This study evaluated the effect on burglary arrest rates when using statistically derived behavioral profiles for burglary offenses and offenders in active police investigations. To do this, an experiment was conducted where one police agency that used the profiles was compared with three matched police agencies that did not. Burglary arrest rates were studied 4 years before and 1 year after the profile was implemented. Results show that the arrest rates for the treated agency increased by 3 times as compared with the control agencies. The interaction effect between treatment/control agency and pretest/posttest arrest rates was significant, showing that the experimental intervention had an effect, after controlling for pre-existing differences between the agencies. These findings on the utility of offender profiling, the first to be derived from an experiment conducted in active police investigations, suggest that the statistically based behavioral profiles could be a useful tool in increasing arrest rates for police.

Keywords: offender profiling; behavioral analysis; experimental evaluation; policing; burglary

With the exception of the supernatural, unidentified flying objects (UFOs), and other types of paranormal activity, few topics have received such attention and credibility as offender profiling (OP), without a single scientific evaluation conducted to test the true validity and utility of the subject or practice out in the field (Dowden, Bennell, & Bloomfield, 2007). However, unlike the belief in aliens, monsters, and extra-sensory perception, which have generally been limited to certain sub-cultures and not accepted in mainstream society, OP has been embraced and utilized by a substantial proportion of well-respected law enforcement agencies and practitioners across the globe (Dowden et al., 2007; Jackson, van Koppen, & Herbrink, 1993; Kocsis, 2007).

This is not to say that there has been no academic coverage of the practice of OP. To the contrary, there is a wealth of literature available on the topic, though most publications generally discuss what profiling is, when it is useful, and other issues relating to the concepts and assumptions underlying its use in policing (Dowden et al., 2007). In the 30 years following its emergence as a viable crime-fighting tool, OP has become an integral part of law enforcement, and its validity and utility have been widely recognized and endorsed by law enforcement agencies around the world.
after the first article on OP was published in 1976, 129 peer-reviewed articles have been published on the topic, of which only 18 used statistics to evaluate profiling methods or hypotheses in some way (Dowden et al., 2007). In fact, over half of these published studies \((n = 75)\) did not include any form of statistical analysis, and not one study has attempted to assess the true usefulness of profiling for law enforcement by experimentally evaluating the profiles as they are applied in the field (Dowden et al., 2007; Homant & Kennedy, 1998; Snook, Taylor, Gendreau, & Bennell, 2009).

Consequently, the present research aims to fill this gap by conducting an empirical evaluation of established statistical behavioral profiles for burglary as they are used by law enforcement in the United States. Using a multi-agency experimental and multiple evaluative measures, this study will scientifically evaluate, for the first time, the true utility and impact of behavioral profiling as it is used out in the field.

**SCIENTIFICALLY EVALUATING OP**

Although different definitions of OP have been proposed over the years (Dowden et al., 2007), in general, OP is defined as a technique for identifying the major personality, behavioral, and demographic characteristics of an offender based on an analysis of the crimes he or she has committed (Douglas, Ressler, Burgess, & Hartman, 1986). Many law enforcement agencies across the world already request profiling services on difficult cases, either from an FBI profiler, trained law enforcement agent, profiler in private practice, or an academic consultant trained in psychology, profiling, or behavioral analysis (Trager & Brewster, 2001). While profiling may be used to prioritize and focus law enforcement’s investigation of offenders, a profile does not constitute trial evidence or justify an arrest of an offender; probable cause is still needed to legally arrest a suspect in any given case (State v. Stevens, 2001). Also, unlike many other fields where evaluation research is being conducted, OP is a field in which the underlying mechanisms and theoretical framework are still being developed and understood. In fact, determining standardized criteria to evaluate the success of OP is something still hotly debated in the field (for details, see the discussion in Bennell, Jones, Taylor, & Snook, 2006, p. 346; Kocsis, 2006, p. 462).

In short, some researchers believe that measuring a profiler’s ability to accurately predict traits of an offender (even if it does not help solve the case) is the best measure of OP success, while others think helping police to identify and ultimately arrest the perpetrator, or even simply aid police in an investigation (even if the case remains unsolved) are the most accurate measures of profiling’s success (Kocsis, 2003). These differences in definitions (even in terms of profiling accuracy vs. utility) have a major impact on any evaluation of OP. Furthermore, profiles may be created for a variety of crimes using different methods and processes, which could also lead to differing levels of success among the various profiles. For example, some profiles are created using rigorous statistical methods and quantitative data, much like the burglary profiles evaluated in the present study, while others are created by profilers who believe that OP is “more of an art than a science” and rely instead upon their “experience, intuition, and educated guesswork” to develop a profile (Geberth, 1990, p. 492; Pinizzotto, 1984, p. 33).

Clearly, these issues tend to make the evaluation of OP much more difficult; the results of one study, using profiles created in a specific manner and evaluated using a particular criterion of success, may differ from the results of a study testing different types of profiles.
or using a different measure of success. Consequently, most studies on profiling are conducted to test the mechanisms and processes underlying OP (see, for example, Canter, Bennell, Alison, & Reddy, 2003; Canter & Fritzon, 1998; Kocsis, Irwin, & Hayes, 1998; Salfati & Canter, 1999) but do not validate the accuracy or usefulness of the profiles after they are implemented.

Still, a few studies have aimed to evaluate the success of OP, though these studies often use indirect, and sometimes very weak, analytical methods and evaluation designs. For instance, several studies have evaluated profiling’s utility by asking police officers how “beneficial” and “helpful” they felt the profiles were during their investigations (see Pinizzotto, 1984). Other studies determined accuracy by comparing the “abilities” of professional profilers to various non-profiler groups in correctly predicting or selecting an offender from a line-up using details of a realistic but fictitious case (see, for example, Kocsis, Irwin, Hayes, & Nunn, 2000; Pinizzotto & Finkel, 1990).

While there are certainly benefits to each of the evaluation methods commonly used in OP, the most accurate and reliable means of evaluating any practice or program is achieved by utilizing a scientific, experimental approach. However, before the methods and details of the present study are described, a more descriptive overview of the previous evaluation approaches is presented.

