

Commissioning, interpreting and making use of evidence on mental health promotion and mental disorder prevention: an everyday primer



Pre-print proof version; publication in press



Eva Jané-Llopis
Heinz Katschnig
David McDaid
Kristian Wahlbeck

**Commissioning, interpreting and making use
of evidence on mental health promotion and
mental disorder prevention:
an everyday primer**

Eva Jané-Llopis
Heinz Katschnig
David McDaid
Kristian Wahlbeck

This booklet was authored on behalf of the Taskforce on Evidence of the
European Commission Mental Health Working Party

Printing of the booklet was made possible by a generous grant from the Directorate
General of Health, Portugal, during the Portuguese EU Presidency, 2007

Many of the technical terms used in the primer have been explained in detail in the
Glossary section. In the text, these terms are indicated by an asterisk (*) when they
first appear

Authorship is presented in alphabetical order.
All authors have contributed equally to the development of this publication

This publication should be quoted:

Jané-Llopis E., Katschnig H., McDaid D., and Wahlbeck K. Commissioning,
interpreting and making use of evidence on mental health promotion and mental
disorder prevention: an everyday primer. Lisbon: 2007

The information and views set out in this publication are those of the authors and
do not (necessarily) reflect the official opinion/views of the European Communities.

Cover design by Sjoerd van Alst

Preface

Positive mental health* is crucial in today's society so as to stimulate growth and development and to contribute to prosperity, solidarity, social justice and increased quality of life across Europe¹. The increasing burden of mental disorders and poor mental health for individuals, families, society and the economy of Europe calls for action to prevent mental ill health and to promote mental health and well-being^{1,2}.

During a meeting of the European Commission Public Health Programme Mental Health Working Party¹ a task force on evidence was created, to produce a paper, this primer, which would provide an analysis of what is understood by evidence-based interventions, a critical overview of evidence*, and an appraisal of intervention evaluation of mental health promotion (MHP) and mental disorder prevention (MDP).

The aim of this primer is to support decision makers in assessing and evaluating available evidence, identifying potential biases, and supporting informed decision-making processes for the implementation of mental health promotion and mental disorder prevention. Section 1 presents some basic definitions of prevention and promotion in mental health and introduces concepts of bias* and generalizability*. Section 2 tackles what types of outcomes are needed in evaluating interventions. Section 3 looks at the different approaches to evaluation*. It highlights the role of quantitative studies, including experimental trials, underlines the role of complementary research methodologies, often called qualitative evaluation, and their importance in complementing quantitative research. It finally introduces economic evaluation* indicating how this can help address the needs of policy makers. Section 4 continues with how different results from different studies may be integrated and interpreted. Finally, section 5 concludes by suggesting some factors to consider when commissioning, conducting and presenting the results of research and how to strengthen the evidence base.

In addition to clarifying concepts on evidence for promotion and prevention in mental health, the goal of this primer is also to encourage and emphasise the need

¹ For more information on the EC Mental Health Working Party, see http://www.europa.eu.int/comm/health/ph_information/implement/wp/mental/mental_en.htm

for using the most appropriate evidence and/or evaluation methods to match the question being asked, to inform on the need for evidence-based assessment, and to help interpret the results of interventions, while highlighting the key issues that should be taken into account when reporting results of a given intervention. The primer stresses the contexts in which the use of evidence can be crucial, such as in providing information to feed decisions on the adoption*, adaptation*, and large scale implementation of a given intervention or strategy. Finally, it identifies guidelines (see “Suggested Reading” at the end of the document) that can support decision makers to judge or appraise the different types of evidence available for such interventions.

In the production of this document we are especially grateful to Mr Jürgen Schefflein, European Commission, Directorate of Health and Consumer Protection (SANCO) for his support during the writing process, to Dr Maria João Heitor dos Santos, Portuguese National Health Institute, for her technical assistance in the production of this brief and supporting the edition and printing of this publication and to the support of the Directorate General of Health, Portugal. Further acknowledgements are specified at the end of the document.

This primer is a short brief, written in everyday language, summarising key areas to take into account when making decisions for implementation. It is based on fully referenced technical scientific reports and publications intended for the specialist reader (see “Suggested Reading”).

We are aware, that decision-makers will always have to live with some degree of uncertainty. Nevertheless, we hope this primer will be useful in highlighting the strengths and limitations of different approaches to evaluation in order to improve public mental health across Europe.

Jämijärvi, July 10, 2007

Eva Jané-Llopis
World Health
Organization
Regional Office
for Europe

Heinz Katschnig
Ludwig Boltzmann
Institute for Social
Psychiatry and
Medical University
of Vienna
Austria

David McDaid
London School of
Economics and
Political Science
UK

Kristian Wahlbeck
STAKES and
Vaasa Hospital
District
Finland

Contents

Preface	3
Summary	7
1. Why evidence-based promotion and prevention in mental health?	11
2. What outcomes are of interest and how can they be assessed?	17
3. Study designs for the evaluation of population mental health interventions	23
4. From evidence to decision-making	35
5. Conclusions: Towards strengthening the evidence base for decision-making on promotion and prevention in mental health	39
Suggested Reading	49
Glossary	51
Acknowledgements	59

Summary

This guide is for policy makers and practitioners who are either commissioning research, are developing a public mental health strategy, or are choosing an intervention for implementation and are confronted by conflicting sources of information. It outlines criteria for good evidence on interventions for mental health promotion (MHP) and mental disorder prevention (MDP) and aims to support the critical appraisal of existing evidence and informed decision making.

The profound burden and costs of the health, social and economic impacts of mental disorders necessitate public mental health actions, not only to treat, but also to prevent disorders and to promote positive mental well-being. Decision-makers need robust evidence to support good decisions when designing strategies, as do professionals when selecting interventions for implementation. They need to consider not only whether something actually works, but in what circumstances and at what cost. Many interventions may be culturally sensitive, and need to be evaluated separately for each setting, culture or region. To accommodate the complexity of public health interventions, evidence should be considered in its broad terms.

Without evidence of effectiveness it is difficult to make a case for investment in mental health. Moreover in the absence of good evidence there is a danger that policies and practices may be introduced that are either harmful, wasteful, or both. This is particularly true of complex interventions in public health and health promotion, which can be delivered across many different sectors and are not usually subject to the same mandatory requirements for evaluation as health care treatments.

When assessing the evidence for an intervention, the pooled evidence from different sources and studies should be scrutinised, through a stepwise critical appraisal of studies and the evidence they entail.

First, the study methodology should be assessed for the risk of systematic bias, i.e. the risk that the study's outcomes are a result of something else than the intervention under evaluation. Hierarchies of study categories have been constructed to support the assessment of risk of bias. The risk of selection bias, i.e. that the intervention group and the control group differ from each other at baseline and thus are not comparable, is minimised in the well performed randomised controlled trial (RCT). To avoid assessment biases, the study outcomes should be measured independently whenever possible, regardless of whether the study is a

RCT or not. In MHP and MDP, RCTs are not always feasible, and, ordered according to increasing risk of systematic bias, non-randomised controlled trials, cohort studies, case-control studies, cross-sectional studies and ecological studies also contribute to the evidence.

Second, even in the case of high quality study methodology, every study needs to be assessed for applicability to the real world, i.e. its generalizability. Evidence from a well-performed RCT may offer no support whatsoever to decision-making if the population of the study differs from the target population, if the intervention cannot be replicated, or if the outcome measured lacks relevance. In many cases, cohort studies or ecological studies are performed in more “real-life” circumstances than RCTs. Pragmatic RCTs, also known as, naturalistic RCTs, are an effort to combine the advantages of RCT (low risk of selection bias) with the advantages of observational studies, but may not always be feasible in MHP and MDP.

Third, qualitative studies offer a complementary approach to generating evidence which sometimes has been overlooked. Quantitative and qualitative approaches are not mutually exclusive, and qualitative studies answer many questions which cannot be answered by quantitative studies.

Fourth, evidence needs to be assessed in combination with cost data. It is important not only to look at the evidence on the effectiveness of promotion and prevention strategies but also their resource implications.

The complex process of assessing existing evidence for MHP and MDP can be supported by the use of guidelines for critical appraisal of studies, developed by different groups. This document presents an example of a matrix that matches types of research questions that want to be answered with corresponding types of evaluative studies that will most likely be able to generate that answer accurately.

Lastly, in each case, transferability issues of the available evidence needs to be considered when making decisions about going to scale in different settings, cultures or regions. Many interventions in MHP and MDP are culturally sensitive, and need to be evaluated separately in each setting, culture or region.

The primer concludes that there is sufficient evidence for promotion and prevention in mental health; quantitative and qualitative studies are available that can provide answers to the questions that we might want to know. Considering and weighing evidence by appraising how it has been generated can help informed decision making. However there is still room to, in general, improve the quality of the research designs applied to evaluate the efficacy and effectiveness of interventions for prevention and promotion in mental health. These interventions should include long-term follow-ups to give sufficient time for interventions to show effect and to provide an accurate estimation of the duration of effects. It is important to note that proven efficacy or effectiveness is no guarantee that programmes or policies will work similarly in different cultural or economic environments. New studies should focus on identifying the mechanisms and processes of adaptation and reinvention without losing initial efficacy.

There are many guides to the generation and use of evidence. Our aims in writing this short primer were to set out as clearly as possible the strengths and limitations of different methods of evaluation, to consider potential bias in study results, to examine how generalisable results may be to real world settings, to describe how quantitative and qualitative approaches should be combined to enhance the use of information, and to suggest how to strengthen the evidence base of mental health promotion and mental disorder prevention at the European level.

Key Points

- Developing and implementing policies without evidence-based knowledge may be harmful and wasteful
- Different research methodologies each have their own strengths and weaknesses, and answer different types of questions
- Systematic reviews of available evidence should be carried out before starting new studies to avoid duplication of efforts
- 'Would it work' questions can only be answered by quantitative research methodologies
- In whatever situation apply always the highest possible standards of research designs to answer the question of intervention effectiveness
- Qualitative methodologies are best suited to answering questions on satisfaction with services, identifying public attitudes, and helping to identify effective ingredients of interventions and user-defined outcomes
- Research on the cost-effectiveness of interventions is important in guiding health and non-health resourcing decisions
- There are well constructed guidelines available on how to conduct studies and report research results which commissioners could oblige evaluators to adhere to
- Irrespectively of having good evidence, when informing decisions about implementation and going to scale, always consider for each case generalizability and transferability of the available evidence, in different settings, cultures or regions
- All programme implementation should include evaluation, encompassing both process and outcome indicators

1. Why evidence-based promotion and prevention in mental health?

To implement mental health promotion and mental disorder prevention interventions it is crucial that policy makers and practitioners have sufficient information. Typical questions asked include “What types of interventions can be implemented?”, “Do they generally work and what effects could we expect?”, “Can one of these interventions potentially also work in my country or given setting?”, and given that this were true, “What would be the implications if we would use the same resources on some other intervention?” Evidence can help answer many of these questions.

What is evidence?

Evidence is commonly referred to as information with the aim to confirm a fact, offering proof that a certain statement reflects the actual reality, or the opposite, that a statement conflicts with the truth. The word evidence is used from different perspectives, ranging from testimony of expert witnesses in court to complex experimental research.

