AGENDA

THE CENTER FOR EVIDENCE-BASED CRIME POLICY

and

THE COCHRANE COLLABORATION COLLEGE FOR POLICY

at

GEORGE MASON UNIVERSITY

present an

EVIDENCE SYNTHESIS WORKSHOP

August 2012

www.cebcp.org

Evidence Synthesis Workshop

August 13, 2012: INNOVATION HALL 105, GEORGE MASON UNIVERSITY

8:30 AM	Registration and check in
9:00 AM	Welcome and introductions.
	So you want to do a systematic review? Developing policy-relevant research questions. *David Wilson and Charlotte Gill (CEBCP-George Mason University)*
9:30 AM	Introduction to systematic reviews: Searching the literature and retrieving studies. *Charlotte Gill (CEBCP-George Mason University)*
10:05 AM	*5 minute break*
10:10 AM	Introduction to systematic reviews: Developing a coding scheme. *Charlotte Gill (CEBCP-George Mason University)*
10:45 AM	*15 minute break*
11:00 AM	Introduction to meta-analysis: Calculating effect sizes and conducting basic analyses. *David Wilson (CEBCP-George Mason University)*
12:00 PM	LUNCH (sponsored by the CEBCP and the Cochrane College for Policy)
1:00 PM	Conducting rapid reviews: Timely evidence for public policy. *Catherine Gallagher (Cochrane College for Policy, George Mason University) and Chantelle Garritty (Ottawa Hospital Research Institute)*
2:30 PM	*Both workshops will join together in Innovation Hall 103 for the closing Keynote address (please see the main symposium agenda).*

TABLE OF CONTENTS

Organizations in attendance	2
Presenters	3
So you want to do a systematic review?	4
Searching the literature and retrieving studies	9
Developing a coding scheme	18
Conducting meta-analyses for advancing practice	29
Coding protocol example	46
Coding manual example	55
Systematic review resources	61
Reading list	63
Meta-analysis: David B. Wilson	64

Acknowledgments

All slides and notes are adapted from Campbell Collaboration training materials produced by Karianne Hammerstrøm, Eamonn Noonan, Terri Pigott, Jeff Valentine, Sandra Jo Wilson, David B. Wilson and others.

Coding protocol and manual examples taken from Charlotte Gill and Jordan Hyatt, *Probation Intensity Effects on Probationers' Criminal Conduct.* Campbell Systematic Reviews. http://campbellcollaboration.org/lib/project/140/

ORGANIZATIONS IN ATTENDANCE

AdvoCare, Inc.
American Bar Association
Ares Consulting
Association of Prosecuting Attorneys
Chakrabarti Habi Education Academy
Councilmember Marc Elrich
DC Metropolitan Police Department
Department Of Youth Rehabilitation Services
Development Services Group
Education Development Center
Fairfax County Juvenile and Domestic Relations District Court
Fairfax County Police Department
Fox Valley Technical College
George Mason University
George Mason University - CEBCP
George Mason University - Cochrane Collaboration College for Policy
Howard University
International Association of Chiefs of Police
International Initiative for Impact Evaluation (3ie)
MITRE
National Sheriff's Association
Netherlands Institute for the Study of Crime and Law Enforcement
Offender Aid and Restoration (OAR)
Ottawa Hospital Research Institute
Prince George's County Police Department
Shady Grove Hospital
University of Chicago Crime Lab
University of Pennsylvania
University of the District of Columbia
US Environmental Protection Agency
US Government Accountability Office
USDA - Center for Nutrition Policy and Promotion
USDA - Evidence Analysis Library Division
USDOJ - Faith Based and Neighborhood Partnerships
USDOJ - National Institute of Justice
USDOJ - Office of Justice Programs
USDOJ - Office of the Assistant Attorney General
VA Department of Criminal Justice Services
Vera Institute of Justice
Westat
World Bank

PRESENTERS

Catherine Gallagher, George Mason University Cochrane Collaboration College for Policy. Professor Gallagher focuses on improving the intersection between health care and justice agencies to better meet the needs of high-risk populations and the public health of their larger communities. Her work on justice-involved adolescents has appeared in the Journal of the American Academy of Child and Adolescent Psychiatry, the Journal of Adolescent Health, Social Science and Medicine and Pediatrics. In addition to her primary research, she develops, monitors and analyzes national statistical programs and provides federal agencies with policy guidance, routinely working with the Office of Juvenile Justice and Delinquency Prevention, the Centers for Disease Control, the Agency for Health Care Research and Quality, the Office of the Surgeon General, and the U.S. Bureau of the Census. She led the epidemiological and legal research efforts behind a joint-agency Federal Initiative on Juvenile Justice Health, and currently serves on the Campbell Collaboration's Crime and Justice Steering Group.

Charlotte Gill, George Mason University. Dr. Charlotte Gill is a senior research associate in the CEBCP and co-Director of the Research Program on Systematic Reviews. She completed her Ph.D. at the Jerry Lee Center of Criminology, University of Pennsylvania in 2010. She also holds an M.Phil in Criminology and an M.A. in Law from the University of Cambridge, United Kingdom. Dr. Gill's research interests include evidence-based crime prevention, juvenile justice, place-based criminology, program evaluation, quantitative methods, and research synthesis. She has been involved in conducting several randomized trials, including a study of low-intensity probation supervision in Philadelphia and experiments in restorative justice with the Metropolitan and Thames Valley police services in the United Kingdom. Dr. Gill also serves as the managing editor of the Campbell Collaboration Crime and Justice Group.

David B. Wilson, George Mason University. David B. Wilson, Ph.D., is Professor and Chair of the Department of Criminology, Law and Society at George Mason University and co-Director of the Research Programs on Systematic Reviews and Criminal Justice Policy. His Ph.D. is in applied social psychology from Claremont Graduate University. His research interests are the effectiveness of offender rehabilitation and crime prevention efforts, program evaluation methodology, and meta-analysis. His researched has included a broad range of topics, including the effectiveness of juvenile delinquency interventions, school-based prevention programs, correctional boot-camps, court-mandated batterer intervention programs, and drug-courts; the effects of sugar on children's behavior; and the effects of alcohol on violent behavior. He is an associate editor of the Journal of Experimental Criminology, a consulting editor for Psychological Bulletin, editor of the Campbell Collaboration Crime and Justice Group and a Campbell Collaboration Steering Group member, and was awarded the Marcia Guttentag Award for Early Promise as an Evaluator by the American Evaluation Association.

So You Want To Do a Systematic Review?

Developing Policy-Relevant Research Questions

David B. Wilson, PhD
Charlotte Gill, PhD
George Mason University

Evidence Synthesis Workshop
August 13, 2012

Overview

- Campbell Collaboration
- What is a systematic review?
- Why do a systematic review?
- Starting your review
- Stages of a review
- Problem formulation

Campbell Collaboration

- Named for Donald T. Campbell (1916-1996)
- "Reforms as Experiments" *Am. Psych. 24, 409-29 (April 1969)*
  - Experimental approach to social policy
- Voluntary network of scholars dedicated to preparing, disseminating, updating systematic reviews
- Modeled after Cochrane Collaboration in health care
- Crime & justice, education, social welfare, international development
- Methods and users groups
- Online library of systematic reviews

`www.campbellcollaboration.org/library.php`

What is a systematic review?

- Synthesis of all empirical evidence on a specific research question using explicit, systematic methods to minimize bias
  - Clearly stated objectives and pre-defined eligibility criteria
  - Explicit, reproducible methodology
  - Comprehensive search for studies
  - Assessment of validity of study findings
  - Systematic presentation and synthesis of study characteristics and findings

`http://www.cochrane-handbook.org`

Why do a systematic review?

- Combine findings from multiple studies
- Address gaps in knowledge
- Provide directions for future research
- Inform policy
- Apply scientific methods
- Organize and accumulate knowledge
- (some limitations!)
  - not enough good primary research
  - not generalizable to other settings
  - can be overly technical
  - challenges of translation

Starting your review

- You will need…
  - review team
  - expertise in topic
  - expertise in searching for studies
  - expertise in data analysis
  - guidance and support (Campbell Collaboration)
  - time
  - money
  - detailed protocol (plan) for transparency

Campbell Collaboration process

- Review and approval of title proposal
- Title proposal published in library
- Protocol peer reviewed by content experts, methods experts and information retrieval expert; editorial assistance; approval by steering committee
- Protocol published in library
- Review peer reviewed by content and methods experts; editorial assistance; approval by steering committee
- Review published in library

`http://www.campbellcollaboration.org/systematic_reviews/index.php`

Stages of completing a review

- Problem formulation
  - Clarifying questions
  - Explicit inclusion/exclusion criteria
- Data collection
  - Literature search
  - Coding studies
- Data evaluation
  - Apply criteria
  - Assessing study quality
- Data analysis and interpretation
  - Combining effect sizes
  - Interpreting results
- Report preparation
  - Narrative, statistics, graphs, tables
  - High level of detail for transparency

Problem formulation

- Goals:
  - Specify hypothesis of interest
  - Specify the evidence that is relevant to this hypothesis
- Scope of review:
  - Narrow questions to test effect of specific treatment
  - Broad questions for generating new knowledge (e.g. common elements of effective programs)
  - Not just about interventions—trends, diagnostic tests, risk factors…
  - Not limited to randomized trials/quantitative data

Research questions

- Effects of x intervention on y outcomes for z population/problem
  - Variations: comparative effects of x_1 and x_2
- How does x_1 relate to x_2 for population z
  - Variations: differential effects between z_1 and z_2
- Is Test A or Test B a better predictor of y?
  - Variations: Which test is a better predictor in z_1 vs z_2 populations

Steps in problem formulation

- Determine relevant conceptual and operational definitions
  - What terminology is used for the intervention/concept (especially across different fields)
  - Combining 'apples and oranges'
    - Only combine measures that examine the same underlying construct

Steps in problem formulation

- Set review parameters—PICOS
  - Populations/participants
  - Interventions
  - Comparison/counterfactual
    - Type of comparison group (other effective/ineffective treatments; treatment as usual; no treatment) has implications for interpretation
  - Outcomes
  - Study designs
- Title development along PICOS lines

[intervention] for [outcome] in [population]

