

NEW GUIDELINES OF CAYT REPOSITORY OF IMPACT STUDIES ON SERVICES AND PROGRAMMES THAT SUPPORT THE DEVELOPMENT OF YOUNG PEOPLE

1. Why good evidence is important

“Evidence-based” practice – meaning “best practice” or “with well-supported evidence” – is a crucial element in policy development and the implementation of programmes in the prevention field. When selecting prevention programmes for young people, policy makers, practitioners and health and education professionals need easy access to reliable and independently validated information. It is therefore necessary to clearly establish “what works,” as well as what counts as good evidence. That means drafting well-defined standards in order to classify the levels of evidence-based research. Intervention programmes showing relevant effectiveness and rigorous methodologies have to be highlighted within the community of (evidence-based) practice within the prevention field.

To this end, CAYT’s primary aim is to provide education and prevention practitioners with evidence of what has proved – or is promising – to be of good practice; to highlight those programmes showing high effectiveness and rigorous evidence.

While doing this, however, we are acknowledging the ongoing and controversial debate around how to identify effective intervention programmes. Many different standards have been used by various organisations and academic bodies to classify preventive interventions, and quite often it has been difficult to establish a shared agreement upon them. This is the reason why we have decided to expand our quality of evidence grade system and to look further when considering what standards should be adhered to in prevention research.

2. CAYT’s standards for good evidence

A well supported evidence-based intervention programme is usually comprised of two components: a strong magnitude of impact, along with a fair and rigorous methodological approach (Nation et al., 2003). In other words, there has to be causal relationship between implementation of the programme and outcomes of the intervention.

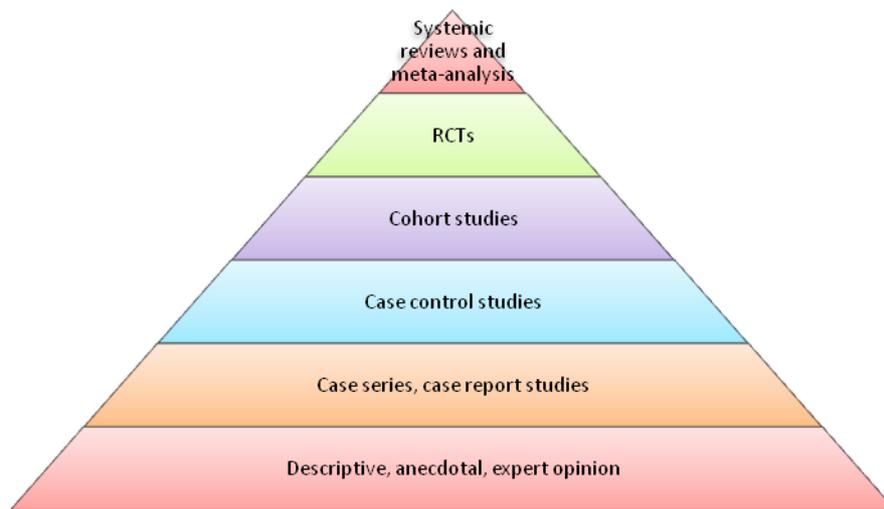
Measuring and classifying the impact of an intervention, according to the existing scoring system, is relatively objective (see table.1). On the other hand, when it comes to classifying the rigour and types of evidence, the methodological complexity and variables of research design have to be considered, making it more difficult to fairly assign an evidence grade to the evaluation of programmes.

A good representation of the hierarchy of evidence types is provided by the pyramid of level of evidence (figure 1). Common standards of evidence normally look at quantity, quality, level and design of the studies considered (Society for Prevention Research, 2005). Following these criteria, at the bottom of the pyramid there are descriptive studies, while at the top of this classification there are Randomised Controlled Trials. RCTs are seen as the ‘gold standard’ of research with the most reliable type of evaluation.

One of the founding premises of CAYT is to place more emphasis on the lower part of the pyramid. This is to allow the huge body of less rigorous studies to be classified and recognised in a fair way alongside those further up the

hierarchy. CAYT aims to create a ready-to-use evidence-based framework for prevention practitioners that includes promising interventions on all levels.

Figure 1: Level of evidence pyramid



3. RCTs and the development of an additional grade for level of evidence

Although it is important not to restrict analyses to RCTs, it is also crucial to define the appropriateness and reliability of these types of studies. Some methodological limitations and ethical risks can affect RCTs, and when assessing the quality of evidence within programmes using RCTs, some additional criteria have to be applied. For this reason, we considered appropriate expanding our scoring system and include a seventh grade on the level of evidence.

The first requirement for assuring the reliability of evidence provided through RCTs is recording the same or similar result each time a study is repeated with different populations or groups, while minimising biases linked to procedures of the study (NICE). From this we made a distinction between the sixth and seventh grade on our scale: a programme that has been evaluated through various RCTs opposed to just one. We also included important criteria which have to be met when conducting an RCT and against which we will assess their level of evidence:

- I. *Independent evaluation*: Whether the evaluation of the study has been conducted by an external and independent evaluator;
- II. *Transferability and generalisability*: This is the degree to which the results of a trial can go beyond the initial study. Generalisability refers to the relevance of the programme to a definable group or population;
- III. *Statistical power of the analysis*: Reflects the most appropriate study design and methodology to answer the research question;
- IV. *Minimum bias*: Refers to the measures taken by the researchers to eliminate biases.

