TAKE STOCK OF CRIMINAL PROFILING:
A Narrative Review and Meta-Analysis

BRENT SNOOK
JOSEPH EASTWOOD
Memorial University of Newfoundland

PAUL GENDREAU
University of New Brunswick

CLAIRE GOGGIN
St. Thomas University

RICHARD M. CULLEN
Memorial University of Newfoundland

The use of criminal profiling (CP) in criminal investigations has continued to increase despite scant empirical evidence that it is effective. To take stock of the CP field, a narrative review and a 2-part meta-analysis of the published CP literature were conducted. Narrative review results suggest that the CP literature rests largely on commonsense justifications. Results from the 1st meta-analysis indicate that self-labeled profiler/experienced-investigator groups did not outperform comparison groups in predicting offenders’ cognitive processes, physical attributes, offense behaviors, or social habits and history, although they were marginally better at predicting overall offender characteristics. Results of the 2nd meta-analysis indicate that self-labeled profilers were not significantly better at predicting offense behaviors, but outperformed comparison groups when predicting overall offender characteristics, cognitive processes, physical attributes, and social history and habits. Methodological shortcomings of the data and the implications of these findings for the practical utility of CP are discussed.

Keywords: criminal profiling; meta-analysis; predictive validity; psychological profiling

Criminal profiling (CP) is the practice of inferring personality, behavioral, and demographic characteristics of criminals based on crime scene evidence (Douglas, Ressler, Burgess, & Hartman, 1986). The frequency with which CP has been used in criminal investigations, as well as the volume of literature addressing this topic, has grown steadily over the past 30 years (Copson, 1995; Egger, 1999; Wilson, Lincoln, & Kocsis, 1997; Witkin, 1996), and profiling techniques are now commonplace within police investigations worldwide (Homant & Kennedy, 1998). This upward trend has occurred in the absence of a well-defined profiling framework and cumulated empirical knowledge in support of CP. Some researchers (e.g., Grubin, 1995; Hicks & Sales, 2006; Muller, 2000; Wilson et al., 1997) have cautioned that CP is growing in popularity in the absence of compelling scientific evidence that it “works” (i.e., is a reliable, valid, or useful tool for assisting with the

AUTHORS’ NOTE: Support for the research reported in this article was provided to the first author by the Natural Sciences and Engineering Research Council of Canada. Correspondence concerning this article should be addressed to Brent Snook, Psychology Department, Science Building, Memorial University of Newfoundland, St. John’s, Newfoundland, Canada, A1B 3X9; e-mail: bsnook@play.psych.mun.ca.

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 34, No. 4, April 2007 437-453
DOI: 10.1177/0093854806296925
© 2007 American Association for Correctional and Forensic Psychology
identification and apprehension of criminals). Given that the effect of profiling on criminal investigations is unknown, the goal of the current article is to conduct a systematic review of the literature to determine the scientific credibility of CP, which in turn will inform the utility of this particular investigative technique for practitioners.

Constructing a profile of an unknown perpetrator typically involves three stages (Annon, 1995; Ault & Reese, 1980; Douglas et al., 1986; Homant & Kennedy, 1998). First, police officers collect crime scene data and forward it to a profiler; second, the profiler conducts an analysis of the crime scene data; and third, the profiler provides predictions about the type of individual likely to have committed the crime in question. The processes that profilers use in analyzing crime scene data can be classified as either “clinical” or “statistical” in nature. Clinically oriented techniques incorporate aspects of the profilers’ intuition, knowledge, experience, and training to generate predictions (e.g., Boon, 1997; Douglas & Olshaker, 1995, 1997; Ressler & Schachtman, 1992; Turvey, 1999; West, 2000). By contrast, statistically oriented predictions are based upon descriptive and inferential statistical models derived from an analysis of characteristics of offenders who have previously committed similar types of crime (e.g., Canter, 2004; Farrington & Lambert, 1997; Keppel & Weis, 1993; Salfati, 2000).

Published accounts testify to the prolific growth in the utilization of CP techniques. Between 1971 and 1981, the FBI provided profiling assistance on 192 occasions (Pinizzotto, 1984). Just a few years later, Douglas and Burgess (1986) indicated that FBI profilers had been asked to assist with 600 criminal investigations per year. More recent accounts indicate that CP was applied by 12 FBI profilers in approximately 1,000 cases per year (Witkin, 1996). Police officers in the United Kingdom have also incorporated CP into their investigations with greater frequency. Copson (1995), for instance, reported that 29 profilers were responsible for providing 242 instances of profiling advice between 1981 and 1994, with the use of CP increasing steadily during that period. Although we do not have an exact estimate of CP prevalence elsewhere, the use of CP has been documented in a variety of countries including Sweden, Finland, Germany, Canada, and The Netherlands (see Ásgard, 1998; Case Analysis Unit, 1998; Clark, 2002; Jackson, Herbrink, & van Koppen, 1997).