Probation intensity effects on probationers' criminal conduct

Steps in problem formulation

- Theory/logic model
  - Helps to describe connections between interventions and outcomes
  - Help to decide which outcomes to use ('proximal' and 'distal' outcomes—immediate or long term)
  - Help to narrow down questions
  - Create your own logic model or use existing ones

Steps in problem formulation

- Set inclusion/exclusion criteria
  - Explicit, operational criteria are required for all systematic reviews
  - PICOS helps set criteria—parameters of intervention, population, study designs etc.
  - Other contextual criteria
    - Geography
    - Language
    - Timeframe
  - Use to screen studies for coding

Elements of a good research question

- Specific
- Answerable
- Measurable constructs
- Practical and relevant for policy/practice
- Logical—based on theory/logic model
- Empirical—can be answered with observable evidence

"SAMPLE"

Searching the Literature and Retrieving Studies

Charlotte Gill, PhD
George Mason University

Evidence Synthesis Workshop
August 13, 2012

Overview

- Elements of a systematic search
- Preparing to search
- Developing a search strategy
- Grey literature
- Information retrieval and management
- Documenting the search

Elements of a systematic search

- *Systematic* search is key to a *systematic* review!
- Research study analogy: sampling and enrollment phase
- Systematic strategy
- Transparent reporting
- Comprehensive in scope
- Goal: uncover all relevant studies that meet inclusion criteria

Preparing to search: Be systematic

- Identify sources
- Develop a search strategy
- Consider time period
- Consider different disciplines
- Construct search terms
- Remember PICOS

…more on this later

Preparing to search: Be transparent

- Search strategy should be transparent and replicable—document <u>everything</u>
  - Each information source, dates covered, dates last searched
  - Full search strategy: keywords, limits, variations for different search engines
  - Flow chart of information through the search and screening process (later)

Preparing to search: Be comprehensive

- Sensitive searches: finding any potentially relevant studies
- Specific searches: finding studies that clearly meet eligibility criteria
- Initial search should aim for more sensitivity than specificity—save the screening for later

Preparing to search: Be comprehensive

- Consult with information specialist (librarian, Campbell trial search coordinator)
- Keyword searches of multiple electronic databases
- Searches of organization websites, special registers, reference lists, listservs
  - "snowball" sampling
- Personal contacts with expertise in the field
- Hand searching of selected journals
- Team approach to screening

Developing a systematic search strategy

- Most searches start with electronic databases
- Mostly focus on published literature
- Search multiple databases in all related disciplines to achieve sensitivity
- Aim for a wide variety of search terms
- Consult a librarian
- Systematic Review Resources page in your workbook includes links to lists of common bibliographic databases

Some common databases

- Criminal Justice Abstracts
- ERIC
- National Criminal Justice Reference Service
- ProQuest
- PsycINFO
- PubMed (Medline)
- Sage Full Text Collections (Criminology, Sociology, Education, Psychology)
- Social Science Citation Index
- Social Science Research Network
- Sociological Abstracts
- Web of Science

(note: most require institutional subscription)

Developing a search strategy

- What are the key concepts to be searched?
- How are these represented in each discipline?
- What are the related terms?
- How are these concepts represented in the controlled vocabulary within each database?
  - Some databases include a thesaurus—alphabetical listing of database descriptors—to help find related terms

Constructing search terms

- What terms should be searched as descriptors or free text?
- What Boolean operators should be used?
- When should truncated terms be used?
  - *arrest* will find arrest, arrested, rearrest, re-arrest…
- What limits should be set?
  - time period
  - publication type…

Using Boolean operators

- AND
  - combines different concepts
  - Both terms must be present to identify a record
    - intensive AND probation
- OR
  - searches for related terms and synonyms
  - either term may be present to identify a record
    - probation OR supervision
- NOT
  - ensures second term will not appear in results
    - adult NOT juvenile

Constructing search terms with PICOS

P	I	C	O	S
adult juvenile male female…	probation supervision intensive frequent ISP…	incarceration …	recidivism arrest conviction offending drug use…	evaluation experiment trial…

OR

AND

(probation* OR supervis* OR case*) AND
(intens* OR frequen* OR ratio OR ISP) AND
(recidiv* OR *arrest* OR *convict* OR *offend*)

Note: may need to modify search string for different databases

Modifying search terms: Examples

Database	Search Date	Total Hits	Search String
Criminal Justice Abstracts	7/11/09	114	probation* AND supervis* AND (intens* OR frequen* OR ratio) AND case* AND (recidiv* OR [arrest* OR re*arrest*] OR [convict* OR re*convict*])
Urban Institute	7/20/09	43	Search, any word forms, proximity = page 1. probation+supervision+case+intensive+recidivism *[15 hits]* 2. probation+supervision+case+intensive+arrest *[12 hits]* 3. probation+supervision+case+intensive+convict *[1 hit]* 4. probation+supervision+case+frequency+recidivism *[4 hits]* 5. probation+supervision+case+frequency+arrest *[4 hits]* 6. probation+supervision+case+frequency+convict *[1 hit]* 7. probation+supervision+case+ratio+recidivism *[2 hits]* 8. probation+supervision+case+ratio+arrest *[4 hits]* 9. probation+supervision+case+ratio+convict *[0 hits]*
Diss. Abstracts	7/27/09	1682	TEXT(probation*) AND TEXT(supervis*) AND TEXT(case*) AND TEXT(intens* OR frequen* OR ratio) AND TEXT(recidiv* OR (arrest* OR rearrest* OR re-arrest*) OR (convict* OR reconvict* OR re-convict*))

Searching for study designs?

- Some reviewers use search terms for evaluation designs/experiments ("S" of PICOS)
- I recommend <u>not</u> using this strategy
- Researchers are not consistent in describing methodology
- Even among randomized experiments, methodological terms not always included in keywords or abstracts
- Several studies have found that up to 67% of relevant studies might be missed when searches capture specific study designs

Grey literature

- Studies that are not commercially published or available through traditional sources
- Examples: technical reports, dissertations
- Failure to search for grey literature can create publication bias
  - Statistically significant findings more likely to be published
- Searching for grey literature increases search sensitivity

Identifying grey literature

- Contact researchers in the field
- Post on forums and listservs
- Search websites of organizations that conduct and use research; contact them
- Search reference sections of relevant studies and reviews for additional references
- Search electronic databases of unpublished and in-progress research
  - ERIC; Rutgers library database; NCJRS
- Citation searches in Google Scholar
  - Find works by author or that cite a reference

Google Scholar: Some cautions

- Generic search engines (Google Scholar) increase likelihood of finding grey literature but *significantly* increase 'false positives'
- Automatically sorted by relevance, so some scholars review only the first 500 or 1,000 hits
- Boolean search is not intuitive (can be done in advanced search, but search strings don't work well)

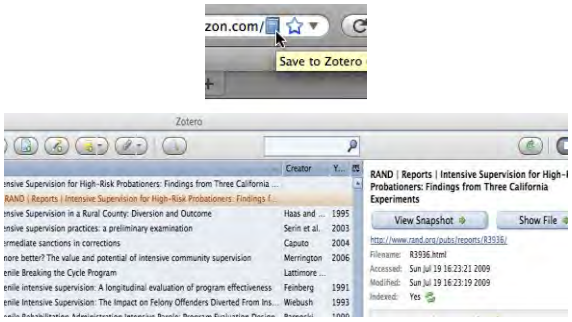
Information retrieval process

- Preliminary searches
  - Define concepts and research question
  - Use standard reference tools and broad searches for key studies and reviews
  - Recommended as a 'scoping' phase before proceeding with review
- Main searches
  - Identify primary studies through searches of online databases, websites, hand searches etc
- Final searches
  - Towards end of review process, refine search terms and update if needed

Information management

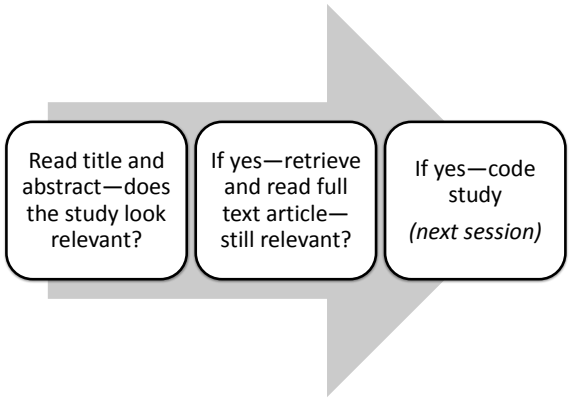
- Export search results
  - Save as Text/save to EndNote, BibTeX etc
  - Zotero: direct import from supported websites
- Import into bibliographic management software
  - RefWorks, EndNote, Zotero…
- Edit bibliographic database
  - Add notes
  - Delete duplicates
  - File in preparation for coding (irrelevant, screen etc)

Information management: Zotero



http://www.zotero.org

Finalizing the 'study sample'



Documenting the process

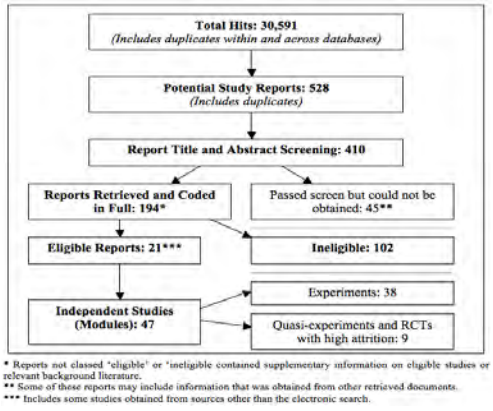
- Remember: search process should be
  - systematic
  - transparent
  - comprehensive…

    …and *replicable*
- Crucial to document all stages of process so you or others can replicate review
  - Search strategies for each database
  - Date searched
  - Studies found
  - Decision rules

Documenting the search strategy

Database	Search Date	Total Hits	Search String
Criminal Justice Abstracts	7/11/09	114	probation* AND supervis* AND (intens* OR frequen* OR ratio) AND case* AND (recidiv* OR [arrest* OR re*arrest*] OR [convict* OR re*convict*])
Urban Institute	7/20/09	43	Search, any word forms, proximity = page 1. probation+supervision+case+intensive+recidivism *[15 hits]* 2. probation+supervision+case+intensive+arrest *[12 hits]* 3. probation+supervision+case+intensive+convict *[1 hit]* 4. probation+supervision+case+frequency+recidivism *[4 hits]* 5. probation+supervision+case+frequency+arrest *[4 hits]* 6. probation+supervision+case+frequency+convict *[1 hit]* 7. probation+supervision+case+ratio+recidivism *[2 hits]* 8. probation+supervision+case+ratio+arrest *[4 hits]* 9. probation+supervision+case+ratio+convict *[0 hits]*
Diss. Abstracts	7/27/09	1682	TEXT(probation*) AND TEXT(supervis*) AND TEXT(case*) AND TEXT(intens* OR frequen* OR ratio) AND TEXT(recidiv* OR (arrest* OR rearrest* OR re-arrest*) OR (convict* OR reconvict* OR re-convict*))

Documenting the sample identification



* Reports not classed 'eligible' or 'ineligible contained supplementary information on eligible studies or relevant background literature.
** Some of these reports may include information that was obtained from other retrieved documents.
*** Includes some studies obtained from sources other than the electronic search.