A starting point of the evidence debate in the health field is *Evidence Based Medicine** (EBM). Growing concern about the use of treatment methods not based on state-of-the-art knowledge led to the creation of EBM³. In EBM, individual studies on a specific topic are critically appraised* in respect to how trustworthy or free of bias they are, and their results are synthesized (usually by systematic reviews* and meta-analyses*), with the findings then cast into evidence-based practice guidelines. EBM emphasises the need to generate knowledge through controlled empirical research that can provide the most unbiased results, and considers the Randomised Controlled Trial (RCT)*, as the gold standard evaluation method to generate reliable evidence.

*Evidence Based Public Health** follows the same principles as EBM, but includes a larger variety of evaluation methods that can capture the idiosyncrasies of the social context and the nature of public health interventions. Because of the complex nature of social interventions, evidence in public health may be best achieved by using both experimental and non-experimental methods. A frequent criticism has

been the insufficient resources available to fund large scale experimental studies for complex public health interventions, which are essential for obtaining reliable evidence.

Regrettably, the word 'evidence' is used in the mental health promotion – mental disorder prevention field to refer to anything, from the results of a rigorous research study to the views of the general public. In addition to “evidence” and “evidence-based” being vague terms frequently used rather loosely, too often in this field, any intervention that has been subject to the most marginal of evaluations may be considered to be “evidence-based”, or often also wrongly named “best practice”.

This primer identifies the research methods of gathering information, their strengths and limitations, and how these result in different types of evidence that are suited only to answering particular questions.

Choosing a topic for promotion or prevention in mental health

Mental health promotion and mental disorder prevention interventions can be implemented across different settings (e.g., school, workplace, community), for different age groups (children, adolescents), or for different population groups (e.g., migrants, pregnant women). Some interventions target the whole population (e.g., an increase in alcohol tax, a media campaign to promote mental health), some target population subgroups at higher risk for developing a mental disorder (e.g., children of parents with a mental illness), and some target high-risk individuals who have symptoms of a mental disorder but not a disorder at that time (e.g., a school intervention for children with symptoms of depression).

Decision makers are firstly faced with identifying for whom, where and what type of interventions would be needed to improve public mental health. After undertaking the appropriate assessment of the situation and needs of a given country or region, the topic for intervention is chosen. This process is beyond the scope of this primer, where the starting point begins once the topics for intervention have been identified.

The Scenario boxes present throughout this primer one hypothetical example of an intervention throughout the paper.

Scenario 1. A school-based intervention to promote mental health

A community in a region where there is mixture of populations from different ethnic backgrounds and levels of socio-economic status decides to support the development and implementation of a school-based mental health promotion programme for their children and adolescents.

The aim of the programme is to enhance social competence and self-esteem of the children, strengthening positive peer relations along with creating increased understanding of cultural differences, tolerance, and development of social networks, to empower the children and support their positive development.

To achieve these aims, the teachers in the school are trained to teach social, emotional and cognitive skills to the children, in an interactive and supportive environment. To ensure that the children will internalize the skills taught, the overall school context is also changed to help children to adjust to the school environment and stimulate opportunities to generalise the social skills learning.

Promotion or prevention?

This school programme described in the Scenario box is a typical example of a mental health promotion intervention. However, the terms “promotion” and “prevention” are understood in different ways, and while some believe them to be synonymous, others conceptualise them as completely different. Facts 1 presents more detailed definitions, following the proposal that within a broad mental health promotion strategy, prevention of mental disorders could be one of its aims and outcomes⁴. When deciding what type of intervention to implement, the choice between prevention, promotion or both is important as it will impact on decisions concerning evaluation methods and outcomes.

Facts 1. Mental health promotion and mental disorder prevention

Mental health promotion (MHP) implies the creation of individual, social and environmental conditions that are empowering and enable optimal health and development. Such initiatives involve individuals in the process of achieving positive mental health and enhancing quality of life. It is an enabling process, done by, with and for the people.

Mental disorder prevention (MDP) aims at reducing occurrence, frequency, and re-occurrence of mental disorders, the time spent with symptoms, or the risk for a mental illness, preventing or delaying their occurrence and also decreasing their impact in the affected person, their families and society.

Scenario 2. A school-based intervention to prevent mental disorders

In our school example, a mental disorder prevention programme could focus instead on identifying children who have for example a large number of depressive symptoms but who do not meet diagnostic criteria for depression. Only these children would be involved for instance in a cognitive behavioural and problem-solving-skills intervention to increase their understanding and recognition of emotions and improve self-esteem, while creating a peer support group. This specific intervention, could be a component of the more comprehensive intervention described previously in Scenario box 1, or a stand alone intervention in a school where there is no decision to implement a universal mental health promotion programme.

Efficacy, effectiveness and cost-effectiveness

But, before deciding for implementation, is there evidence that prevention and promotion interventions work? Recent publications of the World Health Organization^{4,5,6,7} and the International Union for Health Promotion and Education⁸ have reviewed the state of art on what works in mental health promotion and mental disorder prevention, underlining that these interventions bring about positive health, social and economic outcomes. However these reviews and a recent overview of action across 30 European countries⁹, also underline how many interventions implemented across countries have not been evaluated, and therefore there is no knowledge of whether they have or have not had an impact on improving mental health and/or preventing mental health problems. For example, a review of 197 different promotion and prevention programmes in mental health for

children aged 0 to 6 that were identified as representing “best practice” across 17 European countries, concluded that only 11% could show any evidence of their efficacy*¹⁰.

To decide on what to implement, it is crucial to have information on what works, what are the effects expected, and in what contexts it can actually work. So evaluation of promotion and prevention interventions is essential; without such evaluation there remains a danger that effective interventions may be overlooked, and, perhaps more worryingly, resources may be invested in strategies and interventions that may not be that effective or even harmful. Evidence can be provided by efficacy, effectiveness* and cost-effectiveness* evaluations. Often these terms are used interchangeably in everyday language but they in fact refer to different things (Facts 2).

Facts 2. Efficacy, effectiveness and cost-effectiveness

Efficacy studies answer the question “*Will it work at all?*” Such studies tell whether an intervention can work under ideal conditions and are crucial to evaluating the impact of an intervention. Because these evaluations are undertaken under ideal conditions or in an experimental setting, results of efficacy studies cannot be globally generalised to real world settings.

Effectiveness studies answer the question “*Does it work in the real world?*” In all areas of health, but especially for promotion and prevention, such studies are crucial as they consider whether the findings of efficacy studies (in ideal conditions) can indeed be replicated in a variety of contexts and settings.

Cost-effectiveness studies answer the question “*Do the benefits of the intervention represent good value for money?*” Cost effectiveness or economic evaluation is crucial in decision-making. Resources are limited and the collection of information not only on effectiveness, but also on both the costs and resource consequences of interventions plays a key role in the area of health promotion and public health.

Issues to consider in appraisal of studies

The evidence generated through evaluation needs thereafter to be assessed. Evidence has to be interpreted and weighted, being aware that different sources of evidence have different strengths and limitations. Syntheses and meta-analyses that pool the research results of different studies provide stronger information on whether a type of intervention is repeatedly effective.

The assessment of evaluations is crucial, but alone will not be sufficient to guide decision-making. In addition, to whether an intervention works, other key issues related to evidence - such as the application of ethical standards or the real impact of an intervention on the population - will be essential in the decision-making process. Critical appraisal is used to assess the quality of evaluations and can help assess the strengths and limitations of study findings¹¹, while taking into account all the other aspects essential to assess the value of the evidence. So it looks at the appropriateness of using a specific study design to evaluate the effects of an intervention and how likely is it that the results might be susceptible to bias. But critical appraisal goes beyond the quality of study designs; for example, it can look at the magnitude of effectiveness, meaning how important or clinically significant a given result is in its context; assess the credibility of the study, for instance “is the

study population relevant to the wider population for whom the intervention is suggested?"; determine how complete a study is – for instance "does it look at outcomes of relevance* to all stakeholders?", "has it looked at costs and cost effectiveness?", "have negative results been interpreted appropriately?"; and consider the transferability of a study, "does it provide enough information on the intervention and the context in which it was delivered?".

Summary points

All the concepts presented in this section are related to evidence-based public health because the decision-making process and the implementation of public health policies, programmes and practices require good evidence on feasibility, efficacy, effectiveness, and cost-effectiveness. Decisions will also be informed by many other factors including the acceptability of a policy to a target population. There will be ethical and political considerations to be made, for example, as to whether policy makers may be willing to sacrifice some absolute gain in health in order to reduce inequalities in health status by focusing interventions on specific vulnerable population groups. Some of these issues are discussed in this document but it is important to note that there is no such thing as "value free evidence"; decisions will always be informed by various values and perspectives.

2. What outcomes are of interest and how can they be assessed?

The previous section made the case for public mental health interventions. Once the decision to act is made, it is important to clarify which impacts of mental health promotion and mental disorder prevention programmes are desired and how they should be measured in the context of evaluation. Transparent reporting of how the outcomes have been chosen is essential for assessing the value of the study. This section sets out some key issues to consider when determining whether the measurement of results in any particular study is relevant to a specific policy question.

How can outcomes be measured?

Mental health promotion and mental disorder prevention can have an impact on many domains in addition to mental well-being. Examples include changes in quality of life, change in use of services, changes in behaviour or attitudes, change in life expectancy, socio-economic impact, and even political change. When evaluating mental health promotion and mental disorder prevention programmes, these outcomes can be measured at the population level, group level (e.g., prisons, schools or ethnic groups) or individual level.

Furthermore, outcomes can be derived from self-rating (e.g., by depression rating scales or questionnaires measuring attitudes) or by external rating (e.g., by family members or independent external assessors).

Outcomes may be assessed at a single point in time (a.k.a. "cross-sectional") or they may reflect a period of time, such as the year preceding the assessment.

Thus evaluative mental health research faces a spectrum of possible outcomes to choose from.

Scenario 3. Measuring the effects of a school-based intervention

In assessing the impact of the school-based mental health intervention, one has to choose between measures focusing on the well-being of individual pupils (e.g., their mental well-being, experience of bullying, drug use or symptoms of distress) and/or social level outcomes (such as school level of pupils' absence, the school's level of educational achievement, or school climate). Long term follow-up outcomes could include employment, crime or even death rates.

Are outcomes predefined?

Ideally, in evaluative studies, appropriate outcome measures are chosen during the study development phase and according to a research hypothesis that matches the policy question. It is important to realise that convenient and readily available outcomes are not necessarily the most important or relevant ones¹². One issue to be cautious of is a study where a vast amount of data has been collected and only certain outcomes are selected and reported after the end of the study. This may seriously bias the findings, because the researcher may choose to present the positive results only and to hide any negative results. All predefined outcomes should be reported, i.e., it is essential that also negative results are reported.

Are all relevant stakeholders involved in choice of outcomes?

In the process of choosing outcomes the interests of different stakeholders need to be taken into account¹³. A well-designed evaluation takes into account the perspectives of the intervention target group, the public, policy makers, professionals, and other stakeholder groups. Involvement of stakeholders in research design and choice of outcomes may help to perform research that is relevant and this also facilitates adoption of study results by the stakeholders. It is important to involve the target group in the evaluation (i.e., targeted pupils, and not only teachers, should be informants in retrieval of data on impact of a school-based intervention).

Scenario 4. Stakeholder views on important effects of school-based intervention

Different stakeholders may have different expectations on a school-based intervention. Teachers may focus on educational achievement, pupils may expect improvements in school climate, and administrators may focus on enhanced productivity. Improvements in mental well-being (and therefore a reduction in the need for curative services) may be of key importance to service providers. Another factor may be cost, while this should not be seen as an outcome, ideally, it can be examined alongside any chosen outcomes in an economic evaluation

In the end, the choice of outcome cannot avoid including a value judgement. When assessing an evaluative study, the judgement of whether policy-relevant outcomes have been used is ultimately based on societal values.

