

Evidence for evidence- based practice: Research and reality in juvenile justice

Mark W. Lipsey
Vanderbilt University

Jerry Lee Lecture
Campbell Collaboration Colloquium
Vancouver, May 2008



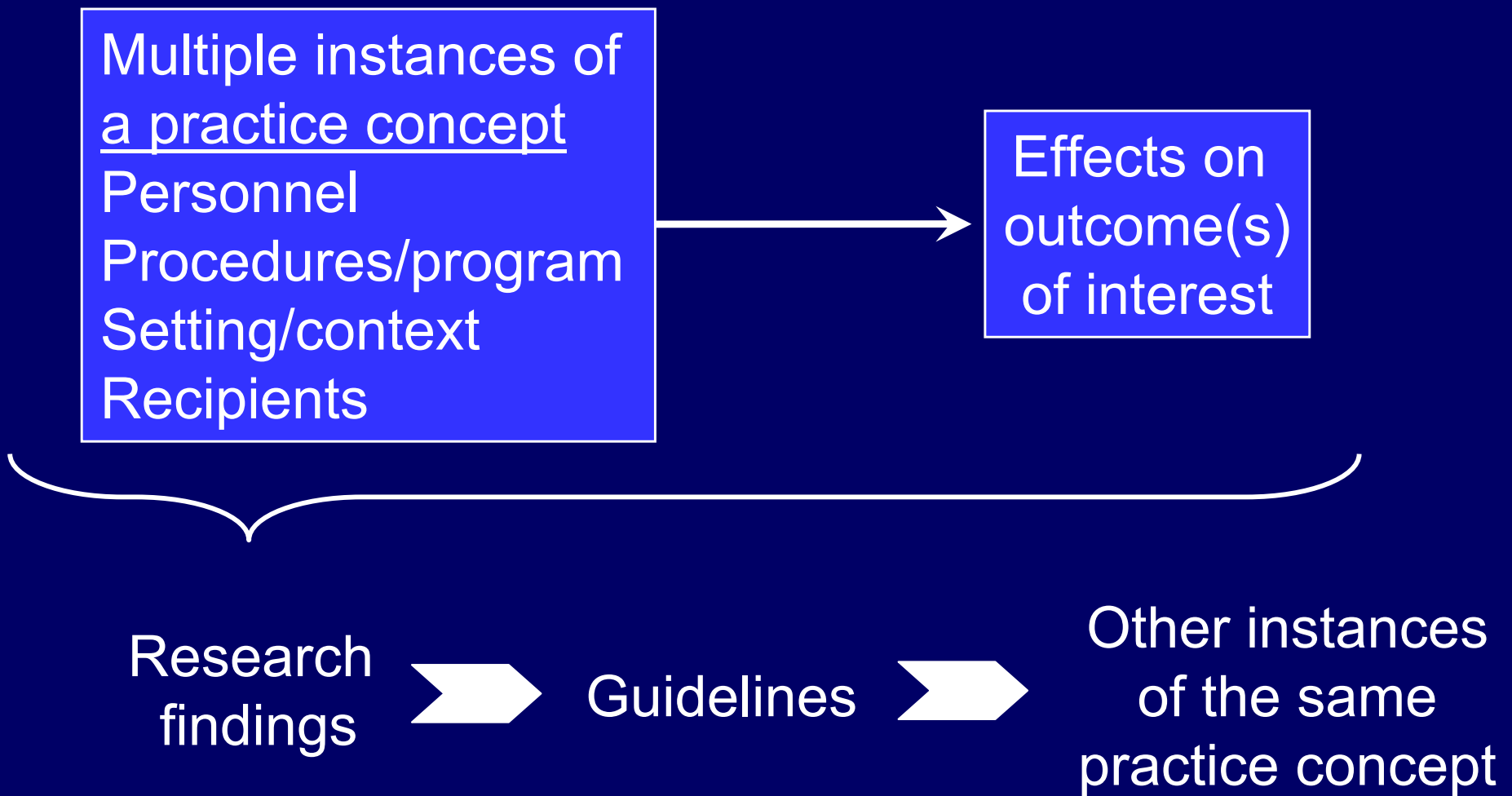
Our point of departure

“The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services.”

(From the C2 website)

This statement affirms our faith that using research to guide practice will make practice more effective ... **but do we have any evidence that this is true?**

Applying research to practice



Applying research to

Multiple instances of
a practice concept
Personnel
Procedures/program
Setting/context
Recipients

Generalize along
dimensions of
similarity– the ‘same’
practice concept

Focus on
Outcome(s)
of interest

Research
findings



Guidelines



Other instances
of the same
practice concept

Applying research to practice

Multiple instances of
a practice concept
Personnel
Procedures/program
Context

If narrow &
specific ...

