The Limits of Evaluation

Alfred Uhl

Green's Law:
"Anything is possible if you don't know
what you're talking about!"
(Arthur Bloch, 1985)

This text has been published in:
"Evaluation – A Key Tool for Improving Drug Prevention"
EMCDDA Scientific Monograph Series #5,
European Monitoring Centre for Drugs and Drug Addiction
Lisbon, September 2000
ISBN 92-9168-105-9

Introduction

Societies have always been confronted with the fact that individuals suffer from severe physical, psychical and social problems, and that these problems usually have an impact on other individuals and society. Societies are responsible for taking appropriate measures – but in what specific areas? and what are appropriate measures?

At first sight the answer seems simple. Most people agree that intervention should start before problems arise, that the focus should be on helping and convincing individuals rather than on coercive measures, that the private sphere of individuals and basic rights should not be violated and that measures should be as effective and efficient as possible; in other words: one should concentrate on primary prevention, use educational approaches, evaluate measures and base decisions on scientific evidence.

If we look into the matter more closely, we realise that things are far more complicated. There is an abundance of relevant problems to focus on, but we have only limited resources for prevention and evaluation and we have to decide on priorities. Very often we lack a sound empirical or theoretical basis to build strategic decisions on. We are aware that more specific research is needed, but at the same time realise that some very interesting and important research questions are beyond economic and/or epistemological feasibility. Commonly we are confronted with phenomena that are hard to assess reliably, with a long latency period between intervention and outcome, with low problem incidence, with a large variety of uncontrollable simultaneous influences and with important contextual factors changing rapidly over time and varying from region to region.

Currently we are observing an increasing demand for prevention and evaluation. This boom correlates with high expectations concerning the feasibility of prevention and evaluation. Many experts are fascinated by the increasing practical importance of their fields but feel uneasy about the growing gap between rising expectations and what can actually be achieved. This leads to a fundamental conflict how to proceed. The dilemma can be expressed by three problematic positions:

- anti-scientism
- deliberate ignorance concerning methodology
- cynical opportunism

The first position, "anti-scientism", is characterised by a rejection of quantitative science and basic research logic. The position is essentially irrational and fits nicely into contemporary alternative and esoteric trends. It is commonly quite popular among persons working in the practical field, who feel threatened by the demand to evaluate, who are not trained in the basic logic of research methodology and who are aware that many expensive evaluation projects do not produce any conclusive evidence anyway. The proponents of this position claim to be in favour of ”process evaluation” and ”qualitative research”, but what they aim at is not process evaluation and qualitative research, but a semantic trick to legitimise all forms of unsystematic and subjective data collection which are not suitable for assessing the value of certain approaches. Not wishing to be misinterpreted, I want to stress my conviction that ”process evaluation” and ”qualitative research” – if defined and understood properly – are central pillars of research and evaluation; nothing is further from my mind than shedding a negative light on these concepts.

The second position is ”deliberate ignorance concerning methodology”. The proponents of this position put much emphasis on experimental designs, inference statistics and objectivity, but they systematically avoid dealing with research limitations, statistical artefacts and methodology. They are commonly confronted with critics who reject their ”scientific” results based on practical experience and common sense. In this context I sometimes think that some researchers take the metaphorical expression ”blindness” as a synonym for ”scientific objectivity” far too literally.

The third position, ”cynical opportunism”, characterises those researchers and evaluators who realise that much of what is commonly produced as ”scientific evidence” is not conclusive at all, but who stick to common and widely accepted research strategies for pragmatic reasons – simply not knowing a sensible alternative. In a certain sense this understandable position comes close to fraud, but since virtually every critical researcher who has been working in the field for some time – including myself – has to accept that some of his own published work might be classified into this category, even rigorous critics of the current state of research commonly hesitate to apply such harsh labels. Is this cautious approach justified? I do not think so. We should not let inevitable mistakes in our past prevent us from naming problems explicitly and we can learn from them how to improve our trade. If, in line with the

above mentioned three problematic positions, we resort uncritically to broadly used fashions, repeat prevailing convictions, create the impression that we are able to cope with any tasks, ignore fundamental uncertainties, misuse the ambiguity of terminology knowingly and neglect basic methodological problems, we can attract more customers for the time being, but in the long run we risk the credibility of our profession.