As the prevalence of CP has grown over the past three decades, there has been a concomitant increase in the volume of published literature on the topic. Reviews of the CP literature have most often outlined its developmental history, described the various theoretical approaches that profilers use, and commented on the need for future research (Egger, 1999; Grubin, 1995; Homant & Kennedy, 1998; Muller, 2000; Wilson et al., 1997). The authors of these reviews appear to have reached the consensus that, notwithstanding deficiencies in the empirical literature regarding its predictive accuracy, profiling works or, at least, has the potential to work. As a consequence, readers of this literature may be inclined to believe that CP is a valuable addition to the standardized investigative repertoire. The sheer volume of scholarly and media attention accorded to CP might also contribute to this belief. Thus, a critical examination of the current status of CP is timely. As a first step toward that end, we conducted a systematic narrative review of the published CP literature and a meta-analysis of the extant experimental studies of profiler accuracy.

Using a classification framework adapted from Gendreau, Goggin, Cullen, and Paparozzi (2002), the narrative review classified CP articles according to whether the authors used commonsense or empirical-based arguments. The genesis of the Gendreau et al. model was based on an analysis of how practical, or “bad,” common sense (as opposed to “good”
common sense and as outlined in the left-hand side of Table 1) can lead to serious errors in judgment in the field of criminal justice policy (see also Latessa, Cullen, & Gendreau, 2002). For the remainder of the article we avoid the pejorative label and refer to it simply as common sense. The application of this model to the narrative review was particularly relevant to an evaluation of the criminal profiling literature because it was hoped that the conclusions of such an analysis would make a useful contribution to the ongoing debate regarding the status of CP as “art” (based on experience and intuition) or “science” (based on empirical research that generates falsifiable hypotheses), and, ultimately, speak to its utility as a criminal investigative tool. Given the acknowledgment by some researchers that there is a shortage of empirical evidence substantiating profiling techniques (e.g., Kocsis, 2004; Kocsis, Irwin, Hayes, & Nunn, 2000), it was anticipated that the CP literature contained a considerable volume of commonsense rationales.

The second step in our analysis of the validity of CP involved a meta-analysis of empirical studies examining the accuracy of criminal profilers. Despite the two broad types of profiling processes mentioned previously, there is no consensus about who can be called a profiler or regulatory body that grants professional profiling designations (see Kocsis, 2004). For the purpose of his studies, Kocsis defined a profiler as anyone who labeled himself or herself a profiler and had engaged in the practice of constructing a profile for a criminal investigation. Alternatively, Hazelwood, Ressler, Depue, and Douglas (1995) argued that only individuals who have considerable investigative experience should be considered profilers. Given the competing definitions and lack of agreement about who may be designated a profiler, we conducted two quantitative analyses. First, we examined the accuracy of self-labeled profiler/experienced-investigator groups vis-à-vis those of naïve groups, such as students and psychologists (i.e., people who did not have any profiling or investigative experience). Second, we compared the accuracy of self-labeled criminal profilers with all other comparison groups (e.g., students, detectives).

TABLE 1: Summary of the Commonsense and Empirical Rationales Used in Coding Criminal Profiling Articles

<table>
<thead>
<tr>
<th>Commonsense</th>
<th>Empirical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources of knowledge</td>
<td></td>
</tr>
<tr>
<td>Qualitative: based on authority, testimonials, anecdotes, and intuition.</td>
<td>Quantitative: based on evidence from scientific literature.</td>
</tr>
<tr>
<td>Analytical process</td>
<td></td>
</tr>
<tr>
<td>Post hoc ergo propter hoc, availability heuristic, hindsight bias, self-serving bias, and illusory correlates.</td>
<td>Data collection from case histories, correlational studies, surveys, quasi- and experimental studies.</td>
</tr>
<tr>
<td>Integration of evidence</td>
<td></td>
</tr>
<tr>
<td>Simply “tell it like it is” and “what everybody knows” statements, explanation by naming, exceptions prove the rule, and idiographic focus.</td>
<td>Causality is complex, results described in probabilistic terms, expectations that the theory will be revised, acknowledgement of covariation consequences, and nomothetic focus.</td>
</tr>
</tbody>
</table>

Note: Adapted from Gendreau et al. (2002).
STUDY 1: NARRATIVE REVIEW

METHOD

Sample articles. Potential studies for inclusion in the narrative review were located through an electronic search of PsycINFO and Criminal Justice Abstracts databases using the keywords criminal profiling, psychological profiling, and offender profiling. Citations in the reference sections of the obtained articles were also checked for possible inclusion in the review. Eligible articles in the narrative review were peer-reviewed journal articles, book chapters, magazine articles, research reports, and published conference papers; all of which had CP as their primary focus.