Are the outcome measures relevant and important to people?

Outcomes can focus on specific details or be general in nature. For most people, general outcomes, such as quality of life and level of functioning, are more relevant than specific outcomes (e.g. a psychiatric symptom). Preferably, outcome

measures should reflect what matters to people.

Effectiveness and cost-effectiveness of mental health interventions and policies can be more easily compared to effectiveness of interventions in other areas of the health sector or beyond if a generic outcome measure common to all can be used.

Are final or intermediate outcomes used?

It is not always easy to measure how successful mental health promotion and mental disorder prevention programmes have been in reaching their goals. One reason is that with public health and mental health promotion interventions, it may often take many years before final health outcomes may be seen. In prevention research the goal is the prevention of a negative event. These events, e.g. suicide, may be very rare and thus seldom occur among people in the study. In these cases researchers may have little alternative than to use some intermediate or proxy outcome, which are known to be or assumed to be linked to final outcomes.

Scenario 5. Final and intermediate outcomes of school-based intervention

For example, in a suicide-prevention school-based intervention, ideally the key outcome is likely to be the (decreased) rate of completed suicides. As the actual number of suicides is often very small, attempted suicide, suicidal ideation, treatment referrals, etc., might be used as intermediate outcome measures.

However, intermediate outcomes introduce a considerable degree of uncertainty in study interpretation if they are used as proxy measures (e.g., attempted suicide may be rather a cry for help and not related to suicides). Generally speaking final outcomes give more valid results, and intermediate outcomes should be avoided if possible.

Have possibilities to detect harm been taken into account?

Mental health promotion and mental disorder prevention programmes can also have harmful effects. This is easily overlooked. Even intuitively beneficial interventions, e.g., preventive psychological debriefing for general population victims of traumatic events to prevent posttraumatic stress disorder can turn out to be harmful when evaluated^{14,15}. Ideally equal weight should be given to the assessment and reporting of possible harm as to the assessment of possible benefit of the intervention, in order to enable a harm vs. benefit assessment.

Which instruments are used to measure outcomes?

In evaluative mental health research, the use of rating scales is widespread. Scales are used to measure a wide spectrum of symptoms, behaviours, perceptions and attitudes. Most of the scales used may not have been tested sufficiently to make sure respondents use them in a consistent manner. Rating scales often fail to capture what the intervention ultimately aims to achieve. Many truly relevant endpoints - such as mortality, employment, or ability for independent living - do not suffer from the problems seen in scales, and are thus preferable if appropriate for the study.

Does responsiveness of measure fit to outcome of interest?

Many scales measuring mental health provide scores that are difficult to interpret and it is not clear how much these scores should change in order to matter to people. If an outcome measure is very sensitive, i.e., a change in outcomes is very easily recorded, then a statistically significant difference can be seen even when the change is so small that it has no real impact. Conversely choosing too rare or insensitive an outcome (which cannot be affected by the intervention or policy change) may result in a falsely negative study¹⁶.

Scenario 6. Impact of choice of scale and outcome measurement in a school-based intervention

Pupils' depressive symptoms may be measured by a sensitive rating scale. Each pupil provides a self-rating, and it may for example turn out that the average rating is statistically significantly better in schools or classes that received the intervention than in comparison schools or classes. If a sensitive scale is used even a small difference between targeted and non-targeted schools or classes can be statistically significant (especially if many schools participate in the study). Whether the finding is relevant in real life cannot be judged from the statistical significance* alone, but one needs to consider also the size of the difference between targeted and non-targeted groups.

Are outcomes assessed independently?

Data collection should preferably be performed by assessors independent of those commissioning or performing the study. Sometimes independently collected data on outcomes can be taken from population surveys, routine register data and official statistics. In any study with a control or comparison group, to minimise any potential bias the independent assessors should ideally not know who is in the control or in the intervention group. This is called "blinded" assessment. If assessment is not blinded, then assessment bias* may occur, i.e., the assessors might be inclined to rate outcomes in the desired direction (usually in favour of the new "experimental" intervention).

Is the follow-up long enough?

A too short follow-up period will miss a lack of sustainability or a "sleeping effect", i.e., an effect which is detectable only after a longer period of time, and will overestimate short-term non-sustainable effects. Many evaluation studies of mental health promotion and mental disorder prevention are too short. Long-term follow-up of a school intervention can for instance be possible through the use of census or register data (such as health or criminal records).

Facts 3. Crucial issues in the assessment of a study's choice of outcomes

Definition of outcomes before start of evaluation

Stakeholders' involvement in outcome selection

Use of outcomes relevant to people

Final outcomes offer more certainty than intermediate outcomes

Possibilities to detect harm

Choice of measurement method: hard facts preferable to scales

Fit of outcome measurement's responsiveness to change

Independent measurement of outcomes

Provisions for long-term follow-up

3. Which study designs can be used for the evaluation of population mental health interventions?

After discussing in the previous section mental health promotion and mental disorder prevention activities and their outcomes, this section deals with how to properly discover whether an intervention works by using the least biased (i.e., the most error free) study design, and also whether an intervention gives value for money.

The randomised controlled trial (RCT) has long been regarded as the “gold standard” for acquiring knowledge about the efficacy and effectiveness of medical interventions, which is documented by the fact that it always ranks on top of so-called “hierarchies of evidence” (next to systematic reviews and meta-analyses) (e.g., Cochrane Collaboration, see “Suggested Reading”). The reason for this high appreciation in medicine is that, by random allocation of patients to different interventions, the risk of bias i.e., the probability of error or of getting wrong results due to selection is minimized. This and other quality characteristics are usually referred to as the “internal validity” of a study.

However, while the logic of the RCT undoubtedly justifies this high ranking, the weakness of many such trials is the lack of generalizability of their results (usually referred to as “external validity” *), i.e., the questionable applicability* of results to the “real world”. This is because most trials are carried out on narrowly defined and highly selected populations and under specific circumstances¹⁷. Also, when evaluating complex psychosocial interventions such as mental health promotion and mental disorder prevention, RCTs are not always feasible and may even not be appropriate, since it might be impractical to find a control group* or not feasible to allocate participants randomly.

These doubts about the appropriateness and feasibility of the RCT have led to increasingly complex debates about other study designs within the quantitative approach, and about qualitative approaches as alternative or supplementary approaches. Also, and not the least important for decision makers, the issue of costs in relation to effectiveness are more and more coming into the foreground. Choosing the most appropriate study design has therefore become an important issue and the following three sections will discuss the advantages and disadvantages of these different types of studies.

However, before doing this it has to be stressed that, regardless of which study design and approach is used, the proper conduct of the intervention (i.e., avoiding *implementation bias**) and the valid and reliable collection of data (i.e., avoiding *assessment bias*) are indispensable pre-requisites for any successful study. A badly conducted RCT might deliver less valid results than a well conducted observational study*.

3.1. Quantitative studies: Classical approaches to studying the impact of interventions and types of study designs

Quantitative studies have the highest reputation in science; in fact, usually quantification is regarded as the essence of science. In the planning and conducting of quantitative studies a clear pattern is followed which has to be laid down in a study protocol.

Participants in intervention studies can be individual persons, but also, for instance, schools or whole communities. In a first step the participants have to be determined by inclusion and exclusion criteria and the procedures for selecting participants have to be laid down. In addition the interventions to be applied to participants have to be defined and standardized, and those giving the intervention have to be trained. Baseline and outcome measures and the corresponding instruments for assessing them have to be chosen, and the time frame of the study, together with the pattern of assessments, have to be determined. Also, those who collect the information have to be trained (in order to obtain a high reliability in data collection). Finally, the statistical methods for analysis have to be defined. The most relevant issue, however, is the choice of the study design¹⁸.

The most basic distinction is, whether a) investigators assign an intervention to the study participants or influence the way the intervention is administered, or b) whether they only study the effect of interventions or programmes which are carried out anyhow. The first type of study is called *experimental**, the second *observational*.



Figure 1. Types of Quantitative Studies 1: Experimental vs. observational

Experimental studies

If the investigator assigns an intervention there are two possibilities – to do the study without or with a control group. Especially in prevention and promotion research, out of practical reasons there is often no control group and the effect of the intervention is just documented in a “pre-post-design”, i.e. in comparing measures before with those after the intervention. Since this does not rule out that

the effect might have also occurred without the intervention, this is a weak study design.

The more appropriate approach for answering the question whether a specific intervention works is the *controlled trial**, i.e., a type of study where a group of people or of organizations/communities (the “intervention group”) is receiving an intervention, while a control group is receiving a different or no intervention. The choice of control intervention (nothing or an alternative intervention) is important, as this can also influence results.

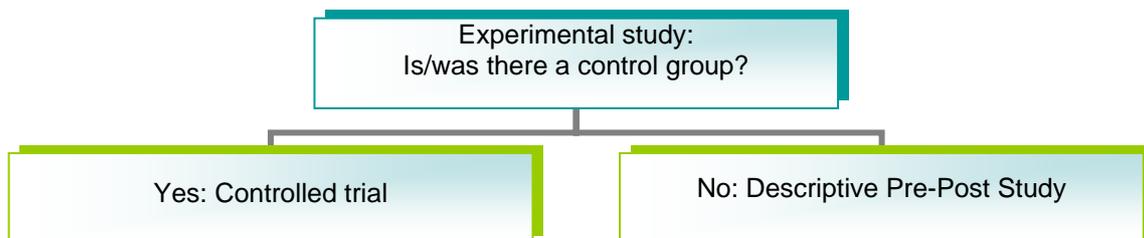


Figure 2. Experimental studies 1: Controlled vs. not controlled studies

The randomised controlled trial (RCT), a specific form of controlled study, is regarded as the optimal experimental study. In RCTs, participants (e.g., individuals, school classes) are allocated at random to either group. Randomisation avoids selection bias*, i.e., it avoids that selected individuals receive the intervention and others receive a different or no intervention. If study groups are not formed by randomization, it may happen that one group (e.g. the intervention group) consists mostly of girls and the control group mostly of boys. In this case one might not be able to decide whether it was the intervention which produced the desired effect or whether gender was decisive.

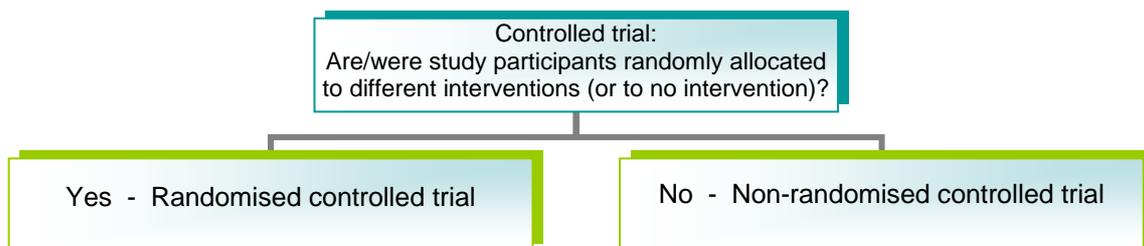


Figure 3. Experimental studies 2: Randomised vs. Non-randomised controlled trials

Assessors should ideally not know (be “blinded”) to which group the participant has been allocated. In prevention and promotion programmes it is almost impossible that study participants are not aware whether they get the intervention or not, because of the nature of psycho-social interventions. Therefore RCTs can still be subject to a degree of bias, as participants' and evaluators' awareness might, for example, influence interpretation of results (“assessment bias”).