Developing a Coding Scheme

Charlotte Gill, PhD
George Mason University

Evidence Synthesis Workshop
August 13, 2012

Overview

- Eligibility criteria and screening
- Development of coding protocol and database
- Coding your sample
- Common mistakes

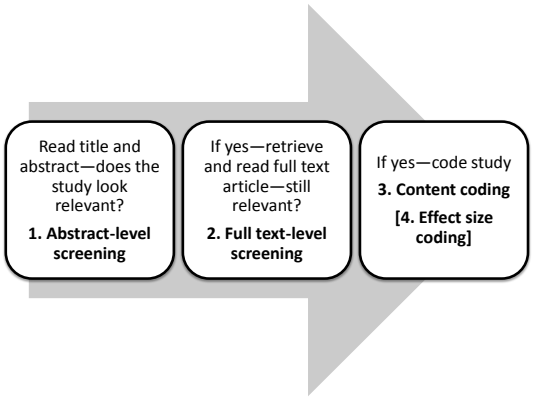
What is coding?

- Extraction of data from primary studies
- Research study analogy: coding serves two purposes:
  - Detailed eligibility screening
  - Data collection phase
- Analogous to surveys or interviews—coding should 'ask questions' of studies to determine eligibility and record data

Why code?

- Account of studies included in the review
- Identify what's missing
- Identify characteristics of PICOS
- Obtain information for data analysis (effect size data)
- Identify variables that may account for different findings across different primary studies

Four levels of coding



Abstract-level eligibility screening

- Read title and abstract: does the study look relevant?
  - This can be done directly in Zotero/other reference management database
  - Use notes and folders ('irrelevant,' 'potentially eligible,' 'code'…)
- If relevant, obtain full text
- Exclude obviously irrelevant studies but do not assume abstract will contain full information
- When in doubt, double code
  - At least 2 trained coders working independently
  - Document decision rules

Full text-level eligibility screening

- Develop eligibility screening form with criteria (part of protocol)
- Link together multiple reports from same study before screening
- Complete form for all studies (whether or not eligible)
- Double code (at least a sample)

Eligibility screening example

B4. The study evaluates an intensive probation or parole program involving increased supervision by probation officers in a reduced caseload, or low-intensity probation (increased caseload, less supervision).

1. Yes 0. No

B5. A difference in probation intensity between the treatment and comparison groups, as evidenced by a change in caseload size, ratio of clients to officers, or other control measures, is a key component of the overall program.

1. Yes 0. No

B6. The study includes a comparison group receiving 'standard probation,' not comprised of dropouts from ISP/low intensity, or other supervision by probation officer (not incarcerated controls). Study design may be experimental or quasi-experimental, but not a one-group research design.

1. Yes 0. No

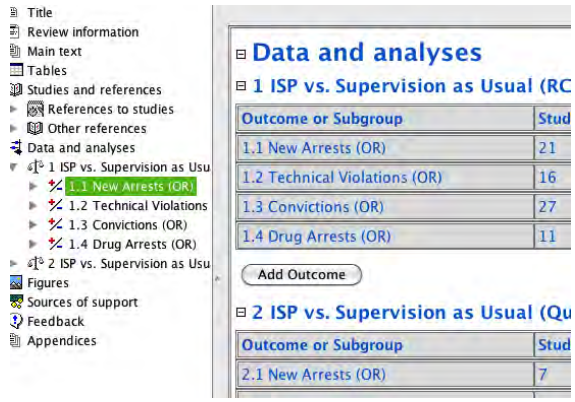
B7. The study includes a post-program measure of criminal behavior (arrest, conviction) or technical violation of probation/parole – may be official or self-reported and dichotomous or continuous.

1. Yes 0. No

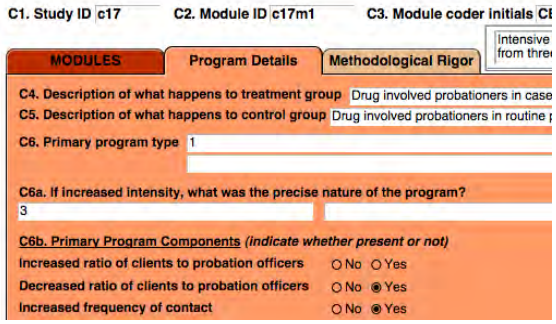
Record keeping

- Develop database for screening and coding
  - RevMan (free but can only code limited info)
  - Create your own relational database (MS Access, FileMaker Pro…)
  - MS Excel
- Database should be structured around eligibility and coding protocol
- Screening and coding can be done directly into database
  - Advantages: save time on data entry; easy export for analysis
  - Disadvantages: multiple copies of files; possible time/cost investment on front end

Record keeping examples: RevMan



Record keeping examples: In-house database using FileMaker Pro



Eligibility screening results

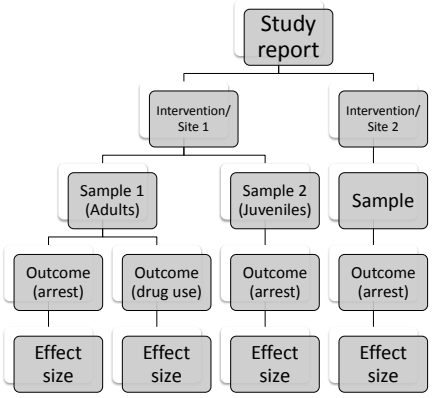
- A set of studies eligible for coding
- An account of ineligible studies and reasons for exclusion
  - Campbell/Cochrane Collaboration reviews often include table of ineligible studies in an appendix

Study Reference	Reason for Exclusion
Smith (1984) Alabama prison option: supervised intensive restitution program. *Fed. Prob., 48,* 32-35.	No comparison group. No crime outcomes.

Content and Effect Size Coding

- Develop a coding manual including:
  - Setting, study context, authors, publication date, publication type…
  - Study methods and appraisal of methodological quality
  - Program/intervention description
  - Participant/sample description
  - Outcomes
  - Findings and effect sizes *for each outcome*
- Coding manual should be available in paper and electronic format
- Coding manual instructions can be incorporated into coding database

Hierarchical data structure



Designing the database

- Coding protocol should be divided into
  - Studies
  - Interventions/sites/modules
  - Sample (participants)
  - Outcomes
  - Effect sizes
- Two methods for coding
  - Flat file (1 record/row per study)
  - Hierarchical structure
- Relational database can help to structure coded data appropriately

Flat file design

Study ID	Int. Type	ES1	DV1	ES2	DV2	ES3	DV3
22	2	0.77	3				
23	2	0.77	3				
31	1	-0.1	5	-0.05	5	-0.2	11
36	2	0.94	3				
40	1	0.96	11				
82	1	0.29	11				

Hierarchical file design

Study level data file

Study ID	Year	Int. Type	TxN
22	1994	2	24
23	2010	1	30

Effect size level data file

Study ID	ES ID	Outcome Type	ES
22	1	1	-0.39
22	2	1	0
22	3	1	0.09
23	1	2	-1.05
23	2	4	0.34

Study context coding

- Setting, study context, authors, publication date and type
  - Multiple publications; 'study' vs. 'report'
  - Geographic/national setting; language
  - Publication type, potential publication bias
  - Publication date vs. study date
  - Research, demonstration, practice studies

Study context coding example

A9. Publication type:

1. Book
2. Book chapter
3. Peer-reviewed journal article
4. Government report (federal)
5. Government report (state/local)
6. Unpublished (e.g., dissertation, technical report, conference paper)
8. Other:

Study method coding

- Methodological quality appraisal may be used in reporting results (e.g. report effects separately for RCTs and quasi-experiments; moderator analysis)
  - Do NOT weight results by study quality scores!
- Several methods available
  - Cochrane Risk of Bias framework
    - Focus on identifying sources of bias
    - Not always appropriate for social science research
  - GRADE system
  - Methodological quality checklists
    - Many exist; all have advantages and disadvantages
  - Direct coding of methodological characteristics

GRADE system

- Quality of evidence across trials
- Outcome-specific
- Considers
  - sparse data
  - consistency of results across trials
  - study designs
  - reporting bias
  - potential confounding
- See `http://www.gradeworkinggroup.org`
- Software available at `http://www.ims.cochrane.org/revman/gradepro`

Direct coding of methods

- Basic research design
  - Nature of assignment to conditions
  - Attrition, crossover, dropout, other changes to assignment
  - Nature of control condition
  - Multiple intervention and/or control groups
- Design quality
  - Initial and final comparability of groups (pre-test scores)
  - Treatment-control contrasts: contamination, blinding
- Other aspects (depends on specific topic)
  - Fidelity, monitoring of implementation
  - Training of data collectors
  - Statistical controls for group differences
  - Handling of missing data

Study method coding example

C23. Overall methodology rating

1. Comparison group lacks demonstrated comparability to treatment group
2. Comparison between 2+ groups, one with and one without the intervention
3. Comparison between program group and one or more control groups, controlling for other factors, or nonequivalent comparison group is only slightly different from program group, or randomized controlled trial with high attrition
4. Random assignment and analysis of comparable program and comparison groups, including controls for attrition

Intervention coding

- General program type
- Specific program elements
- Treatments etc received by comparison group
- Treatment implementation issues
  - Integrity, fidelity
  - Amount, length, frequency, 'dosage'
- Goal: differentiate across studies

Intervention coding example

C6a. If increased intensity, what was the precise nature of the program?