Limited domain of
generalization

Few studies;
expect little
variation

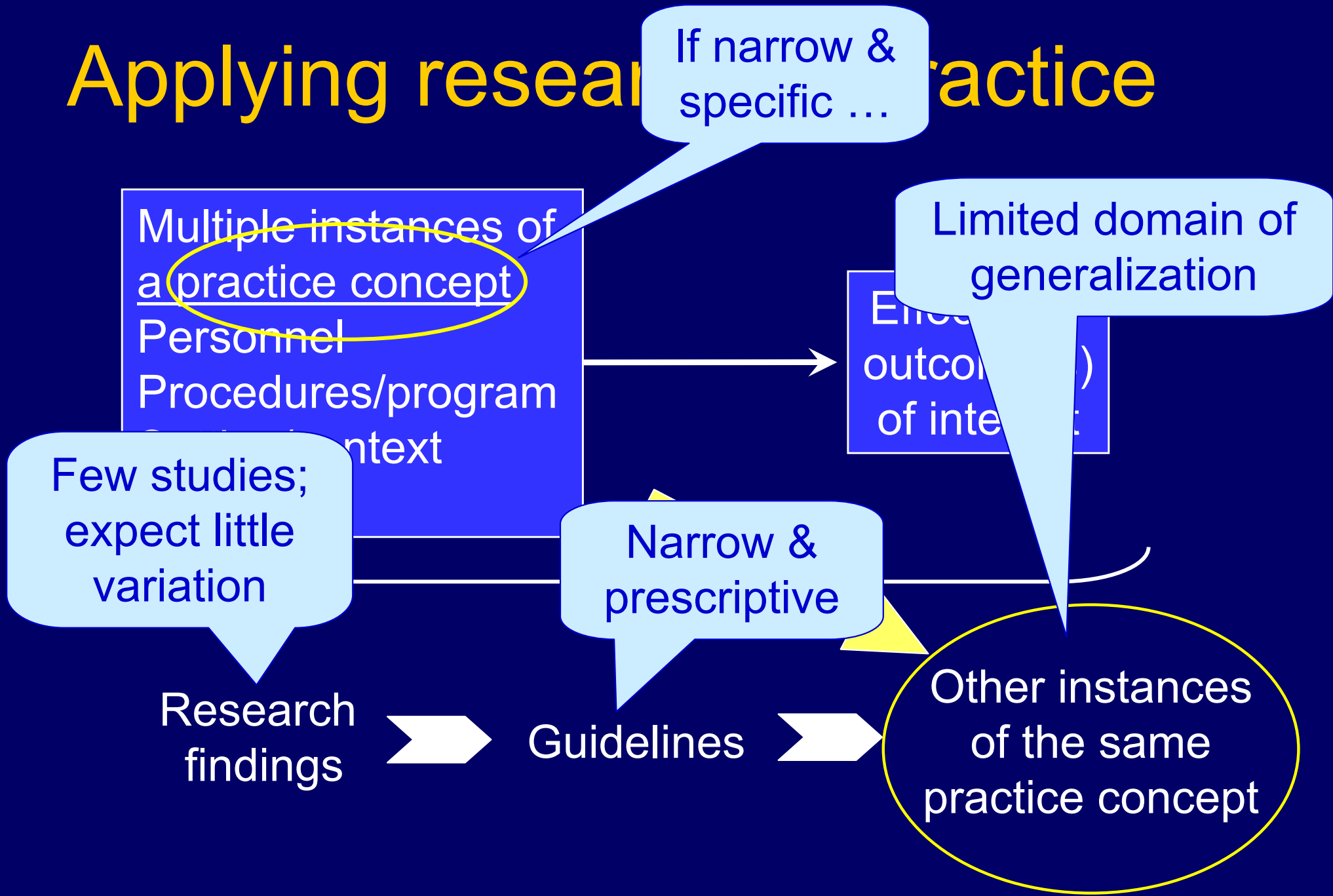
Narrow &
prescriptive

Research
findings

Guidelines

Other instances
of the same
practice concept

Encouraged
outcomes
(
of interest



Applying research to practice

Multiple instances of
a practice concept
Personnel
Procedures/program
Context

Many studies;
expect much
variation

Research
findings



Guidelines



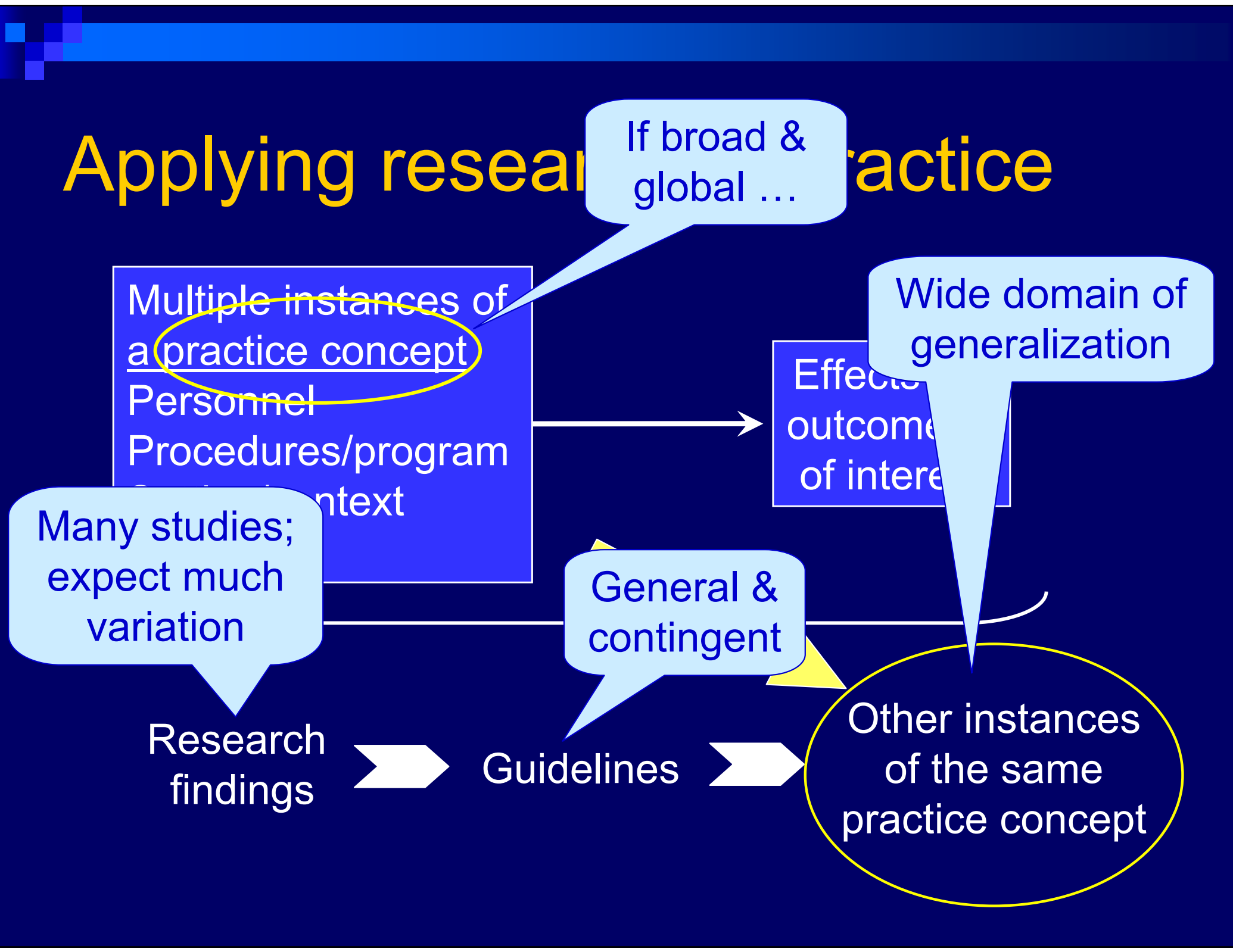
Other instances
of the same
practice concept

If broad &
global ...

Wide domain of
generalization

General &
contingent

Effects
outcome
of interest





Practice concept: Operating procedure of a specific program (SOP)

Example: Conflict resolution program in Franklin Middle School, Columbus Ohio

Research:

- Evaluation study of the effects of that particular program (rarely more than one study)

Generalizes to:

- That same program as it continues unchanged
- Practice using similar procedures and similar personnel with similar students in similar settings

Practice concept: Local SOP (continued)

Research synthesis:

- None; primary study and rarely any replications
- If replications, expect little variation in effects

Practice guidelines: [Follow the SOP]

- Affirmation (or not) of positive effects from the SOP that defines the practice
- Local operational variants or recipient characteristics associated with better outcomes

Practice concept: Intervention protocol

Examples: Olweus Bullying Prevention Program,
Functional Family Therapy, Bright Beginnings
Literacy Curriculum

Research:

- Evaluation studies of implementations of that protocol with the outcomes of interest

Generalizes to:

- Other implementations of that protocol
- But effects may differ with local variations in fidelity, personnel, recipients, and setting

Practice concept: Protocol (continued)

Research synthesis:

- Focus on ES mean, variation in fidelity
- Moderators for setting, sample characteristics
- Relatively few studies typically available

Practice guidelines: [Follow the protocol]

- Affirmation (or not) of positive mean effects from the protocol that defines the practice
- Contingent on fidelity of implementation
- Settings, types of recipients for which evidence of positive effects



Evidence-based practice currently defined around intervention protocols

Lists of model programs, e.g.:

- Blueprints for Violence Prevention
- National Registry of Evidence-based Programs and Practices (NREPP)
- Helping America's Youth (HAYS) community guide
- OJJDP Model Programs Guide

Practice concept: Generic intervention

Examples: Interpersonal skills training, family therapy, school-based conflict resolution

Research:

- Evaluation studies of programs of that type with the outcomes of interest

Generalizes to:

- Other implementations of that program type
- But effects will differ by SOP or protocol as well as local variations in implementation, personnel, recipients, and setting

Practice concept: Generic intervention (continued)

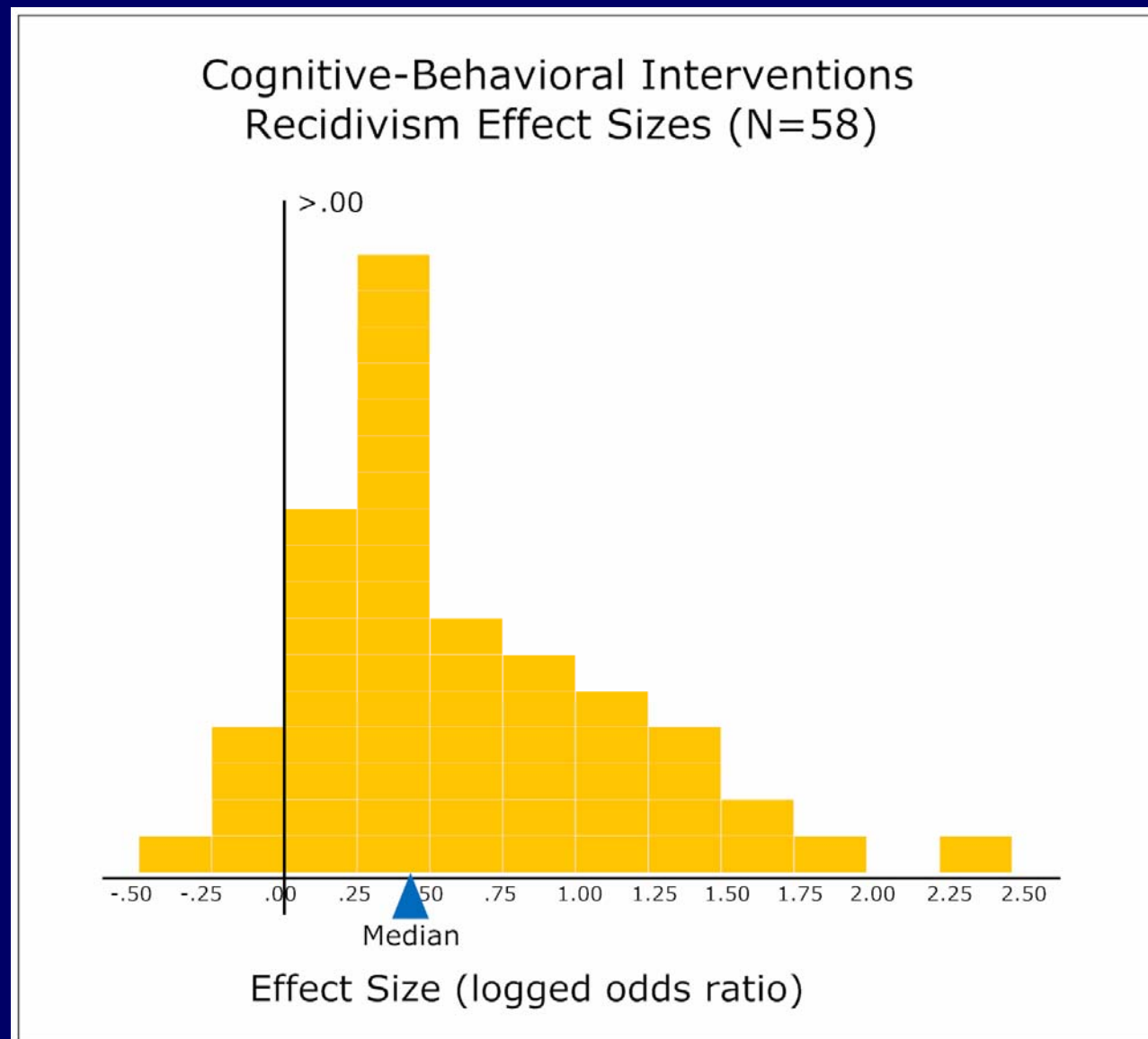
Research synthesis:

- Little focus on ES mean
- Moderators for intervention variants, setting, implementation quality, sample characteristics, etc.
- May be many studies available

