

Alcohol prevention and evaluation in the era of evidence based practice – the need for a systematic approach to evaluation

BO SANDBERG

ABSTRACT

AIMS – This article focuses on the shortcomings of experimental outcome evaluations. **MATERIAL** – The study uses two studies of the alcohol prevention program Prime for Life (PFL) to illustrate problems associated with a 'one dimensional' goal focus and how implementation may affect the efficiency of an intervention. **RESULTS** – As evidence based practice often fails to acknowledge the importance of context and implementation quality in its quest to find 'what works', it is argued in the article that a wider perspective on efficiency is needed in evaluations of alcohol prevention. **CONSLUSIONS** – To be able to find out what works and during what circumstances, evaluators need to look for a wider set of traits that constitute effective interventions. **KEYWORDS** – Alcohol prevention, evidence based practice, program evaluation, Prime for Life, Sweden

Submitted Initial 09.04.2010

Review completed 13.08.2010

Final version accepted 13.01.2011

Introduction

The overall objective of Swedish alcohol policy is to lower the total alcohol consumption (prop. 2005/06:30). Since the middle of the 1990s the Swedish national alcohol policy has been through major changes and the possibility to use the traditional Swedish control measures, such as high taxes, sales monopoly, and limited import quotas, have been reduced. Since the membership in the European Community, the Swedish government has emphasized the need for new strategies regarding alcohol policy in Sweden (prop. 1993/94:150, appendix 5). A sign of this shift is the movement from national efforts towards a vigorous promotion of the development of preventive measures at a municipal level and the active financial support in establishing 'Local Prevention

Coordinators' (Karlsson & Tigerstedt 2004). A political orientation towards using 'evidence based practice' (EBP) as a model for choosing interventions can be observed in the process of selection and appliance of new strategies (Blom & Morén 2003). Social science should according to EBP contribute to social policy and practice "... [B]y testing the effectiveness of interventions aimed at dealing with specific problems prior to their general application" (Tilley 2000, 2).

In this article, I will first shortly describe experimental evaluations and their role in an era of evidence based practices, and the need for a broader program evaluation approach that recognizes a wider set of qualities constituting 'efficient interventions'. To illustrate some issues of a narrow out-

come focus in evaluating alcohol prevention initiatives, I will use empirical data from two quasi-experimental outcome evaluations of the alcohol prevention program, *Prime for Life*. The outcomes of this specific program on alcohol use are not the focus of this article. The evaluations are used as examples in a discussion about outcome evaluation and are used to illustrate the importance of implementation quality when it comes to alcohol prevention. Consequently, the outcome results from the evaluations illustrated throughout this article will only be presented briefly.¹

Experimental outcome evaluation and alcohol prevention in the era of evidence based practices

As a way to deal with inefficiency and lack of research based policies, EBP is spreading within the field of social policy in Sweden and other countries. There is a desire to bridge the gap between research and practice in the field of social work. A comprehensive national project has been launched in Sweden to create a framework for the integration of practice, education and research with the objective of stimulating interventions studies and implementing scientifically documented methods (see Anderberg & Dahlberg 2005, 71; Dir 2007:161; FHI 2008; FHI 2009; Kriminalvården 2007). Also the National Board of Health has defined its mission 'to pull social work practice out of the world of opinions and untested knowledge and introduce it into the world of evidence and awareness of what works (...)' (Sundell et al. 2009, 1).²

EBP is closely linked to the use of experimental outcome evaluation designs to

assess the efficiency of interventions. Alcohol prevention should, according to this perspective, consist of strategies and interventions that have been tested and 'proven' efficient in outcome evaluations, preferably set up as randomized controlled experiments. Public management in Sweden has since the 1980s been geared towards goal-oriented thinking, competition, outsourcing and privatization and as a consequence there has been an increased focus on control and auditing to make sure that public services are effective (Reeder 2010). In search for "efficiency" in social policy, ideas about evidence have been adopted from evidence based medicine (EBM). The use of scientific methods that produce outcomes valid under certain controlled circumstances, has become standard for evaluations of social interventions.

When evaluating social interventions there is a need to consider some important limitations of randomized controlled experiments. First, a randomized experiment may restrict other types of activities for both the experiment and the control group. A randomized experimental design limits the choices for those implementing an intervention, as the evaluator wants maximum control over the experiment (Sandberg & Faugert 2007). Secondly, randomly assigning individuals, groups or activities to experiment and control groups can be difficult, and un-ethical; it confines the possibility of participants and stakeholders to choose what they wish to participate in (Lilja et al. 2004). It can also be unethical not to offer all clients a specific treatment if there are valid reasons to believe that this treatment is better than no intervention at all (Dahmström 1996). And thirdly, in many circumstances ran-

dom allocation is not useful for instance when the unit of study is a community rather than an individual (Tilley 2000). In community interventions a randomization is also likely to undermine the social and collective mechanisms that are essential to planning and implementing a 'community effort' (Hope 2005).

Meta-analysis is of great importance in EBP; studies and interventions (for example randomized controlled experiments) must meet certain quality criteria in order to be included in the analysis. Effect Sizes (ES) are used to compare different studies and to calculate the average effect of the type of intervention studied. ES expresses the difference between two groups, independent of sample size and without using statistical significance. This means that different programs using a common scale can be compared regardless of the size of the studies; ES is the meat of meta-analysis.