**PRIOR EVALUATIONS OF OP**

**CONSUMER SATISFACTION SURVEYS**

The most common, but most subjective, method of evaluating the success of OP has been through the use of “consumer satisfaction surveys” given to law enforcement agencies that used profiles in past investigations. These studies gauge how satisfied police felt after using profiles in their investigations, with the assumption that valid and useful profiles would translate into greater levels of police satisfaction, and that invalid and ineffective results would lead to dissatisfaction and police no longer wanting to use profiles in their investigations (Kocsis & Palermo, 2007).

Results of consumer satisfaction studies indicate that police generally believe that profiles benefit their investigations in some way, though the exact nature and extent of the benefit remains unclear (see, for example, Copson, 1995; Jackson et al., 1993; Pinizzotto, 1984; Snook, Taylor, & Bennell, 2007; Trager & Brewster, 2001). For instance, in one of the best-known studies of this type, all 192 profiles created by the FBI from 1971 to 1981 were evaluated by the police who used the profiles. Seventy-seven percent of police departments reported that the FBI profiles significantly helped their investigations, although fewer than half of all cases involving a profile were ultimately solved ($n = 88$; Pinizzotto, 1984). Also, the survey showed that while 17% of police stated that the profiles directly aided in the identification of a suspect, another 17% felt that the profiles “were not useful at all” (Pinizzotto, 1984).

In the United Kingdom, Copson (1995) found that 83% of police felt the profiles they received were operationally useful and added that the profiles aided their understanding of the case (61%), opened new leads or lines of inquiry (16%), or actually helped to solve the case (14%). This new information on why profiles are deemed useful to police provided the first insight into how OP plays a role in actual cases, and more importantly, influences police decision-making in investigations. A more recent study in the United States found
that 63% of surveyed officers felt that profiles assisted in their investigations, and 38% stated that profiles directly assisted in the identification of a suspect (Trager & Brewster, 2001). However, a quarter of police said that the profiles had actually hindered their investigation in some way (Trager & Brewster, 2001).

The most recent study of this type, conducted in 2007 by Snook, Taylor, and Bennell using 51 police officers in major crime divisions across Canada, found that about 94% of the officers felt that profilers help solve cases, 92% thought that profiles are valuable to criminal investigations, and 40% reported that a profile directly helped solve a case. Furthermore, 65% of the officers reported that the profiles made a significant positive contribution to the investigation, with 14% stating the profile helped focus an investigation, better understand the suspect (10%), provide interrogation strategies (3%), or confirm the officer’s feelings on the case (32%; Snook, Taylor, & Bennell, 2007). About 52% of officers felt the profiles opened new leads in cases, and 74% felt the profiler was accurate in predicting the characteristics of the offender (Snook, Taylor, & Bennell, 2007).

While these consumer satisfaction studies on police perception of profiles are generally positive across nations and time, there is still a methodological flaw in using consumer satisfaction to infer the accuracy of OP, as police satisfaction with profiles does not necessarily provide proof of the true accuracy of OP (Jeffers, 1991; Kocsis & Palermo, 2007; State v. Stevens, 2001; U.S. House of Representatives, 1990). Alison, Smith, and Morgan (2003) have helped demonstrate this; the researchers provided police with two separate profiles and asked each group to rate each profile’s level of accuracy at face value. Although only one of the profiles was genuine and the other was fake, both profiles received nearly equal assessments of perceived accuracy by both police groups (Alison et al., 2003). Two additional studies found a positive linear relationship between an individual’s initial belief in profiling and the subsequent perceived accuracy of a profile (Kocsis & Hayes, 2004; Kocsis & Heller, 2004). A follow-up study showed that an individual’s feelings about the accuracy of OP may also be a function of the professional standing of the profiler and an individual’s own beliefs about profiling (Kocsis & Middledorp, 2004). In other words, the more someone believes in the validity of profiling or is impressed by the credentials of the profiler, the more correct and useful the profiles may appear to be (Kocsis & Middledorp, 2004). This concern was voiced by Campbell (1976), who stated that police may be more seduced by the credentials of the profiler than the profiles themselves (Petherick, 2006). Finally, it has been suggested that the police may view OP as more accurate due to the “Barnum Effect,” named after P. T. Barnum’s circus trick where intentionally ambiguous details are technically accurate, as they are so ambiguous they fit almost every possible case (see Snook, Cullen, Bennell, Taylor, and Gendreau, 2008, for a detailed discussion of these and other limitations of profiling).

COMPARATIVE TESTS OF PROFILER ABILITIES

The previously described findings and concerns have led some researchers to examine the abilities of profiling from yet another approach, this time comparing the actual and perceived accuracy of profilers to non-profiler groups. The first study of this kind was conducted by Pinizzotto and Finkel (1990), where participants from various backgrounds (psychologists, detectives, college students, and FBI profilers) were asked to predict the characteristics of an offender in murder and rape cases in a written “profile” and select the
offender from a multiple-choice “line-up” of varied suspect trait descriptions. Accuracy was determined by each participant’s ability to correctly predict or identify the traits of the true offenders for each offense. Results showed that the FBI profilers were by far the most accurate in choosing the true offender from a line-up of suspects, as they were correct 100% of the time. However, the detectives were most accurate in predicting traits of offenders in the homicide case, and had the most correct responses when the rape and homicide cases were combined (Pinizzotto & Finkel, 1990).

Following this study, a series of related projects were undertaken by Richard Kocsis and his colleagues to further evaluate the accuracy of profilers when compared with groups possessing what FBI profilers have stated are the most “useful” traits of a profiler (Hazelwood, Ressler, Depue, & Douglas, 1995), such as field experience (police), understanding human behavior (psychologists), aptitude for logical reasoning (undergraduate science majors), and pure intuition (psychics; Kocsis, 2004, 2007; Kocsis, Hayes, & Irwin, 2002; Kocsis et al., 2000; Kocsis & Middledorp, 2004).