Procedure. Studies that met the inclusion criteria were first coded along the following dimensions: article characteristics (e.g., year of publication and location of research), principal author characteristics (e.g., gender, qualifications, discipline, whether a criminal profiler, theoretical approach, and number of CP publications), and author perspective on CP (e.g., opinion of CP, whether CP is an art or science, and beliefs regarding profiler accuracy/usefulness).3

Using the classification framework adapted from Gendreau et al. (2002), sample CP articles were classified as to whether authors used commonsense or empirically based arguments in interpreting phenomena. Articles were evaluated along the following criteria: (a) sources of knowledge (e.g., qualitative vs. quantitative), (b) analytical processes (e.g., hindsight bias vs. experimental), and (c) integration of evidence (e.g., idiographic vs. nomothetic focus). Gendreau et al.’s original classification framework included 25 specific categories of commonsense (k = 14) and empirical (k = 11) rationales, the majority of which were suitable for inclusion in the present review (see Table 1; see also Kimble, 1994; Matlin, 1998; Myers, 1996). One addition and one deletion were made to the classification scheme for the current research purpose. For the narrative review, and in light of the first author’s familiarity with the CP literature, it was deemed appropriate to add post hoc ergo propter hoc reasoning as a supplemental item to the analytical process criterion in the commonsense category.4 The anchoring heuristic was not included as a commonsense rationale for the criminal profiling literature as it could not be operationally defined to facilitate coding. Table 1 contains the final list of rationales used in the narrative review.

Interrater reliability. Agreement of the coding of the commonsense/empirical criteria was assessed by having the fifth author independently code 13 (10%) randomly selected articles. The agreement of coding for commonsense and empirical rationales, measured using Yeaton and Wortman’s (1993) statistic, was 68% for source of knowledge, 85% for analytical processes, and 72% for integration of evidence. Coding agreement was 69% for article characteristics, 90% for principal author characteristics, and 69% for author perspective. Overall coder agreement was 76%.

Confidence intervals and effect size calculations. Confidence intervals (CIs) were defined as the plausible values for the population parameter μ (Wilkinson & The Task Force on Statistical Inference, American Psychological Association, 1999). Where appropriate, we considered substantive interpretations of values, including the upper and lower limits of a CI, and compared these with the mean (see Cumming & Finch, 2005). Of particular concern
was the width of the CIs, which indicate the precision of the estimate of \( \mu \); wider CIs indicate greater uncertainty in this regard. The judgment of the degree of width leading to a conclusion of uncertainty depends on what researchers in the field define as relevant (Smithson, 2003). For the purpose of this analysis, CIs with a width greater than .10 were defined as imprecise, thereby suggesting that replication of the obtained results is required.

The common language (CL) effect size indicator was used to compare the commonsense and empirical ratings of articles based on article characteristics, principal author characteristics, and author perspective (McGraw & Wong, 1992). The CL effect size indicator is a practical statistic that converts an effect size into a probability that a score sampled from Group A will be larger than one sampled from Group B. To illustrate, assume articles with a clinical profiling orientation (Group A) contain an average of 5.2 \((SD = 3.3)\) commonsense rationales while articles with a statistical profiling orientation (Group B) have an average of 1.6 \((SD = 1.4)\). The mean of B is subtracted from A and then divided by the square of the pooled standard deviations. The resultant value of 3.6 is treated as a \( z \) value, and a table of normal curve values is used to interpret its magnitude. In this example, \( z = -1.00 \), which corresponds to a value of .84 or 84% of the area under the normal curve. Such a CL value would indicate that articles with a clinical profiling orientation use more commonsense rationales than articles with a statistical profiling orientation 84% of the time.

**RESULTS**

A total of 130 CP articles were reviewed for the narrative review. This literature contained a large amount of commonsense type sources of knowledge and analytical processes.