Scenario 7. Randomisation in a school-based intervention

Concerning our school example a RCT would imply: (1) to choose individual pupils (or whole school classes) for inclusion in the study according to predefined inclusion and exclusion criteria; (2) to randomise either the pupils or the classes (“cluster randomisation”*) to an intervention programme or to no intervention; and, (3) to assess outcome after implementation at a predetermined time period.

Observational studies: Cohort and case-control studies

If experimental studies are not practical or feasible, *observational designs* are used as the next best type of study. Here usual practices are observed and studied, either prospectively by a *cohort study** or retrospectively by a *case-control study**.

In observational studies, interventions that are being delivered in their usual settings are studied. Observational studies are prone to selection, confounding* and assessment bias, and thus appear lower down the evaluation hierarchy in relation to RCTs. Nevertheless, for specific questions (e.g., studying side effects* or long-term harmful outcomes) a well conducted observational study may be better than a small and poorly designed and conducted RCT. In some circumstances, observational studies may be the only practical possibility to answer specific research questions. However, caution is justified. For example, for years observational studies had shown that women who took hormone replacement therapy (HRT) were less likely to develop heart disease, yet recent RCTs suggest the opposite¹⁹.

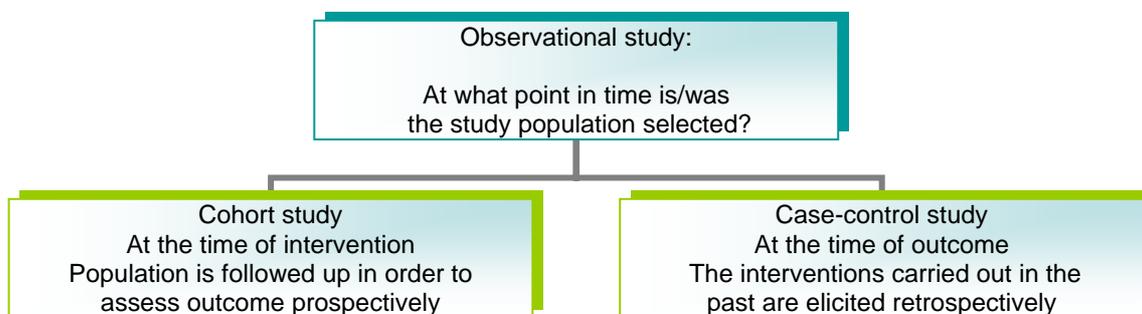


Figure 4. Observational studies: Cohort and Case Control Studies

If the study begins with an intervention (which is administered routinely in its usual setting) and follows participants over a period of time prospectively in order to measure outcomes, then it is called a *cohort study* (also referred to as prospective or longitudinal study). If there is no control group then it is a descriptive cohort study. If a control group, who did not receive any intervention or a different one, is also followed up prospectively then this is called an analytical cohort study (at times also referred to as quasi-experimental* study), since this approach can, to a certain degree, rule out that changes over time obtained with the intervention would have occurred also without an intervention. However, since by their nature such studies cannot use randomization, other factors than the intervention itself might account for the differences between the intervention and control group. This problem can

only partly be coped with by “matching” the control group with the intervention group according to major variables such as gender, age, social class and the like.

Scenario 8. School-based cohort study

Concerning our school example a cohort study would look into pupils, classes or schools where mental health promotion or mental disorder prevention programmes have been carried out and would follow them over time in order to assess an outcome after a specific time period. For purposes of comparison, also pupils, classes or schools on whom no intervention has been applied could be followed in order to assess possible changes in the outcome variables.

If the study begins with an outcome (e.g., selects persons who are depressed and others who are not depressed) and looks back in time to see whether and which individuals have been exposed to an intervention or to which type of intervention, this is called a case-control study. Because of their retrospective nature, case control studies are prone to recollection bias (where prospective cohort studies are not) and are placed lower on hierarchies of evidence. Case control studies have a control group by definition and the task is to make the intervention group and the control group as similar as possible (by a procedure called 'matching') in order to avoid wrong conclusions (that the results are caused by the intervention while they are due to some confounding variable) – much in the same way as described above for cohort study with control groups (where also selection and confounding bias may occur).

Scenario 9. School-based case -control study

Concerning our school example a group of adolescents or young adults who are depressed at this present time, are compared with a group of adolescents who are not depressed in respect to whether they had taken part in mental health promotion or mental disorder prevention programme while they were still attending school.

Where do “Pre-Post Designs” and “Interrupted time series analyses” fit into the above system?

Sometimes one reads that a study follows a “*pre-post design*”. First, the term pre-post can be understood in a very general sense – every intervention study* has to measure the status of the participants before and after the intervention. However, as already mentioned this term is specifically used for experimental studies without a control group, where a baseline measure is taken first (“pre”-intervention), an intervention is carried out, and finally an outcome measure is taken for the whole group of participants after the end of the intervention (“post”-intervention). As previously stated such studies are of little value, since changes might have occurred also without an intervention – and therefore controlled trials were advocated.

What can make a difference is the number of such assessments, both pre- and post intervention. For instance, it would be useful for experimental studies to repeatedly assess outcome measures over a prolonged period of time after the end of the intervention, in order to assess sustainability of effects obtained.

In what is called *Interrupted time series analyses**, the number of such assessments both before and after the intervention is large, so that time trends

which might have occurred anyhow can be captured. Usually, this approach is not used in experimental studies (although it would be in principle possible), but in observational studies, and preferably for indicators measuring the status of whole groups (e.g., the number of suicides in an area, the number of bullying cases in schools). Mostly these are already available data, where a naturally occurring intervention is being analysed for a possible influence on the size of these indicators over the long term (e.g., for analysing the question, whether the detoxification of coal gas had an influence on suicide rates).

Is it enough to have well-designed and conducted quantitative studies?

The ultimate aim of evaluation research is the application of the findings in the real world, i.e., outside a research setting. For this purpose, judging the quality of a study by its ability to minimise the risk of bias is not enough. Even well-designed and well-conducted studies may not supply the evidence needed, simply because study populations may be highly selective and often do not correspond to the real world population to whom the study results should become applicable. One factor which is unique to the RCT is that potential study participants refuse participation because they do not feel comfortable with the idea of being randomly allocated to an intervention. Also, research settings are often specialised (e.g., concerning funding) and, at times it may simply be uneconomical to replicate in real world settings those interventions that have proven successful in a study. The awareness of this problem has led to the development of so-called “pragmatic RCTs”, i.e., studies performed in real-life naturalistic conditions. Generally speaking, however, it has to be acknowledged, while RCTs are very powerful concerning their 'internal validity', the relevance of their results for the real world, i.e., their 'external validity' is not guaranteed.

Observational studies, while having a lower 'internal validity', i.e., being more prone to bias, study 'usual practice' and may therefore be more relevant to the real world than RCTs performed under ideal conditions. In other words, observational studies can contribute to evaluate the 'effectiveness' of an intervention, which is to illustrate whether an intervention works in a real world setting or not.

There are therefore limitations as well as strengths associated with different study designs. What does this mean for policy makers? One approach may be that used in England where a “Pilot evidence grading scheme for public health interventions” assesses evidence not only in terms of the quality of the study design but also in terms of a separate hierarchy of “corroboration”, referring to the questions whether an intervention works and matters in the real world²⁰. Finally, it has to be stated that there may also be instances where evidence on effectiveness is so obvious that conducting a trial is not a good use of resources.

3.2. Qualitative methods: complementary methods of evaluation

Qualitative studies can help us to understand what factors impact on the successful implementation of interventions in different settings and contexts. They can also help us to identify the satisfaction, potential additional outcomes of interest and appropriateness of interventions for different sectors of society. They do not provide information on effectiveness, but can be used to complement the findings of quantitative research.

Qualitative research can very loosely be characterised as “*the systematic collection, organisation, and interpretation of textual material derived from talk or observation. They are used in the exploration of meanings of social phenomena as experienced by individuals themselves, in their natural context*”²¹. The term qualitative can refer to a very broad range of evaluation methods. It is not our intention to describe these in detail here although some examples are outlined for illustrative purposes.

What is the role of qualitative research?

What qualitative research cannot do is answer the question, “Does an intervention work?” For this purpose quantitative research is needed. Qualitative research methods however play an essential complementary role alongside quantitative methods of evaluation.

Qualitative research can help initially in the design of quantitative research studies. For instance they may be used to help identify potential outcomes of interest or explore potential hypotheses to help in the framing of future large scale intervention studies, in effect acting as a precursor to quantitative evaluation.

Qualitative research can also help inform the debate over the generalisation of findings from experimental studies. While it is essential to know whether an intervention works, this may not be enough. It is also useful to know in what settings it works, for whom, and under what circumstances. Qualitative techniques can help provide insights into why apparently effective interventions do not work in specific settings. This is of particular importance when looking at mental health promotion and mental disorder prevention programmes. Qualitative research can also be used to inform assessment of intermediate outcome in the process of delivering an intervention. This may be particularly useful for studies with a long time frame.

Scenario 10. Use of qualitative methods for school-based intervention

In the case of the school-based intervention, a systematic review of the literature may indicate that mental health promotion interventions in schools are highly effective. However what if most of this evidence comes from studies conducted in the United States? For instance the organisational structure of the education system or differences in culture might impact on the willingness of children, and/or their parents, to participate in a same school-based programme in a European setting. Perhaps the attitudes and training of the professionals required to deliver the intervention may also differ. If the intervention is delivered in exactly the same way as in the United States, perhaps it will not appear to be successful. Qualitative analysis may identify how the intervention may need to be adapted to be successfully used in another setting.

The intervention may have very different levels of effectiveness in different schools. Perhaps in one school, where there is a high proportion of students whose first language is not that of the country, little success can be seen? Qualitative methods might also help to identify barriers (other than language) that impede successful implementation. Perhaps those implementing the school-based programme have not followed implementation guidance correctly. Again qualitative analysis can look at the processes by which interventions are delivered to see how closely they have followed such guidance.

In summary, the success or failure of an intervention, for which we already have good evidence of impact, may relate to issues about the target group/individual or the contexts and circumstances of the intervention. If the benefits of programmes are to be maximised, then they should either be targeted at those groups for whom they are likely to be successful, and/or adapted to reach other groups. Many programmes may fail, not because the intervention has not worked, but because the intervention has not actually been delivered as intended (e.g., it has been poorly reproduced, or has not reached enough people or has only been partially implemented)²². Qualitative techniques can help identify these factors so that policy makers can adapt effective interventions to improve their chances of success in their own system.

Facts 4. Some qualitative research methods

In-depth interviews:

Face-to-face conversations with the purpose of exploring issues or topics in depth. They may be based on a loose structure of open-ended questions. They aim to get below the surface of the topic discussed, explore individual perspectives in more detail, and uncover new areas or ideas not anticipated at the outset. They are useful for exploring experiences or behaviour, opinions or beliefs, feelings, or knowledge on a subject.

Focus groups:

A method of group interview that explicitly includes and uses group interaction. The group members are encouraged to talk to one another: asking questions, exchanging anecdotes and commenting on each other's experiences and points of view. It is particularly useful for exploring group experiences (e.g., difficulties in obtaining employment) or knowledge (e.g., services to help obtain employment) and can be used to examine not only what people think but also why they think that way.

Observational qualitative studies:

The systematic observation of organisational settings, team behaviour, and interactions allows researchers to uncover everyday behaviour rather than only relying on interview accounts. It can help identify what really happens in particular settings and in the formative evaluation of new services.

