# "Evidence Based Principles in Sociological Studies."

**Article** · January 2016

**1 author:**

Emirhan Darcan
Universität Bern
**28** PUBLICATIONS   **6** CITATIONS

**Some of the authors of this publication are also working on these related projects:**

Project   Autonomous Weapon Systems (AWS) View project

Project   "Career Advancement for Refugee Researchers in Europe (CARe)" View project

# Kriminoloji

# Dergisi

Turkish Journal of Criminology
and Criminal Justice

# Kriminoloji

# Dergisi

Turkish Journal of Criminology
and Criminal Justice

# Evidence Based Principles
# in Sociological Studies

## Sosyolojik Çalışmalarda
## Delil Temelli Prensipler

Emirhan DARCAN [*]

**Abstract**

In the first part of this study the general context of the principles regarding evidence based will be discussed in detail, following that based on Evidence Based Research the efficacy of the interventions will be discussed and tabulated.

The study reveals some facts related following questions. How can replications of a particular sociological program improve our understanding of its effectiveness? What is meta-analysis? What are its strengths and weaknesses for synthesizing knowledge across evaluation studies? What techniques can be used to promote the quality of "real world" program replications? Do they ensure program success? How are evidence-based principles different from evidence-based programs? Do the two approaches complement each other?

**Key Words:** *Evidence-Based Principles, Replications, Meta-Analysis.*

**Özet**

Çalışmanın ilk kısmında delil temelli prensiplerin genel hatlarına yer verilmeye çalışılacaktır. Sosyolojik çalışmaların etkinliğinde delil temelli araştırma konusu tartışılacaktır.

Bu çalışmada delil temelli araştırmalar ile ilgili bağzı sorulara cevap bulmaya ve açığa çıkartmaya çalışmaktadır. Bu sorular sırasıyla şöyledir. Herhangi bir sosyolojik programın başka bir şekilde tekraren araştırılması bizim programın etkinliğini nasıl arttırmaktadır? Toplu çözümleme nedir? Değerlendirme çalışmalarında bilginin sentezlenmesine yönelik güçlü ve zayıf yönleri nelerdir? Gerçek yaşama uyarlanacak çalışmalarda hangi teknikler kullanılır? Bu teknikler programın

---

[*]   Sosyolog, edarcan23@gmail.com

başarısını sağlar mı? Delil temelli prensiplerin delil temelli program-
lardan farkı nedir? Bu iki yaklaşım birbirilerini tamamlarlar mı?

**Anahtar Kelimeler:** *Delil temelli prensipler, Uyarlama, Toplu çözümleme*

## 1. How can replications of a particular program improve our understanding of its effectiveness?

Before we can address the reasons why replications of a particular
program improve our understanding of its effectiveness, it is important
to briefly describe how such effectiveness is evaluated in the first place.
By doing so we can identify what are the questions that cannot be
answered by doing a single evaluation of a program, and how
replicating the program in different contexts and with different
populations can help us to understand its effectiveness.

The first time that we evaluate a particular program the main goal is
to determine if the treatment applied is effective. That is, we want to
determine if the relationship between the independent and the
dependent variables is causal and not due to the effect of third variables.
To do so, multiple research designs can be used, but many scholars
(Sherman, Farrington and Weisburd, among others) consider
randomized control trials to be the "gold standard" due to the fact that
they have the highest internal validity of any known design (Petrosino et
al, 2003). One of the most important advantages of this type of design is
that any changes in the dependent variable can be linked unambiguously
to the intervention and that it addresses most threats to internal validity.
However, some of the most important shortcomings of randomized
control trials are that not every program in criminal justice is suitable for
randomization, that in order to have some statistical power a large
sample is needed and that very often randomized control trials are
carried out on very specific subjects in a very specific setting in a
particular time frame, narrowing intensively the generalizability of the
findings. Furthermore, experiments give us information about if a
program works or if it does not work, but it cannot explain the "black
box": why does the program work?