Figure 1 shows the percentage of articles that used commonsense (i.e., clear bars) or empirical (i.e., shaded bars) arguments, as well as the 95% CI about each mean percentage. In terms of sources of knowledge, the widths of the CIs were all greater than .10. It can be seen that anecdotal arguments were the most frequently endorsed knowledge source (60%) with a CI of 52%–68%. This was followed by testimonials (45%, CI = 37%–54%), authority (42%, CI = 33%–50%), and use of scientific evidence (42%, CI = 33%–50%). Intuition was the least commonly used source of knowledge (23%, CI = 16%–30%).

Commonsense rationales also dominated among analytical processes, with the widths of all of the CIs greater than .10. Considering the 25 possible comparisons between commonsense versus empirical rationales, 68% of the CIs of the former did not overlap with the CIs of the latter.

As for integration of evidence, the recognition that causality is complex and the expectation that theory will be revised were at least 15% higher than the most frequent commonsense rationales. In total, considering the 25 possible comparisons between empirical versus commonsense rationales, 72% of the CIs for the former were higher and did not overlap with the others. In all but one case (i.e., everybody knows), the CIs were greater than .10.

Based on ratings of each of the criteria in Table 1, commonsense and empirical scores were calculated for each article. The mean frequency for use of commonsense or empirical rationales across all articles was 4.23 \((SD = 3.28, CI = 3.66–4.80)\) and 3.39 \((SD = 2.25, CI = 3.00–3.78)\), respectively. The CL statistic indicated that commonsense arguments were used more than empirical arguments 58% of the time.

We also examined the use of commonsense and empirical rationales as a function of article characteristics, principal author characteristics, and author’s perspective. Reported next
Figure 1: Incidence of Commonsense and Empirical Rationales in Criminal Profiling Articles (n = 130)
are those variables that were reported by at least 20% of the articles and exhibited at least a 65% difference in terms of the CL statistic.

The following article characteristics and principal author characteristics were associated with a greater use of common sense in the range of 72%–84%. These characteristics included articles that were clinical in orientation, published before 1990, emanated from the United States, and written by law enforcement professionals. In contrast, empirical arguments were more frequently used (i.e., range: 69%–93% of the time) by articles that were statistically oriented, published after 2000, appeared in refereed journals, came from outside the United States, and authored by academics.

The following types of author perspective were associated with a greater use of common sense (i.e., range: 72%–81%): expressed a positive opinion of CP, regarded CP as useful or accurate, and labeled profilers as experts. On the other hand, empirical arguments were used more frequently (i.e., range: 69%–77% of the time) in articles that challenged the utility of CP (i.e., articles that had a negative, mixed, or unspecified conclusion about the accuracy of profiles, the expertise of profilers, and the usefulness of profiling).

DISCUSSION

The evidence generated in this study indicated that commonsense rationales have flourished in the CP literature. Even if one were to view the glass as half full in this regard, and focus on the lower limits of the CIs (keeping in mind their distance from the mean), the percentage estimates of the frequency of some of the commonsense rationales for sources of knowledge and analytical processes are of concern if CP is to be considered a scientific domain.

Those who would promote CP as a scientific practice may point to the reality that commonsense rationales are those that are most available to the authors, as the research literature is in its infancy from an empirical perspective, which makes it difficult to substantiate validity arguments with empirical sources. These individuals might also draw attention to the fact that authors appear to have recognized the importance of empirical rationales in attempting to integrate evidence. Admittedly, endorsement of causality is complex and expectation that theory will be revised were noted by authors as components of the process of integrating evidence in slightly more than half of the reviewed articles. Nevertheless, commonsense ways of integrating evidence (e.g., “tell it like it is”, exceptions prove the rule, and idiographic focus) appear, in our opinion, disconcertingly frequently (percentages ranged from 8% to 41%; see Figure 1, Panel C).

The use of commonsense and empirical rationales varied strikingly according to the article characteristics and the author’s perspective. Articles with a clinical orientation, and those that were authored by law enforcement officers, were more likely to contain commonsense rationales and less likely to contain scientific rationales than those that had a statistical orientation and were written by academics, respectively. Similarly, articles in which the authors were more favorable toward profiling contained more commonsense and less empirical rationales than those that were less favorable. It was intriguing to discover a greater proliferation of commonsense rationales in articles coming from the United States (> 70% vs. the United Kingdom and other countries). Regardless of the reasons underlying this overemphasis on common sense and the relative shortage of empirical support in the CP literature, there is a notable incongruity between CP’s lack of empirical foundation and the degree of support for the field, as expressed by the authors of CP articles.
Granted the foregoing, we emphasize the need for replication of the results presented herein. Even though interrater agreement was adequate in our view, given the subjective nature of some of the coding tasks involved, we recognize the need for others to replicate our coding decisions and encourage them to do so. In addition, the widths of the CIs were, in our estimation, of such a magnitude as to limit definitive conclusions. One would be more favorable toward mean results that are more closely bounded by their respective CIs. Finally, we acknowledge that the literature search was not exhaustive. We may have overlooked some studies and, given the rapid expansion of the CP literature, we expect that quite a few additional studies will soon be available for extending and replicating the results of this study. We look forward to further refinement of the conclusions reached in this review.

