

CEBHA+

Evidence-based Public Health



POCKET GUIDE

INTRODUCTION

The Collaboration for Evidence-based Healthcare and Public Health in Africa (CEBHA+) aims to build long-term capacity and infrastructure for evidence-based healthcare and public health in sub-Saharan Africa. Within this network, we developed a five-day workshop on Evidence-based Public Health (EBPH) that aims to introduce the principles of EBPH and focuses on finding, appraising, interpreting and applying best evidence to public health questions relevant to the African setting. This EBPH pocket guide is primarily intended for participants attending the EBPH 5-day workshop. It mirrors the content covered during the workshop, focusing on using evidence to inform public health decision-making. We also provide useful links to additional resources and a glossary of relevant terms.

DEFINITION OF EVIDENCE-BASED PUBLIC HEALTH (EBPH)

“ the conscientious, explicit and judicious use of current best evidence in making decisions about the care of communities and populations in the domain of health protection, disease prevention, health maintenance and improvement (health promotion)”¹ by “integrating evidence from scientific research and practice to improve the health of the target population”, as well as engaging the community in decision-making².

¹Jenicek (1997). Epidemiology, evidenced-based medicine, and evidence-based public health. *J Epidemiol* 7(4):187-97

²Vanagas G, Bala M, Lhachimi SK. Evidence-Based Public Health 2017. *Biomed Res Int*. 2017;2017:2607397. doi:10.1155/2017/2607397

CONTENTS

EBPH Steps	04
Step 1: Formulate a clear question.....	06
Step 2: Find the evidence to answer your question.....	10
Step 3: Critically appraise and interpret the evidence.....	14
Step 4: Apply the findings in your context.....	24
Step 5: Evaluate the process.....	27
Glossary	28
Useful resources	34
Acknowledgements	36

EBPH STEPS



Let's imagine that you are a senior researcher at the University of Rwanda who has been asked to serve as a temporary advisor to the Ministry of Health. The prevalence of smoking among the adult male population, at 12.4%, is still high, and is linked to almost 3% of all deaths. Currently, smoking bans are in place across Rwanda, including a ban on smoking in public places, and a decrease in overall smoking prevalence has been observed. Additionally, workplace educational interventions on smoking cessation are common, but they have not been effective in reducing smoking rates. The Ministry of Health wants you to help develop a plan for reducing the prevalence of smoking among the adult population.

You recently attended a conference, and heard a lecture on the impact of workplace educational interventions on smoking cessation that include the promotion of physical activity, and you wonder whether this is more effective than the current workplace educational interventions that do not include aspects on physical activity. You want to explore whether this may be a good option for Rwanda.

Throughout this booklet, you will apply the steps of EBPH to decide whether or not you will advise the Ministry of Health to encourage the implementation of interventions which enhance workplace educational interventions by promoting the uptake of physical activity in Rwanda.

All of the scenarios and evidence described below are fictive, and have been created for this pocket guide.

Formulate a clear question

In formulating a question, it is important to clearly think through all relevant components of that question to ensure that your question is clear and precise. In thinking about your task of consulting for the Ministry of Health, you can use the PICO tool to formulate the relevant question:

PICO/PECO (TS) TOOL

P Population	Adult smokers
I/E Intervention/ Exposure	Workplace educational interventions and promotion of physical activity
C Comparison	Educational interventions
O Outcome	Smoking prevalence
(T) Time frame	[e.g., min. 24 months of follow-up]
(S) Study design	[e.g., Intervention studies, such as randomised controlled trials (RCTs), cluster RCTs, interrupted time series (ITS) studies, or systematic reviews of intervention studies]

Example intervention question:

How effective are interventions that promote smoking cessation through physical activity in addition to workplace educational interventions compared to workplace educational interventions alone in reducing smoking prevalence among adult smokers?

TYPES OF QUESTIONS

Depending on the type of question you have, different study designs will be more appropriate to answer it. In each case, a **systematic review** of all available studies is better than an individual study to inform public health decisions.

TYPE OF QUESTION	IDEAL STUDY DESIGN*
Treatment/Prevention (intervention) <i>Are workplace educational interventions effective in reducing smoking?</i>	(Cluster-) Randomised Controlled Trial (RCT)
Screening for disease or Risk factors <i>Does lung cancer screening lead to increased five-year survival among elderly smokers?</i>	RCT
Risk/Aetiology <i>Does smoking increase the risk of breast cancer among women?</i>	Cohort or case-control study
Diagnosis <i>Can spirometry be used in rural settings to diagnose lung cancer?</i>	Cross-sectional study
Prognosis and Incidence <i>What is the five-year survival rate of smokers after first cardiac arrest?</i>	Cohort study
Prevalence <i>What is the nation-wide prevalence of smoking among school-going teenagers?</i>	Cross-sectional study

*Sometimes in public health, conducting the theoretically ideal study design is not possible. For example, it may not be ethical or feasible to randomize large populations, thus for some intervention questions you may not find an RCT or a cluster RCT. In such cases other study designs, such as quasi-experimental studies (e.g. controlled before-after (CBA) studies ITS studies, and others), or observational studies (e.g. Cohorts), may provide valid sources of evidence.