1. 'Front door' prison diversion (probation instead of prison)
2. 'Backdoor' prison diversion (early release from prison)
3. Enhanced probation
4. Enhanced parole
5. Enhanced probation and parole
8. Other:

Participant coding

- Participants/clients/sample
  - Data at aggregate level (characteristics of entire sample)
  - Mean age/age range
  - Gender mix
  - Racial/ethnic mix
  - Average risk
  - Special groups

Participant coding example

D13. Approximate gender description of sample:

1. All male (>90%)
2. More males than females (60-90% male)
3. Roughly equal males and females
4. More females than males (60-90% female)
5. All female (>90%)
9. Can't tell

Outcome coding

- Construct measured
- Measure/operationalization used
- Source of information
- Composite or single indicator
- Scale: dichotomous, count, discrete ordinal, continuous
- Reliability and validity
- Time of measurement (e.g. relative to treatment)

Outcome coding example

E7. Recidivism construct represented by this measure:

1. Arrest
2. Charge
3. Conviction
4. Technical violation
5. Probation revocation
6. Incarceration
8. Other:

Effect size coding

- Findings
  - Sample size
  - Outcome value
  - Outcome metric
  - Statistical significance and test used
  - Calculate effect size when possible
  - Code data on which computations are based
- More on this next session

Effect size coding example

Effect size data – all effects

F13. Treatment group sample size for this ES:
F14. Comparison group sample size for this ES:

Effect size data – continuous outcomes

F15. Treatment group mean:
F16. Comparison group mean:
F17. Are the above means adjusted?
1. Yes 0. No
F18. Treatment group standard deviation:
F19. Comparison group standard deviation:

Extracting the data

- Double coding
  - Ideally: double code all studies
  - Can also double code a sample
  - Inter-rater reliability (e.g. Cohen's Kappa)
  - Training sample: use to clarify disagreements and refine coding protocol
    - Key decisions: inclusion/exclusion criteria, key characteristics, risk of bias, effect size coding

Common mistakes

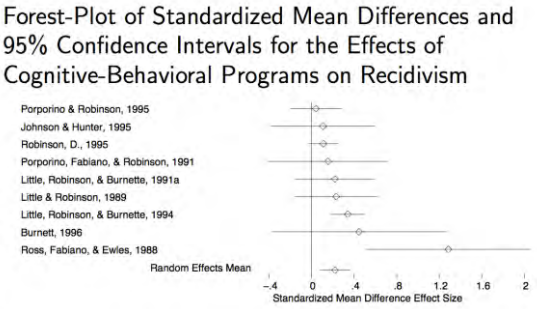
- Too many coding items (risk of spurious results)
- Too many subjective coding items
- Coding two reports from the same study as two different studies
- Including non-independent samples as separate studies
  - Includes multiple independent treatment samples compared to the same control group
- Coder drift; inadequate training
- Failure to ask questions

Conducting Meta-Analyses for Advancing Practice

David B. Wilson, PhD
George Mason University

Evidence Synthesis Workshop
August 13, 2012

The End-Game



Overview

- Historical background
- Logic of Meta-analysis
- Effect sizes
  - What are they
  - Common types
- Basic analysis
  - Mean effect size
  - Confidence interval
  - Homogeneity analysis
- Random-effects versus Fixed-effect model
- Moderator analysis
- Publication Bias
- A note about software

A Great Debate

- Eysenck 1952: Psychotherapy doesn't work
- Dizzying array of mixed results followed
- Glass (with Smith) average results from 375 studies
- Glass coined the term meta-analysis

Deep Roots

- Pearson (1904): averaged correlations between inoculation for typhoid fever and mortality
- Fisher (1944): independent studies individually may not be significant, yet the aggregate seem improbable
- W. G. Cochran (1953): developed methods of averaging means across studies
- A. Wicker (1967) average correlations between attitudes and behavior
- Concurrent with Smith and Glass (1977) were
  - Hunter and Schmidt (1977) *Validity generalization*
  - Rosenthal and Rubin (1978) *Interpersonal expectancy effects*

Logic of Meta-analysis

- Narrative review methods:
  - Focuses on statistical significance
  - Lacks transparency and replicability
- Weakness of statistical significance:
  - Significant effect is a strong conclusion
  - Non-significant effect is a weak conclusion
  - How do you balance a collection of significant and non-significant effects?

Logic of Meta-analysis

- Meta-analysis:
  - Focuses on direction and magnitude of effect
  - Approaches task as a research endeavor
  - Examines pattern of evidence across studies
    - Average effect
    - Consistency of effects
    - Relationship between study features and effects

Some Preliminaries

- A meta-analysis should adopt systematic review methods
  - Comprehensive search for all relevant studies
  - Explicit inclusion/exclusion criteria
  - Systematic and reliable coding

Effect Size

- Encodes relationship of interest into a common index
- Must be:
  - comparable across studies
  - independent of sample size
  - have a computable standard error
- Many different effect size indexes
- Multiple methods of computing each
- Most common:
  - Correlation coefficient (*r*)
  - Standardized mean difference (*d* or *g*)
  - Odds-ratio and Risk-ratio

Computing Effect Sizes

- Must compute effect size from information provided
  - Conversions from other statistics
    - t-test
    - p-value
    - descriptive statistics
    - etc.
  - Manipulation of data
    - Collapsing across subgroups
    - adding "drop-outs" back into the treatment condition
  - Some conversions better than others (algebraic equivalents; rough approximations)
- Some studies simply do not provide necessary information

Standardized Mean Difference

- Fundamental relationship:
  - Group contrast
  - Continuous dependent variable
- Logic: scaling effects based on standard deviation
- Definitional equation:

$$ES_{sm} = \frac{\bar{X}_1 - \bar{X}_2}{s_{pooled}}$$

Standardized Mean Difference

- Based on a t-test

$$ES_{sm} = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

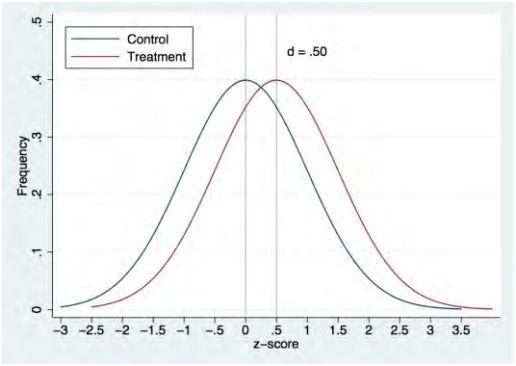
- Based on a correlation

$$ES_{sm} = \frac{2r}{\sqrt{1 - r^2}}$$

- Based on 2 by 2 table (dichotomous outcome; logit method)

$$ES_{sm} = \ln\left(\frac{ad}{bc}\right)\frac{\sqrt{3}}{\pi}$$

Visual Example of *d*-type Effect Size



Methods of Computing *d*

- Lots of methods of computing *d*
- Goal is to reproduce what you would get with means, standard deviations, and sample sizes
- Some methods are straightforward

Computing *d* from a *t*-test

Formula for *d*:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Formula for *t*:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{n_1 + n_2}{n_1 n_2}}}$$

Therefore:

$$d = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

Online Effect Size Calculator

http://gunston.gmu.edu/cebcp/EffectSizeCalculator/index.html

Correlation as Effect Size

- Fundamental relationship:
  - Two inherently continuous constructs
- Correlation "comes" standardized

$$ES_r = r$$

- Example: Relationship between GRE scores and performance in graduate school

Odds-Ratio

- Fundamental relationship:
  - Group contrast
  - Dichotomous dependent variable
- Data can be represented in a 2 by 2 contingency table

	Success	Failure
Treatment Group	*a*	*b*
Control Group	*c*	*d*

- Odds-ratio effect size computed as:

$$ES_{OR} = \frac{ad}{bc}$$

Basics of Meta-Analysis

- Goal:
  - Describe the distribution, including its mean
  - Establish a confidence interval around the mean
  - Test that the mean differs from zero
  - Test whether studies tell a consistent story (are homogeneous)
  - Explore the relationship between study features and effect size

Determining the Mean Effect Size

- Problem: some effect sizes are more accurate than others
- What we need is an index of precision
- Standard error is a direct measure of precision
- Hedges and Olkin solution:
  - Weight by the inverse variance
  - Provides a statistical basis for:
    - Standard error of the mean effect size
    - Confidence intervals
    - Homogeneity testing

Some Preliminary Transformations

- Small sample size bias correction for the standardized mean differences:

$$ES'_{sm} = \left(1 - \frac{3}{4N - 9}\right) ES_{sm}$$

- Fisher's Z_r transform of correlations (ES_r):

$$ES_{Z_r} = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

- Log transform of Odds-ratios ES_{OR} (also for Risk-ratio):

$$ES_{\ln(OR)} = \log(ES_{OR})$$

Inverse Variance Weights

- Standardized mean difference ES_{sm}:

$$se_{sm} = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{ES_{sm}^2}{2(n_1 + n_2)}}$$

- Correlation ES_r (actually, the Fisher's Z_r):

$$se_r = \frac{1}{\sqrt{n-3}}$$

- Odds-ratio ES_{OR} (actually, the logged odds-ratio)

$$se_{OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Inverse Variance Weights

- Inverse variance weight *w*:

$$w = \frac{1}{se^2}$$

Almost ready

- At this point, we have for each study:
  - An effect size
  - An inverse variance weight
- Problem: statistical models assume independence
- Only include one effect size per study (or independent sample)
- Multiple analyses for different subsets of independent effects
  - Different outcome constructs
  - Different time periods

Inverse Variance Weighted Mean

Meta-analytic mean effect size is:

$$\overline{ES} = \frac{\Sigma w_i ES_i}{\Sigma w_i}$$

where ES_i is the effect size for each study (i) and w_i is the inverse variance weight

Standard error of the mean effect size is:

$$se_{\overline{ES}} = \frac{1}{\Sigma w_i}$$

Some Basic Inferential Statistics

Confidence intervals can be constructed in the usual manner:

$$\overline{ES}_{lower} = \overline{ES} - se_{\overline{ES}} 1.96$$
$$\overline{ES}_{upper} = \overline{ES} + se_{\overline{ES}} 1.96$$

And a z-test can be performed as:

$$z = \frac{\overline{ES}}{se_{\overline{ES}}}$$

An Example: Cognitive-Behavioral Programs for Adult Offenders

Author	Sample Size	*d* Effect Size	*w*
Burnett, 1996	60	0.45	5.684
Johnson & Hunter, 1995	98	0.11	15.934
Little & Robinson, 1989	180	0.23	25.049
Little et al 1991	152	0.22	27.852
Little et al 1994	1,381	0.34	150.485
Porporino & Robinson, 1995	757	0.04	65.529
Porporino et al 1991	63	0.16	11.953
Robinson, D., 1995	2,125	0.11	187.177
Ross et al 1988	45	1.28	6.441

Note: These studies are a subset of studies included in Wilson et al. (2005) and represent two specific treatment programs (Moral Reconation and Reasoning and Rehabilitation) and studies that were randomized or used high quality quasi-experimental designs.