In the literature there are some general concerns regarding meta-analysis that should be noticed. Systematic reviews are typically based on studies from many countries; the generalizations [based on a sample that is applicable to a population] observed in an experiment's population cannot automatically be transferred to another population. This means that results from systematic reviews in one country, might have limited scientific value as guidance for interventions in another country (Lilja et al. 2004). Publication bias, i.e. when published research is systematically unrepresentative of the completed studies, can severely affect the results of systematic reviews (Dickersin 2005; Rothstein 2007). Meta-analysis has received renewed attention with the popularity of EBP, since

systematic reviewers claim that meta-analysis is a more valid way to review research than for example narrative review (Egger et al. 2000). Also, reporting of outcomes within studies is often incomplete and biased due to selective withholding of statistically non-significant results in papers (Chan & Altman 2005). When calculating the average ES of studies in a meta-analysis, there is a risk of skewed results caused by randomized experiments with inadequate quality (Petrosino 2003; Sutton et al. 2001). In a systematic review, only a fraction of the gross number of all evaluations and studies on a specific type of intervention published will be included. For example, in a study on treatment of alcohol- and drug-problems made by the Swedish Council on Technology Assessment in Health Care (SBU), 478 articles regarding treatment of risky alcohol consumption were identified. Out of these 27 (6 %) were randomized controlled studies that were included in the study (SBU 2001). This is obviously a strength as well as a weakness; in a meta-analysis a lot of research findings are ignored because they did not meet the inclusion criteria. At the same time the strict protocol creates a solid foundation for generalizations based upon methodologically similar studies.

The need to widen the perspective on outcomes and efficient interventions

EBP can be seen as a desire to simplify and to create 'order out of chaos' in extremely complex social practices (Patton 2000). However, reality bears little resemblance to 'what works' and there are many barriers that affect the transfer of knowledge from the experts to managers and practi-

tioners (Gendreau 1996). Since the 1980s, multiple systematic reviews of alcohol interventions have been made, concluding that programs and policies that are targeted at the individual, especially school-based alcohol and drug prevention programs, have either indeterminate or limited effects (Babor et al. 2010; Baker et al. 2006; Cuijpers 2002; Foxcroft et al. 2003; WHO 2004; WHO 2009). Perhaps surprising to some, many of these programs are still in use. A fundamental problem with a narrow outcome focus is the lack of need for the evaluator to understand how an intervention program works in order to estimate its net effect. 'A focus on what works alone produces a 'one dimensional' approach to measuring outcomes that concentrates on one variable at the expense of many other independent variables' (Cook 2006, 10). Alcohol behavior is dependent on numerous factors such as psychological aspects, genetic factors, environmental variables and lifestyle. It is not likely that one given intervention can successfully target all the different aspects that help explain the level of alcohol consumption in a group. Outcome evaluators seldom recognize that the most significant result of an alcohol prevention program could be its ability '(...) to cultivate understanding and support for alcohol policies, and (---) to motivate those who are at risk for hazardous or harmful alcohol use to seek help' (WHO 2009, 54).

Even a small set of process oriented data in an outcome evaluation can contribute to a critical understanding of what happens during implementation of a program and help us understand why even evidence based practices may fail to deliver desired outcomes. There is a need to promote evaluations that not only generate evidence on

what works, but also generate evidence on what circumstances various programs require in order to function (Sanderson 2000) so that the gap between 'evidence' and 'reality' can be bridged. In the end of this article I will exemplify how this can be made by using a systematic evaluation approach where quantitative outcome evaluations are combined with designs that evaluate the theoretical frameworks of interventions, study the quality of their implementation, *and* recognize different 'stakes' in social policy. First, however, I will shortly introduce the Prime for Life program and the two evaluations that will be used to further explore some problems associated with narrow outcome evaluations that consequently become problems for the evidence based practice movement.

The Prime for Life program

Prime for Life (PFL) is an alcohol prevention program based on the 'Lifestyle Risk Reduction Model' developed by the Prevention Research Institute in Lexington, Kentucky (U.S.A). PFL is designed to challenge beliefs and attitudes that contribute to high-risk alcohol and drug use. PFL is delivered by a protocol that emphasizes the role of empathy, specific guidance regarding personal choices, the management of resistance, support for change, and a plan for success. Based on results from outcome evaluations of the PFL program used as intervention for court-referred impaired driving offenders in the U.S, PFL is included in the Substance Abuse and Mental Health Services Administration's registry of evidence-based programs and practices (NREPP 2009). In Sweden, the program has mostly been used (and evalu-

ated) as primary or secondary alcohol prevention in schools, at workplaces and in municipalities as part of their local alcohol prevention work.

Evaluations of the PFL program

Centre for Social Research on Alcohol and Drugs (SoRAD) in Sweden was asked to evaluate the effects of the PFL program at the Department of Hospitality, Culinary Arts and Meal Science at Örebro University and within the Swedish Armed Forces. The stakeholders explicitly wanted to find out the effect of the program on participants' alcohol consumption. For the Swedish Military Headquarters, it was a matter of being able to justify the cost of the intervention as part of their alcohol policy by finding out if it affected the personnel's alcohol consumption. At Örebro University, the intervention was initiated and financed by the National Alcohol Committee,³ which was interested in knowing if the program could help reduce alcohol consumption among university students.