There is definitely an alternative to these three positions. We can and should aim at informed and critical efforts to find lasting solutions. If we express the inherent uncertainties encountered in daily prevention and evaluation, put the finger on weak spots, reject tasks that are not feasible because of economic and/or epistemological restraints, are precise in terminology and do not avoid methodological problems, we risk disappointing potential customers, but in the long run improve our profession and contribute to a sound foundation for a good and lasting reputation.

Some of us, frustrated by the complexity of our task and tempted by the need to tender for projects for economic reasons, may sympathise with one of the three mentioned problematic positions, but we should rigorously reject them. I am convinced we may be quite optimistic in spite of the difficulties we encounter. Primary prevention and evaluation of primary prevention is a challenging task, but if we really understand our profession, despite all its limitations, there are numerous promising approaches, sensible options and solutions available to prevent us becoming depressed or cynical.

The COST-A6 Project ”Evaluation of Primary Prevention in the Field of Illicit Drugs”

In 1992 the European Commission commissioned a project on evaluation of actions related to problems with illicit drugs within the COST-A6 action (**C**o**o**pération **S**cientifique et **T**echnologique). At the initial project meeting the task was split up into 5 subtasks, each to be carried out by an independent work group. When WG II (Evaluation of Primary Prevention in the Field of Illicit Drugs) met initially two facts became immediately apparent:

- The list of properly evaluated primary prevention projects in Europe was rather short.
- Much more work was needed to base the project on a sound theoretical basis.

WG II felt that the terminology used in prevention and evaluation was far too ambiguous and vague for a systematic discussion, and that some central methodological problems inherent to the field were usually not adequately considered and analysed. As a result a consensus study on ”Definitions–Concepts–Problems” was initiated using the Delphi Approach (Lindstone & Turoff, 1975).

Most concepts referred to in this article were developed in this Study. The report (Uhl, 1998) is available free via the Internet (http://www.api.or.at/lbisucht.htm). Since limited space in this article does not allow the author to go into details, the reader is referred to the original source for more specific details.

Ambiguity of central terms and concepts

The inevitable ambiguity and vagueness of our language can be used – and commonly is regularly used – as a powerful tool to mislead others without having to deceive them directly, and to circumvent insoluble problems by reinterpreting central terms. A sensible, consistent and honest dialogue requires precise and unequivocal terminology. It does not matter whether words have different meanings as long as we are aware of this and can state precisely which meaning we refer to. Logic tells us that definitions are conventions and cannot be true or false. We cannot prevent others from inventing and using different terminologies. What we can do is to identify relevant definitions of scientific terms, to demand that others state what concepts they refer to, to try to convince others to drop inadequate interpretations and to suggest standardisation of terminology.

Everyday vs. scientific interpretation

The everyday meaning of ”prevention” refers exclusively to actions taken to prevent a problem before it occurs.

Scientists using the term ”prevention” usually understand it in a much broader sense including ”primary”, ”secondary” and ”tertiary prevention”, as defined initially by Caplan (1964). The COST-A6 WG II suggested splitting ”tertiary prevention” into two distinct subcategories resulting in:

- **Primary Prevention** to prevent problem onset.
- **Secondary Prevention** to intervene if a problem is likely to occur (prevention in high-risk groups) and/or in not yet fully manifested problems (prevention of problem manifestation).
- **Tertiary Prevention–Type A** to deal with fully manifested problems (prevention of further harm in addicts).
- **Tertiary prevention–Type B** or **Quaternary Prevention** to prevent problems from recurring once successfully treated (relapse prevention).

”Evaluation” according to the everyday sense is the process of determining whether a technique or strategy is of any value. Central questions are:

- ”Does it work?” and
- ”Is it ethically justifiable?”