**STUDY 2: META-ANALYSIS**

**METHOD**

**Sample studies.** Eligible studies for Part 1 of the meta-analysis were those that compared the accuracy of self-labeled profiler/experienced-investigator groups with different comparison groups and met the following criteria:

- The study used an experimental scenario (i.e., contained an independent variable) in which both the crime and the criminal were known to the experimenter.
- The study compared the predictive accuracy of an experimental group consisting of profilers or groups with investigative experience (e.g., detectives) to one or more comparison groups (e.g., students).
- The study reported statistical information regarding the relation between the predictor in question and the dependent variable that could be converted into a common effect size ($r$).

A search of the CP literature revealed 10 potential studies, of which 4 met the above criteria (i.e., Kocsis, 2004; Kocsis, Hayes, & Irwin, 2002; Kocsis, Middledorp, & Try, 2005; Pinizzotto & Finkel, 1990). Among the excluded articles, two contained no usable data (Kocsis, 2003b; Jackson, van den Eshof, & de Kleuver, 1997); one was a review of two other studies included in the sample (Kocsis, 2003a); one used psychics rather than profilers as the experimental group (Reiser, Ludwig, Saxe, & Wagner, 1979); one used a design that involved two groups of participants ascertaining the accuracy of two contrasting profiles when presented with the characteristics of the real criminal (Alison, Smith, & Morgan, 2002); and one did not provide adequate statistical information (Kocsis et al., 2000).

In Part 2 of the meta-analysis, inclusion Criterion 2 changed. That is, the experimental group consisted of only self-labeled profilers and the comparison groups consisted of all other groups (e.g., detectives, students). Using this classification, four studies met the selection criteria (i.e., Kocsis, 2004; Kocsis et al., 2000, 2005; Pinizzotto & Finkel, 1990).5

**Procedure.** Studies were coded along the following three dimensions: (a) **Characteristics**—first author’s qualifications, discipline, and number of publications in the area; theoretical orientation; first author’s gender; research location; publication type; publication year; type of profiling task used; and whether any author is a profiler. (b) **Sample demographics**—composition of comparison groups, size, and mean age of experimental and comparison groups; quantity and type of experimental stimuli. (c) **Outcome data**—mean number of overall accurate predictions was coded in each study (i.e., overall offender).
In addition, Kocsis et al. (2000), Kocsis et al. (2002), Kocsis (2004), and Kocsis et al. (2005) disaggregated their overall offender measure into the following four submeasures: (a) cognitive processes (e.g., motive, whether the offender was comfortable in the area where the offenses took place, whether the offender exhibited remorse), (b) physical characteristics (e.g., gender, age, ethnic background, hair color, facial hair), (c) offense behaviors (e.g., whether the offender took precautions to protect his or her identity, whether the offender removed items from the crime scene), and (d) social history and habits (e.g., marital status, level of education, alcohol consumption, military experience). Each of these submeasures was individually coded.\(^6\)

**Interrater reliability.** Agreement of the coding of the variables in the three dimensions above was assessed for the five studies involved in both meta-analyses by having the fifth author independently code all of the studies. Coding agreement, measured using Yeaton and Wortman’s (1993) statistic, was 98% for study characteristics, 99% for method characteristics, and 98% for outcome data.

**CIs and ES calculations.** As we were primarily concerned with practical rather than statistical significance (Kirk, 1996) in this study, the use of 95% CIs was emphasized. For the purpose of this analysis, CIs with a width greater than .10 were defined as imprecise. As it is extremely rare to find $\mu = 0$ in psychological research (Hunter & Schmidt, 2004; Schmidt, 1996), we chose a more parsimonious interpretation of CIs that contain zero. That is, all CI values, including zero, were accepted as plausible. Finally, we made reference to significance testing only when a CI included zero or when comparing the CIs of two different means. A CI that includes zero indicates that $p > .05$ and when CIs do not overlap, or barely touch, $p < .01$. Conversely, $p > .05$ when the overlap is about half the average margin of error (Cumming & Finch, 2005; Schenker & Gentleman, 2001).