Due to the nature and the characteristics of the programs that are
evaluated, quasi-experiments are the norm in criminal justice. Although
quasi-experiments per se do not rule out all threats to internal validity,
different elements can be introduced in every particular design in order

to increase the internal validity of the study. In these cases, it is more difficult to override the influences of third variables in order to state that a program is effective, but quasi-experiments have some strengths and advantages over experiments: they are more practical, easier to set up, very often they are more cost-effective, they rule out certain ethical issues that may arise in randomized control trials and, most important, have a higher external validity. Although they place a higher emphasis in the "black box", very often quasi-experiments cannot answer the question of why and how a particular program works (Farrington and Welsh, 2005).

Once we have tested that the cause precedes the effect, that there is a correlation between the variables and when we have overridden the influences of third variables we can say that the program was effective. The question now is: to what extent this causal relationship holds over variations in persons, settings, treatments, and outcomes? This is where replication can help us to better understand the effectiveness of the program.

Shadish, Cook and Campbell (2002) identify up to five different threats to external validity. They are five possible reasons why the findings of a study might not hold over changes in some of the elements of the design. The first is the change in the units of analysis: if we conduct a study with college students, will we find the same results if other type of population (i.e. housewives) had been selected? Does it matter if the subjects in our study are male or female? Do variations in the racial or ethnic composition of the sample affect the results? Do high-risk offenders respond differently to treatment than low-risk offenders? The second threat is related to the change in the setting, since an effect found in one kind of setting may not hold if other kinds of settings were to be used. A very clear example in the area of criminal justice is the difference between the programs that are carried out in prison or in the community. The context in both cases is completely different, and so are the relationship between the subjects and the people that implement the program and the motivation to complete the program, among others. The third threat to external validity refers to the variation of the outcome observations. For example, if we use different measures of recidivism, our findings might be subject to change. One example could be intensive supervision programs: if we measure recidivism by aggregating technical violations and arrests for new offenses, our findings will show an increase in recidivism. If, on the contrary, we leave out technical

violations and we just measure the number of arrests or convictions for new offenses, we could find that an increased supervision does reduce recidivism, although it results in more technical violations. The fourth threat is interaction of the causal relationship over treatment variations: an effect found with one treatment variation might not hold with other variations of that treatment, or when that treatment is combined with other treatments, or when only part of that treatment is used. There is no need to explain that completing only part of the treatment might have some effect in the findings about the effectiveness of the program. The hypothesis is that any treatment will be much more effective on the completers than on those who drop-out of the program. In the same sense, variations in parts of the treatment might also affect its effectiveness, as we might be adding or removing one of the sections most related to the success of the program. Finally, the last threat to external validity is context-dependent mediation: an explanatory mediator of a causal relationship in one context may not mediate in another context. We will refer to this last threat later.

How can replications of a particular program improve our understanding of its effectiveness? Replication with different units of analysis, treatment variations, outcomes and settings gives us information about the context in which a particular program works, and under which circumstances it does not work. It helps us to understand what might be the underlying mechanisms that make a certain intervention successful among a population, and why the same exact program does not work in another context (Braga, 1999). The most famous example of this process is the Minneapolis Domestic Violence Experiment conducted by Sherman in the early 1980s to evaluate the effectiveness of various police responses to domestic violence (Eck, 1997). The first evaluation of the program showed such a great success of mandatory arrest, that after the study was made available to the public numerous other states and law enforcement agencies changed their policies in order to adopt that measure. However, when the study was replicated some years later in several other cities, the evaluation found mixed results, with three studies finding that offenders who were arrested experienced higher levels of recidivism. As a result, further analysis showed that the measure of mandatory arrest worked well when the arrestees were employed, but that the same strategy produced the opposite results when the men were unemployed.

In summary, replication not only tells us if the study results would hold over variations of the context and, therefore, our findings are generalizable to other populations and settings, it also helps us in our analysis of the "black box": how and why does a particular program work? As we have seen before, experiments and quasi-experiments do not address this issue specifically, and replication provides us with further information that can help to answer that question. By analyzing the different findings of several replications of a program, we can study what are the characteristics that can be found in the more successful studies, which are not present in the least successful, and therefore infer what the underlying mechanisms at play are.