TYPE OF QUESTION	IDEAL STUDY DESIGN*
<p>Experiences/ Motivations/ Perceptions</p> <p><i>What are barriers to smoking cessation among smokers who do not quit after receiving a workplace educational intervention?</i></p>	<p>Qualitative study</p>

Considering that you are interested in the effectiveness of an intervention, you realise that a RCT would be the best study design to answer your question, and that a systematic review of RCTs would be better than a single RCT.

*Sometimes in public health, conducting the theoretically ideal study design is not possible. For example, it may not be ethical or feasible to randomize large populations, thus for some intervention questions you may not find an RCT or a cluster RCT. In such cases other study designs, such as quasi-experimental studies (e.g. controlled before-after (CBA) studies ITS studies, and others), or observational studies (e.g. Cohorts), may provide valid sources of evidence.

Find the evidence to answer your question

Now that you have a clear question, you can go about searching for appropriate evidence to answer the question. There are a wide range of electronic databases that index medical research, but you narrow the list down to the following four, because you know each of them is relevant for evidence on public health interventions.

DATABASE	DESCRIPTION	ACCESS
MEDLINE (via PubMed)	Journal citations for life sciences and biomedicine from around the world. Includes systematic reviews and primary studies.	 https://www.ncbi.nlm.nih.gov/pubmed
Cochrane Library	Collection of databases in medicine and healthcare provided by Cochrane and other organizations. Includes systematic reviews and RCTs.	 https://www.cochranelibrary.com/

McMaster Health Evidence	Collection of quality-rated systematic reviews evaluating the effectiveness of public health interventions.	 https://www.healthevidence.org/
PDQ-Evidence	Evidence informing decisions about health systems. It links together systematic reviews, broad syntheses of reviews and the included primary studies, thus creating a network of evidence.	 https://www.pdq-evidence.org/

TIPS FOR BUILDING A PUBMED SEARCH STRATEGY

Combine search terms related to PICOTS components in the following ways:

X AND Y	Studies with both X and Y
X OR Y	Studies with either X or Y
(X OR Y) AND (E OR F)	Studies with both (X or Y) and (E or F)
X NOT Y	Studies with only X
Trunca*	Studies with 'Truncation' or 'Truncated' or 'Trunca', etc.
MeSH terms	Studies indexed with specific medical subject headings

Given that both RCTs and systematic reviews would be relevant to your question, you decide to search MEDLINE using PubMed. Using the tips above, you develop the following search strategy, based on the PICO components identified in Step 1:

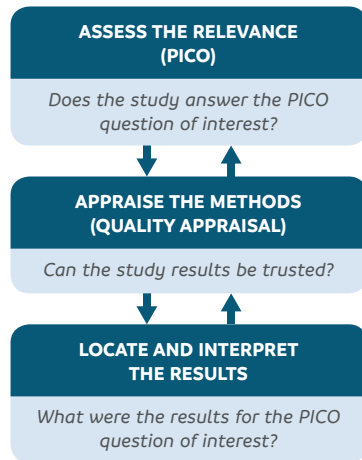
EXAMPLE SEARCH STRATEGY

Population	"Tobacco smokers" OR smok* OR "Tobacco Smoking"[Mesh]
	AND
Intervention	"physical activity" OR sport* OR exercis*
	AND
Comparison	"education" OR "training" or "inform*"
	AND
Outcomes³	"Smoking cessation" OR "Smoking cessation" [MeSH] OR quit* OR stop*
	AND
Study design	"Systematic review" OR Review OR "Systematic Review"[Publication Type] OR "Randomized controlled trial"[Publication Type] OR "RCT"

³Adding outcomes to a search strategy can restrict the number of records retrieved. Therefore, depending on your question, you might want to consider omitting keywords linked to the outcome in your search.

Critically appraise and interpret the evidence

After searching in PubMed, it is time to sift through the search output to see if you can find something that may help answer your question. Once you have identified a suitable study, you need to make sure that it is relevant and trustworthy, and that you understand the results.



In **ASSESSING THE RELEVANCE** of public health evidence, we should compare the PICO elements of the studies identified with those of our question of interest; if these match on all aspects the study results can be considered applicable.