Results of these studies suggest that profilers generally surpass all other groups when it comes to the accuracy in predicting offenders using crime scene information, although the profilers also demonstrated the highest degree of statistical variance among all the groups in the studies (Kocsis, 2006). Furthermore, when the level of absolute rather than relative accuracy is considered, results of these studies seem quite different. For instance, Kocsis and colleagues (2000) found that profilers were the most accurate when compared with psychologists, students, police, and psychics. However, the profilers in the study averaged a 46% total accuracy rate, compared with a 40% average for the students and a 38% average accuracy rate for the psychics in the sample (Kocsis et al., 2000). In other words, being the “most accurate” profiler group does not necessarily indicate a high level of profiling accuracy as these studies only evaluate the relative abilities of various groups of people, and no direct measurement of profiling’s actual impact on police investigations ever takes place (Bennell et al., 2006). A meta-analysis assessing the Kocsis study series was conducted by Brent Snook and colleagues (2007), which found that while profilers were found to slightly outperform the non-profilers, the confidence intervals surrounding the point estimates for the profilers versus non-profilers accuracy rates were so wide that little information on the comparative accuracy of these groups could actually be determined.

Still, even if a measure of absolute accuracy of profilers were available, it would not be possible to determine how useful the profiles might be when applied for their true purpose: helping police identify offenders in unsolved investigations. This issue was brought to light in the Coals to Newcastle project carried out by the British Home Office in the 1990s (Gudjonsson & Copson, 1997). In this study, 184 profiled crimes were retrospectively analyzed to determine how many times the profiles successfully led to the identification of a criminal. While there is no information regarding exactly how profiles led to an arrest for the cases solved in this study, results indicated that the predictions made by profilers regarding an offender were accurate approximately 66% of the time; however, the profiles directly led to an arrest in just 5 of the 184 cases (Gudjonsson & Copson, 1997). In other words, even though the profiles had a 66% accuracy rate in terms of correctly predicting the actual traits of offenders, the retrospective analysis conducted in the Coals to Newcastle study showed that there was only a 2.7% success rate when the profiles were actually applied in the field (Gudjonsson & Copson, 1997). To our knowledge, no additional studies on the direct impact and utility of profiles in the field have ever been conducted.
Evaluating the Utility of Profiling in the Field

While this overview of past evaluations of OP is not exhaustive, it does demonstrate that despite numerous attempts to evaluate the accuracy and utility of profiling, the field remains relatively close to where it started in terms of scientifically assessing how profiles impact “real world” outcomes and perform in actual active police investigations. In fact, in their article titled, “On the Need for Scientific Experimentation in the Criminal Profiling Field: A Reply to Dern and Colleagues,” Snook et al. (2009) stated that “we applaud any attempts to train profilers on best practices and to evaluate those practices . . . we need evidence that (criminal profiling) works” (p. 1092). The authors also stated,

Although Dern and colleagues (2009) are undoubtedly correct to suggest that “scientifically evaluating the actual application of case-analytical methods in practical police work will always be a difficult task” (p. 1089) . . . we hope this will not continue to be the case with (criminal profiling). Advances in social science mean that complex processes can be assessed. (p. 1093)

Snook and colleagues were pioneers on the scientific evaluation of OP, as they strongly encouraged both academics and practitioners to conduct and publish experimental evaluations on the effectiveness of OP in the field. Unfortunately, as there is a great deal of difficulty in implementing the research design needed to properly conduct such an evaluation, the experimental research that they called for has yet to take place.

However, the aim of the present research is to follow Snook et al.’s (2009) recommendation by conducting an experimental evaluation on the effect of offender profiles when applied in active police investigations. Specifically, we will test new evidence-based profiles for burglary in the United States by comparing the burglary arrest rates for a police department using the profiles to the arrest rates of matched agencies using only their standard investigative tools. Although this research may have little direct relevance to the true accuracy of profiling or the utility of profiles created using other methods and processes, it will establish a baseline measure for the investigative utility of OP and be the first study to scientifically evaluate a profile utilized in police investigations using an experimental design.

Method

While there are several methods of evaluating the effect of a given treatment on an outcome, it is well-established that the most reliable and valid method of doing so is through the use of an experimental design (Farrington & Welsh, 2006). Experiments come in two types: randomized controlled trials (RCTs) and non-randomized experiments. Both types attempt to model what would have happened to an experimental group if the treatment had not been applied (the counterfactual), to scientifically determine whether a program or treatment had an effect on a given outcome. Although random assignment of participants is essential to establishing definitive causality within an evaluation, as randomization eliminates most threats to internal validity (Lösel, 2008), non-randomized experiments rely upon many key design features and statistical analyses to test for effect size and may also yield strong causal inferences. Furthermore, as RCTs are often difficult or impossible to implement in social science settings because of the ethical and practical challenges associated
with randomization, non-randomized experimental designs are far more common in social science research (Farrington & Welsh, 2006; Lösel, 2008).

The most common non-randomized experiment is the nonequivalent group design, where pre- and posttest measures of an outcome are determined for both the treatment and control groups. Oftentimes, this design will include (a) additional covariates to ensure sufficient matching between the conditions, (b) multiple pre- and post-treatment measures to get a more accurate effect size measure, (c) multiple treatment or control groups for better comparisons, or (d) some combination of all three of these. While the lack of randomization prevents causality from being definitively established in the nonequivalent group design, the inclusion of these additional measures adds significant credibility to the results by addressing threats to internal validity.

Although random assignment was not feasible in this study, due to difficulties associated with securing multi-agency access and cooperation, travel costs, and time limitations, a nonequivalent group experimental design was used to eliminate many confounds and mimic randomization as far as possible using statistical controls and analyses. Specifically, multiple measurements of the outcome measure (burglary arrest rate) were collected from participating agencies for the periods before and after the implementation of the experimental treatment (OP). Furthermore, several control variables were collected from each participating agency to deal with selection effects and statistically equalize the treatment and control conditions in future analyses. Each of these measures, and their operationalization, are discussed in the sections to follow.

PRE- AND POSTTEST MEASURES

To measure the success of the statistical behavioral profiles for burglary when applied in the field, the number of burglaries cleared by arrest in each police jurisdiction in the periods before and after the burglary profiles was used as the outcome measure. This item was operationalized using semi-annual (i.e., every 6 months) burglary arrest rates for each agency participating in the study for both the pre- and post-test evaluations.