The studies examined change in alcohol consumption, with assessments taken at baseline two weeks before and 12 months after the program.⁴ Only those individuals participating in both the pre- and post test were included in the evaluation studies. The evaluations had a quasi-experimental design and measured alcohol consumption by using the first three questions in the Alcohol Use Disorder Test (AUDIT, see Babor et al. 2001). Risky alcohol consumers were identified by using the AUDIT-scale and 'risk awareness' was measured by using a set of nine questions derived from the PFL curriculum.

At Örebro University, the PFL program was implemented as a voluntary course

for students at the Department of Hospitality, Culinary Arts and Meal Science in the spring of 2005. 62 out of 220 students at the department attended the program, out of which 44 students (71 %) participated in the outcome evaluation. Their results were compared to results from a control group (N=354) within the same university. Also a process evaluation consisting of various interviews with department staff and instructors as well as students was performed.

In 1999, the commander-in-chief decided that the PFL program would be implemented for all personnel throughout the Swedish Armed Forces. At the point of the evaluation, the Swedish Armed Forces had about 20 000 civilian and military employees at 18 regiments geographically spread over the whole country. Six regiments ready to start their training were chosen as the experiment group. Out of 885 participants in the PFL program in 2005, 446 individuals (55 %) participated in the study. The control group (N=515) consisted of individuals at nine regiments who had not implemented the program during the evaluation period. Data in the evaluation was collected through questionnaires administered by staff consultants at the experiment and control regiments.

Evaluation results in short

Outcome results from both evaluations indicated a large impact on the risk awareness among PFL participants (Sandberg 2006; 2007a; 2007b). The effects of the program on alcohol behavior were small or trivial.⁵ The results verified previous Swedish evaluations of the program in similar settings (Sjölund & Andréasson 2004; Hallgren et al. 2009). An evaluator

with an 'evidence perspective' focusing on the outcome of the program would most likely settle with the conclusion that the PFL intervention as a mechanism only provides a limited support for behavioral change when it comes to alcohol consumption.

The PFL evaluations focused on reporting the outcome of the program in terms of behavioral changes in participant's alcohol consumption. However, supplement data and a deeper analysis gives us a better understanding of the contexts within which the program and the evaluations were executed. The evaluation of PFL at Örebro University showed that the implementation of the program had not been satisfying; The Alcohol Committee and the university management had failed to gain approval for the program among staff and instructors at the department, which caused discontent for the PFL program and complicated the scheduling and implementation of the program. This may have affected the selection of students participating in the program as well as the efficiency of the PFL program. Similarly, data from the evaluation of PFL within the Swedish Armed Forces indicated that the implementation of the program had been difficult, most likely due to the impact of a major re-organization of the Swedish Armed Forces during the evaluation period. The organization was not capable of reaching their own target goals regarding the PFL training or collecting the data for the evaluation study. The latter contributed to low response rates and a systematic bias of the data that made the evaluation results less reliable (see Sandberg 2007b for further details).

Problems associated with a narrow set of goals and outcomes

Scientific outcome evaluations typically use either political goals or desired outcomes (usually defined by the researcher) to assess the worth and value of an activity. A problem with goal oriented evaluations is that evaluators seldom critically discuss the goals that they evaluate; political goals are often somewhat vague and unrealistic (Patton 1997; Vedung 1998) and thus ethically inapplicable as guidance for an evaluation. On the other hand, when scientists design evaluation studies of alcohol prevention, change of alcohol behavior often becomes the preferred outcome. As a result, *other outcomes* that might be important for policy makers and stakeholders, but are of less interest to the researcher, may be ignored. Most crucial, obviously, is the failure to recognize that the program might have other outcomes that potentially cause more harm than good (McCord 2003).

A program like PFL can contribute to alcohol policy formation and implementation. Prior to the implementation of the PFL program within the Swedish Armed Forces, 32 percent of those in the experiment group (N=446) and 36 percent in the control group (N=508) claimed that they were fairly or very familiar with the organization's alcohol policy. There was no statistical difference between the two groups. However, at the 12 month follow up, 54 percent in the experiment group compared to 42 percent in the control group claimed that they were fairly or very familiar with the alcohol policy. The increase was statistically significant for the experiment group ($p=0,000$) but not for the control group, and there was now

also a statistically significant difference between the experiment and the control group ($p=0,000$). From a policy perspective then, an alcohol prevention program can help create a shared perception of risks associated with alcohol such as occupational risks in an organization, increase awareness of alcohol policies and reduce the risk of organizational and individual harm caused by alcohol consumption. These and other changes that are not directly associated with the level of alcohol consumption could lead to increased cooperation between stakeholders, a better functioning administration and care of clients, improvements of people's health or living conditions, enhancement in documentation procedures and less bureaucracy in an organization for example. However, such outcome measures are seldom likely to be reported within the scope of a 'one-dimensional' experimental outcome evaluation of alcohol prevention, and we do not know much about what the long-term impact of these and other outcomes might have on participants' future alcohol behavior.