You initially find a RCT, Stamatis et al. 2009, that seems relevant to your question, as it examines the effectiveness of the Be Active Worksites Programme. This program expands the standard support for smoking cessation among employees, which includes the provision of educational materials on the negative health effects of smoking, to include opportunities for regular physical activity. But just because it is relevant, it does not mean you can trust the results; for this you will need to think about and appraise the study's quality.

APPRAISING THE QUALITY of a study involves assessing how the study was conducted, i.e. the extent that the applied methods avoid systematic errors (bias) in the results.

The **Cochrane EPOC risk of bias tool**⁴ is one of many tools that can be used to appraise the quality of **primary studies**. The version of the tool seen below can be used to appraise the quality of RCTs, cluster RCTs, non-randomized controlled trials, and controlled before-after studies:

- Was the allocation sequence adequately generated?
- Was the allocation adequately concealed?
- Were baseline outcome measurements similar?
- Were baseline characteristics similar?
- Were incomplete outcome data adequately addressed?
- Was knowledge of the allocated interventions adequately prevented during the study?
- Was the study adequately protected against contamination?
- Was the study free from selective outcome reporting?
- Was the study free from other risks of bias?

⁴ Cochrane Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews. EPOC Resources for review authors, 2017. Available at: <https://epoc.cochrane.org/resources/epoc-resources-review-authors>

In this trial by Stamatis et al. 2009, 1631 smoking employees working in a large factory were randomly allocated using a computer random number generator to receive either the Be Active Worksites Programme immediately (intervention) or at a later point in time (control). Participants in both groups additionally received standard educational materials, highlighting the negative health effects of smoking. After randomisation, 841 individuals, received the intervention, while 790 individuals, received the control. The primary outcome of interest was 30-day smoking cessation – not having smoked in the previous 30 days – among former smokers. The study also assessed the number of cigarettes each individual smoked per day as a secondary outcome. Outcomes as well as participant characteristics were obtained from all 1631 employees through a survey conducted both at baseline and follow-up, 9 months after the implementation of the intervention. Relevant participant characteristics and smoking rates were balanced across groups after randomisation.

Using the Cochrane EPOC risk of bias tool, we identify multiple potential sources of bias for this study.

- Knowledge of the allocated intervention was not prevented (and would indeed be infeasible for this type of intervention); given that outcomes were collected through a survey and based upon participant recall, whether the participant received the intervention or not may have influenced how a given participant responded; for example, participants having received the intervention may be more likely than those not having received the intervention to underreport on smoking, if they feel this is the more desirable answer.
- Given that all participants were employees and worked at the same factory, some contamination likely took place, as those who did not receive the intervention likely heard about or observed parts of the programme; this may have biased the results.

Although the study quality is not perfect, you decide that the results of this RCT may still be trustworthy, and that you wish to continue on to interpret the study results.

OTHER TOOLS TO ASSESS THE QUALITY OF PRIMARY STUDIES:

- **The revised Cochrane risk of bias tool for randomized controlled trials (RoB 2.0):**
<https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>
- **Risk of Bias in Non-randomized Studies – of interventions (ROBINS-I):**
<https://sites.google.com/site/riskofbiastool/welcome/home>

INTERPRETING THE RESULTS involves critically reading the results section, including the text, tables and figures, locating and making sense of the results that address the PICO of interest. Often not all of the results reported in a publication address the relevant PICO.

In comparative studies, data are generally compared using specific effect estimates. Outcome data are often expressed in either dichotomous form (e.g. ceased smoking, yes or no) or continuous form (e.g. number of cigarettes smoked, 0-40 cigarettes/day).

In the RCT, Stamatis et al. 2009, you identify and need to interpret the following results. Regarding the primary outcome – 30-day smoking cessation: at follow-up, of those who received the intervention, 159 had not smoked in the previous 30 days. In the control group, 122 had not smoked in the previous 30 days.