Less important are questions such as ”How does it work?” or ”Why does it work?”

Quite different is the scientific definition of evaluation circumscribing a heterogeneous class of concepts ranging from documentation and description via hypotheses generating exploratory research to hypotheses testing confirmatory research.

If we understand ”prevention” and ”evaluation” according to their everyday meaning, many promising approaches towards reducing social and individual problems are not ”prevention”, and evaluation is commonly not feasible because of economic and/or epistemological restraints. If we understand both terms in the scientific sense, we may include a wide range of possible interventions and practically all projects can somehow be evaluated. What we offer may be very different from what our customers expect – but according to established scientific standards it is difficult to argue that this is not prevention or evaluation.

The following section describes some common scientific classification systems for ”evaluation” and some newer concepts developed by the COST-A6 WG II.

POI-Classification (Data Dimension)

This popular classification (e.g. Clayton and Cattarello, 1991) is based on the kind of data collected.

- **Process evaluation** is concerned with systematic assessment of the process during programme execution.
- **Outcome evaluation** is characterised by the question: ”Could the objectives of the programme be attained?”
- **Impact evaluation** addresses the question: ”Did any effects occur that were not explicitly planned?”

SPO-Classification (Data Dimension)

This classification by Donabedian (1980), developed in the context of quality assurance, is also based on the kind of data collected.

- ”**S**tructural data” describe structural conditions, e.g. place of intervention, qualification of programme executors, characteristics of target persons, etc.
- ”**P**rocess data” describe parameters of programme execution.
- ”**O**utcome data” describe intervention effects.

Context evaluation (Data Dimension)

Prevention programmes are developed for certain target groups under specific historic and cultural conditions (context). Context variables such as attitudes, knowledge, experiences and fashions vary from culture to culture and are subject to rapid change. Programmes proving effective under certain contextual conditions may turn out ineffective or even counterproductive in other conditions. Only close monitoring of context can guarantee that context differences and changes are identified and lead to necessary adaptations.

FS-Classification (State-of-the-Programme Dimension)

This classification by Scriven (1967) refers to the developmental state of the programme under scrutiny.

- **Formative evaluation** happens in the formative phase, i.e. while a programme is still being developed.
- **Summative evaluation** happens in the summative phase; i.e. after a programme has been finalised.

4-Phase model of programme development

The COST-A6 WG II suggested extending Scriven's FS-classification to a 4-phase model of programme development

- **Phase 1** = "**concept phase**" = "**preformative phase**"
  is characterised by the development of a prevention concept. This purely reflective phase precedes the "formative phase", starting with the intention to develop a new programme and ending with the first preliminary draft.
- **Phase 2** = "**development phase**" = "**formative phase**"
  is characterised by forming the programme based on observation and small trials until a final version without obvious shortcomings evolves.
- **Phase 3** = "**testing phase**" = "**first summative phase**"
  aims at confirming the effectiveness of the final version – ideally involving experimental or quasi-experimental designs. If such a "global assessment of effectiveness" – is not feasible, it is usually at least possible to test parts of the causal model behind the prevention approach (partial proof of effectiveness) or to derive effectiveness based on existing empirical evidence (historic deduction of effectiveness).
- **Phase 4** = "**routine phase**" = "**second summative phase**"
  is characterised by the routine application of the tested final version. In this last phase – since effectiveness has already been established – the main emphasis in this phase is on quality assurance.

DEC-Classification (methodological dimension)

A third way to classify evaluative research is "descriptive vs. exploratory vs. confirmatory evaluation"; a concept related to the kind of conclusions researchers may legitimately draw on epistemological grounds (methodological dimension).

Descriptive evaluation is a synonym for collecting and recording data, documenting phenomena, categorising them and summarising the findings, without directly aiming at the formulation of new hypotheses and theories.