To calculate the burglary arrest rate for each department, the burglary incidence and arrest figures were collected for each participating agency; the total number of cases where an offender was arrested was divided by the total number of offenses that occurred in the same time frame. It is important to note that the burglary arrest rate, rather than the burglary clearance rate, was utilized as the before and after outcome measure in this experiment. This is because the arrest rate is calculated using only the number of offenders identified and apprehended for offenses (with the requisite legal evidence and probable cause needed for police to arrest the offender), while the clearance rate includes other “exceptional” means of closing cases, such as if the victim recants the charges, a lack of evidence to pursue the charges, or if the offender dies and the state drops the charges. Although the term arrest rate is sometimes used interchangeably with clearance rate, the small but important technical difference between these two figures could have an impact on the outcome. Therefore, the more stringent and precise terminology (i.e., cases cleared only by arrest) is used in this research to better measure the profiles’ impact on police investigations.

All pretest measures, including current burglary arrest rates, burglary incidence rates, and prior arrest rates, were collected semi-annually for each participating agency from January 1, 2008, until December 31, 2011, when the experimental period began. The same
posttest measures were collected for a 1-year follow-up period, from January 1, 2012, to December 31, 2012, for each department participating in the experiment.

TREATMENT AND CONTROL GROUPS

The treatment and control groups in this research are four police departments in the state of Florida, selected based on their similarity on several key factors. The criteria we used to match the treatment and control agencies include the similarity of the agencies such as the agency location (e.g., touristic or coastal), size, current burglary arrest rate, and crime rate similarity of the jurisdictions, in terms of the burglary incidence rate.

Using these criteria, the four police departments were selected based on closely matching features from the full population of police agencies in the state of Florida, though all relevant information was collected for each department year between 2008 and 2012 to statistically control for pretest similarities and differences in analyses. While the ideal experiment would utilize randomization to select participating agencies and determine which agency receives the treatment and which are controls, as is often the case with applied research, the opportunity for randomization was not possible in this project. In this case, it would be difficult, if not impossible, to randomly assign enough police departments to get the benefits of randomization in equaling experimental units on all measured and unmeasured variables. Therefore, as one agency was the most time- and cost-efficient location to conduct the training and evaluation needed in this evaluation, it was selected as the agency to receive the profiling treatment while the others were designated as control conditions. A summary of characteristics of each participating agency, renamed for anonymity, is shown in Table 1.

THE EXPERIMENTAL TREATMENT

The experimental treatment in this study was the implementation of the Statistical Patterns of Offending Typology (SPOT) for burglary developed by Fox and Farrington (2012). In short, Fox and Farrington utilized a Latent Class Analysis (LCA) to assess the presence of statistically related patterns of offending behaviors and offender traits within 405 randomly selected, solved burglaries that took place in Florida between 2008 and 2009. The offense reports and arrest records for each burglary were collected and coded for analysis, as these provided critical information on the behaviors observed at the crime scenes, and were used to statistically classify the burglaries into different styles of offenses. The criminal history records and demographic features for each offender responsible for the solved burglaries were also collected, to classify burglars into sub-types that were shown to be linked to the burglary offense styles. Specifically, four behavioral offending patterns

---

**TABLE 1: Test Characteristics of Participating Police Departments**

<table>
<thead>
<tr>
<th>Department</th>
<th>Location</th>
<th>Officers to citizens</th>
<th>Burglary rate</th>
<th>Burglary arrest rate (%)</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Central coast</td>
<td>1:186</td>
<td>131</td>
<td>11.9</td>
<td>Treatment</td>
</tr>
<tr>
<td>B</td>
<td>Central</td>
<td>1:313</td>
<td>148</td>
<td>12.8</td>
<td>Control</td>
</tr>
<tr>
<td>C</td>
<td>Central coast</td>
<td>1:239</td>
<td>102</td>
<td>15.9</td>
<td>Control</td>
</tr>
<tr>
<td>D</td>
<td>Gulf coast</td>
<td>1:188</td>
<td>94</td>
<td>18.7</td>
<td>Control</td>
</tr>
</tbody>
</table>

*Note. Characteristics calculated using the average values for the 4 years preceding the start of the experiment.*
were identified in the LCA, and were labeled as follows: (a) organized, (b) disorganized, (c) opportunistic, and (d) interpersonal style offenses. Each of the offense styles are committed by burglars with a unique set of traits and criminal histories. And in each case, the offending history and personality features of the criminal generally reflect the key behavioral features witnessed at the crime scene.

The SPOT behavioral analysis technique was selected for evaluation in the present study as it is one of the few investigative typologies, or “profiles,” developed utilizing a rigorous and objective statistical analysis of crimes committed in the United States. Furthermore, as the profiles were created for a high volume and difficult-to-solve crime such as burglary, the opportunity for the behavioral profile to be frequently and widely utilized by police is substantially increased, as is the ability to measure improvement in the burglary clearance rates, which are notoriously low across the United States (Vaughn, DeLisi, Beaver, & Howard, 2008). Additional information on the SPOT method, how it was developed, opportunities for use, and the limitations associated with the typology are discussed in detail in Fox and Farrington (2012). A chart illustrating the details of the burglary profiles is provided in Table 2.

Using the offense–offender links identified for each of the four behavioral profiles—the organized, disorganized, opportunistic, and interpersonal style burglaries—department-wide training sessions, “how to guides,” and one-on-one field training sessions were conducted over a 3-week time frame for all officers and property crime detectives in the treatment agency. This training consisted of a seminar on the purpose and uses of the burglary profiles, an in-depth explanation and description of the burglary profiles, the intended

<table>
<thead>
<tr>
<th>Offense style</th>
<th>Offense description</th>
<th>Offender description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunistic</td>
<td>Unlawful entry—Entry left open</td>
<td>Young offenders Adolescent onset</td>
</tr>
<tr>
<td></td>
<td>No preparation or tools</td>
<td>Short criminal career Low offending frequency Do not know victim</td>
</tr>
<tr>
<td></td>
<td>Unoccupied residence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low value items stolen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Little evidence left behind</td>
<td></td>
</tr>
<tr>
<td>Organized</td>
<td>Clean but forced entry Tools brought to scene No evidence left behind High value items stolen that often require fence/network</td>
<td>Older offenders Adolescent onset High offending frequency Limited versatility—Prior arrests for theft/burglary</td>
</tr>
<tr>
<td>Disorganized</td>
<td>Forced entry Scene left in disarray Tools and/or evidence left</td>
<td>Young offenders Early onset Long criminal career High offending frequency</td>
</tr>
<tr>
<td></td>
<td>Low value or no items stolen</td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td>Occupied residence Target is victim—Not objects Attempted, threatened, or committed violence at scene Personal items stolen</td>
<td>Adult aged Late criminal onset Solo offender Have a car Single/not cohabiting</td>
</tr>
</tbody>
</table>
use of the profiles in investigations, and the limitations of the profiles so no expectations or legal boundaries were violated in this experiment. Field training on how to apply the profiles was also provided to the detectives so they would know what elements to look for at each crime scene to correctly identify a burglary as a specific style of offense. Specific examples of each burglary profile (i.e., an organized burglary scene, disorganized scene, etc.) were also illustrated to the detectives so they could recognize and classify similar scenes by identifying the key characteristics, and determine the most common traits found among offenders committing each style of crime.