Action research:

Emphasises the participation of those being researched in the evaluation. Five key factors have been identified: flexible planning – content and direction not determined at the outset but evolve over time; iterative cycle – problems identified, action taken, change evaluated etc; subjective meaning – meanings attached to situation by those involved are included to empower service users; simultaneous improvement – by promoting change; unique context – social context in which the research is being conducted must be considered.

As with quantitative methods of evaluation, guidelines on how the qualitative methods should be conducted and how their results should be reported are now available (see “Suggested Reading”).

Scenario 11. Using qualitative methods in evaluating the success of school-based intervention programmes

Evaluators spend time observing interaction between pupils and with teachers following the introduction of the school-based programme. They also use observation to assess how guidelines on implementation of the programme are being followed by teachers, educational psychologists and other school-based staff. In addition to looking at ways of communicating, they might also look for non-verbal signals such as body language, which might indicate disinterest or enthusiasm for the new programme.

A series of one-to-one in-depth interviews with key staff take place to obtain insights on their attitudes towards the school, children and the intervention. Focus groups for children are used to see how they feel about the new programme and whether they think it is relevant. Both staff and children might be asked for suggestions on how the delivery of the intervention could be improved. A postal or telephone questionnaire might collect parental perceptions of the impact of the programme on behaviour.

The results of these qualitative methods, could suggest in our example, that the materials and guidance for the school were too US orientated and the language needed to be adapted to the local context. Parents also felt that some of the school-based programme duplicated other activities. The dominant view in focus groups of children was that they enjoyed the programme but felt too much time was spent talking about gang culture when the real problem was bullying. These responses were used to adapt the programme to aid in implementation.

3.3. Economic evaluation: building the base for policy decisions

It is important to understand what resources are required to deliver promotion and prevention interventions in mental health and what their economic impact is. Scarcity is endemic; decision makers have to choose between alternative uses of limited budgets and human resources. Economic evaluation can potentially be a very useful aid to such decision-making.

In its simplest form, economic evaluation compares the costs and outcomes of two or more interventions. Consider our school-based programme in Figure 5. Here the decision is whether to invest in this new programme or to make no additional investment.

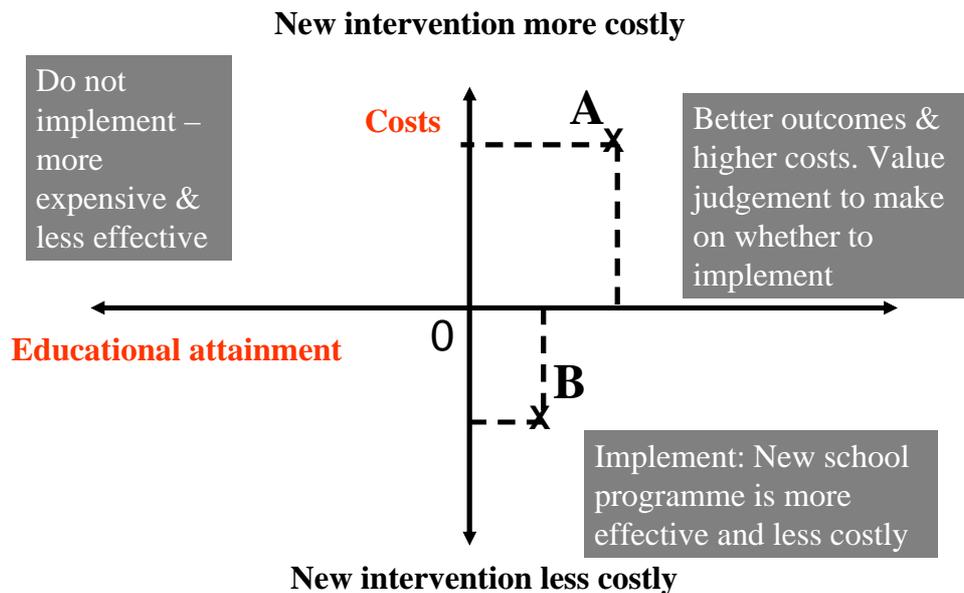


Figure 5. Assessing the cost effectiveness of a school-based intervention

Point B indicates that the new school intervention is both more effective and less costly than the current arrangements in schools (e.g., other existing mental health promotion programmes, other types of support, or no action at all). In these circumstances the task for the decision-maker looks quite straightforward: recommend wider use of the new programme. Perhaps the school programme is at point A: it produces better outcomes but costs more than the existing interventions or than nothing being implemented at this stage. The decision now is more complex, because a trade-off is needed: are the better outcomes worth the higher costs? This ultimately is a value judgement; societies will differ over how much they are willing to spend to obtain these better outcomes. Moreover decisions should never be made on the basis of cost alone, they will also need to take account of other factors such as equity*; for instance decision-makers may be willing to invest more in a school programme that reaches disadvantaged groups.

Types of economic evaluation

A number of different economic evaluation techniques can be used (see “Suggested Reading” for available guides), while all estimate costs in the same way. These include not only the monetary costs of a programme, e.g., professional training, staff time, use of buildings and equipment etc., but also other non-monetary inputs such as the contribution of unpaid volunteers. In addition to the immediate costs of delivering an intervention, it may also be important to look at the impact on future costs. For a school programme, what will be the impact on the future need for mental health services or on services in other sectors? What might be the impact on lifetime career prospects or on time spent in paid employment or in unpaid voluntary activities?

The techniques differ in how they measure outcomes. The simplest method, cost minimisation analysis, on the basis of existing evidence, assumes that effectiveness is identical for all options; the evaluation thus concentrates on finding the least costly.

The most frequently used technique, cost effectiveness analysis* (CEA), compares costs with a *single* outcome dimension, such as the level of educational achievement, rate of suicide, level of mental well-being, rate of return to employment or a symptom-specific measure. Focusing on just one measure of effectiveness may not always be helpful to decision-makers. Cost consequence analysis* (CCA), which is similar to CEA, compares costs with a *range* of outcomes for interventions, without making any judgement as to which outcome is of more importance. This is left to decision-makers.

Unlike CEA or CCA, cost utility analysis* uses a common measure of outcomes, potentially allowing decision-makers to compare the merits of investing in mental health promotion, for instance, with other public health interventions or health care treatments. Outcomes are measured in what economists call 'utilities', this simply is a measure of satisfaction that individuals attach to a specific health status, or health outcome. Common examples are the Quality Adjusted Life Year* (QALY) and the Disability Adjusted Life Year* (DALY). Time spent in states of poor health have a lower value than time spent in perfect health.

In cost benefit analysis* (CBA) all costs and outcomes are valued in monetary terms. If benefits exceed costs, then the school programme should be provided. With two or more alternatives, the intervention with the greatest net benefit would be the most favourable. There are different ways of putting a monetary value on effectiveness outcomes. One way is to survey the public, asking how much they would be willing to pay for the school-based programme. To help with this question individuals would be provided with information on the potential outcomes (and negative consequences) that could be achieved. There are challenges in obtaining meaningful values from cost-benefit analysis, especially if the public have negative attitudes towards mental health problems. Its use is however increasing and there are some CBA studies for mental health promotion or mental disorder prevention. CBA is widely used in sectors such as transport and education. Using CBA enables the decision-maker to compare investing in mental health promotion or mental disorder prevention with investing in any area of the economy.

4. From evidence to decision-making

The features of different approaches to evaluation and some of their strengths and limitations have been briefly described in previous sections. Here two further issues are considered: first, which types of study can be most helpful to addressing specific questions of policy makers and other stakeholders? Second, whether the strength of evidence on the efficacy-effectiveness of any programme can be strengthened by synthesising the results of a number of previous evaluations.*

Choosing the right study design to answer different policy relevant questions

In the previous sections we have highlighted some of the key questions that policy makers may be interested in. These include the most fundamental “can this work?” (efficacy) and “will it work in real life conditions?”, or will it be better than something else under real life conditions (effectiveness). Other key questions may concern value for money in investing scarce resources in an intervention and acceptability - will the public and/or service users be willing to use the proposed services? We may also be interested in the impact of an intervention on the level of health inequalities between different groups or on the potential side-effects and consequences of an intervention - will it do more good than harm?

These are just some of the questions that may need to be answered to improve our understanding of whether an intervention should be introduced into policy and practice. Others can include understanding why a specific mental health promotion/mental disorder prevention intervention or comparison intervention were chosen. How was the intervention delivered? And why was a specific outcome measure chosen? How relevant are the results of a specific study for the population as a whole or for other target groups such as service users, employers, schools or families? Are users, providers, and other stakeholders satisfied with the intervention?

These questions cannot all be answered by one study design. How can we know which study designs are useful for which questions? One helpful approach is presented in Table 1, which is a matrix of study designs that assesses their ability to answer different key questions. The example here is related to services for children but the findings are equally relevant to all areas of research.

Table 1. An example of a typology of evidence (example refers to social interventions in children. (Adopted from Muir Gray²³, taken from Petticrew & Roberts, 2003²⁴).

Research question	Qualitative research	Survey	Case-control studies	Cohort studies	RCTs	Quasi-experimental evaluations	Non-experimental evaluations	Systematic reviews
<i>Effectiveness</i> Does it work? Does doing this work better than doing that?				+	++	+		+++
<i>Process of service delivery</i> How does it work?	++	+					+	+++
<i>Salience</i> Does it matter?	++	++						+++
<i>Safety</i> Will it do more good than harm?	+		+	+	++	+	+	+++
<i>Acceptability</i> Will children/parents be willing to or want to take up the service offered?	++	+				+	+	+++
<i>Cost effectiveness</i> Is it worth buying this service?						++		+++
<i>Appropriateness</i> Is this the right service for these children?	++	++						++
<i>Satisfaction with the service</i> Are users, providers and other stakeholders satisfied with the service?	++	++	+	+				+

There are different strengths and limitations for different methodological approaches. The more '+'s a methodology has, the better suited it is for answering a specific type of question (as indicated on the left side of the table). The table illustrates that for questions such as appropriateness of interventions, user satisfaction with services, the process of service delivery and acceptability, qualitative research or surveys may have a more important role to play than experimental studies. If we want to answer the question does it work – the most suitable approach is to pool the findings of several RCTs through systematic review/meta-analysis.

Pooling evidence: systematic reviews and meta-analyses

We have already emphasised that prior to commissioning any new evaluation it is important first to identify what is already known about the efficacy and/or effectiveness of an intervention. One way of doing this is by referring to published systematic literature reviews (e.g., those of the Cochrane or Campbell Collaborations; see “Suggested Reading”) or in the absence of such reviews commissioning such a systematic review.

Systematic reviews are a method of making sense of and summarising large amounts of study-derived information. They contribute not only to answering the

question “Does it work?”, but also to answering other questions related to quantitative outcomes and/or qualitative information, if such studies were undertaken. They can also help identify gaps in knowledge where little is known.

There are few studies that have results so generalisable that we should accept their findings outright. Evaluation results need to be put in their context and be compared with the results of studies in similar populations. It is not uncommon that two similar studies reach sharply different conclusions due to differences in the context.

Thus, it makes sense to bring together information from all relevant studies. But one should be cautious in simply conducting literature reviews: unless these are undertaken in a systematic fashion that methodically trawls through many potential sources of information, there is a strong danger that the findings of a review may be biased²⁵. This does not imply that there is no value in an overview written by an expert, but such a non-systematic literature review may have the specific perspective of its author. The systematic review uses a set of scientific methods to explicitly limit this bias by attempting to include all relevant studies that meet specified inclusion criteria in order to answer a particular question. Systematic reviews are widely used to inform policy making in many parts of the world.