An Example: Cognitive-Behavioral Programs for Adult Offenders

Stata output from "meanes.ado"

```
. meanes es_calc [w=w]
(analytic weights assumed)
Version 2005.05.23 of meanes.ado

No. of obs  =       9                              Homogeneity Analysis
Minimum obs =   .0421                                        Q =      14.19
Maximum obs =   1.285                                       df =         8
Weighted SD =   0.169                                        p =    0.07695
------------------------------------------------------------------------------
                |    Mean    -95%CI    +95%CI        SE         Z         P
----------------+-------------------------------------------------------------
Fixed effect    | 0.20458   0.11659   0.29258   0.04490   4.55671   0.00001
Random effect 1 | 0.22173   0.07911   0.36435   0.07277   3.04718   0.00231
Random effect 2 | 0.21192   0.09744   0.32640   0.05841   3.62808   0.00029
------------------------------------------------------------------------------
1 Random effects variance component (method of moments)   = 0.01685
2 Random effects variance component (full information ML) = 0.00624
```

Homogeneity Testing

- Homogeneity analysis tests whether the assumption that all of the effect sizes are estimating the same population mean is a reasonable assumption.
- If homogeneity is rejected, the distribution of effect sizes is assumed to be heterogeneous.
  - Single mean ES not a good descriptor of the distribution
  - There are real between study differences, that is, studies estimate different population mean effect sizes
  - Three options:
    - model between study differences
    - fit a random effects model
    - do both

Homogeneity *Q* Statistic

- *Q* is simply a weighted sums-of-squares:

$$Q = \sum w_i (ES_i - \overline{ES})^2$$

- There are easier computational formulas:

$$Q = \sum w_i ES_i^2 - \frac{(\Sigma w_i ES_i)^2}{\Sigma w_i}$$

- It is distributed as a chi-square with $k - 1$ degrees of freedom, where k is the number of effect sizes

Alternative to *Q*

- *Q* is statistically under-powered when the number of studies is low and when the sample size within the studies is low
- $I^2 = 100\% \times \frac{Q - df}{Q}$
- Larger values of I^2, the more heterogeneity
- 75%: large heterogeneity
- 50%: moderate heterogeneity
- 25%: low heterogeneity

Random versus Fixed Effects Models

- Fixed effects model assumes:
  - there is one true population effect that all studies are estimating
  - all of the variability between effect sizes is due to sampling error
- Random effects model assumes:
  - there are multiple (i.e., a distribution) of population effects that the studies are estimating
  - variability between effect sizes is due to sampling error + variability in the population of effects

Fixed versus Random: Which to Use?

- A random-effects model becomes a fixed-effect model when distribution is homogeneous
- Assumptions of fixed effects model rarely plausible
  - Consequence: standard error that is too small; confidence intervals that are too narrow
- Bottom-line: Use the random-effects model

Computing a Random Effects Model

- Fixed effects model: weights are a function of sampling error
- Random effects model: weights are a function of sampling error + study level variability
- Thus, we need a new set of weights

Computing a Random Effects Model

- First, compute τ^2 (random effects variance component):

$$\tau^2 = \frac{Q - df_Q}{\Sigma w_i - \frac{\Sigma w_i^2}{\Sigma w_i}}$$

- Second, re-compute the inverse variance weights:

$$w_i = \frac{1}{se_i^2 + \tau^2}$$

- Third, re-compute meta-analytic results using new weight

Moderator Analysis

- Modeling between study variability
  - Categorical models (analogous to a one-way ANOVA)
  - Regression models
- Fixed and random effects version of each (latter often called “mixed” models)

Analog to the ANOVA

- Useful for a single categorical independent variable
- Produce a separate mean effect size for each category
- Recall that *Q* is a sum-of-squares
- The total sum-of-squares (*Q*) can be partitioned
  - Variability between groups ($Q_{between}$)
  - Residual variability within groups (Q_{within})

Analog to the ANOVA

- $Q_{between}$ analogous to an *F*-test between means
- Q_{within} assesses whether residual distribution homogeneous
- Note: in a random effects (mixed effects) version of this, the Q_{within} is not meaningful

Analog to the ANOVA

Experimental versus Quasi-experimental Studies in the Domestic Violence

Meta-Analytic Analog to the One-way ANOVA, Mixed Effects Model

Source	Q	df	P
Between	0.0072	1	0.93227
Within	15.2630	12	0.22737
Total	15.2702	13	0.29079

Descriptive Fixed Effects Meta-Analytic Results by: random

random	Mean	St. Er.	[95% Conf.	Int.]	z	P>|z|	k
0	.285	.17366	-.05493	.62579	1.6437	0.10025	7
1	.264	.18284	-.09435	.62236	1.4439	0.14876	7
Total	.275	.12591	.02848	.52206	2.1862	0.02880	14

Random effects variance component (via iterative max. likelihood) = .1513908
Standard error of random effects variance component = .0818776

Meta-analytic Regression

- Conceptually identical to multiple regression
  - Effect size is the dependent variable
  - Study moderator variables are the independent variables
- Can handle multiple variables simultaneously
- Don't use standard OLS regression procedures (even if weighted)
- Must use specialized software

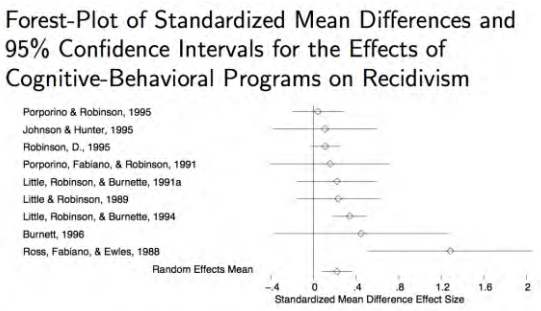
Meta-analytic Regression

```
*****  Inverse Variance Weighted Regression  *****
*****  Random Intercept, Fixed Slopes Model  *****

------- Descriptives -------
      Mean ES     R-Square            k
        .1483        .2225      38.0000

------- Homogeneity Analysis -------
                    Q            df            p
Model         14.7731       3.0000        .0020
Residual      51.6274      34.0000        .0269
Total         66.4005      37.0000        .0021

------- Regression Coefficients -------
               B       SE  -95% CI  +95% CI        Z        P     Beta
Constant   -.6752    .2392  -1.1439   -.2065  -2.8233    .0048    .0000
RANDOM      .0729    .0834   -.0905    .2363    .8746    .3818    .1107
TXVAR1      .3790    .1438    .0972    .6608   2.6364    .0084    .3264
TXVAR2      .1986    .0821    .0378    .3595   2.4204    .0155    .3091

------- Method of Moments Random Effects Variance Component -------
v      =   .04715
```

Forest Plots

- Visual representation of results
- Row for each study that shows
  - study label
  - sample size; may include other information
  - effect size (dot, square, diamond)
  - confidence interval (horizontal line)
- Row for the overall mean results
  - effect size (dot, square, diamond)
  - confidence interval (horizontal line)

Example Forest-Plot



Publication Selection Bias

- Statistically significant effects are more likely to be published than nonsignificant effects
- Important threat to the validity of meta-analysis (and any other method of reviewing studies)
- Search for and include unpublished studies that meet eligibility criteria
- Examine difference between published and unpublished studies

Publication Selection Bias

- Statistical approaches to assessing publication bias
  - Funnel plot: Scatterplot of effect size against standard error of effect size
  - Trim-and-fill method (Tweedie and Duvall)

Comments on Software

- Specialized software
  - RevMan (developed for Cochrane)
  - CMA – Comprehensive Meta-analysis
- Add-ons for Common Statistical Programs
  - Stata – lots of macros available
  - SPSS – macros available from my website
  - SAS – macros available from my website
  - R – lots of procedures available

Comments on Software

- For computing effect sizes
  - CMA
  - RevMan
  - ES Calculator by Wilson
    - `http://www.campbellcollaboration.org/resources/effect_size_input.php`
    - `http://gunston.gmu.edu/cebcp/EffectSizeCalculator/index.html`

Question of When to do a Meta-analysis

- Minimum number of studies
- Design Similarity
- Design Quality
- Heterogeneity
- Multiple meta-analysis as part of a review

Final Comments

- Methods continue to advance
- Methods for analyzing dependent effect sizes actively advancing
- Common errors
  - Incorrectly computing effect sizes
  - Not recognizing situations where effect sizes can be computed
  - Using fixed-effect models
  - Not using moderator analysis to compare mean effect sizes for study subsets
  - Focusing too much on statistical significance and not size and direction of effect

PROBATION INTENSITY EFFECTS ON
PROBATIONERS' CRIMINAL CONDUCT
SYSTEMATIC REVIEW CODING SHEETS

A. STUDY LEVEL CODING SHEET

Instructions: one study level coding sheet to be used per study. If the study is reported in multiple documents, use the primary publication as the study identifier and list other document numbers below.