Additional training was given to the sergeant of the property crimes unit, the police chief, the deputy chief, and the captains in the treatment group to oversee the continued use of the burglary profiles. The crime analysts were briefed on how to apply the profiles to their electronic databases to generate suspects and limit potential leads, once detectives classified a crime scene into a specific offense style. For example, the burglary profiles suggest that disorganized burglaries are statistically more likely to be committed by young adult offenders who began their criminal careers very early in life (often in childhood), commit a high rate of offenses ranging from vandalism and drug offenses to more serious crimes like violence and burglary, and likely do not have any connection to the victim. Using the details of the profile, the crime analysts would search the database of past offenders who fit these criteria and draw up a list of potential suspects for the detectives to consider pursuing for additional investigation. As always, additional evidence constituting probable cause must be established in each case for an arrest warrant from a judge to be legally issued against a suspect. Although the profiles are not considered legal evidence, and therefore cannot lead to the arrest of a suspect without sufficient hard evidence to constitute probable cause as required in all other arrests, the profiles are meant to benefit the police by prioritizing the most probable types of suspects for a certain style of offense, as shown in prior statistical analyses of offender traits and offending behaviors. That way, police will spend their time first investigating the offenders who are statistically most likely to have committed a certain burglary style and will not waste precious time looking into offenders who are statistically unlikely to have committed the offense.

After the SPOT training was completed on January 27, 2012, the experimental period began. Burglary arrest rates were then collected from each participating agency for approximately 1 year, until December 31, 2012.

ANALYTICAL APPROACH

To evaluate the impact of the experimental treatment (i.e., the burglary profiles) on the burglary arrest rates for the treatment and control conditions while controlling for other factors, an ANCOVA and conditional multivariate regression were conducted in this study. Both of these methods are general linear models used to assess the effect of a treatment on an outcome measure. While these methods are very similar, the ANCOVA is able to statistically adjust the means of the main dependent variable (in this case, burglary arrest rates) to what they would be if all groups were equal on the covariates. By controlling for all pre-existing differences between groups included in the ANCOVA, the analysis significantly reduces within-group error variances, limits threats to internal validity, increases statistical power (i.e., the ability to find a significant difference between groups when one exists), and has shown to be a highly effective tool in removing unaccounted for variation between
cases when randomization is not possible (Tabachnick & Fidell, 2007). Of course, even with the use of several covariates, no statistical technique can ever fully equate groups and eliminate all selection bias in the same way as randomization could. Nevertheless, when randomization is not possible, multivariate regression and ANCOVA are the most useful statistical methods of addressing selection effects in experimental designs (Tabachnick & Fidell, 2007).

In the present analyses, the outcome variable is the police department’s burglary arrest rate for each 6-month period throughout the 5-year experiment \((n = 10)\). The unit of analysis is therefore the bi-annual period for which data were collected \((n = 10 \text{ per agency})\), resulting in a total sample size of 40 for all four agencies. The main predictor variables are the department’s experimental condition \((1 = \text{control}, 2 = \text{treatment})\) and the evaluation testing status \((1 = \text{pretest}, 2 = \text{posttest})\). The average arrest rate prior to the experiment and the burglary incidence rate are included as covariates in the models. To directly measure the effect of the burglary profiles on the outcome while holding all other predictors and covariates constant, the interaction term treatment condition by testing status is also included in the analyses.

To use the regression and ANCOVA analyses, checks for multicollinearity among the predictor variables were conducted, with no strong and significant correlations found among the main variables and covariates, indicating independence among the measures (Stevens, 2002). To ensure that the assumption of equal variances in the sampled data was not violated, a Levene’s test of variance homogeneity among all main effect variables was conducted. The test showed a non-significant outcome \((\text{Levene} = 1.53, p = .227)\), meaning that the null hypothesis of equal variances cannot be rejected, and the analyses may be used.

Finally, to determine whether any significant differences exist among the control and treatment conditions at the start of the experiment, a series of \(t\) tests were conducted, with the results indicating no significant differences between conditions due to burglary incidence rates, officer to citizen ratio, or the department’s semi-annual burglary arrest rate for the 4 years prior to the experiment. This suggests that the treatment and control conditions are well-matched based on these criteria; thus, the threats to internal validity due to selection effects are minimized. Results of the \(t\) tests are shown in Table 3.

### RESULTS

To determine how burglary arrest rates differed between the treatment and control conditions before and after the profiles were implemented, the burglary arrest rates for the treatment and control departments in the pre- and posttest periods were collected and analyzed.

Table 4 shows that, while the average burglary arrest rates for the treatment and control agencies were initially very similar, the treatment agency actually started off slightly lower
than the control agencies (11.3% vs. 15.9%, respectively). However, 1 year after the profiles were implemented, the treatment agency tripled their arrest rate from the pretest to the posttest period, while the control arrest rate had slightly decreased (30.1% vs. 10.9%, respectively). This difference in the posttest burglary arrest rates for the treatment and control conditions yields a $t$ test odds ratio of 3.52 (95% CI = [1.64, 7.52]), suggesting that the treatment agency was more than 3.5 times more likely to close an unsolved burglary than the agencies in the control condition after the burglary profiles were implemented. The changes in burglary arrest rates between conditions and across testing periods of the profiling experiment are illustrated in Figure 1.

While these results suggest that the profiles had a sizable effect on the burglary arrest rates in the treatment department after the behavioral analysis technique was implemented, this initial analysis does not control for the effect of the relevant covariates, which can also influence the department’s burglary arrest rates. Therefore, ANCOVA and multivariate regression analyses were conducted to assess the impact of the burglary profiles above and beyond the effect of the other covariates in the model.