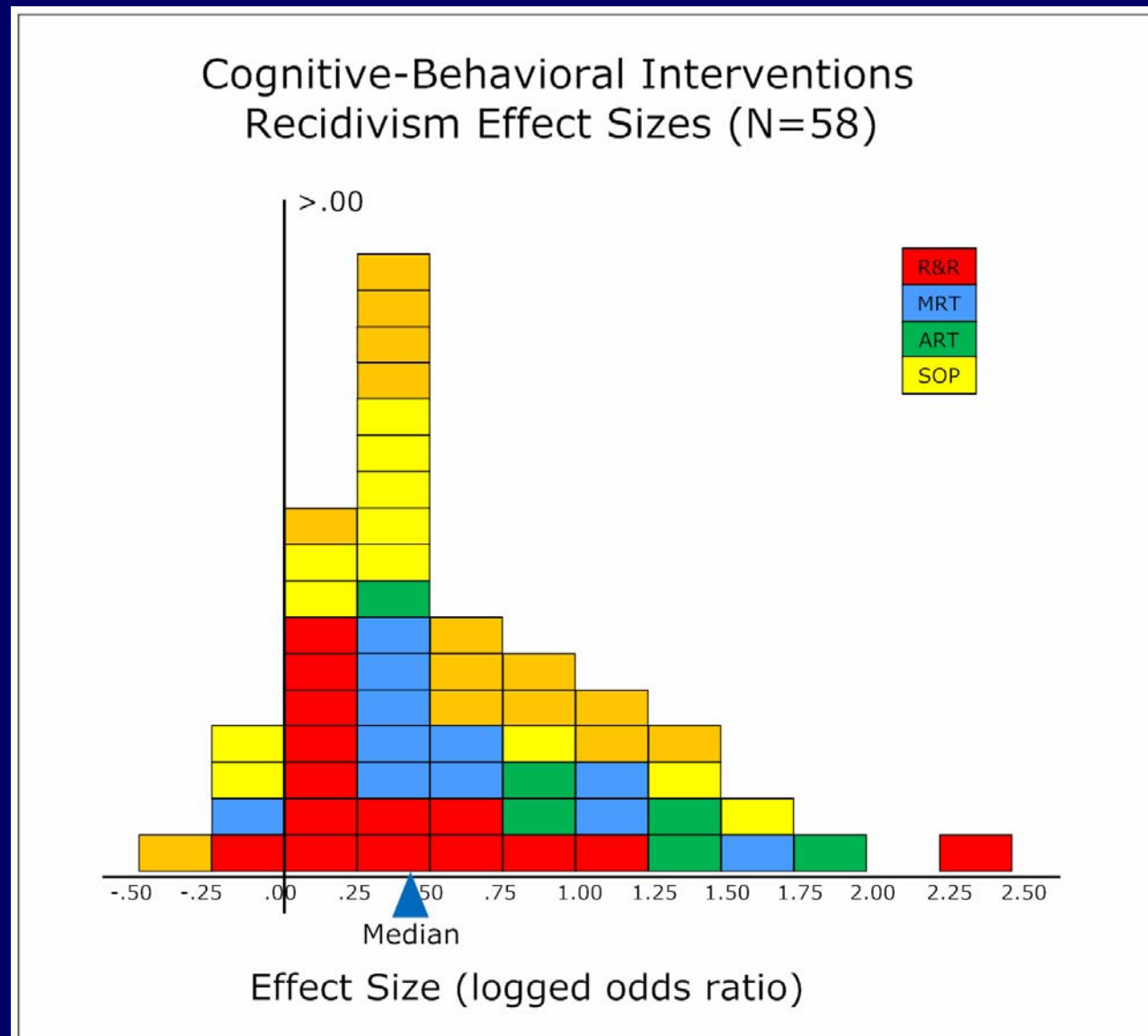
Practice guidelines: [Follow the principles]

- Affirmation (or not) of positive effects from some variants of the intervention
- Settings, types of recipients for which evidence of positive effects for those variants