Exploratory evaluation goes beyond mere description. Exploratory research ranges from collecting basic information in rather unexplored scientific areas to the hypothesis-driven development of new models and theories. There are no strict rules concerning procedures in exploratory studies. Basically anything with a chance of giving a better insight into relevant phenomena is possible and legitimate, as long as it is made explicitly clear that results of the exploratory phase are not final in any sense.

Confirmatory evaluation is not concerned with the discovery of new phenomena and/or the formulation of new hypotheses but with proving existing ones. Confirmatory evaluation uses the principles of probability theory and inductive statistics to discriminate substantial effects from irrelevant chance

effects. If feasible on epistemological and economic grounds, exploratory results should eventually be tested in confirmatory studies.

IE-Classification (evaluator dimension)

In the course of programme development and application it naturally makes a lot of difference who organises and directs evaluation. This can be done by programme developers and or programme staff (**internal evaluation**) or by external experts (**external evaluation**).

DSME-Classification

The COST-A6 WG II suggested integrating all classification concepts mentioned till now into a comprehensive 4-dimensional classification concept

- **data dimension**
  - structural data
  - process data
  - outcome data (=explicitly expected effects)
  - impact data (=not explicitly expected effects)
  - context data
- **state-of-programme dimension**
  - development phase = preformative phase
  - development phase = formative phase
  - testing phase = first summative phase
  - routine phase = second summative phase
- **methodological dimension**
  - descriptive approach
  - exploratory approach
  - confirmatory approach
- **evaluator dimension**
  - internal evaluation
  - external evaluation

The DSME-Classification, even though it integrates most scientifically established concepts to classify evaluation, is quite useful for describing evaluation but by no means complete or sufficient. In the following section several important varieties of evaluation – from need assessment through efficiency evaluation – will be mentioned and defined to round up the picture.

Needs assessment

Central in developing or implementing specific prevention programmes is ”need assessment”, i.e. to assess whether there is a need to become active in a specific problem area, and whether existing needs are already covered sufficiently by existing programmes and/or services.

Ethical evaluation

Personal convictions of evaluators naturally have a strong impact on any evaluation. Evaluators confronted with programmes usually judge immediately whether the techniques and strategies suggested are ethically acceptable to them – if they correspond to their personal values. It is sensible to make ethical evaluation an explicit topic.

Historical evaluation

A common form of evaluation is an expertise based on personal experience and knowledge of research. Since this process makes use of existing (=historical) data, the label ”historical evaluation” is appropriate.

Methodological evaluation

Methodological evaluation is concerned with the question whether empirical research is based on an appropriate epistemological and statistical basis.

Feasibility evaluation

If a programme is probably not feasible under real life conditions, it makes little sense to spend much effort on proving effectiveness. It is commonly possible to identify lack of feasibility without elaborate and expensive research designs in the formative phase of programme development.

Monitoring of unexpected adverse side effects

Monitoring of unexpected adverse side effects is a central aspect in every phase of evaluation. The number of potential problem areas is almost unlimited and new adverse side effects may arise if contextual conditions change.

Effectiveness evaluation

Effectiveness evaluation aims at proving that predefined goals (specific changes in the primary efficacy variables) can be achieved. If economic or methodological limitations render it impossible to design methodologically adequate experimental or quasi-experimental studies to provide global proof of effectiveness, it is usually at least possible to test parts of the causal model behind the prevention approach (partial proof of effectiveness). In some cases effectiveness can legitimately be derived from empirically based theory (historical deduction of effectiveness = historical evaluation). If we have strong evidence that a certain strategy works, we do not have to test it over and over again.

Quality assurance (QA)

The COST-A6 WG II recommended to define QA as ”evaluation of implementation quality under routine conditions”, i.e. after a programme has been implemented. The idea is to make sure that the level of programme execution does not decline. QA can be organised internally by the programme staff (quality management = QM) and externally by independent evaluators (quality control = QC).

Efficiency evaluation

Efficiency evaluation, not to be confused with effectiveness evaluation, includes cost-effectiveness analyses (CEA) and cost-benefit analyses (CBA).