Table 5 presents the results of the ANCOVA, which was used to assess the impact of the experimental treatment on the burglary arrest rates by statistically controlling for other relevant factors.

Results of the ANCOVA suggest that an agency’s prior arrest rate for burglary is a significant predictor of future burglary arrest rates, even after controlling for all other factors...
The main effects for experimental condition (i.e., using the burglary profiles) and the testing status (i.e., pre- or posttest period) were also found to be strong and significant predictors of the burglary arrest rate in the model (Treatment: $F = 15.26, p < .001$, Testing: $F = 17.18, p < .001$). However, the largest effect size in the ANCOVA model comes from the experimental interaction term, which had nearly 1.5 times more effect on burglary arrest rates than the experimental condition or testing status alone ($F = 23.76, p < .001$). As the ANCOVA is also able to provide additional information on the effect size and power, it was found that the condition by testing interaction had the highest effect size and power of all the variables in the model. Specifically, the interaction effect had a partial eta squared ($\eta^2$) of .411, which indicates that the use of the burglary profiles had a large effect on the burglary arrest rates during the field experiment (Stevens, 2002). The power associated with the condition by testing interaction is .997, which is also very high, according to Cohen (1988).

To further explore the impact of the experimental treatment on the burglary arrest rates, conditional multivariate regressions were also conducted. Table 6 presents the results of the regressions, which were conducted in separate pre- and posttest models to more thoroughly assess and compare how the experimental treatment impacted burglary arrest rates in each period.

**Note.** Dependent variable = burglary arrest rates; $R^2 = .545$; adjusted $R^2 = .479$; $F = 8.16; p < .0001; n = 40$; $\eta^2$ = partial eta squared, SS = Type III sum of squares. *$p < .05$. **$p < .01$. ***$p < .001$.

### Table 5: ANCOVA on the Effect of Burglary Profiles on Burglary Arrest Rates

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F$</th>
<th>$\eta^2$</th>
<th>Power</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary rate</td>
<td>2.84</td>
<td>.077</td>
<td>.374</td>
<td>72.89</td>
</tr>
<tr>
<td>Prior burglary arrest</td>
<td>6.86*</td>
<td>.168</td>
<td>.721</td>
<td>175.97</td>
</tr>
<tr>
<td>Treatment</td>
<td>15.26***</td>
<td>.310</td>
<td>.967</td>
<td>391.44</td>
</tr>
<tr>
<td>Testing period</td>
<td>17.18***</td>
<td>.336</td>
<td>.981</td>
<td>440.88</td>
</tr>
<tr>
<td>Treatment × Testing interaction</td>
<td>23.76***</td>
<td>.411</td>
<td>.997</td>
<td>609.65</td>
</tr>
</tbody>
</table>

### Table 6: Conditional Multiple Regression Models Evaluating on the Effect of Burglary Profiles on Burglary Arrest Rates in the Pre- and Posttest Periods

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1: Pretest</th>
<th></th>
<th>Model 2: Posttest</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b$</td>
<td>$\beta$</td>
<td>$SE$</td>
<td>$t$</td>
</tr>
<tr>
<td>Burglary rate</td>
<td>0.020</td>
<td>.286</td>
<td>0.011</td>
<td>1.76</td>
</tr>
<tr>
<td>Prior burglary arrest</td>
<td>1.10**</td>
<td>.493</td>
<td>0.364</td>
<td>3.02</td>
</tr>
<tr>
<td>Treatment</td>
<td>−0.181</td>
<td>−.015</td>
<td>1.94</td>
<td>−0.93</td>
</tr>
</tbody>
</table>

**Note.** Model 1: $R^2 = .280$; adjusted $R^2 = .203$; $F = 3.64; p < .05; n = 32$; Model 2: $R^2 = .822$; adjusted $R^2 = .689$; $F = 6.18; p < .05; n = 8$. *$p < .05$. **$p < .01$. ***$p < .001$. in the ANCOVA model ($F = 8.16, p < .0001$). The main effects for experimental condition (i.e., using the burglary profiles) and the testing status (i.e., pre- or posttest period) were also found to be strong and significant predictors of the burglary arrest rate in the model (Treatment: $F = 15.26, p < .001$, Testing: $F = 17.18, p < .001$). However, the largest effect size in the ANCOVA model comes from the experimental interaction term, which had nearly 1.5 times more effect on burglary arrest rates than the experimental condition or testing status alone ($F = 23.76, p < .001$). As the ANCOVA is also able to provide additional information on the effect size and power, it was found that the condition by testing interaction had the highest effect size and power of all the variables in the model. Specifically, the interaction effect had a partial eta squared ($\eta^2$) of .411, which indicates that the use of the burglary profiles had a large effect on the burglary arrest rates during the field experiment (Stevens, 2002). The power associated with the condition by testing interaction is .997, which is also very high, according to Cohen (1988).
difference in burglary arrest rates between the treatment and control groups at the onset of the experiment.

The results of Model 2, which specifies the impact of the experimental treatment on burglary arrest rates after the treatment was implemented, are quite different than those found in Model 1 ($F = 6.18, p < .05$). First, prior arrest rate is no longer a significant predictor of an agency’s burglary arrest rate after the treatment was implemented in the experimental condition. Instead, the experimental treatment was found to be a strong and significant predictor of burglary arrest rates in the posttest period, even after controlling for the burglary incidence rate and the agency’s past burglary arrest rate ($b = 20.15$, $p < .01$). Specifically, receiving the experimental treatment was found to increase the raw burglary arrest rates by 20.15%, or more than 266% overall, when controlling for all other factors, as compared with the agencies that did not receive the profiling treatment. This finding suggests that the burglary profiles were significantly and positively related to an increased number of arrests in unsolved burglaries during the year-long posttest period, as compared with otherwise similar departments not using the burglary profiles during the same time frame.