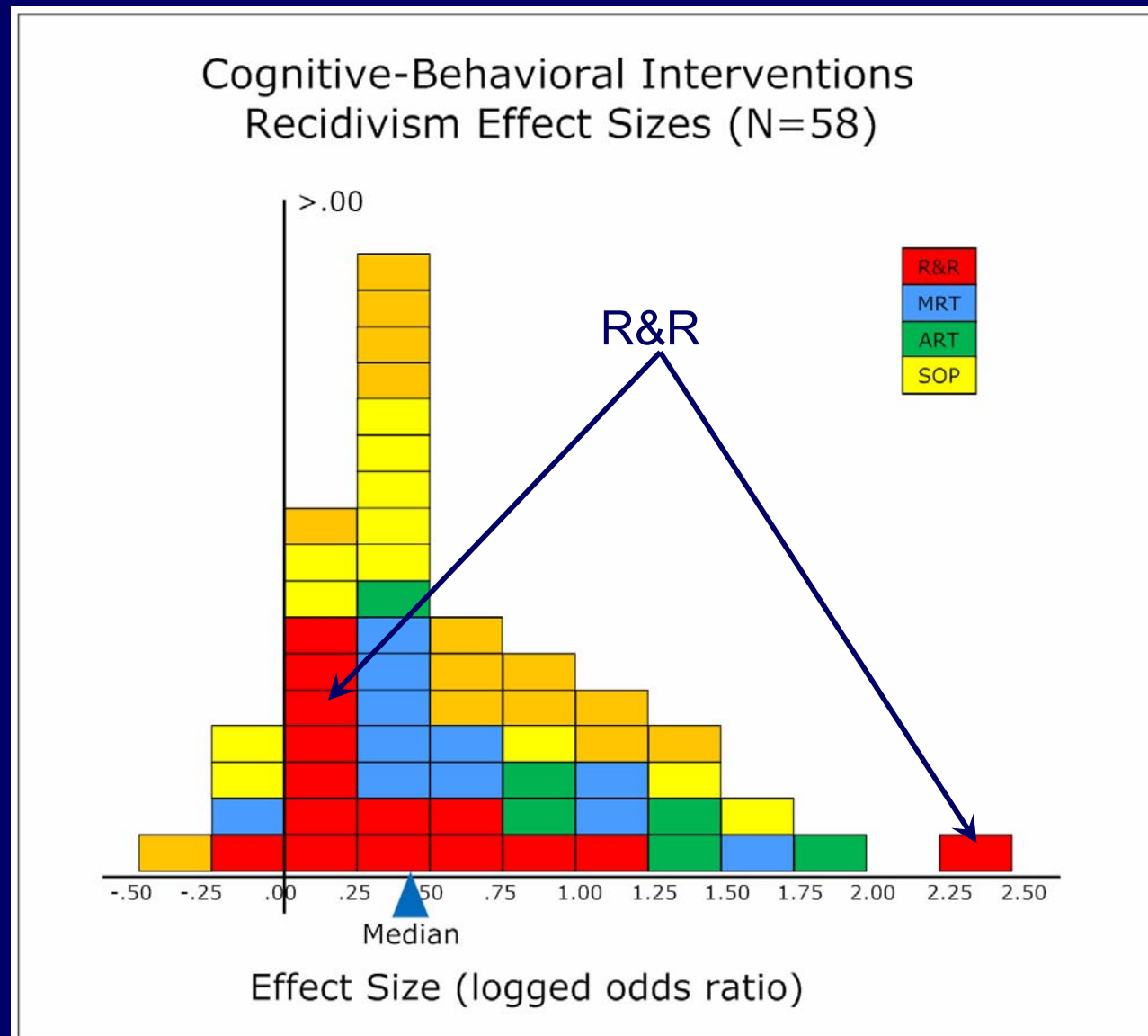
Effects for generic intervention



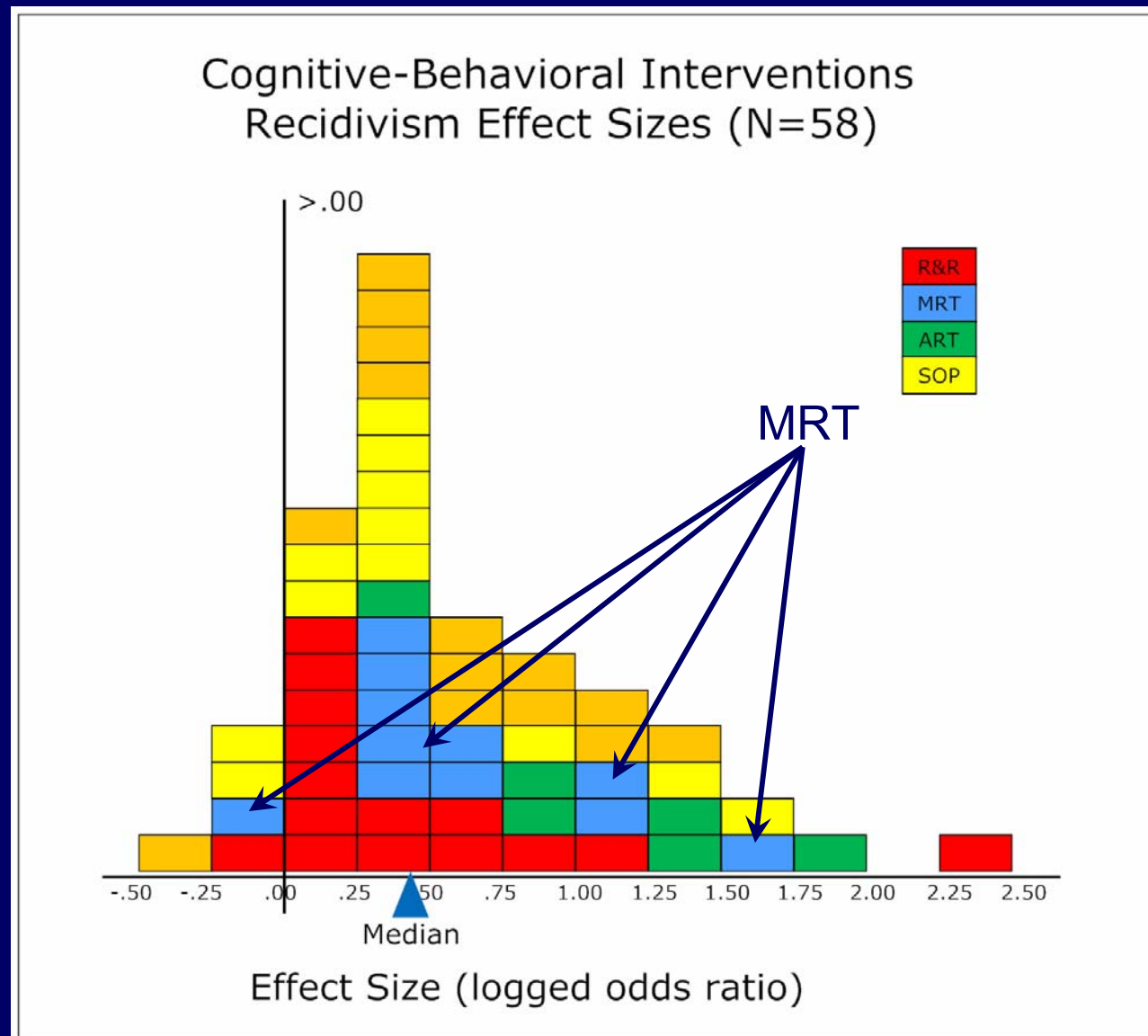
Mix of protocol and SOP programs



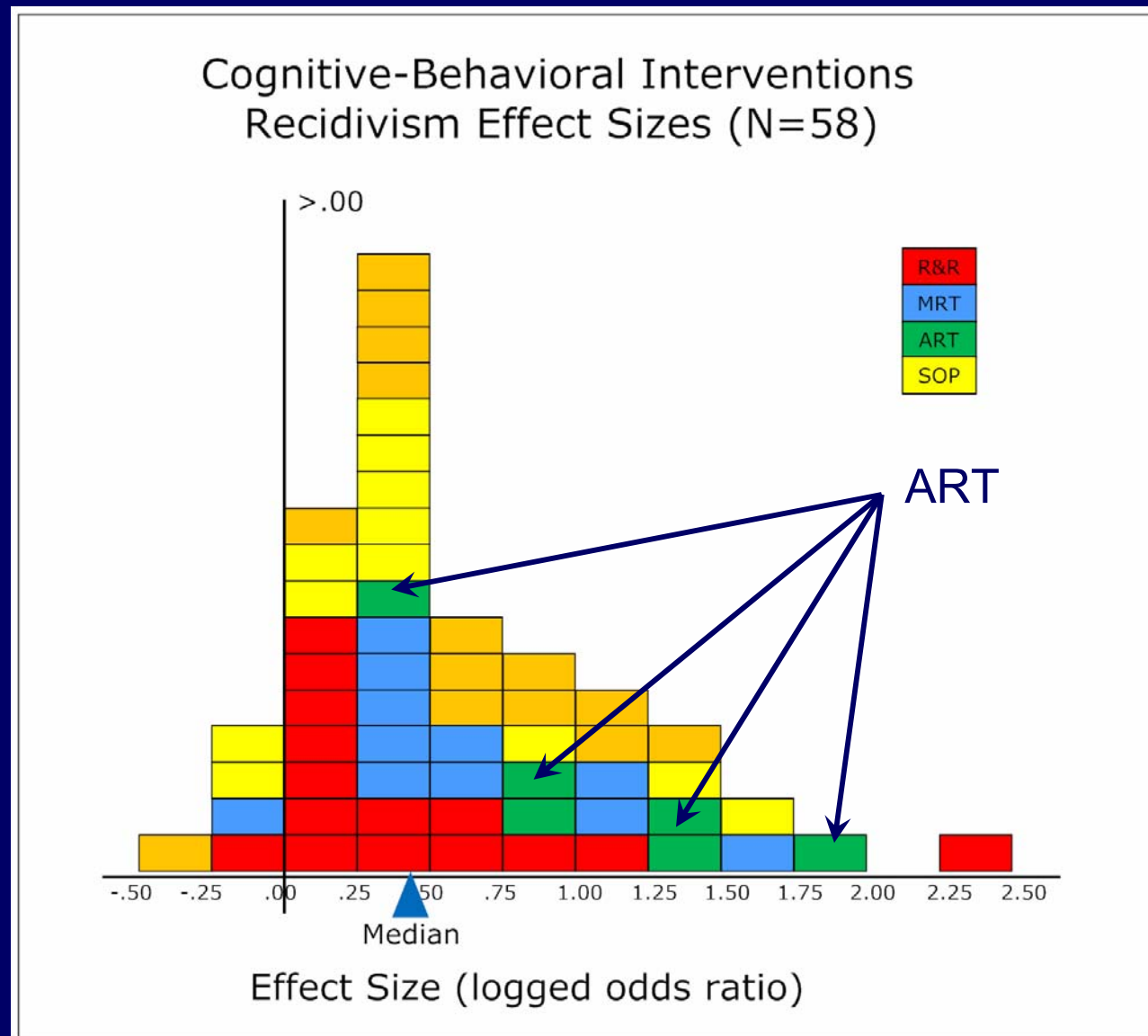
Mix of protocol and SOP programs



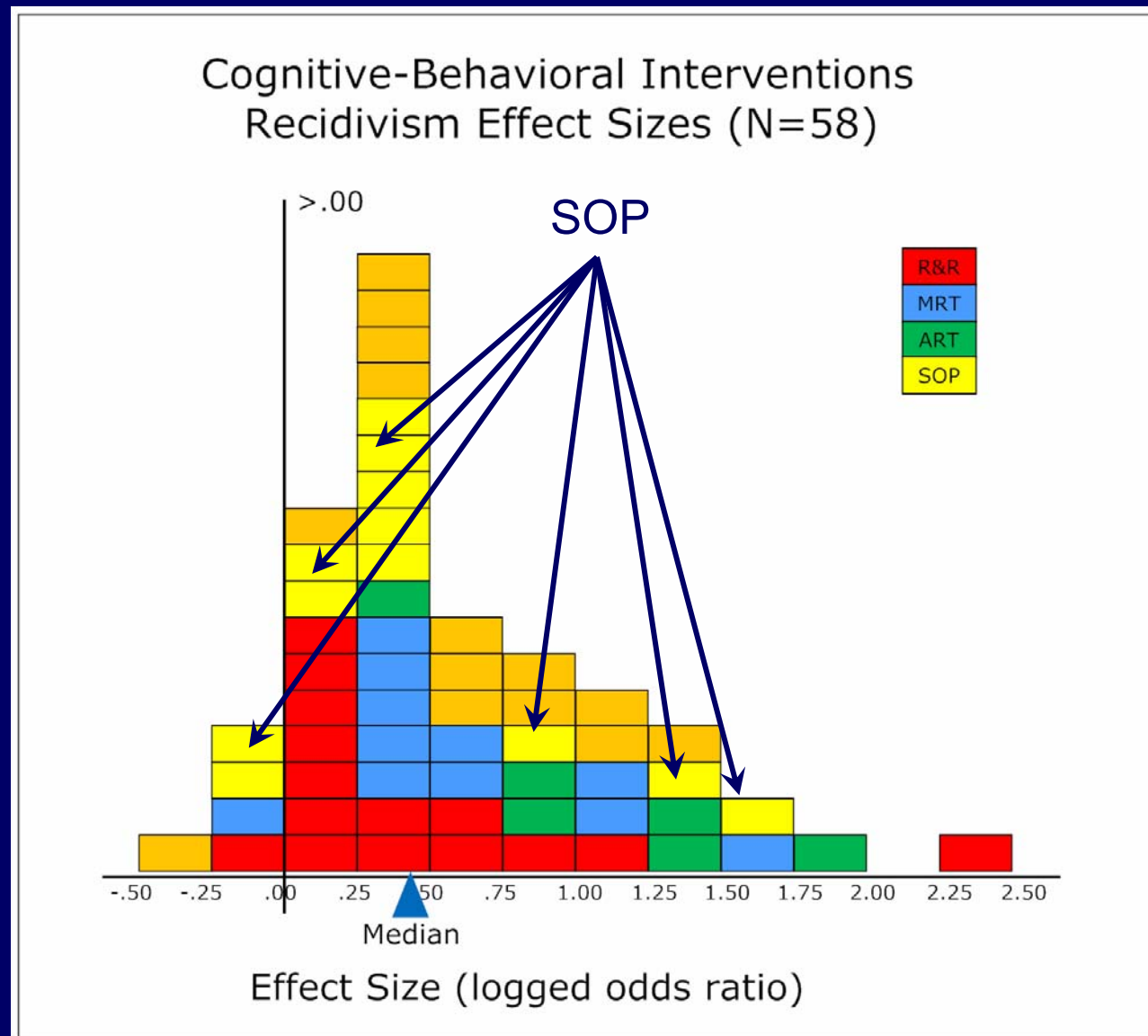
Mix of protocol and SOP programs



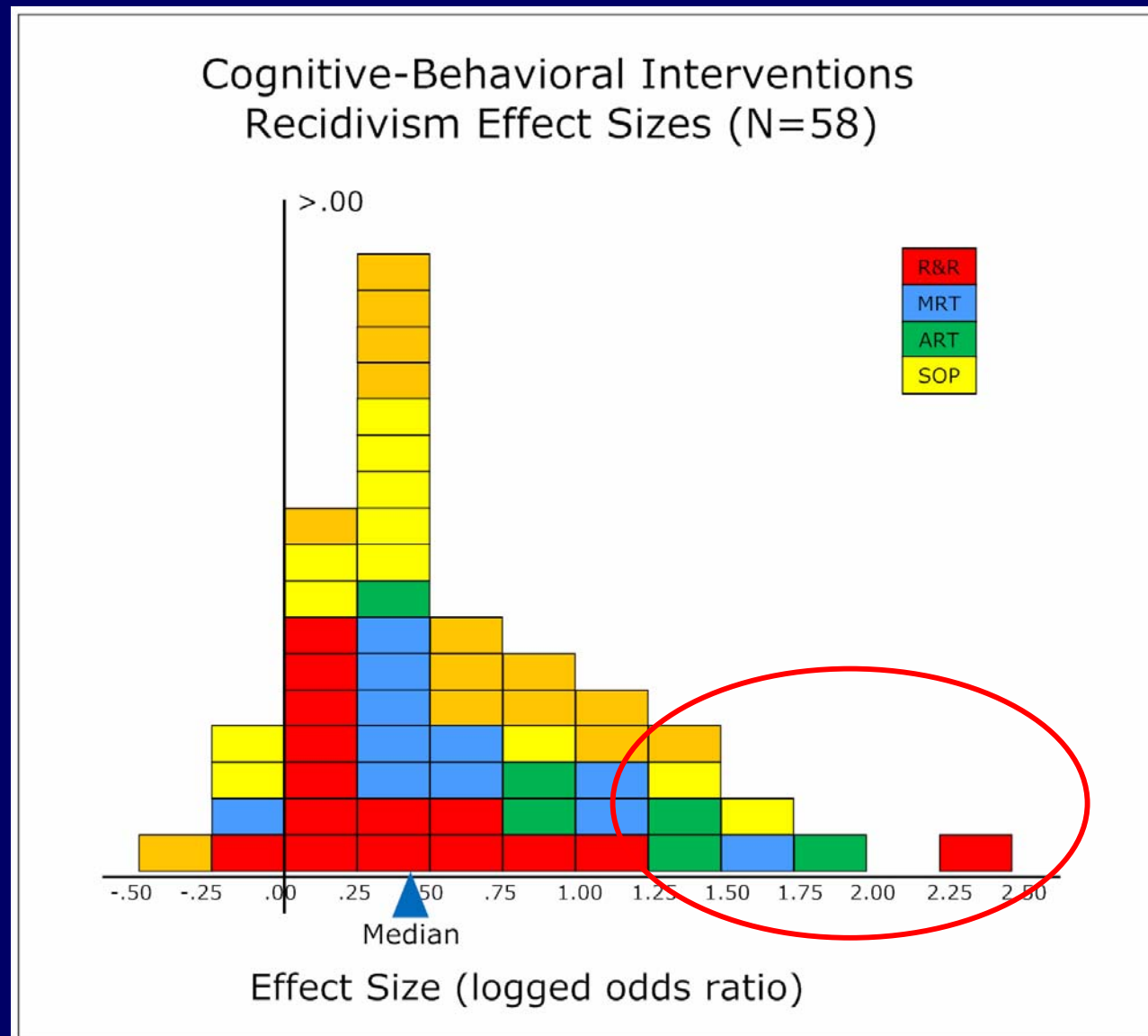
Mix of protocol and SOP programs

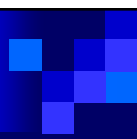


Mix of protocol and SOP programs



Mix of protocol and SOP programs





Further exploration of synthesis and practice guidelines for generic interventions

Interesting because:

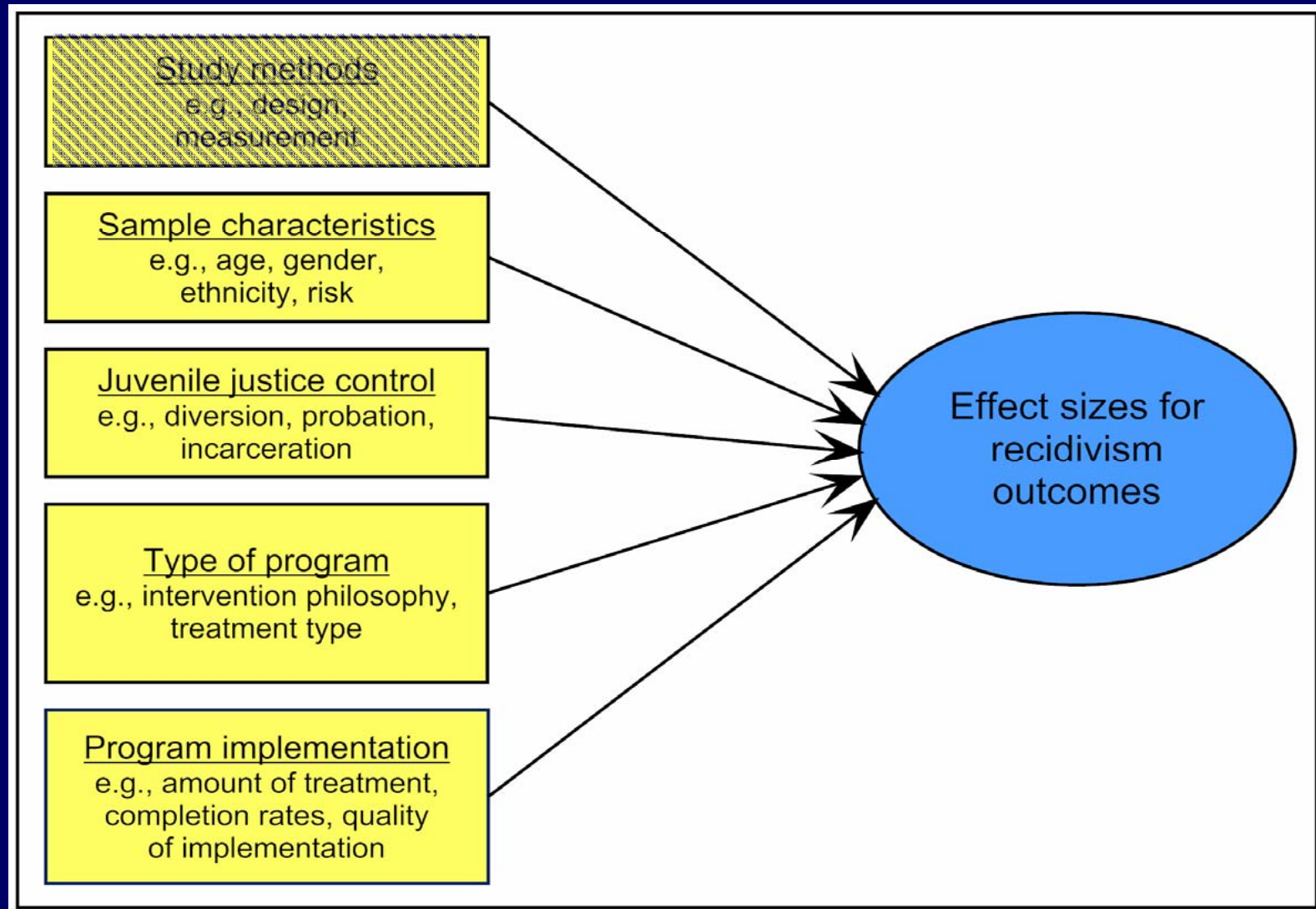
- Can assemble more research on instances of a generic category than for protocols or SOPs
- Results potentially generalize to a broader domain-- including many of the kinds of program used in actual practice
- Even broad evidence-based guidelines may steer programs toward better outcomes



Source data: Meta-Analysis of interventions for juvenile offenders

- 548 independent study samples from 361 primary research reports
- Juveniles aged 12-21 received an intervention intended to have positive effects on their subsequent delinquency
- At least one delinquency outcome; assignment was random or, if not, pretreatment differences were reported or matched
- Conducted in English speaking countries between 1958 and 2002

Effect sizes assumed to be a function of study and program characteristics





Key Findings



Method matters (a lot)

Effect size differences associated with:

- Outcome measurement (type of recidivism, timing)
- Design (initial equivalence; design type)
- Sample size
- Publication source

Used as statistical control variables

Some characteristics of the juveniles matter

	Std coefficient
Mean age	.05
Gender mix	-.07 [†]
Ethnicity	.03
Delinquency risk	.41 [*]
Aggressive history	-.12 [*]

† p < .10 * p < .05

JJ supervision doesn't matter much (with risk, etc. controlled)

Std coefficient

No supervision

Diversion

-.06

Probation/parole

.01

Incarceration

-.08

† p < .10 * p < .05



Program “philosophies” (Group 1)