The question now is if replication gives us some information that cannot be reached through other means, like the use of realistic evaluation. Realistic evaluation is based on the idea that the outcome is the result of the sum of the underlying mechanisms and the context, assuming that every single intervention is context specific. It emphasizes the importance of identifying and analyzing the contextual factors in order to address the question "what works for whom under what circumstances" (Petrosino et al, 2003). This is certainly very similar to the contribution that replication brings to the table, but it is not exactly the same. Through replication we can ultimately infer what the underlying mechanisms of a program are, while when conducting realistic evaluation the hypothesis with the possible underlying mechanisms has to be done in advanced, it needs to be based on theory and some indicators must be identified, in order to be able to evaluate if the underlying mechanisms proposed are in fact the ones that are causing the program to be successful. Replication gives us some "hints" of what might be causing the differences among evaluations, while realistic evaluation assumes beforehand what those differences might be. Does this mean that when we carry out realistic evaluation there is no need to replicate the study? It does not. Replication has still the function of confirming that the identified mechanisms do work for a variety of units of analysis in different settings.

Another important way in which replication of a particular program improves our understanding of its effectiveness is by providing us with the opportunity of accumulating evidence. Some evidence-based clearing houses have protocols requiring multiple successful studies (e.g. programs with a least two successful high-quality evaluations) in order

to be able to consider a program as "evidence-based". An example would be the "Blueprints for violence prevention", which requires promoted "model programs" to have two or more successful evaluations using randomized control trials or quality quasi-experiments (Farrington and Welsh, 2005). Other required factors are that the evaluation of the programs must show sustained declines in violence, delinquency or drug use and that the mediating factors must be analyzed. However, it is not necessary that the replication of a program is successful in order to improve our understanding of the effectiveness of the treatment, since multiple replications of a program can be aggregated in a meta-analysis even if they do not show significant results (Marston and Watts, 2003).

This last aspect is directly related to the discussion about how replication should be carried out. Should the programs be replicated in their integrity, as a whole package, or should only the underlying mechanisms be identified and replicated, while adapting the rest of the program to the very specific context where the replication is carried out? In other words, should evidence-based programs or evidence-based principles be replicated? An evidence based program is a standardized intervention that has somewhat fixed characteristics that can be precisely documented, can (in theory) be reproduced with reasonable fidelity in new contexts and which efficacy has been established by randomized control trials or high-quality quasi-experiments (Lipsey et al, 2007). Again, the "Blueprints for violence prevention" can be used as an example of evidence based programs. On the other hand, evidence based principles are not concerned with standardized programs, are used to develop locally relevant interventions and their characteristics vary according to context (e.g. opportunities, offenders, existing program infrastructure) (Odom et al, 2005). The efficacy of principles emerges across studies, which means that some replication is necessary to determine what such principles are. The choice between replicating evidence based programs or principles is relevant, since it can affect the extent to which replications of a treatment can be accumulated and analyzed collectively (Tilley and Laycock, 2002).

As we have seen, replication can improve our understanding of the effectiveness of a particular program in different ways. First of all, we can use replication to determine to what extent the results of a treatment are generalizable, and if the findings hold over variations in persons, settings, treatments, and outcomes. Second, replicating a study is

important in order to analyze the reasons why a program works and what are the underlying mechanisms that the intervention triggers (the "black box"). Third, replication allows us to accumulate evidence, which in turn makes further analysis of aggregated studies possible, providing us with a deeper understanding of how a treatment works. And finally, evidence-based principles are also identified through the analysis of multiple studies, which provides us with an important tool to grasp what are the core elements that need to be taken into account when replicating a study in a complete different context.

**2. What is meta-analysis? What are its strengths and weaknesses for synthesizing knowledge across evaluation studies?**

Meta-analysis is one of the systematic ways of reviewing existing literature on a topic. As the other systematic reviews, researchers conducting a meta-analysis aim for a complete review of the literature, including published and unpublished studies and positive / negative, significant / non-significant results. A set of criteria is established to determine inclusion of the study (certain methods, sample size) in the review.

In meta-analysis, individual evaluations of a program are used as "subjects", and meta-analysis is the statistical analysis that is conducted on them. The goal is to find an "average effect" of the program through the aggregation and analysis of the individual studies and their effect sizes.