A1. Study ID:_____ studid
A2. Cross-ref document ID:_____ xref1
A3. Cross-ref document ID:_____ xref2
A4. Cross-ref document ID:_____ xref3
A5. Coder initials:_____ coder
A6. Date coded:_____ codate
A7. Title:_____ title
A8. Author(s):_____ author
A9. Publication type: pubtype

1. Book
2. Book chapter
3. Peer-reviewed journal article
4. Government report (federal)
5. Government report (state/local)
6. Unpublished (e.g. dissertation, technical report, conference paper
8. Other _____

A10. Journal ref. (vol., issue):_____ jref
A11. Publication year:_____ pubyr
A12. Date range of **research**:_____ resdate
A13. Country of publication: _____ publoc
A14. Country of study setting:_____ resloc
A15. Number of treatment-comparison contrasts in report: _____ mods
Only independent treatment group samples should be counted; see Instructions for Section B.
*If no comparison group, just complete **B. ELIGIBILITY CHECKLIST**.*
A16. Is the same comparison group used in each contrast? cxlmod

0. No
1. Yes
8. N/A

B. ELIGIBILITY CHECKLIST

B1. First author's last name: _____ elname
B2. Coder initials:_____ coelig
B3. Date eligibility determined:_____ eldate

To be eligible, a study must meet the following criteria. Answer each question with 1=yes, 0=no

B4. The study evaluates an intensive probation or parole program involving increased supervision by probation officers in a reduced caseload, or low-intensity probation (increased caseload, less supervision).
1. Yes 0. No evpro

B5. A difference in probation intensity between the treatment and comparison groups, as evidenced by a change in caseload size, ratio of clients to officers, or other control measures, is a key component of the overall program.
1. Yes 0. No evsep

B6. The study includes a comparison group receiving 'standard probation,' not comprised of dropouts from ISP/low intensity or other supervision by probation officer (not incarcerated controls). Study design may be experimental or quasi-experimental, but not a one-group research design.
1. Yes 0. No evcomp

B7. The study includes a post-program measure of criminal behavior (arrest, conviction) or technical violation of probation/parole – may be official or self-reported and dichotomous or continuous.
1. Yes 0. No evoutc

For documents that do not meet the above criteria, answer the following questions:

B8. Document is not a quantitative evaluation (no data regarding effects of ISP/LIP reported).
1. Yes 0. No evndat

B9. Document is a review article relevant to this project (e.g., references to studies, background information for write-up).
1. Yes 0. No evusef

B10. Document status (circle one): elstat
1. Eligible
0. Not Eligible
9. Relevant Review

Notes:

C. TREATMENT-COMPARISON CODING SHEET

Instructions: if the study reports on multiple treatment-comparison contrasts, or multiple treatments compared to a single comparison group, each contrast should be coded on separate Treatment-Comparison Coding Sheets. Only independent evaluations should be included (i.e. multiple treatment groups should not have overlapping participants).

Identifying Information

C1. Study ID:_____ studid
C2. Module ID:_____ modid
C3. Coder initials:_____ comod

Program Details

C4. Description of what happens to treatment group:_____ txdesc

_____

_____

C5. Description of what happens to comparison group:_____ cxldesc

_____

_____

C6. Primary program type progtype
1. Increase in probation intensity
2. Decrease in probation intensity
8. Other: _____

C6a. If increased intensity, what was the precise nature of the program? progdesc
1. 'Front door' prison diversion (probation instead of prison)
2. 'Back door' prison diversion (early release from prison)
3. Enhanced probation
4. Enhanced parole
5. Enhanced probation and parole
8. Other: _____

C6b. Primary program components (*indicate whether present or not*):

Program increases ratio of clients to probation officers	1. Yes 0. No	progir
Program decreases ratio of clients to probation officers	1. Yes 0. No	progdr
Program increases frequency of contact with probation officer	1. Yes 0. No	progif
Program decreases frequency of contact with probation officer	1. Yes 0. No	progdf
Program increases drug testing requirements	1. Yes 0. No	progidt
Program decreases drug testing requirements	1. Yes 0. No	progddt
Other: 1. Yes 0. No _____		progoth

C6c. If Yes for any of the above, state exact numbers if available (*999 if not*):

Control ratio: _____ / Treatment ratio: _____	racxl/ratx
Control freq: _____ / Treatment freq: _____	frcxl/frtx
Control drug tests: _____ / Treatment drug tests: _____	drcxl/drtx
Other: _____	

C6d. Additional program components (*indicate whether present or not*)

Curfew	1. Yes	0. No	addcomp_curf
Drug treatment	1. Yes	0. No	addcomp_drug
Electronic monitoring	1. Yes	0. No	addcomp_em
Employment program/assistance	1. Yes	0. No	addcomp_empl
Halfway house	1. Yes	0. No	addcomp_hh
Home visits	1. Yes	0. No	addcomp_hv
House arrest	1. Yes	0. No	addcomp_harr
Offense-specific treatment (e.g. sex offender treatment)	1. Yes	0. No	addcomp_offtx
Other treatment	1. Yes	0. No	addcomp_tx
Other 1. Yes 0. No _____			addcomp_oth

C7. What happened to the comparison group? cxltype
1. 'Supervision as usual'
8. Other: _____

C8: Was supervision for treatment group provided by anyone other than probation officer? posup
0. No
1. Yes (explain): _____
9. Don't Know/Can't Tell

C9. Length of intervention in months (weeks/4.3):

Minimum _____	txlmin
Maximum _____	txlmax
Mean _____	txlmn
Fixed (same for all subjects) _____	txlfix

C10. Did the intervention follow a set protocol? txprot
0. No
1. Yes
9. Don't Know/Can't Tell

C11. What supervision philosophy was stated? txphil
1. Control
2. Treatment
3. Hybrid
8. Other _____
9. Don't know/not stated

C12. Did the intervention remain consistent over time? txcons
0. No
1. Yes
9. Don't Know/Can't Tell

Methodological Rigor

C13. Control variables used in statistical analyses to account for initial group differences?
0. No cxlvars
1. Yes

C14. Subject-level matching? matched
0. No
1. Yes

C15. Random assignment to conditions? rassgt
0. No
1. Yes

C16. Measurement of prior criminal involvement? prior
0. No
1. Yes

C17. Rating of initial similarity between treatment and control group: prsim

1	**2**	**3**	**4**	**5**	**6**	**7**

(1 = Nonrandomized; high likelihood of baseline differences between groups or known differences related to future recidivism)
(5 = nonrandomized design with strong evidence of initial equivalence)
(7 = randomized design with large N or small N design with matching)

C18. Was attrition discussed in the report? attrep
0. No
1. Yes

C19. Is there a potential threat to generalizability from overall attrition? attgen
0. No
1. Yes

C20. Is there a potential threat to internal validity from differential attrition? attint
0. No
1. Yes

C21. Did the statistical analysis attempt to control for differential attrition effects? attstat
0. No
1. Yes
9. Don't Know/Can't Tell

C22. Statistical significance testing used? sigtest
0. No
1. Yes

C23. Overall methodology rating methrat

1. Comparison group lacks demonstrated comparability to treatment group
2. Comparison between 2+ groups, one with and one without the intervention
3. Comparison between program group and one or more control groups, controlling for other factors, or
   nonequivalent comparison group is only slightly different from program group, or
   randomized controlled trial with high attrition
4. Random assignment and analysis of comparable program and comparison groups, including controls for attrition

Notes on methodology

_____

_____

D. SAMPLE LEVEL CODING SHEET

Instructions: A study may report results separately for distinct samples (e.g. persons with/without prior arrests). Each distinct sample must have its own coding sheet. The treatment-comparison contrast is the same for the different samples.
Samples should be independent, i.e. no overlapping participants. Some studies report the results broken out by different subgroups (e.g. by gender). Only one of these breakouts can be used – choose the one with the most information, or the one most relevant to the review.

Identifying Information

D1. Study ID:_____ studid
D2. Module ID:_____ modid
D3. Sample ID:_____ sampid
D4. Coder initials:_____ cosamp

Sample Description

D5. Description of treatment group sample:_____ txsamp

_____

_____

D6. Description of comparison group sample:_____ cxlsamp

_____

_____

D7. Total N in treatment group at beginning of study:_____ txn
D8. Total N in comparison group at beginning of study:_____ cxln
Note: D7 + D8 = total sample size prior to attrition. If multiple samples are being coded, the sum across samples must equal the total sample size prior to attrition.

D9. Age range of study participants sampage

1. Adolescent (12-18)
2. Youth (18-21)
3. Adult (21+)
4. Adolescent and youth
5. Youth and adult
6. Adolescent, youth, and adult
8. Other:_____
9. Unspecified/can't tell

D10. Youngest age included in sample (*999 if unknown*):_____ yage
D11. Oldest age included in sample (*999 if unknown*): _____ oage

D12. Exact proportion of males in sample (if known): _____ prmale

D13. Approximate gender description of sample sampgen

1. All male (>90%)
2. More males than females (60-90% male)
3. Roughly equal males and females
4. More females than males (60-90% female)
5. All female (>90%)
9. Can't tell

D14. Race/ethnicity of sample (*999 if unknown*):

% Asian_____ rasian
% Native American___ rnative
% Black_____rblack
% White_____rwhite
% Hispanic_____ rhisp
% Other_____rother

D15. General offender type: offtype

1. Violent and/or person crimes
2. Nonviolent and/or nonperson crimes
3. Mixed: violent/nonviolent
4. Specialized caseload: drugs
5. Specialized caseload: sex offenses
6. Specialized caseload: mental health
7. Specialized: domestic violence
9. Don't know/Can't tell
8. Other _____

D16. Composition of supervised offenders: offcomp

1. All probationers
2. All parolees
3. Probationers and parolees

D16a. If combination of probationers and parolees (*999 if unknown*):

% probation: _____ pcpro % parole: _____ pcpar

D17. Probationer/parolee risk level: offrisk

1. Low risk
2. Medium risk
3. High risk
4. Low and medium risks
5. Medium and high risks
6. All risk levels
7. No risk assessment
9. Don't know/Can't tell
8. Other _____

D18: How was risk determined? riskjmt

1. Statistical model
2. Prior convictions
3. Instant offense
4. Judgment of PO/intake
5. Classification instrument
7. N/A
8. Other:_____
9. Don't know/Can't tell

D19. Probationer/parolee need level: offneed

1. Low need
2. Medium need
3. High need
4. Low and medium need
5. Medium and high need
6. All need levels
7. No need assessment
9. Don't know/Can't tell
8. Other _____

E. DEPENDENT VARIABLE CODING SHEET

Instructions: code each dependent variable reported in the study separately. The same dependent variable measured at multiple times should be coded only once. For non-crime outcomes, code only items E4, E7 and E8.