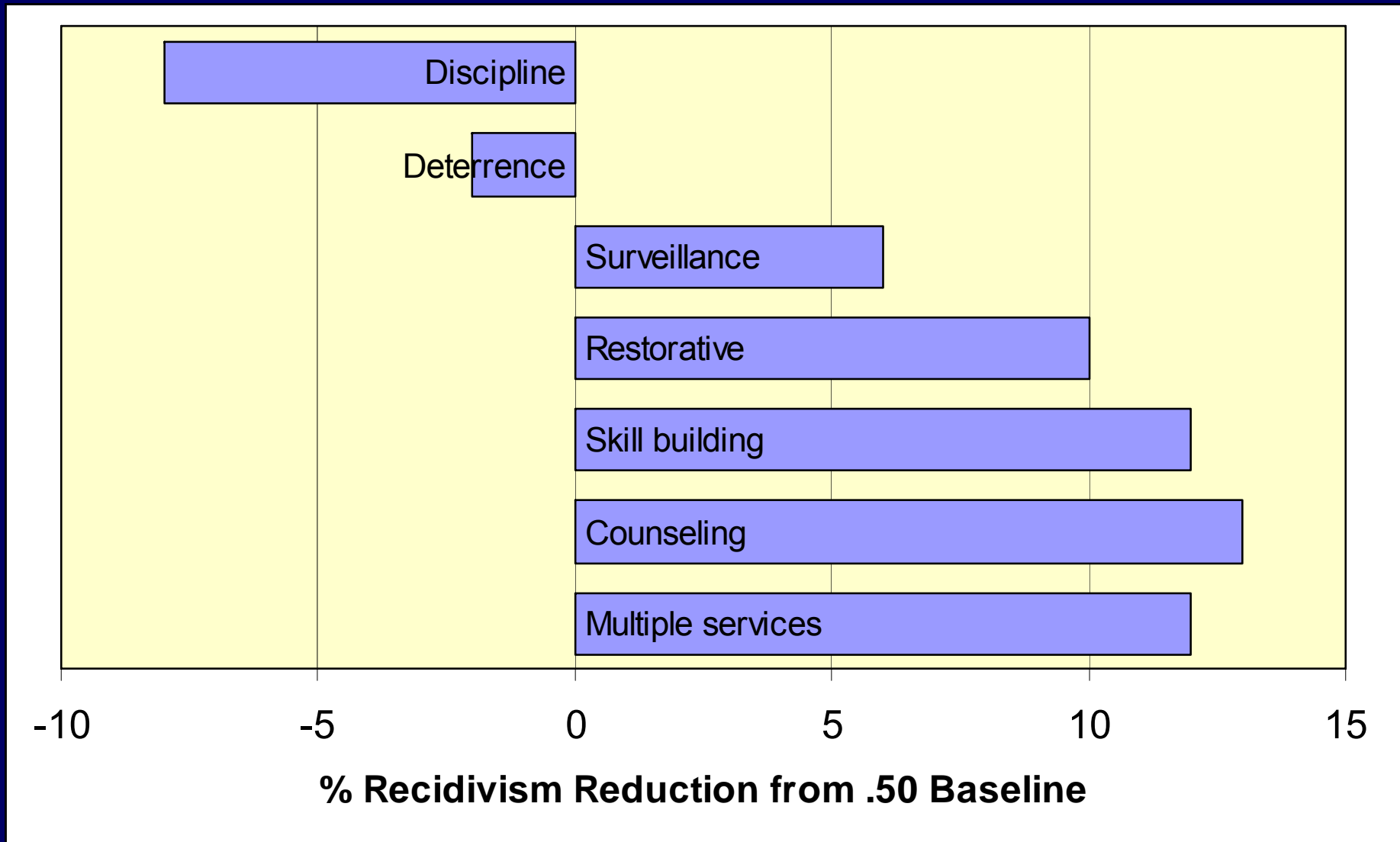
- Discipline: e.g., paramilitary regimens in boot camps
- Deterrence: e.g., prison visitation (Scared Straight)
- Surveillance: e.g., intensive probation or parole.



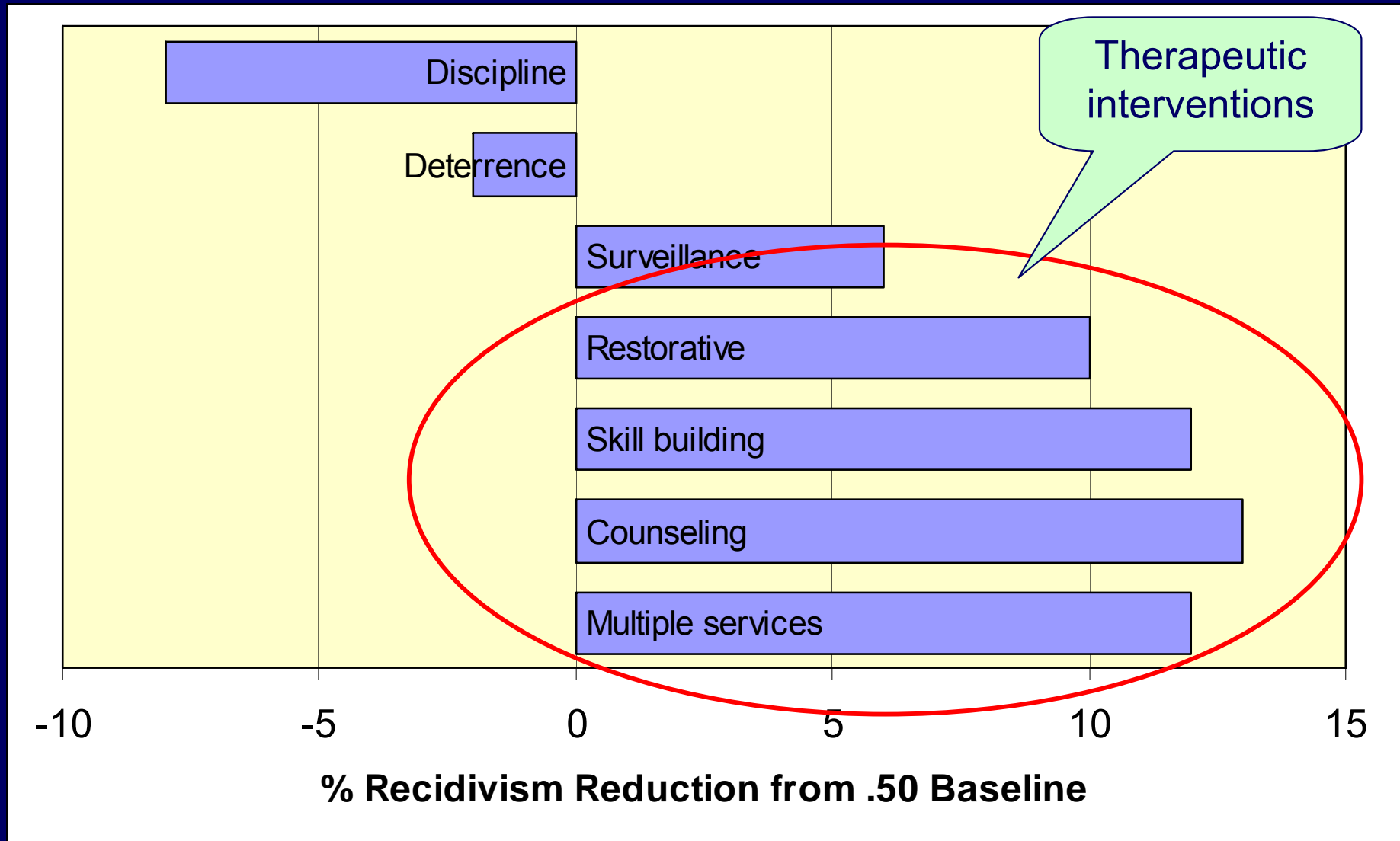
Program “philosophies” (Group 2)

- Restorative: e.g., restitution, mediation
- Skill-building: behavioral, CBT, social skills, challenge, academic, & vocational
- Counseling: individual, mentoring, family, family crisis, group, peer, mixed counseling, mixed with supplementary referral
- Multiple coordinated services: case management, service broker, multimodal regimen.