Finally, the cost of reaching goals is often ignored in outcome evaluations. A study of the evaluations mentioned in the Swedish national budget proposals 2000 and 2002 showed that only ten percent of the evaluations made included some sort of discussion about the cost-efficiency of evaluated interventions and programs (Statskontoret 2002). Even fewer evaluations made efforts to actually calculate their cost efficiency. 'When a prevention program is demonstrably shown to yield positive outcomes, it is normally assumed (...) that the overall benefit is greater than the overall cost, even when this is not em-

pirically demonstrated' (Tilley 2002, 54). However, an intervention can also generate positive results, but at a cost that is politically impossible to justify (Sandberg & Faugert 2007). In the case of PFL, the evaluations indicated that implementation was problematical (causing extra costs) due to contextual factors that will be discussed below.

Problems associated with the quality of implementation

In both of the PFL evaluations, data indicated that the implementation of the program did not go 'according to plan'.

The Alcohol Committee approached the Department of Hospitality, Culinary Arts and Meal Science with a request to pilot the PFL program at their department. In order to gain approval for the program among instructors and administrative personnel, they were offered to participate in a PFL training about six months prior to its implementation on the students. Nine representatives, i.e. 50 percent of the staff at the department, participated in this training. According to the PFL instructors, there was an outspoken negative attitude towards the PFL program among the personnel. In interviews, several representatives of the department articulated a negative experience of the PFL program and a concern that the program would not suit 'their students':

The students will not come back after the first day of training. (---) If they sit there and receive the same information as I did, they will never come back.

Instructors at the department felt that the implementation of the program was

handed out as a top-down 'order' from the university management and that they were forced to adjust their class schedules to fit the PFL program. The problems the staff saw with implementing PFL at their department are exemplified in the following excerpt from a group interview with instructors and staff that participated in the PFL training:

- (...) if the training would have been better, we would not have had so many discussions...
- Then our attitude would have been different
- Yes... then it would not have been so difficult...
- Probably not...

'It feels hard to motivate...to remove my own course elements to fit in this training when I know what it is...' said one of the instructors. The majority of the interviewed instructors also expressed concerns about the teaching style of the PFL representatives.

We are continuously evaluating ourselves and we want to be the best in the business and make sure that what we deliver in class is really top of the notch ... then it felt like, well this PFL thing does not really meet our standards.

The Alcohol Committee had several discussions with the department, the student union and PFL instructors in order to make sure that they could implement the program and to make sure that the students would participate, as there was a concern that the negative attitude among the per-

sonnel might negatively affect student participation in the program. The Alcohol Committee chose to directly inform the students about the PFL program at the annual meeting of the department's student union, to offer free lunch for participants, and to print individual diplomas to all participating students.

Surprisingly, in the two week follow up study of the PFL program, 92 percent of the students (n=40) that participated in the program graded the teaching style of the PFL instructors as 'good' or 'very good', and 98 percent ranked the instructors' knowledge in the field of alcohol prevention as good or very good. Six months after the intervention 30 students that had participated in the program were asked how they viewed the teaching style of the PFL instructors vis-à-vis their regular instructors at the department. One student said:

We are used to having instructors that only talk and know nothing about pedagogy. We get celebrity chefs, people from the business (...) Prime for Life was a lot better. They were experienced lecturers and they knew what they were doing.

An analysis of students at the department that participated in the PFL program (n=36) and students who had participated in the baseline study but not in the actual program (n=34), indicated no statistically significant differences between those who participated in the program and those who did not when it came to how many drinks they normally drank, how often they consumed alcohol and how often they binge drank. 'Our experience had no impact on student participation' said one instructor

and explained that she was confident in that all personnel was neutral and professional when they talked about the program with their students. Still, it can be assumed that in order to achieve long term commitment to alcohol prevention and efficient interventions, efforts have to be perceived as meaningful and beneficial for involved instructors, school administration, teachers *and* students (Lilja et al. 2004).

The evaluation of PFL in the Swedish Armed Forces did not include any systematic collection of data that focused on implementation of the program, but there is some information at hand that gives an idea of the implementation of the program; In 2004 the Swedish parliament approved of a reorganization of the Swedish Armed Forces that included the disbanding of a number of regiments and reduction of its staff (both military and civilian personnel) with approximately 25 percent. Prior to the evaluation, the alcohol and drug coordinators at the military headquarters did not believe this situation would have any impact on program implementation or data collection for the evaluation. However, the implementation of PFL experienced several problems. Both the training of personnel in the PFL program and the collection of data at control regiments proceeded slower than planned. One regimental commander bitterly stated: 'I have other things to prioritize... such as giving people notice.' Written field reports from staff consultants who administrated the questionnaires at regiments also testified that some regiments were unwilling to invest time in planning and implementing the program or the evaluation while dealing with the major changes taking place within the organization. In retro-perspective, the timing

for implementing and evaluating the PFL was probably not the best. However, it can be argued that the need for an alcohol prevention program was perhaps greater than ever within the Swedish Armed Forces; stress due to organizational changes can be a significant contributor to alcohol consumption (Pohorecky 1991) and from the commander-in-chief's perspective, it may have been quite reasonable to insist on continued implementation of the program throughout the re-organization. In any case, it is reasonable to assume that the changes that were taking place within the organization compromised both the quality of the implementation and the evaluation of the program.

Poor project management and implementation precipitates significant waste in both money and effort and may cause client dissatisfaction that may damage long-term goals for alcohol prevention policies; it also has a negative impact on the possibility to reach desired outcomes. This is an aspect that the evidence based practice movement needs to take into serious consideration: It is crucial to gain both a deeper understanding of how social interventions work and what produces outcomes, as well on understanding how efficient interventions should be applied in order to deliver positive results in new settings.