- CBA compares the value of the benefits obtained from a programme with its costs.
- CEA compares the efficiency of approaches aiming at identical goals.

Methodological aspects relevant to evaluating prevention programmes

Some of the methodological problems encountered in the evaluation of primary prevention in the field of illicit drugs can be solved easily with appropriate strategies and moderate efforts, other problems require large efforts – possibly not justified in many specific situations – and still other problems constitute insurmountable research limitations of an economic and/or epistemological nature. A few particular methodological problems will be addressed in the following section.

Exploratory research is presented as confirmatory

The differentiation between exploratory and confirmatory research is of fundamental importance in science. Both forms of research depend heavily on each other and it is not justified to consider either approach as having a higher status. Unfortunately, Popper (1934), while stressing the importance of exploratory research, reserved the attribute ”scientific” for ”confirmatory research”. This caused the fateful perception that ”exploratory research” is inferior to ”confirmatory research” and tempted many scientists into presenting exploratory work as confirmatory. Since journals usually restrict the size of

research articles and commonly do not really care about the epistemological foundations of research, it is not always easy to identify such fundamental shortcomings in scientific articles.

Many variables are treated as ”primary efficacy variables”

The dichotomy primary efficacy variables vs. secondary variables is central to specify outcome variables within confirmatory research projects. Under ideal conditions confirmatory research projects should have only one primary efficacy variable. This can be a specific variable or an index generalising over a set of several variables. If one primary efficacy variable is not feasible, the nominal significance level has to be adjusted to the number of primary efficacy variables or alpha-inflation jeopardises all significance considerations. This phenomenon is well known to scientists, but very few care about the practical implications.

Inadequate surrogate variables

The dichotomy efficacy variables vs. surrogate variables is based on the concept of indirect measurement via causally linked dimensions – a common procedure in science. Surrogate variables are used to assess phenomena in cases where direct assessment of efficacy variables is difficult or impossible. Popular synonyms for surrogate variables are ”intermediate variables”, ”indicator variables” or ”proxy measures”. The use of surrogate variables is justified, given that causal relationships between intermediate variables and efficacy variables have been established through experimental or quasi-experimental research. It is highly questionable ,though, if the causality assumption is based on correlation only. Causality implies correlation but unfortunately the converse is not the case. In practical research we commonly observe that any correlations in line with research expectations are interpreted causally while implausible correlations are not so interpreted.

Short-term, medium-term vs. long-term effects

In designing programmes to prevent problems, we are primarily interested in lasting effects (long-term effects). Short-term effects and medium-term effects may play a role as secondary variables – to explain the mechanisms leading from specific actions to the desired effects. They may also serve as surrogate variables to assess the ultimate problem dimension indirectly. But they are definitely no substitutes for essential long-term effects. Evaluators unable to demonstrate lasting effects commonly imply that short-term effects are relevant per se, but this reasoning is not really convincing.

Knowledge, attitudes, personality characteristics, life skills, etc., as outcome variables

Variables such as knowledge, attitudes, personality characteristics, life skills, etc., commonly play a prominent role in the evaluation of primary prevention outcome and many researchers treat these variables as primary variables to assess effectiveness. A lot of empirical research suggests that changes in most of these variables do not result in relevant behaviour change or problem reduction.

Non-linear relationships

The assumption that an average reduction in problem indicators reduces the overall risk for problematic developments in a cohort is widespread but incorrect. The relationship between indicators and associated problems is commonly u-shaped rather than linear.

- According to the ”self-protection/self-medication hypothesis” (Uhl & Springer, 1996) persons with severe psychological, psychical and/or social problems tend either to avoid alcohol totally (abstainers) or to consume problematic quantities (problem drinkers). If we could successfully reduce the average problem level through primary prevention, we would expect the number of abstainers and the number of problem drinkers to diminish, resulting in more persons who consume alcohol and fewer persons with alcohol problems. A similar relationship has been demonstrated concerning cannabis.