For instance, a review to answer the question “What is known about the effectiveness of school-based mental health promotion programmes?” might be undertaken²⁶. If, by looking at the impact we see consistently from many different studies in different settings that an intervention appears to be effective, then this strengthens the likelihood that our school-based programme will be effective.

It may also be possible to go one step further by using a statistical technique called 'meta-analysis' to pool quantitative evidence from different studies. When the same outcome is recorded in several studies, one overall level of effectiveness can be estimated, to provide an estimate of the clinical significance* (e.g., relevance or the outcome), or how much change is expected from such types of interventions. If this estimate suggests that an intervention is effective, we can have an even higher degree of confidence that the intervention does indeed work compared with the results of a systematic review.

A lack of information from reviews or strong RCTs on the long term impact of interventions does not necessarily mean that no input is available for the decision-making process. One possibility may be to use information obtained from reviews to provide some information on the potential long term impacts and resource consequences of interventions. For instance, in the case of a school-based programme, simple projections or models of population impact drawing on a systematic review of available evidence from published literature or even expert opinion, as well some information on the necessary resources to deliver an intervention and its potential impact on the need for future resources may be estimated. If such projections suggest, for example, that an intervention appears to be cost effective even at a very modest level of effectiveness, then this may strengthen the case for careful piloting and evaluation of an intervention. The uncertainty around such long term outcomes would be reduced as data from evaluations of interventions becomes available.

To summarise:

- Pooling evidence through systematic review/meta-analysis is powerful for answering many questions.

- If we want to know whether interventions work then ideally findings from meta-analyses of RCTs, or systematic reviews of RCTs, are more reliable than those from single RCTs or observational methods.
- Similarly for qualitative evidence, results from systematic reviews (and where possible) meta-analyses of qualitative studies, will provide stronger and more reliable evidence than the results of a single observational qualitative study.
- The findings gained from meta-analyses, systematic reviews or RCTs can be augmented by information from most other types of study design.
- In the absence of robust evidence on efficacy or effectiveness, it may be helpful to model or project the potential effectiveness and/or cost effectiveness of interventions.

5. Conclusions: Towards strengthening the evidence base for decision-making on promotion and prevention in mental health

The previous sections have outlined the need to take into account possible caveats when commissioning, evaluating and appraising evidence on given interventions, and how the results of studies could be interpreted to be useful and supportive to policy making. There are still critical issues to consider both before making decisions for implementation and during the process of implementation itself. Evidence from research studies will be only one of a number of factors taken into account in the decision-making process. Some policy decisions and interventions may be considered worth doing on the basis of social justice, political, ethical, equity issues, reflect public attitudes and the level of resources available, rather than be based on health outcomes alone.

If policy makers wish to facilitate the greater use of evidence in decision-making and to promote the use of the evidence-base to support and validate decision-making for mental health promotion and mental disorder prevention, the following conclusions can help provide relevant, high quality evidence, useful in: 1) commissioning evidence; 2) assessing and using evidence; 3) implementing and generating new valuable evidence.

5.1. Supporting the generation of high quality evaluations

Different policy relevant questions that require evidence include: “Does the intervention work, can it work in my setting?” “What will it cost to deliver and what broad benefits may it convey?” Therefore, when trying to answer such questions it is essential to identify what type of available evidence exists and might be helpful for this purpose.

Don't reinvent the wheel – first make use of existing high quality evidence

Answering these questions does not necessarily mean commissioning new original research studies. The most powerful tool is the rigorous systematic review and (where possible) meta-analysis, as this combines the results from many previous well-designed studies rather than just relying on the results of a single study alone (see section 3 quantitative evaluation). Too often little is done to make use of all

knowledge not only from previous evaluations but also from epidemiological research.

Commission and make use of high quality and broadly based evaluation

Decision-making and commissioning of useful evaluations should be based on high quality studies that use the appropriate research designs to answer each specific question (see Table 1). The evaluation and the related commissioning of evidence along with its interpretation should be broad based, and take into account other factors that will impact on successful implementation. These include the appropriateness and acceptability of an intervention in any one culture or setting, constraints on available human and financial resources, and any difference in the context in which an intervention is to be delivered (see section 3, qualitative evaluation).

When commissioning studies ensure that methodological standards are adhered to for both the conducting and reporting of studies

It is critical to enhance the quality of evaluation. Guidelines have been developed by major international bodies on both the conducting and reporting of most research methodologies (see “Suggested Reading” for further information). It is critical that research funders build in incentives to ensure that high quality studies comply with such guidelines. These guidelines apply as much to high priority studies undertaken in real world conditions (where the evidence base may still be limited), as they do to efficacy studies (see section 3, quantitative evaluation and section 4).

Include long-term monitoring and follow-up

Sometimes the success or failure of promotion interventions cannot be fully determined for a long period of time. All mental health promotion and mental disorder prevention programmes should routinely collect information on long-term health impacts (e.g., development of new cases of depression after a few years of the intervention) as well as social and economic outcomes (e.g., educational attainment, sick leave rates, crime). Interventions need sufficient time to show effect (or lack thereof) and to provide an accurate estimation of the duration of any effects. Knowledge of the duration of effects should help improve the effectiveness of interventions by guiding decisions about when and for how long interventions should be provided. Long-term follow-up can also show the real reach of programme effects and will lead to more convincing advocacy messages to influence the support for interventions.

5.2. Assessing and using evidence

Appraise critically existing evidence

When considering available evidence it is essential to assess the quality of evaluations and the strengths and limitations of study findings, including the appropriateness of using a specific study design to evaluate the effects of an intervention and the likelihood that the results are susceptible to bias. However it is also critical to look for example at the *magnitude* of effectiveness, (how important or clinically significant a given result is in its context); the credibility of the study (is the study relevant to the wider population for whom the intervention is intended); how complete a study is (relevance of outcomes for all stakeholders); or the transferability of a study to a different context of that in which it was delivered (see section 1).

Improve reporting and communication between researchers, policy makers and other stakeholders

Studies need to be transparent in the way that they are reported. The findings need to be presented in a way that makes sense to different audiences including policy makers, professionals, and the general public. For instance traditional statistical outcomes should be transformed into understandable percentages of improvement, which are easier to understand. A range of publications are required; for example technical research publications should always be accompanied by a brief non-technical summary of the findings. Workshops for policy makers and other key stakeholders may also be appropriate to adapt and communicate findings that are understandable, in order to ensure the use of available evidence.

5.3. Supporting evidence based implementation

Engage key stakeholders

During the needs assessment (beyond the scope of this primer) it is important to involve different stakeholders in the process of identifying policy relevant questions to ask and policy priorities. To successfully implement evidence-informed policy it is important to engage key stakeholders by developing a shared vision, clear goals and objectives for a given intervention, considering the different values and acceptability to the general public of a given implementation decision. The goals of a given initiative need to be concrete, attainable, measurable and agreed by all members. An early assessment of participation readiness, such as community readiness, is also crucial in determining the nature and timescale of implementing a new programme.

Look at whether the intervention can be transferred to different countries and cultures

Just because an intervention has been effective in one country or culture, this does not mean that it will necessarily be effective elsewhere. When it is clear that an intervention can work in a new setting, studies should focus on identifying the mechanisms and processes of adaptation and reinvention that can help maintain effectiveness (see section 3 on qualitative evaluation). It is essential to explore the transferability of preventive practices to different cultural situations. Qualitative research methods can be used alongside quantitative research methods to provide essential insights into the processes for successful transferability, adaptation, and innovation.

Support evaluation through creating partnerships between research and practice

One limitation of the available evidence for prevention and promotion in mental health is the lack of evaluation studies of programmes that have already been implemented and sustained in the real world. The creation of partnerships for the implementation and evaluation of new and existing interventions for prevention and promotion between practitioners and research teams should be stimulated. Such collaborative alliances could result in research and practitioners working together in the design, implementation and evaluation of programmes and subsequently increase knowledge of effectiveness in the real world. This may help improve the quality of implemented interventions and generate the further real world evidence that can help in the decision-making process.

Develop an infrastructure to support policy that promotes sustainability

Infrastructures that support mental health promotion and prevention and encourage collaboration within other public health initiatives as well as with other government sectors outside health can help ensure the sustainability of all programmes. It can also help to ensure that a holistic approach to implementation is adopted. For this purpose, resources should be distributed across different priorities: 1) assigning sector responsibilities and supporting capacity building; 2) promoting accountability; 3) funding programme implementation in partnership with evaluation research institutions; 4) funding training and education related to the promotion and prevention of mental health; and 5) stimulating the development of a co-ordinated body of different parties that are involved in mental health promotion programming and policy.

5.4. Before, during and after implementation

Make use of models to help estimate long term impacts

The impacts of some interventions on health and other outcomes may take many years to be realised. In the absence of information on long term outcomes, decision modelling techniques can be a very useful aid to the policy making process. Using available data on short term impacts and costs can be used to estimate long term costs and consequences of different programmes. Data used in models about potential long term impacts can be varied – if an intervention appears to be cost effective using very conservative assumptions this may provide powerful support for investment in promotion and prevention in mental health. Similarly using what economists call threshold analysis, i.e. identifying the level of effectiveness that an intervention must achieve for a given level of resource in order to be considered cost effective, can also be helpful. This has for instance been used to help inform policymakers about the potential cost effectiveness of suicide prevention programmes.

Consider the impact of different policy decisions on mental health

This primer has focused on how to identify, evaluate and implement interventions to improve mental well-being. As part of the policy making processes, it can be important also to consider their health impacts. For instance what might be the impact on mental health of a new urban regeneration scheme? Health impact assessment is a well developed technique for identifying the potential health risks and opportunities associated with different policies. Incorporating health impact assessment (including mental health indicators) into the policy making process can help promote a multi-sectoral approach to the promotion of mental health and well-being.

5.5. Final words

This primer does not claim to have included all existing methods or techniques of evaluation, but we have concentrated on those that may be of most relevance for prevention and promotion in mental health in Europe. The principles for evidence-based decision-making in mental health promotion and prevention are summarised in the box below (Facts 5).

Facts 5. Principles for evidence

- ✓ Thoroughly search for available information to avoid duplication
- ✓ Use high quality available research-based information to answer appropriately questions that need answers
- ✓ Undertake critical assessment (ethical issues, acceptability, resources) to see if it fits with needs
- ✓ Weigh the strengths and limitations of assessed evidence and decide on best course of action or no action

Without evidence of effectiveness it is difficult to make a case for investment in mental health. Moreover in the absence of good evidence there is in fact a danger that inappropriate policies and practices are introduced that may both be harmful and waste scarce resources. However it is important to note that there is no such thing as “value free evidence”; decisions will always be informed by various values and perspectives, and decision makers will always inevitably be faced with certain degrees of uncertainty.