The mixing of high risk and low risk participants

Matching an intervention to the personal characteristics of the participants is another aspect of implementation quality. Such matching is seldom done. Even in treatment of offenders, it is quite common that treatment programs are implemented with a 'one size fits all' approach (Matthews

et al. 2001, 466). University students are commonly described as high risk alcohol consumers (Bullock 2004); approximately 40 percent of the participating students in the PFL evaluation engaged in high risk alcohol consumption (compared to approximately 18 percent in the Swedish Armed Forces).⁶ However, even in a high-risk setting such as the university, it is important to note that the majority of the population is not high risk consumers of alcohol. Since the PFL program is designed for groups that make high-risk choices, it is important to notice that when the program is implemented as primary or secondary prevention, it is not reasonable to expect all participants to potentially change their alcohol behavior as a consequence of the program. On the contrary, a majority of participants might actually learn that their alcohol consumption is of low-risk, and that they have no reason to change their alcohol behavior at all.

'Many programs appear ineffective due to the fact that inappropriate cases mask the success that these programs have with the clients who are appropriate for the services provided' (Matthews et al. 2001, 466– 467). In an outcome evaluation of alcohol prevention, low-risk consumers of alcohol might actually mask the impact of the program when it comes to alcohol consumption. This problem can quite easily be solved by controlling for high and low risk groups in an evaluation; When risk level in alcohol consumption is controlled for, the largest (yet small) effects of PFL are found among high-risk consumers of alcohol (Sandberg 2007a; 2007b). Some evaluators do not report ES unless results are statistically significant. This is problematic as the likeliness to find sta-

tistically significant results is dependent on the size of the population studied. For example, in an evaluation of PFL at high-schools in Stockholm, Sweden, about 13 percent (N=602) of the pupils were identified as high-risk consumers of alcohol. The evaluators noticed a 'major decrease of consumption' between the pre-test and post-test in this group, but this '(...) considerable change does not reach statistical significance (...)' (Sjölund & Andréasson 2004, 31. My translation) and because of this the evaluators did not report the ES for this group.

Controlling for high and low risk groups when analyzing evaluation data can illuminate differences, but do not compensate for the possible impact the mixing of high and low risk participants may have on program implementation and outcome. To my knowledge, there are no specific studies on the impact of alcohol prevention initiatives that do not match interventions with client characteristics. However, in criminal justice programs, it has been shown that failure to match the level of service with the risk of the offenders can have serious consequences. Intensive services and measures of control directed at low-risk offenders can be harmful as these interfere with the generally pro-social lifestyles of the low-risk offenders (Matthews et al. 2001).

Final comments: Why there is a need for a broader approach to 'what works'

As illustrated in this article, evaluations of alcohol prevention with a narrow focus on outcomes and a limited set of variables run the risk of overlooking other results that might be important to stakeholders and

crucial for understanding the theoretical foundations and practical limitations to interventions.

The focus of 'evidence' is usually on changes in alcohol behavior, and reduced alcohol consumption is often viewed as the ultimate outcome of an intervention, even when the intervention is applied in a way that is different from how the program was intended to be used. Consequently, outcome evaluations of alcohol prevention tend mostly to verify previous empirical knowledge on the effects prevention strategies. Other data, however, might give insight to an intervention's ability to increase knowledge of, for example, risks associated with alcohol consumption or awareness of alcohol policies that might motivate and support people in seeking help.

'What works' is just as much a matter of quality of implementation as it is a matter of using interventions that have been proven efficient. In the attempt to generalize, outcome-oriented evaluators often neglect to study the complex reality that distinguishes the evaluated activity. In the case of PFL, previous outcome evaluations of the program validate the conclusion that the intervention has a limited impact on alcohol behavior among participants; however, we do not know much about the quality of the program implementation in previous evaluations. Few published outcome evaluations of interventions refer to formal documentation describing the context, content and delivery of the intervention (Michie et al. 2009).

There is a tendency inherent within what works to assume that there is 'out there' an objective reality to be studied and a full grasp of the context in which

policy interventions are to be applied (Cook 2006, 104)

Identical interventions might work in one setting but not in another (Blom & Morén 2003; Haraschi et al. 1999; Lab 2007; Tilley 2000). In evidence based practice and outcome evaluations, implementation has been relatively neglected (Durlak 1998) and as one consequence there is not enough information on how to replicate evidence based interventions (Dombrowski et al. 2007; Michie et al. 2009; Riley et al. 2008).

There is a need to widen the perspective on evaluation. In conclusion, I will exemplify with a program evaluation perspective used to study programs and interventions used within correctional services. The Canadian evaluation protocol Correctional Program Assessment Inventory (CPAI) consists of more than 150 variables that study treatment programs through multiple data collection methods, such as document studies, interviews and observation studies (Gendreau & Andrews 2001). The idea of this model is to compare a program with the common characteristics of 'effective intervention' programs (derived from literature reviews and meta-analyses). An effective program is viewed as a form of ideal type (Latessa 1999); a benchmark for the intervention studied. Gendreau & Andrews (2001) summarize what constitutes effective programs within six areas:

1. Program implementation and leadership: effective programs are well designed and implemented, based on a sound theoretical model, and have qualified and motivated leadership.
2. Offender/client assessment: effective

programs receive appropriate clients and conduct standardized and objective assessment of risk, need and responsibility factors

3. Treatment/intervention characteristics: effective programs target 'deviance' producing behaviors, use effective behavioral treatment models, deliver services and treatment effectively, and prepare clients to remain or return to the community.
4. Staff characteristics: effective programs have well educated, qualified, experienced, and well supervised staff that have input into the program.
5. Quality assurance and evaluation: effective programs make sure that quality assurance is provided and evaluations are regularly conducted.
6. Other items such as ethical guidelines and levels of support: effective programs are supported both internally and in the community.