Protective factors and risk factors

It is common practice in prevention research to look for variables that correlate with favourable and unfavourable outcomes and to call them ”protective factors” and ”risk factors”. Some of these factors are ”context variables” that cannot be changed (e.g. gender, ethnic background, etc.) and others are ”intermediate variables” such as attitudes, level of education, skills, etc., that can be influenced to a varying degree through appropriate interventions.

”Protective factors” and ”risk factors” can be seen in three functions:

- as ”indicators” to identify high- and low-risk groups, e.g. to identify target groups for specific secondary preventive measures,
- as ”surrogate variables” to assess indirectly primary efficacy variables that are hard or impossible to assess directly.
- as ”starting points” for preventive intervention strategies.

The first of these three functions (indicator) is relatively unproblematic, but the latter two functions (surrogate variables, starting points for interventions) assume causality – a highly problematic assumption if based on correlation studies only.

The problem of heterogeneity

A central implicit assumption behind many statistical procedures is homogeneity of effects; in other words, the assumption that all subjects react more or less homogeneously to interventions under scrutiny. Whenever the homogeneity assumption is grossly violated, statistical analysis based on this assumption may yield highly misleading results. A very illustrative example is the famous Grand Rapid Study (Borkenstein et al., 1964), a milestone in traffic research, relating the effects of alcohol consumption to traffic safety. The data initially seemed to prove that moderate alcohol levels help to reduce traffic accidents. Hurst (1973) reanalysed the data, controlling for heterogeneity (different drinking patterns of drivers) and showed that the published effect was artificial, caused by the fact that abstainers - who are of course always sober -on average cause four times as many accidents as sober non-abstainers. In all homogeneous cohorts the accident risk increased monotonously with higher alcohol levels.

Variation over time and context

Ideal in empirical research are stable causal relationships over time independent from situational factors (context). In prevention this is hardly ever the case. Specific circumstances and the susceptibility to certain preventive approaches vary greatly from culture to culture, from subculture to subculture and from cohort to cohort – and all these factors are subjected to rapidly changing fashions and trends.

Generativity

”Generativity” is a term for the well-known fact that small interventions may generate unpredictable effects evoking further unpredictable effects and so on, until an abundance of unpredictable and non-reproducible consequences have arisen; according to chaos theory (e.g. Steward, 1989) a common phenomenon in nature. In some situations the magnitude of generativity may outweigh systematic prevention effects by far. Generativity should not be confused with unanticipated systematic effects that may be included as expected effects in future evaluation.

Power considerations

We usually only have a realistic chance to prove programme effects

- if problem incidence is high,
- if the effects are massive and/or
- if the sample sizes are large.

If a certain problem in a given population, e.g. the manifestation of problematic drug consumption in the total population, has an incidence of less then 0.1% per year and if a programme reduces problem manifestation by 20% (relative effect) we can expect problem incidence to be reduced from 0.1% to 0.08% within this year. This is equivalent to an absolute effect of 0.02% or 1 out of 5000 persons. In order to have an acceptable chance to prove the intervention effect, we need a total study sample of more than 360 000 persons. Power considerations are commonly neglected in programme evaluation and designs with inadequate sized sample (underpowered designs) are widespread. Publication bias

Underpowered designs increase the likelihood of strong publication bias, caused by the fact that underpowered studies not yielding statistically significant results are hardly ever considered for publication, while underpowered studies containing statistically significant results are readily accepted.

Measurement problems with self-reported consumption

A widespread assessment strategy in prevention evaluation is to use self-reports, a relatively unreliable source of information. The possible magnitude of the bias can be illustrated by a finding from Uhl & Springer (1996) who were confronted with the fact that almost two thirds of several age cohorts who had already admitted illicit drug use decided not to ”remember” previous drug use 10 years later.

Dependency of observational units

It is a common situation in prevention that the observational units are sampled in clusters or groups (e.g. school classes) and treated as independent units anyway, causing an increased rate of significances by mere chance. In practical research the above problem is usually solved by simply ignoring it.