*“Not everything that counts can be counted,
and not everything that can be counted counts”*

Albert Einstein

References

- ¹ European Commission 2005. Green paper. Improving the mental health of the population: Towards a strategy on mental health for the European Union.
http://ec.europa.eu/health/ph_determinants/life_style/mental/green_paper/mental_gp_en.pdf
- ² WHO. Mental Health Action Plan for Europe. 2005.
<http://www.euro.who.int/document/mnh/edoc07.pdf>
- ³ Cochrane A. Effectiveness and efficiency. London: The Nuffield Trust. 1976
- ⁴ World Health Organization (2004). Prevention of mental disorders: Effective interventions and policy options. Summary report. Geneva: World Health Organization 2004.
http://www.who.int/mental_health/evidence/en/prevention_of_mental_disorders_sr.pdf
- ⁵ World Health Organization (2004). Promoting mental health: Concepts - Emerging evidence - Practice. Summary report. Geneva: World Health Organization 2004.
http://www.who.int/mental_health/evidence/en/promoting_mhh.pdf
- ⁶ Herrman H., Saxena S., & Moodie, R (Eds). (2005). Promoting Mental Health: Concepts, Emerging Evidence, and Practice. Geneva: World Health Organization, 2005.
http://www.who.int/mental_health/evidence/MH_Promotion_Book.pdf
- ⁷ Hosman C., Jané-Llopis E., & Saxena S. (Eds). Prevention of Mental Disorders: The Evidence. Oxford: Oxford University Press. (In press).
- ⁸ Jané-Llopis E, Barry M, Hosman C & Patel V (2005) (Eds) The evidence of mental health promotion effectiveness: strategies for action. Promotion and Education, Supplement 2, 9-25.
<http://www.iuhpe.org>
- ⁹ Jané-Llopis E, Anderson P. (Eds) (2006). Mental health promotion and mental disorder prevention across European Member States: a collection of country stories. European Commission 2006.
http://www.gencat.net/salut/imhpa/Du32/html/en/dir1662/dd11714/country_stories.pdf

- ¹⁰ Mental Health Europe. Mental Health Promotion for Children up to 6 Years. 1999. <http://www.mh-europe.org/files/Directory%20Mental%20Health%20Promotion%20Children%20up%20to%206.pdf>
- ¹¹ Rychetnik L, Frommer M, Hawe P, Shiell. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119–127.
- ¹² Gilbody SM, House AO, Sheldon TA. Outcomes measurement in psychiatry. A critical review of outcomes measurement in psychiatric research and practice. York, UK: NHS Centre for Reviews and Dissemination 2003.
- ¹³ Rose D, Thornicroft G, Slade M. Who decides what evidence is? Developing a multiple perspectives paradigm in mental health. *Acta Psych Scand* 2006;113(Suppl. 429):109–114.
- ¹⁴ Van Emmerik AAP, Kamphuis JP, Hulsbosch AM, Emmelkamp PMG. Single-session debriefing after psychological trauma: A meta-analysis. *Lancet* 2002;360:766–771.
- ¹⁵ Rose S, Bisson J, Wessely S. Psychological debriefing for preventing post traumatic stress disorder (PTSD). (Cochrane Review). *The Cochrane Database of Systematic Reviews* 2002, Issue 2. <http://www.mrw.interscience.wiley.com/cochrane/clsysrev/articles/CD000560/frame.html>
- ¹⁶ Althaus D, Hegerl U. The evaluation of suicide prevention activities: State of the art. *World J Biol Psychiatry* 2003;4:156–165.
- ¹⁷ Gilbody SM, Wahlbeck K, Adams CE. Randomised controlled trials in schizophrenia: time for change? *Acta Psych Scand* 2002;105:243–51.
- ¹⁸ Grime DA, Schulz KF: An overview of clinical research: the lay of the land. *Lancet* 2002; 359: 57–61.
- ¹⁹ Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *BMJ* 1997;315:149-153.
- ²⁰ Weightman A, Ellis S, Cullum A, Sander L, Turley R. Grading evidence and recommendations for public health interventions: developing and piloting a framework. London: NICE, 2005
- ²¹ Malterud K, Qualitative research: standards, challenges, and guidelines. *Lancet* 2001; 483-488
- ²² Barry M., Jenkins R. Implementing Mental Health Promotion. 2007. Edinburgh: Churchill Livingstone Elsevier.
- ²³ Gray M. Evidence-based healthcare. London: Churchill Livingstone, 1996.
- ²⁴ Pettricrew M, Roberts H. Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology and Community Health*. 2003; 7:527-529

²⁵ Jadad AR, Cook DJ, Jones A, Klassen TP, Tugwell P, Moher M, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280:278–280.

²⁶ Stewart-Brown SL (2006) What is the evidence on school health promotion in improving health or preventing disease and, specifically what is the effectiveness of the health promoting schools approach. Copenhagen WHO Regional Office for Europe (Health Evidence Network Report; <http://www.euro.who.int/document/e88185.pdf> (accessed March 2006)

Suggested Reading

Byford S, McDaid D, Sefton T. Because it's worth it. A practical guide to conducting economic evaluations in the social welfare field. York: Joseph Rowntree Foundation, York Publications, 2003. Available at: <http://www.jrf.org.uk/bookshop/eBooks/1859351123.pdf>

Campbell Collaboration: <http://www.campbellcollaboration.org>

Cochrane Collaboration: <http://www.cochrane.org>

Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A. Integrative approaches to qualitative and quantitative evidence. NHS Health Development Agency 2004. <http://www.nice.org.uk/download.aspx?o=508055>

GRADE working group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490–1994.

Grime DA, Schulz KF: An overview of clinical research: the lay of the land. *Lancet* 2002, 359: 57–61. (For a more detailed discussion of study types, their internal validity and possible systematic biases, the reader is referred to the whole series of articles published by the authors in the *Lancet* (2002, Volume 359) from 5 January to 16 March 2002.)

Hale J, Cohen D, Ludbrook A, Phillips C, Duffy M, Parry-Langdon N. Moving from evaluation into economic evaluation: a health economics manual for programmes to improve health and well-being. Cardiff: National Assembly for Wales, 2005.

Heller R, Evidence for Population Health. Oxford: Oxford University Press, 2005.

Jadad A, ed. Randomised Controlled Trials. London: BMJ Books, 1998.

Kemm, J.: The limitations of 'evidence-based' public health. *Journal of Evaluative Clinical Practice* 2006;12:319–324.

Lavis J, Robertson D, Woodside J M, McLeod C B, Abelson J. How can research organisations more effectively transfer research knowledge to decision makers? *Milbank Quarterly* 2003;81(2):221–248.

Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomised controlled trials. Consolidated Standards of Reporting Trials. JAMA 1998;279(18):1489–1491.

Moher D, Cook D J, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet 1999;354:1896–1900.

Petticrew M, Roberts H. Systematic reviews in the social sciences. A practical guide. Oxford: Blackwell Publishing, 2005.

Pope C, Mays N (eds). Qualitative research in health care. Third edition. London: BMJ Books, 2006.

Quigley, R., L. den Broeder, P. Furu, A. Bond, B. Cave and R. Bos. Health Impact Assessment International Best Practice Principles. Special Publication Series No. 5. Fargo, USA: International Association for Impact Assessment, 2006.

Rothwell, PM. External validity of randomised controlled trials: “To whom do the results of this study apply?”. Lancet 2005;365:82–93.

Sackett DL, Rosenberg WMC: On the need for evidence-based medicine. Journal of Public Health Medicine 1995;17:330–334.

Glossary

The task force has compiled this short glossary using several sources, among them:

- the Glossary of Cochrane Collaboration terms:
<http://www.cochrane.org/resources/glossary.htm>ⁱⁱ
- the Bandolier EBM Glossary: www.jr2.ox.ac.uk/bandolier/glossary.html
- "A dictionary of epidemiology" by J.M. Last,ⁱⁱⁱ
- "A glossary for evidence-based public health" by L. Rychetnik et al.^{iv}, and
- Mrazek & Haggerty, 1994^v

Adaptation

Modifications of health promotion and prevention programmes to better suit a particular environment or context.

Adoption

The approval and uptake of a health promotion or prevention programme. The rate of adoption is dependent on a programme's acceptability.

Allocative efficiency

Is a technical term used in health economics to assess how well different interventions are helping to maximise possible benefits. The use of resources is allocatively efficient if no change in the way in which resources are distributed could improve the welfare of one individual without reducing the welfare of someone else.

Applicability

See: *Generalizability*

ⁱⁱ Green S, Higgins J, editors. Glossary. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005].
<http://www.cochrane.dk/cochrane/handbook/handbook.htm> (accessed 06 Oct 2006)

ⁱⁱⁱ Last JM, ed. A dictionary of epidemiology (4th ed). Oxford: Oxford University Press, 2001.

^{iv} Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence-based public health. *J Epidemiol Community Health* 2004;58:538-545.

^v Mrazek & Haggerty, 1994. Reducing Risks of Mental Disorders: Frontiers for Preventive Intervention Research. National Academy Press, Washington DC.

Ascertainment bias

See: *Assessment bias*

Assessment bias (synonyms: ascertainment bias, observer bias and detection bias)

This type of bias arises if the knowledge of people's assignment introduces a systematic difference in the process of outcome assessment.

Bias

A systematic error or deviation in results or inferences from the truth. In studies of the effects of health interventions, the main types of bias arise from systematic differences in the groups that are compared (See: *Selection bias*), the intervention that is provided, exposure to other factors apart from the intervention of interest (See *Performance bias*), withdrawals or exclusions of people or groups entered into a study (attrition bias) or how outcomes are assessed (*Assessment bias*). Reviews of studies may also be particularly affected by reporting bias, where a biased subset of all the relevant data is available (modified from Cochrane).

Case-control study

A study that compares people with a specific disorder, problem or outcome of interest (cases) to people from the same population without that disorder, problem or outcome (controls), and which seeks to find associations between the outcome and prior exposure to particular risk factors. This design is particularly useful where the outcome is rare and past exposure can be reliably measured. Case-control studies are usually retrospective, but not always (modified from Cochrane).

Clinical significance

A conclusion that an intervention has an effect that is of real and practical meaning to people's health status.

Clinical trial (synonym: Intervention study)

An experiment to compare the effects of two or more (healthcare) interventions. Clinical trial is an umbrella term for a variety of designs of healthcare trials, including uncontrolled trials, *controlled trials*, and *randomised controlled trials* (modified from Cochrane).

Cluster randomised trial

A subtype of randomised trials in which clusters of individuals (e.g. clinics, families, school classes, geographical areas), rather than individuals themselves, are randomised to different arms (modified from Cochrane).

Cohort study

An *observational study* in which a defined group of people (the cohort) is followed over time. The outcomes of people in subsets of this cohort are compared, to examine people who were exposed or not exposed (or exposed at different levels) to a particular intervention or other factor of interest. A prospective cohort study assembles participants and follows them into the future. A retrospective (or historical) cohort study identifies subjects from past records and follows them from the time of those records to the present. Because subjects are not allocated by the investigator to different interventions or other exposures, adjusted analysis is usually required to minimise the influence of other factors (confounders) (Cochrane).

Confounding bias

Is a systematic deviation from the principle that the different study groups must be treated identically apart from the intervention which is to be evaluated (e.g., people in the intervention group are interviewed frequently, but people not receiving the intervention are interviewed less frequently). This will give a biased estimate of the effect due to the biased study design.

Controlled (clinical) trial

A *clinical trial* that has a *control group*. Such trials are not necessarily randomised (Cochrane).

Control group

1. [In a *controlled trial*:] The arm that acts as a comparator for one or more experimental interventions.
2. [In a *case-control study*:] The group without the disorder or outcome of interest (Cochrane).

Cost-benefit analysis (CBA)

Is an *economic evaluation* where costs and all outcomes of an intervention are valued in monetary terms. This allows potential investments in decisions across different sectors of society to be compared.

Cost-consequences analysis (CCA)

Compares the costs of two or more interventions with differences in a range of intervention specific measures of outcome. It does not make any judgement as to which outcome is of most importance.

Cost-effectiveness analysis (CEA)

Is an *economic evaluation* that compares the costs of two or more interventions with differences in one single intervention specific measure of outcome.