The benefit of an evaluation model such as the CPAI is that it delivers results that can indicate both strengths and weaknesses of a program and provide detailed recommendations for program improvement. It can also help explain, by focusing on the context that is needed to support effective interventions, why an evidence based intervention might fail to produce desired outcomes.

CPAI approaches interventions with multiple methods of data collection such as interviews, document studies, and participating observation. In a study of over 100 correctional programs across the U.S.A., Latessa (1999) highlights some of the most common weaknesses of the programs assessed:

Regarding implementation and leadership (1), programs were often based on strong theoretical models derived from the treatment literature, but designed with little regard for the research on what works for the type of clients they were serving. This can be compared to the strong theoretical foundation of PFL with a specific design for groups that make high-risk choices when it comes to alcohol – and its implementation as primary prevention in high-schools, at work places etc. where the vast majority of participants do not make high-risk choices when it comes to alcohol consumption. An intervention may not be inefficient in itself, but depending on how it is put into action might make it seem so. Further, few programs used standardized instruments to assess the clients risk prior to program participation (2). Even if assessments were made, information was seldom used in the decision-making process or the delivery of the service. Program characteristics (3), turned out to be the least satisfying area: programs were inconsistently implemented, lacked structure and were implemented without any manuals and with few rewards to encourage client participation and compliance. Programs also failed to provide aftercare services or 'booster sessions' for participants. The typical weakness in staff characteristics (4) was staff turnover, lack of training on the interventions and treatments utilized and lack of supervision and assessment of service delivery skills. Standardized assessment of clients using meaningful performance measures (5) where uncommon, and most programs did not track the clients after they had left the program. Observations regarding other areas (6) mostly involved lack of ethical

guidelines for interventions and the lack of external support through, for example, advisory boards.

Even if an intervention is 'evidence based' it may not necessarily be implemented in a way that brings out its full potential. Latessa (1999) further concluded that even though many programs and interventions were based on strong theoretical models and evidence based practices, 'relatively few programs provided services and treatment consistent with the principles of effective intervention' (Latessa 1999, 426).

To bring 'evidence based' policy to a higher level, where efficient interventions can be understood and implemented successfully, we need to widen the perspective on both 'evidence' and how to perform evaluations of alcohol prevention.

Outcome evaluations have to take into consideration a broader array of effects, including stakeholder expectations that use method triangulation, and incorporate multiple aspects of knowledge recognizing the complexity of social policy. The EBP movement can only be successful when it acknowledges the need to understand how and in what circumstances various efforts (might) work and when it provides policy-makers and practitioners with the tools and knowledge on how to make these efforts work.

Declaration of interest None.

Bo Sandberg, researcher
Department of Criminology,
Stockholm University, Sweden
E-mail: bo.sandberg@criminology.su.se

NOTER

- ¹ The outcome results have been presented in SoRAD-publications (Sandberg 2006; 2007a; 2007b).
- ² The National Board of Health cooperates with SFI Campbell to produce reviews concerning the effect of interventions within social work and crime prevention. SFI Campbell is a part of the international Campbell Collaboration that works to improve decision-making through systematic reviews on the effects of interventions within the areas of education, crime and justice, and social welfare.
- ³ The committee was assigned by the government to implement the National Alcohol Policy (prop. 2000/01:20) during 2001-2007. The goal of the National Alcohol Committee was to coordinate and promote initiatives to reduce the harm of alcohol in Sweden.
- ⁴ Both studies included a two week follow up after the PFL program to evaluate the participants' experiences of the intervention.
- ⁵ Results on alcohol consumption were calculated as Effect Sizes using Cohens d. An ES of less than 0.20 is considered trivial, between 0.20 and 0.50 a small effect, 0.50-0.80 a medium effect and 0.80 or higher a large effect of a program (SBU 2001).
- ⁶ High risk consumption was defined by the AUDIT-10 scale (Babor et al. 2001) where 8 points or more was used as the cut-off for risky alcohol consumption for men, and 6 points or more for women (Sandberg 2006; Sandberg 2007b).