These criteria are fundamental to the reliability and appropriateness of a fair and well-conducted RCT. This is because RCTs are vulnerable to many types of biases which can influence the ways in which the results of the programme are interpreted and used. Although the main distinction between the sixth and seventh grade lies in the replication of experimental trials, studies classifying in either of the top two grades of our scale will be assessed against the above criteria.

Appendix: details of impact grades and quality of evidence grades are set out below

Table 1: impact grades

Impact grade	Description
0 (none)	No relationship between the youth service and the outcome in question.
1 (low)	Provision of the youth service may be positively related to one but not all outcomes or just for sub-groups of the target population.
2 (medium)	The youth service has moderate impact on all outcomes and sub-groups or high impact on some outcomes and sub-groups.
3 (high)	The youth service has high impact on all outcomes and sub-groups.

Table 2. Level of evidence grades

Score	Type of study	More Description	Example of a study	How to improve the quality of evidence
0	Basic	Studies that describe the intervention and collect data on activity associated with it.	A study that describes the intervention and states how much it cost or how many hours of services young people received.	Collect some “before and after” data on the outcome of interest for those receiving the intervention. If it is too late for that, collect outcome “after” data for the group receiving the services and try to compare these outcomes with comparable youth using other sources of data.
1	Descriptive, anecdotal, expert opinion	Studies that ask respondents or experts about whether the intervention works.	A study that uses focus groups or expert opinion or indeed surveys those who received the intervention after they received it.	Collect some “before and after” data on the outcome of interest for those receiving the services. If it is too late for that, collect outcome “after” data for the group receiving the services and try to compare these outcomes with comparable youth using other sources of data.
2	Study where a statistical relationship (correlation) between the outcome and receiving services is established	The correlation is observed at a single point in time, outcomes of those who receive the intervention are compared with those who do not get it.	A study that conducts a survey only after the services have been delivered and concludes that youths who received the services responded more positively than those who did not.	This evidence does not allow for the fact that prior to the intervention youths who received the service may have been different from those who did not. Collect some before and after data on the outcome of interest for those receiving the intervention. If it is too late to do that, see if you can compare outcomes for a clearly defined comparison or control group using other “before” data sources, such as administrative data.
3	Study which accounts for when the services were delivered by surveying before and after	This approach compares outcomes before and after an intervention.	A study that conducts a survey before and after the program.	If you have before-after data you can measure the change in a particular outcome after the services were delivered. Try to determine whether you can compare this gain in the outcome for those who received the youth services to the gain for a similar group of youth who did not receive the services. You might use administrative data for this.
4	Study where there is both a before and after evaluation strategy and a clear comparison between groups who do and do not receive the youth services	These studies use comparison groups, also known as control groups.	A study that matches two locations where both individuals and areas are comparable and surveys them before and after the program e.g. pilot studies.	You have most of the data you need. Contact an expert on statistics or econometrics and they will be able to apply various statistical methodologies to improve the robustness of your results e.g. matching methods to define a better control or comparison group. NOTE: this is the minimum level of evaluation quality applied by the Social Research Unit et al (2011), which also stipulates that any such study fulfil various quality criteria.

5	As above but in addition includes statistical modelling to produce better comparison groups and of outcomes to allow for other differences across groups	Study with a before and after evaluation strategy, statistically generated control groups and statistical modelling of outcomes.	A study that uses a statistical method, such as propensity score matching, to ensure that the group receiving the youth services is similar to the comparison group and a statistical model of outcomes (e.g. difference in difference).	Short of a random control trial, this methodology is the most robust. To improve confidence in the results try to collect additional data, perhaps from administrative sources, on the comparison group to determine any differences between them that may have pre dated the intervention.
6	Study where intervention is provided on the basis of individuals being randomly assigned to either the treatment or the control group.	Study that compares results from two independent randomly generated groups (one receiving the intervention and the other not) and uses statistical analysis to determine the programme's effectiveness.	A study which conducts a Randomised Controlled Trial, taking into account the following criteria: <i>i</i>) a fair and independent evaluation has to be conducted; <i>ii</i>) ensuring the transferability and generalisability of the programme; <i>iii</i>) statistical power of the analysis; <i>iv</i>) ensuring minimum bias	The gold standard. It is challenging to run RCTs, with cost, ethical and practical issues arising. Even with RCTs you have to think about how generalisable it is to other situations: for example, if an RCT only looked at a youth service for males, it cannot indicate how well the youth service would do for females.
7	Various studies that evaluate an intervention which has been provided through random allocation at the individual level.	The intervention has been evaluated more than once and its effectiveness is assessed through more than one RCT showing high level of statistical analysis and reporting high quality of evidence	A series of studies which conduct RCTs on a particular intervention programme, taking into account the following criteria: <i>i</i>) a fair and independent evaluation has to be conducted; <i>ii</i>) ensuring the transferability and generalisability of the programme; <i>iii</i>) statistical power of the analysis; <i>iv</i>) ensuring minimum bias	The same challenges of level 6 apply here. To strengthen the evidence, conduct meta-analysis or systematic reviews of RCTs, comparing the results from various studies involving experimental analysis.