Results of the multivariate analysis suggest that the profiles had a significant and positive effect on burglary arrest rates in the posttest period, even after controlling for other influential factors. However, the conditional models do not tell us whether the treatment effects observed in the pre- and posttest periods are statistically different from one another. To test the significance of the experimental treatment across testing periods, the $z$ test used by Paternoster, Brame, Mazerolle, and Piquero (1998) is utilized. The $z$ test compares the unstandardized coefficients of a single variable across two conditional models to determine whether there is a statistical difference among the observed effects (Gibson, Walker, Jennings, & Miller, 2010). A $z$ test result of 1.96 or more, in absolute value, suggests that the observed coefficients significantly vary across the conditional models (Paternoster et al., 1998). The formula for the $z$ test is as follows:

$$z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}}$$

The results of the $z$ test conducted using the $b$ coefficients for the experimental treatment across the pre- and posttest periods shows a strong and significant difference for the treatment effects ($z = -2.88, p < .05$). Specifically, the agency that received the burglary profiling treatment was significantly more likely to have higher burglary arrest rates in the posttest period when controlling for all other factors, as compared with the pretest period.

**DISCUSSION**

The goal of our research was to test the utility of statistically generated burglary profiles in active police investigations, using the first scientific field experiment conducted on the realm of OP. Our multi-agency experiment involved the cooperation of four major police departments in Florida, matched on several key criteria such as burglary incidence rate and arrest rate, jurisdiction size and location, and number of sworn officers. One department, referred to as the treatment condition, received training on the SPOT burglary profiles and how to properly and legally implement them in their ongoing investigations. The remaining police departments served as control groups in the experiment and conducted their
investigations without any use or knowledge of the burglary profiles. After observing each of the departments for a 1-year follow-up period, we evaluated the impact of the profiles on burglary arrest rates using several analytical techniques that statistically controlled for latent differences between the experimental conditions and potential influence of several covariates on the outcome variable.

Results of the experiment showed that the police department using the burglary profiles solved more than 260% more burglaries in the posttest period than the police departments not using the profiles, despite all conditions having nearly identical arrest rates at the start of the experiment. Furthermore, the burglary arrest rate for the treatment department was shown to have increased by more than 260% (a 20% raw increase) in the posttest period as compared with the control group, even after statistically controlling for prior arrest rates and burglary incidence rates among the participating agencies.

While prior studies have utilized experimental designs to test OP, often by comparing the accuracy of various “profiler” groups against each other (see, for example, Kocsis, 2004; Kocsis et al., 2002; Kocsis & Middledorp, 2004), no study has tested the impact of OP in active police investigations using an experimental design. The lack of precedence on this issue makes the interpretation of findings in comparison with past research rather difficult, as the study designs and types of profiles being tested vary so greatly. However, it should be noted that the approximate 20% raw increase in burglary arrest rates that resulted when the treatment group used the profiles in their ongoing investigations falls directly in-between the 17% of police that reported the FBI profiles directly led to the identification of a suspect (Pinizzotto, 1984), and the 38% of police who reported that FBI and non-FBI profiles directly assisted in the identification of a suspect (Trager & Brewster, 2001). In other words, the rise in arrest rates seen for the department using the burglary profiles tested in this field experiment fell within the range of responses police gave regarding their perception of profiling’s utility in past research.

Nevertheless, this is only the first study to conduct an experimental field evaluation of OP, and there are several critical caveats of the study that will be discussed in the limitations section. However, it should be noted that one major benefit of the present research design, along with the ability to evaluate profiling in real-world conditions, is that the external validity (i.e., generalizability) is significantly increased in non-randomized versus randomized experimental designs. As the purpose of this experiment was to apply and test the offender profiles in a realistic setting to determine the level of utility that can be expected if these burglary profiles are implemented in future police investigations, the high level of external validity makes the research design a good match for the intents and purposes of the present study. The high level of external validity also means that the results of the study should be highly applicable to similar agencies using similar profiles, and future experimental research on OP should be directly comparable with the present study. The need for future research on profiling, particularly utilizing experimental field designs, will be discussed in the section to follow.

In short, the results of this study were very positive for the statistically generated burglary profiles and provide the first empirical support that the OP may be a useful and positive tool for investigators to utilize in active field investigations. However, as this is only the first study of its kind on the topic, and there are several important limitations to be considered, it is essential that additional field experiments on OP are conducted to more thoroughly assess the nature and accuracy of the findings seen in the present research.
LIMITATIONS AND FUTURE RESEARCH

Although there were several promising new findings to result from this experimental evaluation of OP, there are also several critical limitations that could impact the interpretation of the present results and the need for future research on this topic. First, as this was not a randomized experiment, it could be argued that possible unmeasured differences between the treatment and control conditions might help to explain these findings. However, as the strongest predictor of future arrest rates—past arrest rates—was statistically controlled for in our analyses and had no impact on the significance of the experimental treatment when included in the statistical models, it is less likely that other unmeasured differences among agencies would have a more significant impact on the outcome. It is also possible that the lack of randomization in the selection of participating agencies and the assignment of agencies to the various conditions may introduce selection effects that could lead researchers to believe that the experimental treatment had an impact on the outcome, when in fact it did not (Cook & Campbell, 1979; Farrington, 1983). But, as the participating agencies were selected based on key similarity criteria which were also used as controls in the analyses, and no variation in any of the covariates were found between agencies before and after the profiles were implemented, the risk of unaccounted for error in this experiment is significantly reduced. Still, it is important to note that even with statistical analysis, control variables, and highly matched conditions, no experimental design can eliminate selection threats entirely without the use of randomization. Also, there may be extraneous variables that have not been accounted for in the analyses that could have subtly mismatched the groups or influenced the outcomes. On the positive side, the lack of randomization reduces the artificiality of the experiment, and increases the external validity and generalizability of the study’s results. However, no causal inferences may be drawn; even with additional statistical controls and matched cases, the lack of randomization prevents any “cause and effect” relationship between the treatment and outcome measures to be made.

Another limitation is that the investigators’ initial feelings about the validity of profiling, or the credentials/professional status of the “profiler,” may have influenced how positively the profiles are perceived (Kocsis & Middledorp, 2004). While issues such as these may influence the officers to pay more, or less, attention to the profiles in practice, potentially leading to variations in the impact of the profiling treatment on the outcome, the experimental design and analyses used in this study makes this unlikely. First, any positive influential effects resulting from the credentials and professional status of “profiler” (who was often younger and had fewer years of law enforcement experience than most of the trainees) would presumably decay to negligible levels in the year-long follow-up period in which the profiles were implemented. Furthermore, training consisted of an in-depth description of the burglary profiles, and how the investigators should accurately, legally, and ethically use the profiles in their own investigations. No clinical profiling consultation on individual cases was conducted. In other words, the officers, detectives, and analysts who received the training served as the case profilers, not the trainer. Finally, as the analyses used in the present study are intended to statistically control for confounds, such as variation in individual differences among the participating agencies, the potential for these threats to internal validity to have made a significant impact on the outcome measure is low.