Cost-utility analysis (CUA)

Is an *economic evaluation* that compares the costs of two or more interventions with differences health related quality of life. The value of the quality of life improvement is measured in units of 'utility', usually expressed by a combined index of the mortality and quality of life effects of an intervention. The best known measure is the *Quality Adjusted Life Year (QALY)*. Using a common measure of outcome allows all potential health related investment decisions to be compared.

Critical appraisal

The process of assessing and interpreting evidence by systematically considering its validity, results, and relevance (Cochrane) to determine whether the findings are valid or credible as a piece of evidence (Rychetnik et al., 2004).

Detection bias

See: *Assessment bias*

Design bias

See: *Implementation bias*

Disability adjusted life year (DALY)

Is a measure of premature deaths and losses due to illnesses and disabilities in a population.

Economic evaluation

Is an *evaluation* that synthesises outcomes with costs using *cost-benefit analysis, cost-consequences analysis, cost-effectiveness analysis, or cost-utility analysis*.

Effectiveness

Or *efficacy* in the real world, is a measure of the extent to which a specific intervention, procedure, regimen or service when deployed in the field in routine circumstances, does what it is intended to do for a specified population (Cochrane, 1999).

Efficacy

Is the extent to which a specific intervention, procedure, regimen or service produces a beneficial result under ideal conditions (Last, 1995). Trials which measure efficacy are sometimes called *explanatory trials*.

Efficiency

Is all about making the best use of available resources within a fixed budget. Society may however be willing to sacrifice the opportunity to maximise efficiency in order to meet other goals such as *equity*. This is known as the equity-efficiency trade-off. See also: *Allocative Efficiency, Technical Efficiency and Equity*.

Epidemiology

The branch of medicine that study the patterns, causes, and control of disease in groups of people.

Equity

Equity can be defined in many, sometimes contradictory ways. The most frequently discussed definitions are having an equal opportunity to access health care for equal need, equal utilisation of health care for equal need, and equity in final health outcomes. Focusing more resources on promoting equity between different sectors of society may mean that these resources are not allocated as efficiently as might be possible. This is known as the equity-efficiency trade-off.

Evaluation

Is a process that attempts to determine as systematically and objectively as possible the relevance, effectiveness and impact of activities in the light of their objectives (Last, 1995).

Evidence

Information that tends to prove a fact. Not limited to the legal sense of the term. Evidence is collected in an orderly way about a health problem and its prevention, treatment or rehabilitation. This information often comes from research. Evidence helps policymakers and other actors understand which interventions work best in different situations (modified from US Centers for Disease Control and Prevention Lung Cancer Glossary^{VI}).

^{VI} <http://www.cdc.gov/cancer/lung/glossary/> (Accessed Oct 20, 2006)

Evidence-based medicine (EBM)

Is the conscientious, explicit and judicious use of current best *evidence* in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research (Sackett^{VII}).

Evidence-based public health

Is a public health endeavour in which there is an informed, explicit and judicious use of evidence that has been derived from any of a variety of science and social science research and evaluation methods (Rychetnik).

Experimental study

In an experimental study the investigator assigns an intervention to the population studied (as opposed to an "observational" study, in which nature and circumstances are allowed to take their course). Experimental studies can be "controlled" (if a control group is used which gets no intervention or a different one), or "descriptive" with a pre-post comparison if no control group is used (referred to as a quasi-experimental study).

Explanatory trial

A trial that measures *efficacy*. It aims to test an intervention in an ideal situation with the full course of the intervention as intended, and use of other interventions may be controlled or restricted (modified from Cochrane).

External validity

See: *Generalizability*

Generalizability (synonym: External validity, Applicability)

The extent to which results provide a correct basis for generalisations to other circumstances (Cochrane).

Implementation bias (synonym: Design bias)

Indicates that actual delivery of intervention deviates from pre-stated protocol or what is intended to be delivered. Implementation bias impacts the validity of evaluation studies.

Indicated prevention

Targets high-risk individuals who are identified as having minimal but detectable signs or symptoms foreshadowing a mental disorder or biological markers indicating predisposition for mental disorder but who do not meet diagnostic criteria for disorder at that time (e.g., school intervention targeting children with symptoms of depression) (Mrazek and Haggerty, 1994). See also: *Selected prevention, Universal prevention*.

Interrupted time series design

A research design that collects observations at multiple time points before and after an intervention (interruption). The design attempts to detect whether the intervention has had an effect significantly greater than the underlying trend (Cochrane).

^{VII} Sackett DL, Rosenberg WM, Gray JA et al. Evidence-based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.

Intervention study

See: *Clinical trial*

Levels of evidence

A hierarchy of study designs that have been grouped according to their susceptibility to bias. The hierarchy indicates which studies should be given most weight in an evaluation where the same question has been examined using different types of study (Rychetnik).

Meta-analysis

The use of statistical techniques in a systematic review to integrate the results of included studies. Sometimes misused as a synonym for *systematic reviews*, where the review includes a meta-analysis (Cochrane).

Non-experimental study

See: *Observational study*

Observational study (synonym: Non-experimental study)

A study in which nature and circumstances are allowed to take their course. Changes or differences in one characteristic (e.g. whether or not people received the intervention of interest) are studied in relation to changes or differences in other(s) (e.g., whether or not they died), without action by the investigator. There is a greater risk of *selection bias* than in *experimental studies* (modified from Cochrane).

Observer bias

See: *Assessment bias*

Outcome evaluation

Refers to the consequent effect of a programme on health outcomes (Rychetnik).
See also: *Process evaluation*.

Positive mental health

Mental health refers to a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contributions to his or her community (WHO, 2001).

Pragmatic trial

A trial that aims to test an intervention policy in a 'real life' situation, when many people may not receive all of the intervention, and may use or be the subject of other interventions as well. This is as opposed to an *explanatory trial*, which is done under ideal conditions and is trying to determine whether an intervention has the ability to make a difference at all (i.e. testing its *efficacy*) (modified from Cochrane).

Primary prevention

Seeks to decrease the number of new cases of a disorder or illness in the population (incidence).

Process evaluation

Is an assessment of the process of programme delivery (Rychetnik). See also: *Outcome evaluation*.

Quality adjusted life year (QALY)

Is a measure of health impact of interventions. One QALY equals an health impact of one extra life year of full health.

Quasi-experimental designs

Studies that do not use random assignment to create the comparison groups; designs include cohort analytic, interrupted time series, and correlational designs. (Cook TD, Campbell DT. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally, 1979).

Randomised controlled trial (RCT)

An experiment in which two or more interventions, possibly including a control intervention or no intervention, are compared by being randomly allocated to units. In most trials the units are individuals but sometimes assignment is to defined groups of individuals (for example, in a household or in a community) (modified from Cochrane).

Relevance

Whether the research is appropriate to the identified review question and whether the study findings are transferable (generalisable) to the population or setting whom the question concerns (Rychetnik).

Significance

In statistics, the probability that a study result has not come about by chance (*in this case the p-values are used, e.g., $p < 0.05$ meaning that only in 5% of all studies carried out with the specific design the result would have come about by chance*).

A study result may be significant in the statistical sense, i.e., in the sense that the probability is high that the result has not come about by chance, but it may still not be significant in the general sense of the word, i.e., it may not be relevant or important, because, for instance, the differences found are too small. See also: *Clinical significance*.

Strength of evidence

Is often assessed on a combination of the study design (level of evidence), study quality (how well it was implemented), and statistical precision (p-value or confidence intervals) (Rychetnik).

Secondary prevention

Seeks to lower the rate of established cases of the disorder or illness in the population (prevalence).

Selection bias

Systematic differences between comparison groups in prognosis or responsiveness to intervention. Random allocation with adequate concealment of allocation protects against selection bias (modified from Cochrane).

Selective prevention

Targets individuals or subgroups of population whose risk of developing a mental disorder is significantly higher than average, as evidenced by biological, psychological or social risk factors (e.g., group interventions for children of mentally ill parents) (Mrazek and Haggerty, 1994). See also: *Indicated prevention, Universal prevention*.

Side effects

Any unintended effect of an intervention. A side effect is not necessarily harmful (modified from Cochrane).

Systematic review (synonym: systematic overview)

A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyse data from the studies that are included in the review. Statistical methods (*meta-analysis*) may or may not be used to analyse and summarise the results of the included studies.

Technical efficiency

The point at which outcomes are maximised for a given level of resources invested in an intervention.

Time series design

A single group research design in which measurements are made at several different times, thereby allowing trends to be detected. An interrupted time series features several measurements both before and after an intervention and is usually more valid than a simple pre-test–post-test design. A multiple time series involves several groups, including a control group (Last, 1995).

Tertiary prevention

Seeks to decrease the amount of disability associated with an existing disorder or illness in the population. See also: *Primary prevention*, *Secondary prevention*.

Universal prevention

Is targeted at the general public or a whole population group that has not been identified on the basis of increased risk (e.g., increasing alcohol tax; a media campaign to promote mental health) (Mrazek and Haggerty, 1994). See also: *Indicated prevention*, *Selective prevention*.

Acknowledgements

The authors want to express their gratitude to all Working Party members who have commented upon earlier drafts of this document.

The authors acknowledge the contribution of the Evidence Task Force of the Mental Health Working Party, nominated by the European Commission Directorate for Public Health and Consumer Affairs. In addition to the authors, the following individuals were members of the task force: Dr Angelo Barbato (Italy), Dr Ricardo Gusmao (as expert, from Portugal), Dr Maria João Heitor dos Santos (Portugal), Dr Karl Kuhn (Germany) and Professor Eleni Petridou (Greece).

Furthermore the authors thank Mr Jürgen Schefflein (European Commission, DG SANCO) for his stimulation and support during the development of this publication.

In the revisions of the document, the authors are especially grateful to Dr Peter Anderson and Dr Martin Elphick for their detailed comments on the final draft of the paper which resulted in further improvements of the text.

The authors are also thankful to Mark Phillips for language revision, to Maiju Seppälä for secretarial assistance and to Sjoerd van Alst for editorial support.

Some support for travel and meetings was received from the MINDFUL and IMHPA projects (co-funded by the European Commission, DG SANCO) and from the Wilhelm and Else Stockmann Foundation, Finland.

Finally, the following host institutions are acknowledged for their continuous support to the authors in this work:

- WHO Regional Office for Europe: Eva Jané-Llopis
- Ludwig Boltzmann Institute for Social Psychiatry and Medical University of Vienna: Heinz Katschnig
- London School of Economics and Political Science: David McDaid
- National Research and Development Centre for Welfare and Health STAKES, Finland and Vaasa Hospital District: Kristian Wahlbeck



Without evidence it is difficult to make a case for investment in mental health. Moreover, in the absence of good evidence there is a danger that policies and practices introduced are either harmful, wasteful, or both.

This primer is for policy makers and practitioners who are developing a public mental health strategy, commissioning, planning or conducting research, or choosing an intervention for implementation, and are confronted by conflicting sources of information.

It highlights steps and choices in evidence-based decision making, helping to assess and evaluate available evidence, identify potential biases, and support informed implementation of mental health promotion and mental disorder prevention.

Some of the key principles include:

- Different research methodologies have their own strengths and weaknesses, and answer different types of questions
- Critical assessment on e.g. ethical issues, acceptability, resources, is necessary to see if choice fits with needs
- Strengths and limitations of assessed evidence should be weighed to decide on best course of action or no action
- Developing and implementing policies without evidence-based knowledge may be harmful and wasteful