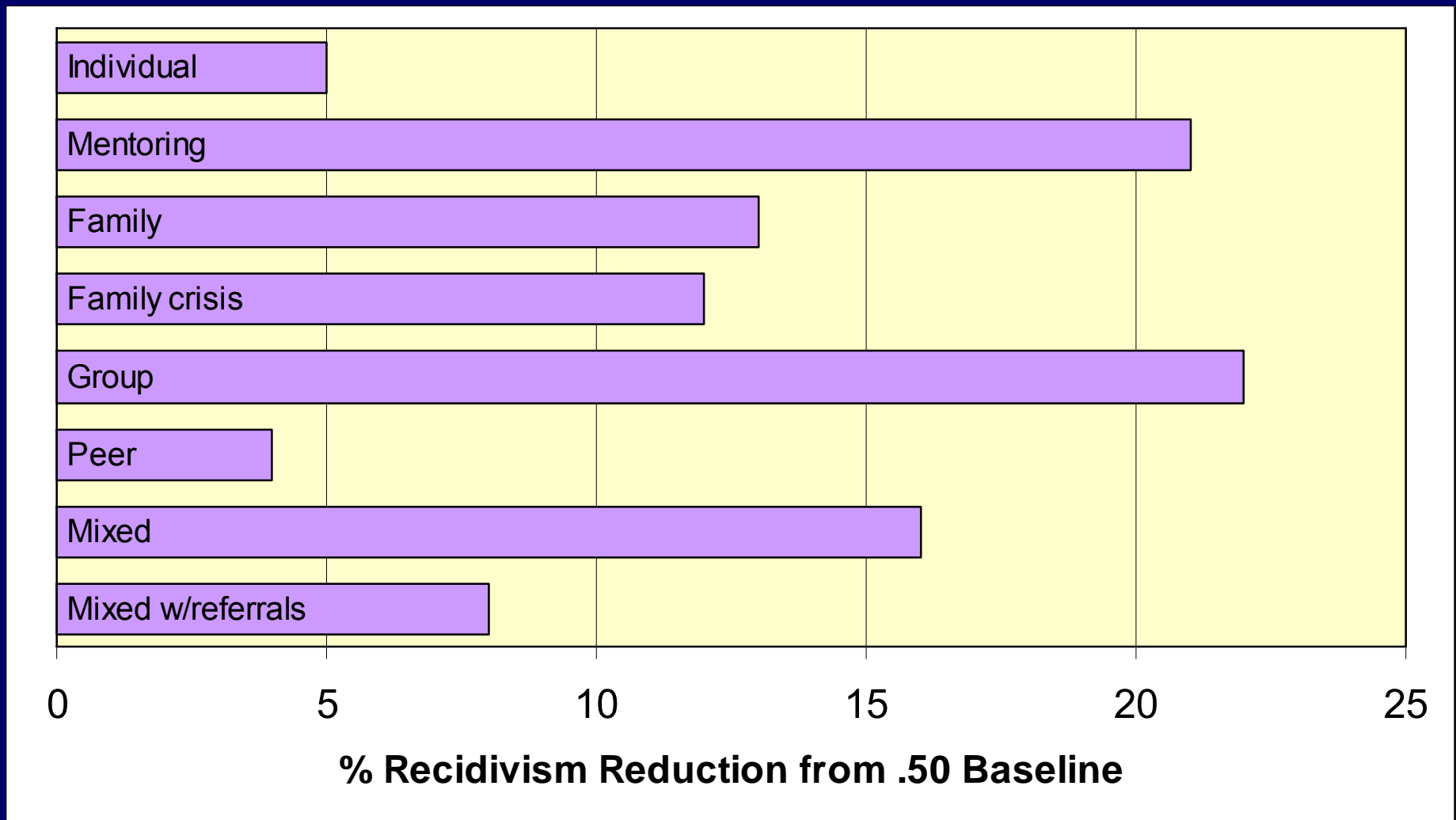
Treatment “philosophy” matters



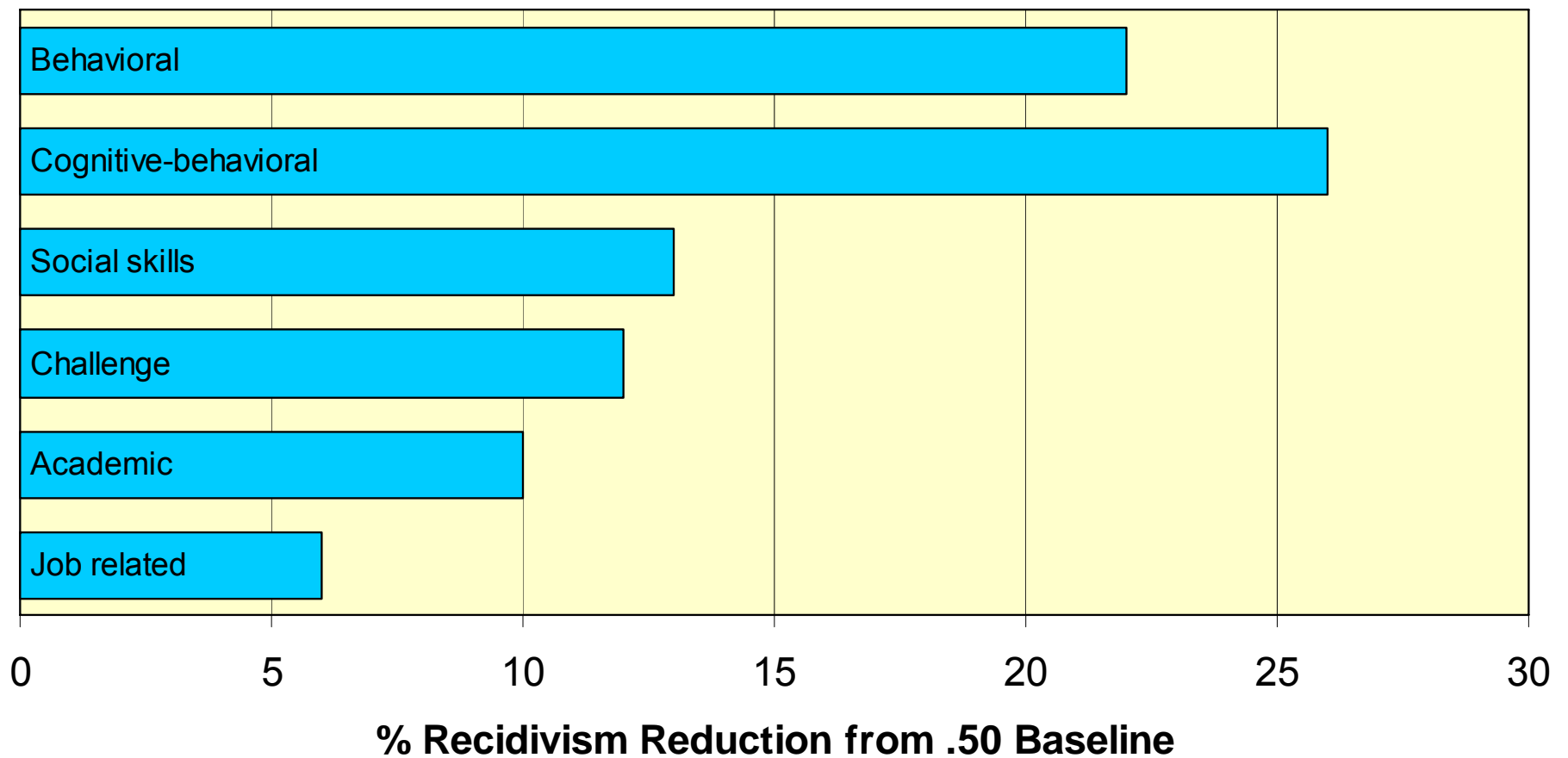
Treatment “philosophy” matters



Type of intervention: Counseling



Type of intervention: Skill-building



Service amount and quality

	Restorative β	Skill Building β	Counseling β	Multiple Services β
Duration of service	-.31	.09	-.02	.11
Total hours of service	.51*	-.03	-.06	-.03
Quality of implementation	.21	.25*	.13†	.18*

† p < .10 * p < .05

Practice principles identified so far

- Focus on high risk cases
- Rely on juvenile justice supervision and surveillance for sanctions and near-term control but not for rehabilitation
- Use interventions with a therapeutic orientation; avoid those with a discipline or deterrence philosophy
- Within a therapeutic category, use the more effective program types
- For a given program type, provide at least the average amount of service to each juvenile
- Deliver that service with high quality.



Operationalizing these practice guidelines

- A rating system for each program type within the therapeutic philosophies
- Pilot projects with the juvenile justice systems of Arizona and North Carolina
- Applied to individual programs based on MIS data about the services they actually provide

Program scoring scheme that assigns points proportionate to the contribution of each factor to recidivism reduction

Standardized Program Evaluation Protocol (SPEP) for Services to Probation Youth

	Possible Points	Received Points
Primary Service:	35	
High average effect service (35 points) Moderate average effect service (25 points) Low average effect service (15 points)		
Supplemental Service:	5	
Qualifying supplemental service used (5 points)		
Treatment Amount:	10	
Duration: % of youth that received target number of weeks of service or more: 0% (0 points) 60% (6 points) 20% (2 points) 80% (8 points) 40% (4 points) 100% (10 points)		
Contact Hours: % of youth that received target hours of service or more: 0% (0 points) 60% (9 points) 20% (3 points) 80% (12 points) 40% (6 points) 100% (15 points)	15	
Treatment Quality:	15	
Rated quality of services delivered: Low (5 points) Medium (10 points) High (15 points)		
Youth Risk Level:	20	
% of youth with the target risk score or higher: 25% (5 points) 75% (15 points) 50% (10 points) 99% (20 points)		
Provider's Total SPEP Score:	100	[INSERT SCORE]