Pearson correlation coefficients ($r$) were calculated for each predictor–criterion relation. When statistics other than $r$ were reported in the study (i.e., $F$, $t$, $p$), the appropriate formulae (see Rosenthal, 1991) were used to convert them to $r$ values. Effect size (ES) magnitudes were assessed by examining the mean $r$ values and their respective 95% CIs for each outcome. The $r$ values were also weighted ($z^+$) by sample size so as to give additional weight to values that came from large samples (Hedges & Olkin, 1985).

The meta-analytic model used here is a fixed effects model; that is, inferences are limited to only those studies included in a meta-analysis and not to those that might have been undiscovered or that will be conducted in the future (with which a random effects model would be appropriate, see Fleiss, 1993; Hedges, 1994). A fixed effects model favors the CP hypothesis in that fixed effects CIs tend to produce CIs that have less uncertainty (i.e., CIs are not as wide) than a random effects model.

Results were also assessed using Rosenthal and Rubin’s (1982) binomial effect size display (BESD). This statistic allows one to examine changes in success rates that are attributable to the predictor variable, assuming a base rate of 50%. It is a useful statistic for demonstrating the practical importance of correlations, particularly those of small magnitude (Hunter & Schmidt, 2004) in that the value of $r$ can be taken at face value. To illustrate in the present context, a value of $r = .30$ would translate into a 30% difference in predictive ability between experimental (e.g., profiler group, $r = .65$) and comparison (e.g., students, $r = .35$) groups.
RESULTS

Reading Table 2 from the left, 30 ESs sampled from 981 participants in predicting overall offender characteristics resulted in a mean \( r = .24 \) (SD = .42). The associated 95% CI was .08–.40. In relation to the four submeasures (\( k = 14, n = 720 \) for each measure), the self-labeled profiler/experienced-investigator groups were slightly better at predicting physical attributes (\( r = .10 \), CI = −.05–.25), but were less accurate than comparison groups in predicting offenders’ cognitive processes (\( r = -.06, CI = -.18–.06 \)) and social habits and history (\( r = -.09, CI = -.25–.07 \)). There was no difference between the groups in predicting offence behaviors (\( r = .00, CI = -.09–.09 \)). The CIs for each of the four submeasures included zero and all were wider than .10.

Results for self-labeled profilers versus comparison groups are displayed in Table 3. Reading from the left, there were 18 ESs sampled from 447 participants in predicting overall offender outcome, resulting in a mean \( r = .32 \) (SD = .45). The associated 95% CI was .10–.54. In predicting cognitive processes, offender physical characteristics and social history and habits, the mean ESs were \( r = .21, .21, \) and .28, respectively, thus indicating a higher mean accuracy rating for the profilers. None of the CIs included zero, but all were wider than .10. Similarly, the obtained value for offense behaviors (\( r = .05 \)) was positive; however, the CI overlapped zero. When \( r \) values were weighted, \( z^\dagger \) for the five predictors was positive, although the CI for offense behaviors contained zero.

DISCUSSION

Predictions by the self-labeled profiler/experienced-investigator groups were not more accurate than those of the comparison groups. First, the point estimates were low. That is, average effect sizes for the former versus the latter were either close to zero or negative in value. Second, CIs were always greater than .10 in width, which should be regarded as a highly tentative result. All but one of the CIs included a substantial range either side of zero. It should be noted that the fixed effects model used in this meta-analysis generates the narrowest estimate of the precision of CI estimates (see Hedges & Vevea, 1998). The only area in which the self-labeled profiler/experienced-investigator groups were more accurate than the comparison groups was in predicting overall offender outcome, as judged by both \( r \) and \( z^\dagger \) (weighted \( r \)). In terms of the BESD, the accuracy rate of profilers was 62% versus 38% for nonprofilers. Self-labeled profilers performed better than the comparison groups across all measures. While this research shows some evidence in favor of profilers, it is tentative. The estimates of

<table>
<thead>
<tr>
<th>Outcome (k)</th>
<th>n</th>
<th>Mean (SD)</th>
<th>95% CI (r)</th>
<th>z*</th>
<th>95% CI (z*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall offender (30)</td>
<td>981</td>
<td>.24 (.42)</td>
<td>.08–.40</td>
<td>.08</td>
<td>.01–.15</td>
</tr>
<tr>
<td>Cognitive processes</td>
<td>720</td>
<td>.06 (.20)</td>
<td>-.18–.06</td>
<td>-.07</td>
<td>-.15–.01</td>
</tr>
<tr>
<td>Physical attributes</td>
<td>720</td>
<td>.10 (.26)</td>
<td>-.05–.25</td>
<td>.13</td>
<td>.05–.21</td>
</tr>
<tr>
<td>Offence behaviors</td>
<td>720</td>
<td>.00 (.16)</td>
<td>-.09–.09</td>
<td>.02</td>
<td>-.06–.10</td>
</tr>
<tr>
<td>History/Habits</td>
<td>720</td>
<td>-.09 (.28)</td>
<td>-.25–.07</td>
<td>-.10</td>
<td>-.18–.02</td>
</tr>
</tbody>
</table>