One of the advantages of meta-analysis is that it allows us to answer the question "does this program work?". It provides a clear, straightforward result that can be easily interpreted and understood not only by researchers, but also by policymakers. Another important advantage is that it allows us to analyze the results of many studies with small sample sizes and often non-significant or weak effects and find a strong average effect. For example, during decades evaluation of programs with a focus on rehabilitation of prisoners had been evaluated, and they showed weak or non-significant effects on reduction of recidivism. However, the use of meta-analysis "rescued" these studies by finding significant results when analyzing them at an aggregated level, which gave some credit to the "rehabilitative ideal" and promoted its return.

However, the use of meta-analysis has also some shortcomings. First of all, not every type of evaluation can be included in it, since it is necessary that the study has been designed as a randomized control trial. This fact has as a consequence that meta-analysis has the same pitfalls as RCTs: we can determine if a program works, but we do not know why it works, and the context in which the program was applied is not taken at all into account. Second, since randomized control trials cannot be conducted on certain policy topics due to several reasons as the impossibility of randomization or the need for a certain sample size to have statistical power, meta-analysis is not an option on many criminal justice policy areas. On the other hand, meta-analysis overcomes to some extent the problem of the lack of external validity of RCTs, since we are aggregating multiple evaluations of a program, which have been carried out in different contexts, and with different types of clients.

### 3. What techniques can be used to promote the quality of "real world" program replications? Do they ensure program success?

According to research conducted by Lipsey, when the pilot project of a program is successful and it is replicated in the "real world", very often such replications show a reduced and limited effectiveness of the program. There are many reasons that can lead to this result, and there are some techniques to overcome these problems in order to promote the quality of "real world" replications.

First of all, pilot programs tend to have a strong leadership and wide resources, both monetary and in terms of staff. The leaders are seen as innovators, and lots of efforts are put into the program. However, when applied to "real world" settings very often the leaders find that some programs are being imposed to them, and that the resources are not sufficient or stable over time. This difficulty can be overcome if stakeholders are brought to the table and a long-term sustainability plan is put in place.

Another common problem is the resistance by the local culture and staff, that oppose to the new program since it translates in a change in the way they worked before or it means an increased effort. This resistance is not usually found in pilot projects, since usually the staff members involved in them are excited to participate in an experiment and a potential innovation and improvement of the system, and they are

often highly motivated and trained volunteers. In order to overcome this problem, training of staff is highly recommended, as it helps to increase not only staff competence, but also motivation and accountability. Another possible technique is to put in place standard operation procedures (SOP) that describes what the new program will look like and how the staff should adapt to it, and guides them in the implementation of the treatment (Lipsey et al, 2007).

A third important threat to the success of "real world" replications of programs it that, as opposed to the pilot project, there is limited (if any) involvement of the experts and the researchers in their implementation. That means that the individuals in charge of the replication have to figure out a lot by themselves, which generates the risk of an inaccurate interpretation of the program. This threat can be addressed by bringing the experts as project leaders or as consultants in the implementation of the program, and by emphasizing the importance of the integrity of the treatment.

Another difficulty that "real world" replications have to face is the variability of the contexts and clients. The pilot project is usually carried out in ideal conditions; it is addressed to the exact population in the adequate context. However, replications of the program have to adapt to new contexts, keeping the core elements of the program (or the principles) and introducing the necessary changes (Tilley and Laycock, 2002). The same must be said regarding the target population: differential characteristics of the population must be accounted for and acknowledged in order to be able to use that information to explain variations in the effectiveness of the treatment.

Finally, very often "real world" implementations of the program are not evaluated or their evaluations are of a very low quality. It is necessary to carry out rigorous evaluations and to ask for feedback about the implementation of the program in order to understand why the effectiveness might change with regard to the pilot project.

## 4. How are evidence-based principles different from evidence-based programs? Do the two approaches complement each other?