Identifying Information

E1. Study ID:_____ studid
E2. Module ID:_____ modid
E3. Sample ID:_____ sampid
E4. Outcome ID:_____ outid
E5. Coder initials:_____ coout

Outcome Information

E6. Outcome label (label used in report):_____ outlab
E7. Recidivism construct represented by this measure rconst
1. Arrest
2. Charge
3. Conviction
4. Technical violation
5. Probation revocation
6. Incarceration
8. Other _____

E8. Offense types included in recidivism measure:

All offenses ('No' for others)	1. Yes	0. No	oall
Drug offenses	1. Yes	0. No	odrug
Person offenses, sexual	1. Yes	0. No	opsx
Person offenses, nonsexual	1. Yes	0. No	opnsx
Person offenses, unspecified	1. Yes	0. No	opuns
Property offenses	1. Yes	0. No	oprop
Weapons offenses	1. Yes	0. No	oweap
Driving offenses	1. Yes	0. No	odriv
Technical or status offenses	1. Yes	0. No	otech
Other 1. Yes 0. No _____			ooth

E9. Measurement scale mscale
1. Dichotomous
2. Trichotomous
3. 4-9 discrete ordinal categories
4. >9 discrete ordinal categories/continuous

E10. Source of data dsrce
1. Self-report
2. Other report (e.g. PO)
3. Official records (police, probation records, court etc.)
8. Other:_____
9. Don't Know/Can't Tell

E11. Length of follow-up period: fulng
1. < 6 months
2. 6-12 months
3. >1, <2 years
4. > 2 years
8. No follow-up
9. Don't Know/Can't Tell

E12. Is cost/benefit data for the program included in the study?
1. Yes
0. No

F. EFFECT SIZE LEVEL CODING SHEET

Instructions: Complete a separate coding sheet for each treatment-comparison contrast for each dependent variable.

Identifying Information

F1. Study ID:_____ studid
F2. Module ID:_____ modid
F3. Sample ID: sampid
F4. Outcome ID:_____ outid
F5. Effect size ID:_____ esid
F6. Coder initials:_____ coes

Effect Size Information

F7. Effect size type estype
1. Baseline (pretest; prior to start of intervention)
2. Post-test (first measurement point, post-intervention)
3. Follow-up (all subsequent measurement points, post-intervention)

F8. Which group does the raw effect favor (ignoring statistical significance)? esdir
1. Treatment group
2. Comparison group
3. Neither (ES = 0)
9. Can't tell (ES cannot be used if this is selected)

F9. Does the investigator report the difference as statistically significant? essig
0. No
1. Yes
8. Not tested
9. Can't tell

F10. If tested, what type of statistical test was used? estest
1. *t* test
2. *F* test
3. χ^2
4. Regression analysis
7. N/A
8. Other:_____
9: Can't tell

F11. Timeframe in months captured by measure (weeks/4.3)
Minimum_____ estmin
Maximum_____ estmax
Mean_____ estmn
Fixed (same for all subjects) _____ estfix

F12. Timeframe in months from end of program to measurement point (weeks/4.3)
Minimum_____ esfumin
Maximum_____ esfumax
Mean_____ esfumn
Fixed (same for all subjects) _____ esfufix

Effect size data – all effects

F13. Treatment group sample size for this ES: _____ estxn
F14. Comparison group sample size for this ES: _____ escxl

Effect size data – continuous outcomes

F15. Treatment group mean: _____ estxmn
F16. Comparison group mean : _____ escxlmn
F17. Are the above means adjusted? 1. Yes 0. No esmadj
F18. Treatment group standard deviation: _____ estxsd
F19. Comparison group standard deviation: _____ escxlsd
F20. Treatment group standard error: _____ estxse
F21. Comparison group standard error: _____ escxlse
F22. *t*-value from an independent *t*-test or square root of *F*-value from a one-way ANOVA with 1d.f. in the numerator (only 2 groups): _____ estval
F23. Exact probability for a *t*-value from an independent *t*-test or *F*-value from a one-way ANOVA with 1d.f. in the numerator: _____ estvalp
F24. Correlation coefficient: _____ escorr

Effect size data – dichotomous outcomes

F25. Number successful in treatment group: _____ estxs
F26. Number successful in comparison group: _____ escxls
F27. Proportion successful in treatment group: _____ estxspr
F28. Proportion successful in comparison group: _____ escxlspr
F29. Are the above proportions adjusted for pretest variables?
1. Yes 0. No espradj
F30. Logged odds ratio: _____ eslogor
F31. Standard error of logged odds ratio: _____ eslorse
F32. Logged odds ratio adjusted? (e.g. from logistic regression)
1. Yes 0. No esloradj
F33. χ^2 value with d.f.=1 (2x2 contingency table): _____ eschisq
F34. Correlation coefficient: _____ esdcorr

Effect size data – hand calculated

F35. Hand calculated *d*-type effect size: _____ eshand
F36. Hand calculated SE of the *d*-type effect size: _____ eshandse

PROBATION INTENSITY EFFECTS ON PROBATIONERS' CRIMINAL CONDUCT
SYSTEMATIC REVIEW CODEBOOK
October 23, 2009

A. STUDY LEVEL CODING SHEET

A1. Study ID

Assign a unique number to each study. If a report presents two independent studies, they should be numbered and coded separately.

A2-4. Cross-ref document IDs

If the study is reported in multiple documents, use the primary publication as the study identifier and list other document numbers below.

A9. Publication type

If two separate reports are being used to code a single study, code the type of the more formally published report.

1. Book
2. Book chapter
3. Peer-reviewed journal article
4. Government report (federal)
5. Government report (state/local)
6. Unpublished (e.g. dissertation, technical report, conference paper)
8. Other

A12. Date range of **research**

Provide the start and end dates for the study where available.

A15. Number of treatment-comparison contrasts in report

Only independent treatment group samples should be counted. They can be contrasted to the same control group.

B. ELIGIBILITY CHECKLIST

B4. The study evaluates an intensive probation or parole program involving increased supervision by probation officers in a reduced caseload, or low-intensity probation (increased caseload, less supervision).

Programs that alter the intensity of supervision (especially intensive probation) usually include many and varied additional components. Eligible studies must involve a change in **supervision** by the probation officer at the probation office as one of the program components. This could be an increase/decrease in the ratio of clients to officers, the frequency of contact per month, or drug testing requirements. Supervision is interpreted as surveillance/control behaviors. Any treatment provided, even by the probation officer, is not considered supervision. Similarly, control by external agencies (such as the private firms that monitor electronic tagging devices) is not considered probation supervision in this study.

B5. A difference in probation intensity between the treatment and comparison groups, as evidenced by a change in caseload size, ratio of clients to officers, or other control measures, is a key component of the overall program.

The change in intensity must be a stated component of the program, not merely incidental to other components and services.

B6. The study included a comparison group receiving 'standard probation,' not comprised of dropouts from ISP/low intensity. Study design may be experimental or quasi-experimental, but not a one-group research design.

Participants receiving increased or decreased supervision should not be contrasted to participants in other criminal justice settings, e.g. prison, or receiving no services/supervision.
Studies should be rated 4 or 5 on the Maryland Scientific Methods Scale (SMS). Rigorous research designs that are not included in the SMS may also be included.

B7. The study includes a post-program measure of criminal behavior (arrest, conviction) or technical violation of probation or parole – may be official or self-reported and dichotomous or continuous.

C. TREATMENT-COMPARISON CODING SHEET

Identifying Information

C2. Module ID

Each distinct treatment-comparison (module) within a study should be given a separate, unique number and linked back to the overall study ID.

Program Details

C6. Primary program type

1. Increase in probation intensity
2. Decrease in probation intensity
8. Other

C6a. If increased intensity, what was the precise nature of the program?

Intensive supervision programs can serve several different purposes. They may be provided as an alternative to imprisonment, or an enhanced sanction for existing probationers. 'Front door' diversion refers to intensive probation programs that operate as a full alternative to those who would otherwise go to prison or jail, whereas 'back door' diversion programs offer intensive supervision to existing prisoners in exchange for early release. Enhanced probation or parole intensifies supervision for offenders already on probation or parole.

1. 'Front door' diversion
2. 'Back door' diversion
3. Enhanced probation
4. Enhanced parole
5. Enhanced probation and parole
8. Other
9. Don't know/can't tell

C6c. If Yes for any of the above, state exact numbers if available

For frequency, state number of contacts per month. If fewer than once a month (e.g. once every 3 months), give the equivalent fraction by month (e.g. 0.33/month).

C7. What happened to the comparison group?

'Supervision as usual' refers to supervision according to the agency's normal standards, without added/decreased intensity.

1. 'Supervision as usual'
8. Other

C9. Length of intervention in months

Divide length in weeks by 4.3 to convert to months.

C11. What supervision philosophy was stated?

Common models of probation supervision usually involve a primarily control/surveillance-oriented approach, a treatment/care-oriented approach, or a hybrid of the two. If the study clearly states which philosophy underpins the intensive/low intensity probation, record it here.

1. Control
2. Treatment
3. Hybrid
8. Other
9. Don't know/not stated

Methodological Rigor

C17. Rating of initial similarity between treatment and control group

Assess initial similarity between treatment and control groups (at baseline) on a scale of 1-7, where 1 is least similar (nonrandomized design; high likelihood of baseline differences between groups or known differences related to future recidivism), 5 indicates a nonrandomized design with strong evidence of initial equivalence, and 7, the highest, indicates a randomized design with a large sample size, or a small sample with matching.