Note: Mean (SD) = mean Pearson correlation coefficient for each predictor with standard deviations in parentheses; 95% CI = confidence intervals about \( r \); \( z^\dagger = [(z) \times (n – 3)] \div (N – 3) \) per predictor, where \( n = \) number of participants per effect size and \( N = \) number of participants per predictor; 95% CI\( _z \) = confidence interval about \( z^\dagger \).

*Indicates that the sample size for each of the subscales is identical as they were taken from three studies written by the same author (all groups in these studies were tested along these predictors).
the effect size were imprecise as the widths of the CIs were extreme (e.g., two to five times the acceptable limit of .10) and one of the CIs included zero. Moreover, in the instances in which the CIs did not include zero, the addition of one more study with a large sample with an effect size of zero would produce more CIs that include zero. Also of note, upon weighting, the magnitude of the mean effect sizes declined in three predictor categories. Even if one were to take the most optimistic stance in favor of profilers and choose the predictive category in which the estimate was the most robust (i.e., overall offender, \( r^+ = .33, CI = .23-.43 \)), the BESD statistic indicates that the profilers’ success rate was 66.5% versus 33.5%. This demonstrates better performance, but not necessarily expert performance.

We contend that, in any field, an “expert” should decisively outperform nonexperts (i.e., lay persons). The practical problem with designating profilers or experienced individuals as experts lies in the fact that their services are requested based on their presumed expertise, which increases the likelihood of their having considerable impact on the direction of a given investigation. Very little is known about the effects of profiling on criminal investigations, particularly on police decision making. Until sound scientific evidence of profilers’ predictive validity becomes available, it would be prudent to assume that CP is as likely to be hazardous (i.e., have unacceptably high false-alarm rates) as it is to be helpful to the criminal investigation process. We suggest a BESD of at least \( r = .60 \) as a possible benchmark, given the legal implications and substantial costs involved in following an investigative lead based on CP hypotheses.

There are two major limitations to the database. First, while credit must be given to one Australian researcher (i.e., Richard Kocsis) for collecting data that are suitable for a meta-analysis, that research has nevertheless been criticized for a range of methodological and conceptual limitations such as the use of multiple choice questionnaires, variation in the length of time given to participants to complete the experiment, and aggregating data to favor the experimental group (Bennell, Jones, Taylor, & Snook, 2006). Second, although the number of participants in the experienced-investigator group (\( n = 74 \)) may be considered sizeable, the self-labeled profiler group was small (\( n = 19 \)). The predictive abilities of a larger number of profilers must be empirically tested before they can seriously be considered as a useful resource for apprehending criminals. Regrettably, there is some indication that profilers are very reluctant to participate in experimental CP studies. For example, Kocsis et al. (2000) reported that only 5 of the more than 40 active profilers whom he contacted agreed to participate in his research.
CONCLUSION

The evidence generated from this research confirms the perceptions of those who have concluded that the CP field relies on weak standards of proof and that profilers do not decisively outperform other groups when predicting the characteristics of an unknown criminal (e.g., Alison, Bennell, Mokros, & Ormerod, 2002; Muller, 2000). Based on the results of the narrative review and meta-analytic reviews presented herein, profiling appears at this juncture to be an extraneous and redundant technique for use in criminal investigations. CP will persist as a pseudoscientific technique until such time as empirical and reproducible studies are conducted on the abilities of large groups of active profilers to predict, with more precision and greater magnitude, the characteristics of offenders.

NOTES

1. Although the scope of CP practice now goes beyond this original definition to include advice on interview strategies, media strategies, prioritizing resources, statement analysis, and so on, we believe that predicting offender characteristics remains the primary goal of CP because all of this additional advice is dependent upon the type of person that the profiler believes committed the crime.

2. Philosophers of science (e.g., McCosh, 1996) defined the reverse of bad common sense—good common sense—to be an empirical or scientific approach for assessing the truth of a matter, because this process is based on careful observation of phenomena and inductive proofs. Philosophers and social scientists have concluded that good common sense is in short supply (e.g., McCosh, 1996; Nisbett & Ross, 1980).