REFERENSER

- Anderberg, M. & Dahlberg, M. (2005): Reflections on How Research and Local Practice Can Evolve Together. NAD publication 47: 71–95
- Babor, T.F. & Biddle-Higgins, J.C. & Saunders J.B. & Monteiro, M.G. (2001): AUDIT: The Alcohol Use Disorders Identification Test. Guidelines for Use in Primary Care. Geneva: World Health Organization
- Babor T.F. & Caetano, R. & Casswell, R. & Edwards, G. & Giesbrecht, N. (2010): Alcohol: no ordinary commodity – research and public policy. Second edition. Oxford: Oxford University Press
- Baker, P. & Young, M. & Thomas, R. (2006): Summary of Evidence: Systematic Reviews on Alcohol Dependence. Queensland Government, Australia
- Blom, B. & Morén, S. (2003): Explaining Human Change – On Generative Mechanism in Social Work Practice. *Journal of Critical Realism* 2 (1): 37-60
- Bullock, S. (2004): Alcohol, Drugs and Student Lifestyle! A Study of The Attitudes, Beliefs and Use of Alcohol and Drugs Among Swedish University Students. SoRAD Research Report 21. Stockholm university, Centre for Social Research on Alcohol and Drugs
- Chan, A-W. & Altman, D.G. (2005): Outcome reporting bias in randomized trials on PubMed: Review of publications and survey of authors. *British Medical Journal* 330, 753-756
- Cook, T.D. (2006): Criminal and Social Justice. London: Sage Publications
- Cuijpers, P. (2002): Effective ingredients of drug prevention programs. A systematic Review. *Addictive Behaviors* 27 (6):1009–1023
- Dahmström, K. (1996): Från datainsamling till rapport – Att göra en statistisk undersökning (From Data Collection to Report – Performing a Statistical Study). Lund: Studentlitteratur
- Dickersin, K. (2005): Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein, H. & Sutton, A. & Borenstein, M. (eds.): *Publication Bias in Meta Analysis: Prevention, Assessment and Adjustments*. Chichester: John Wiley
- Dir. 2007:161 Kommittédirektiv – Socialt råd (Committé Directive for the Social Council). Stockholm: Regeringskansliet
- Dombrowski, S.U. & Sniehotta, F.F. & Avenell, A. & Coyne, J.C. (2007): Towards a cumulative science of behaviour change: Do current conduct and reporting of behavioural interventions fall short of best practice? *Psychology and Health* 22: 869–874 .
- Durlak, J.A. (1998): Why program implementation is important. *Journal of Prevention & Intervention in the Community* 17 (2): 5–18
- Egger, M. & Davey Smith, G. & Altman, D. (eds.) (2000): *Systematic reviews in health care: Meta-analysis in context*. London: BMJ Books
- FHI (2003): Uppföljning av nationella handlingsplanen för att förebygga alkoholskador (A Follow-up of the National Alcohol Prevention Plan). Stockholm: Statens folkhälsoinstitut
- FHI (2008): Effekter av lokalt alkohol- och narkotikaförebyggande arbete (Outcomes of Local Alcohol and Drug Prevention Initiatives). Östersund: Statens Folkhälsoinstitut
- FHI (2009): Det drogförebyggande arbetet i Sverige 2008. Rapport om det förebyggande arbetet mot alkohol-, narkotika-, tobaks- och dopningsproblem (Drog Prevention efforts in Sweden 2008. Report on the Prevention of Alcohol, Narcotics, Tobacco and Drugs). Östersund: Statens Folkhälsoinstitut
- Foxcroft, D. & Ireland, D. & Lister-Sharp D. & Lowe, G. & Breen, R. (2003): Longer-Term Primary Prevention for Alcohol Misuse in Young People: A Systematic Review. *Addiction* 98: 397–411
- Gendreau, P. (1996): Offender rehabilitation: What we know and what needs to be done. *Criminal Justice and Behavior* 23: 144–161
- Gendreau, P. & Andrews, D. A. (2001): *The Correctional Program Assessment Inventory 2000 (CPAI 2000)*. Saint John: University of New Brunswick
- Hallgren, M. & Källmén, H. & Leifman, H.