C23. Overall methodology rating

1. Comparison group lacks demonstrated comparability to treatment group
2. Comparison between 2+ groups, one with and one without the intervention
3. Comparison between program group and one or more control groups, controlling for other factors, or
   nonequivalent comparison group is only slightly different from program group, or
   randomized controlled trial with high attrition
4. Random assignment and analysis of comparable program and comparison groups, including controls for attrition

D. SAMPLE LEVEL CODING SHEET

D9. Age range of study participants

1. Adolescent (12-18)
2. Youth (18-21)
3. Adult (21+)
4. Adolescent and youth
5. Youth and adult
6. Adolescent, youth, and adult
8. Other
9. Unspecified/can't tell

D13. Approximate gender description of sample

1. All male (>90%)
2. More males than females (60-90% male)
3. Roughly equal males and females
4. More females than males (60-90% female)
5. All female (>90%)
9. Can't tell

D15. General offender type

Note the type of offender caseload selected for the program

1. Violent and/or person crimes
2. Nonviolent and/or nonperson crimes
3. Mixed: violent/nonviolent
4. Specialized caseload: drugs
5. Specialized caseload: sex offenses
6. Specialized caseload: mental health
7. Specialized: domestic violence
8. Other
9. Don't know/Can't tell

D16. Composition of supervised offenders

Note supervision status of participants selected for program

1. All probationers
2. All parolees
3. Probationers and parolees

D17. Probationer/parolee risk level

If eligibility/participation was based on offender risk assessment, note risk level here.

1. Low risk
2. Medium risk
3. High risk
4. Low and medium risks
5. Medium and high risks
6. All risk levels
7. No risk assessment
8. Other
9. Don't know/Can't tell

D18: How was risk determined?

If a risk assessment was carried out, on what information was it based?

1. Statistical model
2. Prior convictions
3. Instant offense
4. Judgment of PO/intake
5. Classification instrument
7. N/A
8. Other
9. Don't know/Can't tell

D19. Probationer/parolee need level

Probation/parole supervision and treatment provided in accordance with the 'risk principle' (also known as "RNR" – Risk/Need/Responsivity) is based on assessments of offender need as well as risk. If participation/eligibility was also determined by an assessed need level, record it here.

1. Low need
2. Medium need
3. High need
4. Low and medium need
5. Medium and high need
6. All need levels

7. No need assessment
8. Other
9. Don't know/Can't tell

E. DEPENDENT VARIABLE CODING SHEET

Outcome Information

E6. Outcome label (label used in report)
What name is used in the study to refer to this particular outcome?

E7. Recidivism construct represented by this measure
1. Arrest
2. Charge
3. Conviction
4. Technical violation
5. Probation revocation
6. Incarceration
8. Other

E9. Measurement scale
Trichotomous and discrete ordinal variables will be treated as continuous in the Effect Size section.
1. Dichotomous
2. Trichotomous
3. 4-9 discrete ordinal categories
4. >9 discrete ordinal categories/continuous

E10. Source of data
1. Self-report
2. Other report (e.g. PO)
3. Official records (police, probation records, court etc.)
8. Other
9. Don't Know/Can't Tell

E11. Length of follow-up period
1. < 6 months
2. 6-12 months
3. >1, < 2 years
4. > 2 years
8. No follow up
9. Don't know/Can't tell

E12. Is cost/benefit data for the program included in the study?
State whether or not the report provides any information on the cost of the program relative to alternative/standard procedure. No need to include any further information at this point.

F. EFFECT SIZE LEVEL CODING SHEET

Effect Size Information

F7. Effect size type
1. Baseline (pretest; prior to start of intervention)
2. Post-test (first measurement point, post-intervention)
3. Follow-up (all subsequent measurement points, post-intervention)

F8. Which group does the raw effect favor (ignoring statistical significance)?
1. Treatment group
2. Comparison group
3. Neither (ES = 0)
9. Can't tell (ES cannot be used if this is selected)

F10. If tested, what type of statistical test was used?
1. *t* test
2. *F* test
3. χ^2
4. Regression analysis
7. N/A
8. Other
9: Can't tell

F11. Timeframe in months captured by measure
Divide weeks by 4.3 to get timeframe in months.

F12. Timeframe in months from end of program to measurement point (weeks/4.3)
Divide weeks by 4.3 to get timeframe in months.

Effect size data – all effects

Complete these two questions for all measures of effect.

F13. Treatment group sample size for this ES: _____ estxn
F14. Comparison group sample size for this ES: _____ escxl

Effect size data – continuous outcomes

Include trichotomous and discrete ordinal measures when coding continuous outcomes.
F17. Are the above means adjusted?
Note whether or not other covariates have been controlled for in producing a mean value.
This definition of 'adjusted' also applies to questions F29 and F31.

Effect size data – hand calculated

This section is not needed if the study provides the outcomes needed to calculate effect sizes in the form of odds ratios or standardized mean differences. It only needs to be filled in where effect sizes are imputed from other information (e.g. a chi-square). If this is the case, use David B. Wilson's effect size calculator in Excel to enter the appropriate information and report the value for *d* in F35.

F35. Hand calculated *d*-type effect size: _____ eshand
F36. Hand calculated SE of the *d*-type effect size: _____ eshandse

SYSTEMATIC REVIEW RESOURCES

The Campbell Collaboration. http://www.campbellcollaboration.org

Campbell Systematic Reviews: freely-available, online library of completed reviews and protocols. http://www.campbellcollaboration.org/library.php

Collection of research synthesis tools, including Dave Wilson's Effect Size Calculator. http://gemini.gmu.edu/cebcp/SynthesisTools.html

Description of the Campbell Collaboration process. http://www.campbellcollaboration.org/systematic_reviews/index.php

Templates and policies for producing a Campbell review. http://www.campbellcollaboration.org/research/the_production.php

Expectations and guidance for Campbell systematic review authors. http://www.campbellcollaboration.org/research/expectations_and_guidance.php

Campbell Collaboration Information Retrieval Guide (including commonly searched databases for the social sciences). http://www.campbellcollaboration.org/resources/research/new_information_retrieval_guide.php

List of free online bibliographic databases.(Crime and Justice) http://www.campbellcollaboration.org/resources/links/CCJG_bibliographic_databases.php

List of useful links for systematic reviews and experiments (Crime and Justice focus). http://www.campbellcollaboration.org/resources/links/links_crime_and_justice.php

Centre for Reviews and Dissemination, University of York: Undertakes high quality systematic reviews in health and social care. http://www.york.ac.uk/inst/crd

Cochrane Collaboration: http://cochrane.org

Cochrane Collaboration College for Policy at George Mason University. http://cochrane.gmu.edu

Comprehensive Meta-Analysis: paid software for meta-analysis and graphing from BioStat. http://www.meta-analysis.com

CrimeSolutions.gov: A new website from the Office of Justice Programs that uses rigorous research to determine what works in criminal justice, juvenile justice, and victim services. http://crimesolutions.gov

Don M. Gottfredson Library of Criminal Justice Gray Literature Database at Rutgers University: a resource for identifying unpublished and hard to find studies. http://www.law-library.rutgers.edu/cj/gray

EPPI-Centre: The Evidence for Policy and Practice Information and Coordinating Centre at the University of London carries out systematic reviews and develops review methodology in social science and public policy. http://eppi.ioe.ac.uk/cms

Evidence-Based Policing Matrix, Center for Evidence-Based Crime Policy, GMU. http://gemini.gmu.edu/cebcp/Matrix.html

Meta-Analysis Unit at the University of Murcia, Spain: Information and advice on conducting meta-analysis including databases of meta-analytic studies and methodological resources in English and Spanish. http://www.um.es/metaanalysis

Peabody Research Institute at Vanderbilt University: Under the direction of Mark Lipsey, this center is dedicated to improving programs for children, youth, and families using rigorous evaluation and research synthesis. http://www.peabody.vanderbilt.edu/Peabody_Research_Institute.xml

RevMan: free software for systematic reviews, meta-analysis and graphing provided by the Cochrane Collaboration. http://ims.cochrane.org/revman/

Systematic Reviews: A new journal dedicated to the publication of high-quality systematic reviews and methodological articles on systematic reviews and meta-analysis. http://www.systematicreviewsjournal.com/

Zotero: free information management software, useful for systematic reviews. Developed by the Roy Rosenzweig Center for History and New Media at George Mason University. http://www.zotero.org

READING LIST

Berk, R. (2007). Statistical inference and meta-analysis. *Journal of Experimental Criminology*, *3*(3), 247–270.

Cooper, H. (1998). *Synthesizing research* (3rd ed.). Thousand Oaks, CA: Sage.

Cooper, H., Hedges, L.V., & Valentine, J.C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed). New York, NY: Russell Sage Foundation.

Farrington, D. P., Weisburd, D., & Gill, C. (2011). The Campbell Collaboration Crime and Justice Group: A decade of progress. In C. Smith, S. Zhang, & R. Barberet (Eds.), *Handbook of International Criminology*. Routledge.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Petrosino, A., Farrington, D. P., & Sherman, L. W. (2003). Special report: The Campbell Collaboration Crime and Justice Group: Early development and progress. *Journal of Offender Rehabilitation*, *38*(1), 5–18.

Petrosino, A., Boruch, R., Soydan, H., et al. (2001). Meeting the challenges of evidence-based crime policy: The Campbell Collaboration. *Annals of the American Academy of Political and Social Sciences, 578,* 14–34.

Petrosino, A., Boruch, R., Farrington, D., Sherman, L., & Weisburd, D. (2003). Toward evidence-based criminology and criminal justice: Systematic reviews, the Campbell Collaboration, and the Crime and Justice Group. *International Journal of Comparative Criminology, 3,* 42–61.

Petticrew, M. (2003). Why certain systematic reviews reach uncertain conclusions. *British Medical Journal*, *326*(7392), 756–758.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Oxford: Blackwell.

Pratt, T. C. (2010). Meta-analysis in criminal justice and criminology: What it is, when it's useful, and what to watch out for. *Journal of Criminal Justice Education*, *21*(2), 152–168.

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology*, *4*(1), 61–81.

Wilson, D. B. (2009). Missing a critical piece of the pie: simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology*, *5*(4), 429–440.

Wilson, D. B. (2010). Meta-analysis. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 181–208). New York, NY: Springer. [ATTACHED ON FOLLOWING PAGE]