3. A copy of the coding guide is available from Brent Snook.

4. Post hoc ergo propter hoc is a reasoning fallacy in which it is assumed that because one event (e.g., profiling advice is provided to a police investigation) occurs before another event (e.g., the crime is solved), the first event must have caused the second event to occur (see Gilovich, 1991).

5. Kocsis et al. (2000) was not used in the first analysis because the statistical information regarding each group’s accuracy levels was not provided (i.e., standard deviations for each group were not provided, and p values were not provided for comparisons between each group). However, that study was included in our second analysis because it provided an omnibus measure of accuracy for both the self-labeled profiler and nonprofiler groups, as well as measures of central tendency. Kocsis et al. (2002) was not included in the second analysis because it did not use any self-labeled profiler groups.

6. A copy of the coding guide is available from Brent Snook.

REFERENCES

References marked with one asterisk indicate studies included in the narrative review. References marked with two asterisks indicate studies included in both the narrative and the meta-analytic reviews.


*Fleiss, J. L. (1993). The statistical basis of meta-analysis. Statistical Methods in Medical Research, 2, 121-145 *


---

*Criminal Justice and Behavior (CJB)* is pleased to offer the opportunity to earn continuing education credits online for this article. Please follow the access instructions below. The cost is $9 per unit, and a credit card number is required for payment.

Access for AACFP members:
- Go to the AACFP homepage (www.aa4cfp.org) and log in to the site using your email address
- Click on the Publications heading in the sidebar on the left-hand side of the page or on the Publications icon at the bottom of the page
- Click on the CJB cover image located at the bottom of the page. You will be prompted for your user name and password to log in to the journal
- Once you have logged in to CJB, select the appropriate issue and article for which you wish to take the test, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

Access for institutional subscribers:
- Log in through your library or university portal to access CJB
- Select the appropriate issue and article, then click on the link in the article to be directed to the test (*note: only marked articles are eligible for continuing education credits*)

If you are not a member of the AACFP and your institution does not subscribe to *CJB*, individual articles (and corresponding test links within the articles) are available on a pay-per-view basis by visiting http://cjb.sagepub.com. However, membership in the AACFP provides you with unlimited access to *CJB* and many additional membership benefits! Learn more about becoming a member at www.aa4cfp.org.

Should you have difficulty accessing this article’s test or the website, please contact Eddie Santos for further assistance at eddie.santos@sagepub.com or 805-410-7528.
In the April 2007 issue of *Criminal Justice and Behavior*, in the article titled “Taking Stock of Criminal Profiling: A Narrative Review and Meta-analysis” by Brent Snook, Joseph Eastwood, Paul Gendreau, Claire Goggin, and Richard M. Cullen, Table 2 should have appeared as follows:

<table>
<thead>
<tr>
<th>Outcome (k)</th>
<th>n</th>
<th>Mean (SD)</th>
<th>95% CI</th>
<th>z+</th>
<th>95% CIz+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall offender (30)</td>
<td>981</td>
<td>.24 (.42)</td>
<td>.08–.40</td>
<td>.08</td>
<td>.01–.15</td>
</tr>
<tr>
<td>Cognitive processes (14)</td>
<td>720a</td>
<td>−.06 (.20)</td>
<td>−.18–.06</td>
<td>−.07</td>
<td>−.15–.01</td>
</tr>
<tr>
<td>Physical attributes (14)</td>
<td>720a</td>
<td>.10 (.26)</td>
<td>−.05–.25</td>
<td>.13</td>
<td>.05–.21</td>
</tr>
<tr>
<td>Offence behaviors (14)</td>
<td>720a</td>
<td>.00 (.16)</td>
<td>−.09–.09</td>
<td>.02</td>
<td>−.06–.10</td>
</tr>
<tr>
<td>History/Habits (14)</td>
<td>720a</td>
<td>−.09 (.28)</td>
<td>−.25–.07</td>
<td>−.10</td>
<td>−.18–.02</td>
</tr>
</tbody>
</table>

Note: Mean (SD) = mean Pearson correlation coefficient for each predictor with standard deviations in parentheses; 95% CI = confidence intervals about r; $z^+ = \frac{[z] \times (n - 3)}{(N - 3)}$ per predictor, where $n =$ number of participants per effect size and $N =$ number of participants per predictor; 95% CI$_{z^+} =$ confidence interval about $z^+$.

*a*Indicates that the sample size for each of the subscales is identical as they were taken from three studies written by the same author (all groups in these studies were tested along these predictors).