Summary and Conclusions

Prevention and evaluation of prevention are almost uniformly considered matters of high public priority. It is self-evident that actions to prevent problems before they occur are superior to interventions after problems are manifest. Besides, there is a broad consensus that we should try to assess whether these interventions produce the desired effects and to optimise our strategies. Actions proposed in this context should be well described, feasible, ethically justifiable, effective and cost-efficient and all this is somehow related to the term ”evaluation”.

If we look critically at the present state of prevention and evaluation, we face many fundamental problems. The central terminology is extremely heterogeneous and vague and many persons are not aware of this. Sound scientific evidence concerning which concepts are worth perusing and which concepts are obsolete is scarce. Prevention programmes are commonly evaluated inadequately or not at all. Many promising research questions in this field cannot be tackled because of fundamental economic, technical and epistemological limitations.

In spite of all this it is my conviction that researchers, preventionists and evaluators have the means to improve the state of prevention and evaluation if they are supplied with sufficient funds to do so and if the public does not expect miracles. Real progress is possible if we decide to be more precise in our terminology and conceptions, if we are ready to accept basic methodological limitations rather than ignore them, if we are ready to admit inevitable uncertainties and if we accept that some things just cannot be accomplished at present. We should withstand public pressure to accomplish impossible tasks and we should criticise colleagues who nourish unrealistic expectations because of calculated insincerity or ignorance.

What we need is concerted action to improve the quality of prevention and of evaluation. Important milestones in this endeavour in Europe were the Project ”Handbook Drug Prevention” by the Pompidou Group, the COST-A6 action resulting in a publication on ”Evaluation Research in Regard to Primary Prevention”, the EMCDDA Project to develop practical Guidelines for Prevention and the implementation of the EDDRA Data-Base by EMCCDA.

Literature

Bloch, A.: Murphy's Law Complete. Methuen, London 1985

Borkenstein, R. F.; Crowther, R. F.; Shumate, R. P.; Ziel, W. P.; Zylman, R.: The Role of the Drinking Driver in Traffic Accidents. Indiana Department of Police Administration, Indiana University, Bloomington 1964

Caplan, G. : Principles of Preventive Psychiatry. Basic books, New York 1964

Clayton, R. R., Cattarello, A.: Prevention Intervention Research: Challenges and Opportunities. In: Leukefeld, C. G., Bukovsky, W. J. (ed.): Drug Abuse Prevention Intervention Research: Methodological Issues. NIDA Research Monograph 107, Rockville, 1991

Donabedian, A.: Explorations in Quality assessment and Monitoring. Vol.1. The Definition of Quality and Approaches to its Assessment. Health Administration Press. Ann Arbor, 1980

Hurst, P. M.: Epidemiological Aspects of Alcohol in Driver Crashes and Citations. Journal of Safety Research, 5, 3, 130-147, 1973

Lindstone, H. A. and Turroff, M.: The Delphi method: Techniques and Applications. Addison-Wesley, Massachusetts, 1975

Popper, K. R.: The Logic Of Scientific Discovery, 8th rev. edition. Hutchinson, London, 1975 (original version 1934)

Scriven, M.: The Methodology of Evaluation. In. Tyler, R. W., Gagne, R. M., Scriven, M. (ed.): Perspectives of Curriculum Evaluation. Rand-McNally, Chicago, 1967

Springer, A.; Uhl, A (eds.): Evaluation Research in Regard to Primary Prevention of Drug Abuse. A COST-A6 Publication, European Commission Social Sciences, Brussels, 1998

Steward, I.: Does God Play Dice? The Mathematics of Chaos. Penguin, London, 1989

Uhl, A.; Springer, A.: Studie über den Konsum von Alkohol und psychoaktiven Stoffen in Österreich unter Berücksichtigung problematischer Gebrauchsmuster–Repräsentativerhebung 1993/94. Originalarbeiten, Studien, Forschungsberichte des Bundesministeriums für Gesundheit und Konsumentenschutz, Wien, 1996