SPEP scoring

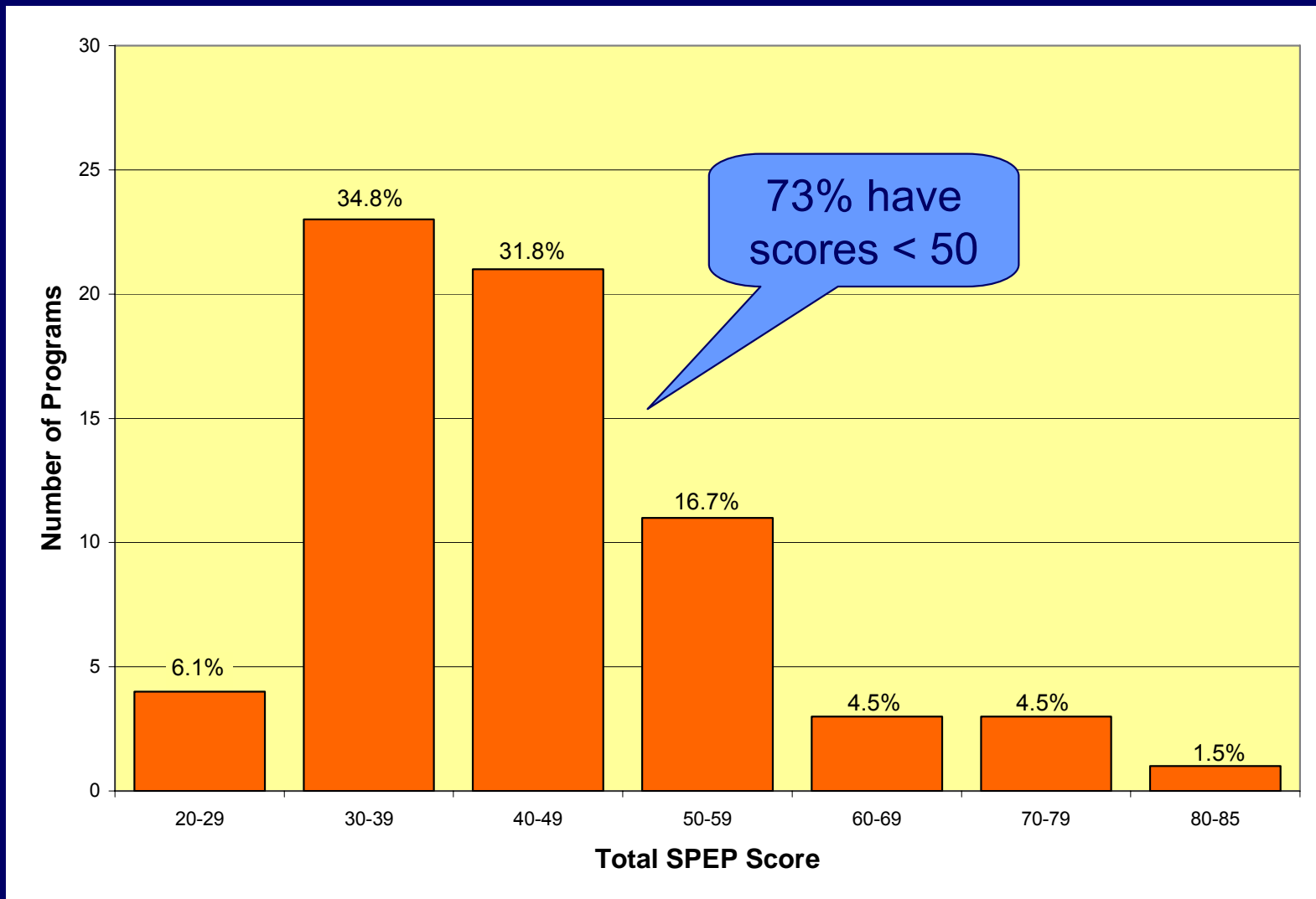
- Based on (a) review and classification of services provided, (b) data on services delivered, and (c) risk scores of clients
- Provides an initial assessment of how closely actual program practice matches “best practice” identified from research
- Gives guidance for program improvements expected to enhance effects on recidivism



Validity study: Does it work?

- Arizona Juvenile Justice Services Division
- Programs provided during 2005-06 to juvenile probationers in five pilot counties
- 1490 juveniles who received services from 66 SPEP rated programs
- 6-month recidivism data on all; 12-month recidivism for most

Distribution of SPEP scores across programs





Logistic regression to predict risk of recidivism based on prior history

Predicted using archival data on:

- number and nature of prior offenses
- risk rating by probation officers
- age, sex, race, county
- number of prior service events

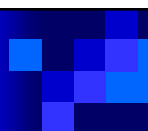
Predictive accuracy

- 6-month recidivism rate was .27; the 12-month rate was .44
- The prediction function correlated .24 with actual 6-month recidivism and .29 with actual 12-month recidivism
- It correctly predicted recidivism for 65% of the cases.

Correlations across service providers of SPEP scores with actual minus predicted recidivism

SPEP Scores	6-Month Recidivism (N=66)	12-Month Recidivism (N=63)
Type of service subtotal	.06	-.19
Amount of service subtotal	-.15	-.30*
Risk subtotal	-.24*	-.13
Total score	-.18	-.29*

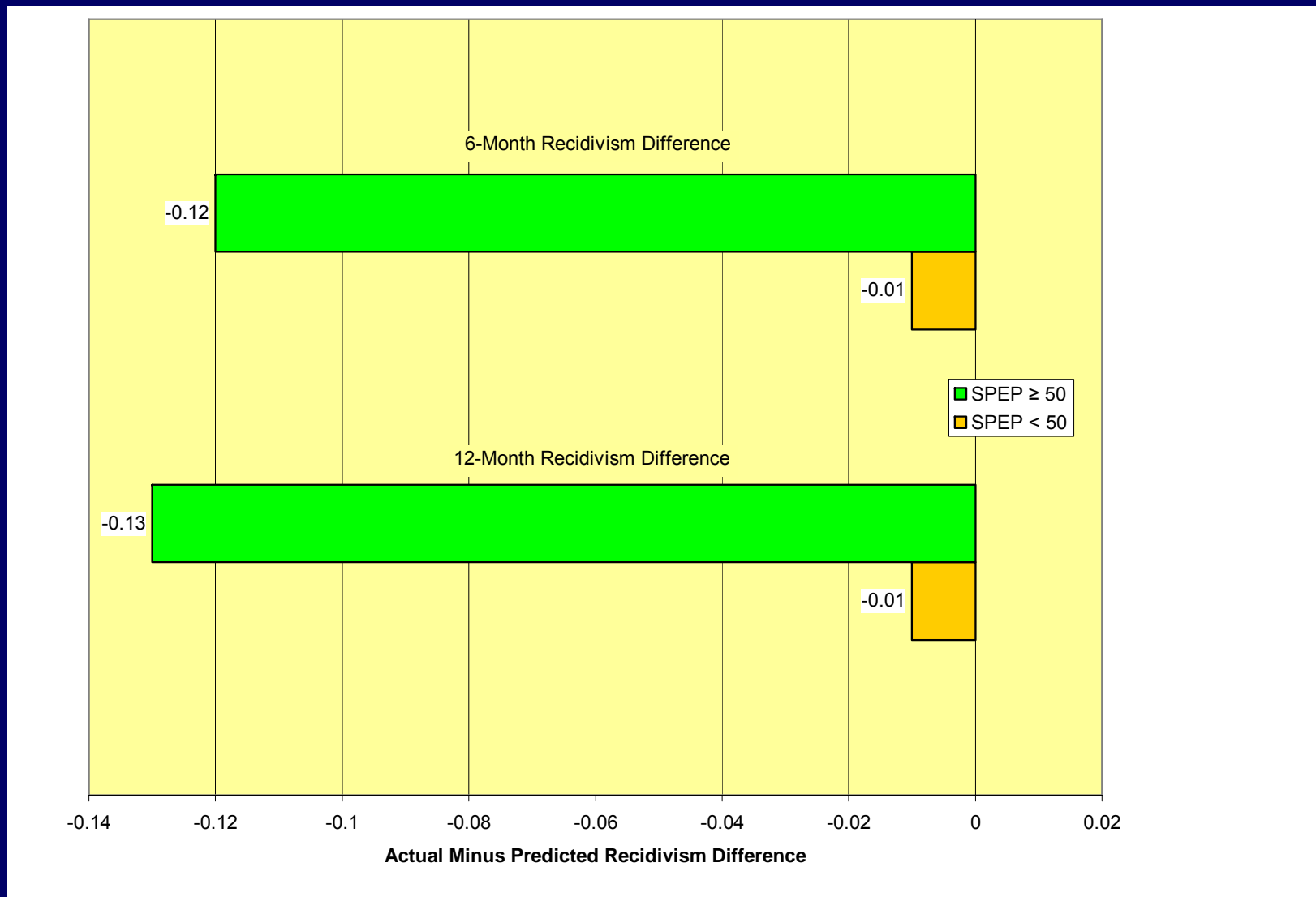
* p<.05



These correlations likely understate the validity of the SPEP ratings for identifying effective programs

- Quality assessments are not yet included in the SPEP ratings
- The range on the SPEP scores is restricted (most programs < 50)

Actual-predicted recidivism for providers with scores ≥ 50 vs. < 50



Recap

- Is evidence-based practice effective?
- ‘Practice’ can be defined narrowly or broadly– SOP, protocol, generic intervention
- Tradeoffs between specificity of practice guidelines and generality of application
- Tradeoffs between specificity of practice, amount of research, and variability of effects
- Advantages to focus on generic interventions rather than SOPs or protocols

Recap (continued)

For the case of interventions with delinquents:

- Major sources of variation in effects are method, risk, implementation quality, treatment philosophy, and treatment type
- Workable 'best practice' guidelines can be developed from such evidence and applied to everyday programs for juvenile offenders
- Programs more in line with those guidelines do indeed show bigger effects on recidivism.



Contact information:
mark.lipsey@vanderbilt.edu

Papers about the meta-analysis work
and related issues can be found at:

www.vanderbilt.edu/cerm