In addition, as the clinical judgments of the investigators using the profiles may theoretically reduce the objectivity of the evidence-based profiles tested in this study, a crucial point of this research was to evaluate how those clinical judgments impact the intended
outcome of OP. In other words, we aimed to assess the impact the profiles actually have when being used by humans in real-life scenarios. Nevertheless, the isolated impact of the clinical judgments and natural human error cannot currently be evaluated in this study beyond assuming that these issues may help explain why an even greater percentage of the burglary cases in the treatment were not solved.

Finally, it is possible that a portion of the arrest rate increase could be attributable to the extra attention that officers in the treatment group gave to burglary cases after receiving the profiling training (i.e., a “Hawthorne effect”) and not necessarily due to the use of the burglary profile. Most research on the Hawthorne effect suggests that when it does occur, the spike in improved results typically decays to nothing after 8 weeks (Clark & Sugrue, 1991; Landsberger, 1958). While this potential limitation is a serious concern to be aware of, again, the prolonged follow-up period in this experiment decreases the risk of the treatment group investigators giving significantly more attention to burglary cases up to 1 year after receiving training. In other words, while it is possible that enthusiasm among investigators may have increased at the start of the experiment, it would have likely faded after a year, thereby normalizing attention paid to burglary but adding only the treatment (i.e., profiling) to the investigations.

Due to these limitations, it is important that this research is both replicated and expanded in the future, by collecting and analyzing additional long-term follow-ups of the departments participating in the current experiment, and ideally by conducting an RCT to establish more convincingly the effect of the profiles on arrest rates in the field. Future studies could also utilize data from a larger sample of police departments and develop and investigate the use of profiles of other types of offenses. Also, in an effort to gain more qualitative information about in which types of cases the profiles were most beneficial, and in what ways the profiles were useful in investigations, future studies should include surveys of officers and detectives to gather this important and insightful information. We also recommend that future experiments include control conditions where increased attention is paid to burglary investigations without use of the burglary profiles, to estimate how much of the treatment effect is attributable to increased attention alone.

CONCLUSION

This study aimed to empirically evaluate the real-world utility of statistically generated profiles for burglary by examining the impact of the behavioral classification tool as it is used in actual police investigations. The field experiment used to determine the utility of profiling in this study is very different from methods utilized in past OP research, such as customer satisfaction surveys or experimental surveys comparing different types of “profilers,” as a scientific field experiment was conducted to determine the impact of the profiles when applied in actual police investigations.

The results of this multi-agency experiment show that, during the 1-year follow-up period, the department using the burglary profiles had a significant increase in burglary arrests compared with the departments not using the profiles, despite having nearly identical arrest rates at the start. Such a considerable gap between the two conditions, as well as the major jump in burglary arrest rates within the department that received the profiling treatment, suggest that the profiles are useful and beneficial during actual police investigations. Although the statistically generated behavioral profile selected for evaluation in this
study is more likely to fare better in the field than profiles created using intuition or subjective methodologies alone, this method may be used as a blueprint to continue the scientific evaluation of profiles of differing types.

Clearly, additional research is still needed to understand the impact of profiling on unsolved cases, and to determine how profiles can be modified to produce even better results in the future. Along with collecting follow-up data from the current experimental agencies, and conducting an RCT with a larger sample to better establish causality between the profiles and arrest rates, future research should aim to answer not only if and how much profiles have an impact on arrest rates but also exactly how and under what circumstances a profile will reliably lead police to the arrest of an offender. It is also important to investigate effects of convictions as well as on arrests.

This study is one of the few that examine OP as it is applied in the field and the only one to do so using an experimental design. Through the continued research and efforts by police agencies, we may be on the road to developing more scientific and better grounded profiles, understanding the impact and implications of profiling, conducting a more quantitative and scientific evaluation of OP, and laying the foundation for a more accurate and useful tool to help police solve more of their most challenging crimes.

NOTES

1. Offender profiles have been utilized in agencies including (but not limited to) the United States (FBI), the United Kingdom (National Crime Agency), the Netherlands (National Police Agency), Canada (Royal Canadian Mounted Police), and hundreds of local and state departments throughout these countries, as well as Finland, Germany, Australia, Singapore, and more.

2. As any potential changes in departmental policy or practice may introduce error to the results of the experiment, at the onset of the study, it was explicitly stated and agreed upon that no changes to departmental policy, procedure, or structure would take place in the treatment group during the experiment. This agreement was upheld, and no modifications other than potential staffing changes took place in the time period. However, these staffing changes would likely be occurring at the control agencies as well.

3. Partial eta squared (\( \eta^2 \)) is the ratio of variance explained in the outcome by each individual predictor variable, while controlling for other covariates in the model, and is calculated by dividing the sum of squares for the effect of interest (SS) by the total SS (Cohen, 1973). Partial eta squared values over 0.17 are considered high, above 0.09 are moderate, and values below 0.09 are considered to be minimal effects (Cohen, 1973).

REFERENCES


**Bryanna Hahn Fox** is an assistant professor in the Department of Criminology and courtesy professor in the Department of Mental Health, Law, and Policy at the University of South Florida. She is a former research fellow in the FBI’s Behavioral Science Unit, and the recipient of the 2013 Excellence in Law Enforcement Research Award from the International Association of Chiefs of Police for her experimental research on statistical behavioral analysis. Her research interests involve the identification of psychological and developmental risk factors for criminal behavior, developing evidence-based policing tools, and conducting experimental field research.

**David P. Farrington**, OBE, is an emeritus professor of psychological criminology and Leverhulme Trust emeritus fellow at the University of Cambridge’s Institute of Criminology. He is the 2013 recipient of the Stockholm prize in criminology for his research in developmental criminology and his work as director of the Cambridge Study in Delinquent Development. In addition to 570 published journal articles and book chapters on criminological and psychological topics, he has published 91 books, monographs, and government publications.