Evidence-based programs are a standardized set of techniques that have been determined to be effective in dealing with a particular problem (Cuijpers, 2002; Haddix et al, 2003; Greenwood, 2008; Green,

2006). Their effectiveness is thoroughly assessed through the use of randomized control trials or high quality quasi-experiments, and the fact that they are standardized makes them very easy to replicate in different settings, as the idea is that the program per se is effective, so it will work no matter the setting in which it is applied. For these reasons, evaluations of evidence-based programs can be easily accumulated, and they provide a straightforward solution to policymakers that seek an answer to a particular problem. Since evidence-based programs are very often "branded" and marketed, they fit with the capitalist ideology of mass production of standardized products, as opposed to the idea of craftsmanship. These programs are "manualized" to ensure a homogeneous implementation across settings (Lipsey et al, 2007). A couple of example of this type of programs would be the Blueprint programs for violence reduction of the University of Colorado, or MST Inc. (Multi-systemic therapy).

Evidence-based principles have very different characteristics (Tilley and Laycock, 2002). They represent a set of tools, a "menu of options" that are available to help understand a problem and design solutions tailored to the local context. They require problem solving techniques, which must be applied after an extensive analysis of the problem has been carried out. Application is heavily based on theory and is context specific, and effectiveness is assessed through multiple applications of the principle in different contexts, which allows for a deep understanding of how that principle works (Green, 2006). With respect to policymakers, they are more complicated to sell because they do not provide "off-the-shelf", "ready-to-use" techniques as evidence-based programs do, and their application requires a previous extensive examination of the problem at hand. A couple of examples of evidence-based principles would be the principles of risk-need-responsivity (Andrews et al, 1990) or the techniques of situational crime prevention (Clarke, Branttingham, etc.).

However, although evidence-based programs and evidence-based principles are very different from one another, they can be combined in different ways. For example, when using situational crime prevention techniques, evidence-based programs can be among the "menu of options" provided as possible responses to a problem (i.e. street lighting, alley-gating, etc.). Or the other way around: evidence-based programs can be applied following the principles of risk-need-responsivity, and

risk assessment instruments can be incorporated to the implementation of evidence-based programs in order to determine the population that the program will target. By using them simultaneously, evidence-based programs and principles complement each other, which helps to overcome the shortcomings that either of them have individually.

### References

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment a meta-analytic investigation. *International journal of offender therapy and comparative criminology*, *50*(1), 88-100.

Braga, A. A., Kennedy, D. M., and Piehl, A. M. (1999) Problem-Oriented Policing and Youth Violence: An Evaluation 'of the Boston Gun Project. Unpublished report to the National Institute of Justice, Washington, DC

Cuijpers, P. (2002). Effective ingredients of school-based drug prevention programs: A systematic review. *Addictive behaviors*, *27*(6), 1009-1023.

Eck, J. (1997) Preventing Crime at Places in Sherman, L., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P. and Bushway, S. *Preventing Crime: What Works, What Doesn't and What's Promising*, A report to the United States Congress. US Departmentof Justice.

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades?. *Journal of Experimental Criminology*, *1*(1), 9-38.

Green, L. W. (2006). Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence?. *American journal of public health*, *96*(3), 406-409.

Greenwood, P. (2008). Prevention and intervention programs for juvenile offenders. *The future of Children*, *18*(2), 185-210.

Haddix, A. C., Teutsch, S. M., & Corso, P. S. (2003). *Prevention effectiveness: a guide to decision analysis and economic evaluation*. Oxford University Press.

Lipsey, M. W., Howell, J. C., & Tidd, S. T. (2007). The Standardized Program Evaluation Protocol (SPEP): A practical approach to evaluating

and improving juvenile justice programs in North Carolina. Final evaluation report. *Center for Evaluation Research and Methodology*.

Marston, G., & Watts, R. (2003). Tampering with the evidence: a critical appraisal of evidence-based policy-making. *The drawing board: An Australian review of public affairs*, *3*(3), 143-163.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional children*, *71*(2), 137-148.

Petrosino, A., Boruch, R. F., Farrington, D. P., Sherman, L. W., & Weisburd, D. (2003). Toward evidence-based criminology and criminal justice: Systematic reviews, the Campbell Collaboration, and the Crime and Justice Group. *International Journal of Comparative Criminology*, *3*(1), 42-61.

Tilley, N., Laycock, G., & Webb, B. (2002). *Working out what to do: Evidence-based crime reduction*. London: Home Office.