along an axis of clinical to research representativeness. However, for the purpose of subgroup analysis a binary variable was created with a cut-off point at the median.

**Estimated treatment effect versus effectiveness/efficacy score.  
The lines show the predicted metaregression line and its 95%CI.**



## EFFICACY-EFFECTIVENESS ANALYSIS: FINDINGS

- There was no significant difference between trials classified as effectiveness and efficacy trials in the effect of brief intervention on the quantity of alcohol consumed, and meta-regression showed no significant relationship between the estimated treatment effect and the efficacy score of the trial.
- This lack of difference may indicate insensitivity in our classification tool.
- In the field of brief alcohol intervention, there has been a growing view that most of the trials to date have been tightly controlled efficacy studies and not particularly representative of routine clinical practice (Babor et al. 2006).  
.....within the context of trial-based evaluation, we feel that the current body of brief alcohol intervention research is applicable to clinical practice. Previous trials have fallen on a continuum from efficacy to effectiveness trials, and the lack of significant difference in outcomes on this dimension suggest that this body of work can inform routine practice.



# PROBLEMS WITH THIS ANALYSIS

- No measure of agreement between raters reported.
- No psychometrics carried out, eg, deletion of items that lower Cronbach's  $\alpha$  or principal components analysis
- No comparisons of effect sizes for individual scale items: 'It is possible that the treatment effect may be related to some of the individual factors which were combined in the efficacy score. However, we did not investigate this as it would have been a *post hoc* analysis, not specified in the protocol.'

Gartlehner *et al.* (2006). A simple and valid tool distinguished efficacy from effectiveness studies. *Journal of Clinical Epidemiology*, 59, 1040-8

- Developed and tested a simple instrument based on 7 criteria of study design to distinguish effectiveness (pragmatic) from efficacy (explanatory) trials.
- Asked directors of 12 EBP Centres to select 6 studies each, 4 considered to be effectiveness trials and 2 efficacy trials.
- Tested proposed criteria against 'gold standard' of selected studies to identify effectiveness studies reliably with minimal false positives (ie, high specificity).
- A cut-off of 6 criteria produced the most desirable balance between sensitivity (0.72) and specificity (0.83).

# Gartlehner et al. criteria to distinguish effectiveness from efficacy trials

- 1. Populations in primary care
- 2. Less stringent eligibility criteria
- 3. Health outcomes
- 4. Long study duration; clinically relevant treatment modalities
- 5. Assessment of adverse events
- 6. Adequate sample size to assess a minimally important difference from a patient perspective
- 7. Intention to Treat analysis

# CONCLUSIONS (1)

- It is a mistake to go straight to effectiveness trials for new forms of SBI intended for different populations in different settings. Such research should begin with foundational research and development studies followed by efficacy trials before large-scale effectiveness trials are mounted.
- To properly interpret the findings of effectiveness studies, especially null findings, it is necessary to ensure that interventions are delivered as intended and as found efficacious.
- Clear criteria are available in the literature to guide progress in movement from efficacy research, through effectiveness research, to dissemination in practice.

## CONCLUSIONS (2)

- In future meta-analyses of alcohol BI trials, more attention should be paid to the development and application of a scale to measure efficacy-effectiveness (or clinical representativeness), including:
  - Theory-based scale construction;
  - Inter-rater reliability testing and reporting;
  - Psychometric refinement;
  - Publication as a topic of interest in its own right.

## CONCLUSIONS (3)

- In relation to the 3 ‘disappointing’ findings:
  - The van Beurden et al trial strongly reinforces what we already suspect – that it is extremely difficult to get health professionals to deliver SBI;
  - The null findings of the SIPS trial cannot be attributed to a failure to translate effects from efficacy trials to real world practice because it seems likely that the majority of trials included in meta-analyses tend to be effectiveness trials;
  - These null findings and those of the PRE-EMPT trial may be due to the lack of fidelity in the implementation of SBI in large, cluster randomised trials.