- & Sjölund, T. & Andréasson, S. (2009): Evaluation of an alcohol risk reduction program (PRIME for Life) in young Swedish military conscripts. *Health Education* 109 (2): 155–168
- Haraschi, T.W et al (1999): Employers and the Employment Option of the New Deal for Young Unemployed People: Employment additionality and its measurement. Institute for Employment Research. University of Warwick, England
- Hope, T. (2005): Pretend it Doesn't Work: The 'Anti-Social' Bias in the Maryland Scientific Methods Scale. *European Journal on Criminal Policy and Research* 11: 275–296
- Karlsson, T. & Tigerstedt, C. (2004): Testing new models in Finnish, Norwegian and Swedish alcohol policies. *Nordic Studies on Alcohol and Drugs* 21(English Supplement): 79–91
- Kriminalvården (2007): Kriminalvårdens FoU-strategi 2008–2010 (Swedish Prison and Probation Service – R&D Strategy 2008–2010). Norrköping: Kriminalvården
- Lab, S.P. (2007): Crime Prevention. Approaches, practices and evaluations. Cincinnati: LexisNexis
- Latessa, E. (1999): What Works in Correctional Intervention? *Southern Illinois University Law Journal* 23: 415–426
- Lilja, J. & Giota, J. & Larsson, S. (2004): Olika synsätt på utvärderingar av preventionsprogram för skolor. (Different Views Upon Evaluations of School-based prevention programmes). *Nordisk alkohol- och narkotikatidskrift* 21(4–5): 297–307
- McCord, J. (2003): Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs. *The ANNALS of the American Academy of Political and Social Science* 587(1): 16-30
- Matthews, B. & Jones Hubbard, D. & Latessa, E. (2001): Making the Next Step: Using Evaluability Assessment to Improve Correctional Programming. *The Prison Journal* 81: 454–472
- Michi, S. & Fixsen, D. & Grimshaw, J. & Eccles, M. (2009): Specifying and reporting complex behavior change interventions: the need for a scientific method. *Implementation Science* 4 (1): 40
- NREPP (2009): Prime for Life Intervention Summary. National Registry of Evidence-based Programs and Practices, U.S.A.
- Patton, M.Q. (1997): Utilization-Focused Evaluation. Newbury Park: Sage
- Patton, M.Q. (2000): A vision of Evaluation that Strengthens Democracy, paper from EES conference in Lausanne October 12 th –14th
- Petrosino, A. (2003): Om och varför? Den potentiella nyttan av att inkludera programteoriutvärderingar i metanalys (If and Why? The Potential Benefit of Including Program Theory Evaluations in Meta Analysis). *Socvet* 2–3–02
- Pohorecky, L.A. (1991): Stress and alcohol interaction: An update of human research. *Alcoholism: Clinical and Experimental Research* 15 (3): 438–459
- Prop. 1994/95:150 Bil.5 Medel för alkohol- och drogpolitiska åtgärder (Appendix 5: Funds for Alcohol and Drug Actions). Stockholm, Regeringskansliet
- Prop. 2000/01:20 Nationell handlingsplan för att förebygga alkoholskador (National Plan of Action for Prevention of Alcohol Harm). Stockholm, Socialdepartementet
- Prop. 2005/06:30 Nationella alkohol- och narkotikahandlingsplaner (National Alcohol and Drug Policies). Stockholm, Socialdepartementet
- Reeder, J. (2010): Det tystade samtalet. Om staten, psykiatri och försöken att undanröja det psykoanalytiska inflytandet (The Silenced Conversation. About the State, Psychiatry and the Efforts to Remove the Psychoanalytical Influence). Stockholm: Norstedts
- Riley, B.L. & MacDonald, J. & Mansi, O. & Kothari, A. & Kurtz, D. & von Tettenborn, L. & Edwards, N. (2008): Is reporting on interventions a weak link in understanding how and why they work? A preliminary exploration using community heart health exemplars. *Implementation Science* 3: 27
- Rothstein, H.R. (2007): Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology* 4: 61–81
- Sanderson, I. (2000): Complexity, Evaluation and Evidence-Based Policy. Paper for Euro-

- pean Evaluation Society, October 12–14th. Lausanne, Switzerland
- Sandberg, B. (2006): Utvärdering av Prime for Life – genomförande samt effekter av alkoholpreventiv utbildning vid Örebro universitet (Evaluation of Prime for Life – Implementation and Outcomes of Alcohol Prevention Training at Örebro University). SoRAD Research report 38. Stockholm university, Centre for Social Research on Alcohol and Drugs
- Sandberg, B. (2007a): Prime for Life – kompletterande analyser samt metodologiska erfarenheter från utvärderingen av Prime for Life vid Örebro universitet (Prime for Life – Complementing Analysis and Methodological Experiences from the Evaluation of Prime for Life at Örebro University). SoRAD Research Report 41. Stockholm university, Centre for Social Research on Alcohol and Drugs
- Sandberg, B. (2007b): Utvärdering av Prime for Life inom Försvarsmakten (Evaluation of Prime for Life in the Swedish Armed Forces). SoRAD Research Report 44. Stockholm university, Centre for Social Research on Alcohol and Drugs
- Sandberg, B. & Faugert, S. (2007): Perspektiv på utvärdering (Perspectives on Evaluation). Lund: Studentlitteratur
- SBU (2001): Behandling av alkohol- och narkotikaproblem. En evidensbaserad kunskapssammanställning (Treatment of Alcohol and Drug Problems. A Systematic Review of Evidence) Statens beredning för medicinsk utvärdering, Stockholm
- Sjölund, T. & Andréasson, S. (2004): PRIME For Life, rapport II. Uppföljning av gymnasieelevers och värnpliktigas alkoholvanor fem månader efter genomgången kurs (Prime for Life, Report II. A Follow-up of High School Pupils' and Military Conscripts' Alcohol Use Five Months after the Program). STAD-Rapport 25. Stockholm förebygger alkohol- och drogproblem, Stockholm
- Statskontoret (2002): Utvärderingar och politik II (Evaluations and Politics II). Statskontoret 2002:21. Stockholm: Statskontoret
- Sundell, K. & Soydan, H. & Tengvald, K. & Anttila, S. (2009): From Opinion-Based to Evidence-Based Social Work: The Swedish Case. *Research on Social Work Practice* 20: 714-722
- Sutton, A. & Abrams, K. & Jones, D. (2001): An illustrated guide to the methods of meta-analysis. *Journal of Evaluation of Clinical Practice* 7 (2): 135–148
- Tilley, N. (2000): Realistic Evaluation: An Overview. Paper presented at the Funding Conference of the Danish Evaluation Society, September 2000
- Tilley, N. (ed.) (2002): Evaluation for Crime Prevention. Monsey: Criminal Justice Press
- Vedung, E. (1998): Utvärdering i politik och förvaltning (Evaluation in Politics and Public Management). Lund: Studentlitteratur
- WHO (2004): What are the most effective and cost-effective interventions in alcohol control? World Health Organization Regional Office for Europe, Copenhagen, Denmark
- WHO (2009): Handbook for action to reduce alcohol-related harm. World Health Organization Regional Office for Europe, Copenhagen, Denmark.