In conceptualising effect estimates for dichotomous outcomes, a 2x2 table is helpful. With this information, we can fill out a 2x2 table as follows:

		Outcome (Smoking cessation)		TOTAL
		YES	NO	
Intervention (Promotion of physical activity)	YES	159 (a)	682 (b)	841 (a+b)
	NO	122 (c)	668 (d)	790 (c+d)
TOTAL		281 (a+c)	1350 (b+d)	1631 (a+b+c+d)

For dichotomous outcomes, typical effect estimates include the **risk ratio (RR)** (also called relative risk) and the **odds ratio (OR)**, these are calculated for the 2x2 table above as follows:

Risk in the intervention group (R_{I I}) Risk in the control group (R_{I C})	$R_{I I} = a/(a+b) = 159/841 = 0.189$ (19%) $R_{I C} = c/(c+d) = 122/790 = 0.154$ (15%)
Risk ratio (RR) (also called relative risk)	$RR = R_{I I}/R_{I C} = 0.189/0.154 = 1.22$
Odds in the intervention group (O_{I I}) Odds in the control group (O_{I C})	$O_{I I} = a/b = 159/682 = 0.23$ $O_{I C} = c/d = 122/668 = 0.18$
Odds ratio (OR)	$OR = O_{I I}/O_{I C} = 0.23/0.18 = 1.28$

As calculated above, the risk of 30-day smoking cessation was 0.189 in the intervention group and 0.154 in the control group. This corresponds to a RR of 1.22 (95% confidence interval: 0.99 to 1.52), i.e. in this sample those receiving the intervention were 22% more likely to cease smoking. This suggests that the intervention may have been effective, but the confidence interval shows considerable uncertainty surrounding this effect.

Regarding the secondary outcome – number of cigarettes smoked per day: participants who received the intervention smoked 11.4 cigarettes per day at follow-up. In the control group, participants smoked 12.3 cigarettes per day at follow-up.

For continuous outcomes, a typical effect estimate includes the **mean difference**; this is simply the absolute difference between the mean values in intervention and control groups, which can be calculated as follows:

Mean in the intervention group (X_{I I})	11.4 cigarettes per day
Mean in the control group (X_{I C})	12.3 cigarettes per day
Mean difference (MD)	$X_{I I} - X_{I C} = 11.4 - 12.3$ $= -0.9$ cigarettes per day

This means intervention group smoked on average 0.9 cigarettes less than the control group (mean difference = -0.9, 95% confidence interval: -1.8 to 0.01). This suggests that the intervention may have been effective, but there is considerable uncertainty surrounding this effect.

Given that you had some concerns regarding the study quality of Stamatis et al. 2009, you decide to have another look at your PubMed results. As the systematic review is the most reliable study design for summarizing and synthesizing evidence of intervention effectiveness, you are especially interested in finding a systematic review.

You identify a systematic review, Choate et al. 2017, assessing whether occupational educational programs lead to smoking cessation. This systematic review looked at, among other categories, interventions which contain an additional physical activity component. It included Stamatis et al. 2009, as well as four other studies with a physical activity component: Gibbs et al. 2013, White et al. 2011, Humphreys et al. 2015, and Wisdom et al. 2015. You thus decide this systematic review is relevant, but now you need to think whether you can trust the results.

AMSTAR 2^{5,6} is a commonly used tool for assessing the quality of public health relevant **systematic reviews**:

- Did the research questions and inclusion criteria for the review include the components of PICO?
- Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?
- Did the review authors explain their selection of the study designs for inclusion in the review?
- Did the review authors use a comprehensive literature search strategy?
- Did the review authors perform study selection in duplicate?
- Did the review authors perform data extraction in duplicate?
- Did the review authors provide a list of excluded studies and justify the exclusions?
- Did the review authors describe the included studies in adequate detail?
- Did the review authors use a satisfactory technique for assessing the risk of bias in individual studies that were included in the review?
- Did the review authors report on the sources of funding for the studies included in the review?
- If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?
- If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?
- Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?
- Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
- If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?
- Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

ANOTHER TOOL TO ASSESS THE QUALITY OF SYSTEMATIC REVIEWS:

- **ROBIS: A tool to assess risk of bias in systematic reviews**
<https://www.bristol.ac.uk/population-health-sciences/projects/robis/>

⁵Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017 Sep 21;358:j4008.

⁶<https://amstar.ca/Amstar-2.php>

In the systematic review, Choate et al. 2017, the PICO of interest as well as the included study designs are well defined. The search strategy appears appropriate and comprehensive, and screening, extraction and the risk of bias were performed in duplicate. The review authors conducted a meta-analysis, in which they assessed and commented on heterogeneity, and assessed the risk of publication bias.

In looking through the AMSTAR 2 tool, however, you realise there are some points in the review at which the results may have been biased:

- In conducting and interpreting the meta-analysis, review authors did not account for the risk of bias of individual studies
- Similarly, in discussing the results of the review, review authors did not consider the risk of bias of individual studies

Nevertheless, given that most of the included studies were low risk of bias, you decide that you can likely trust the results of Choate et al. 2017. The forest plot for this meta-analysis can be seen here:



The forest plot shows a risk ratio and 95% confidence interval for each included study. It also provides a pooled risk ratio, which is a weighted average of the effect estimates from each included study. The pooled risk ratio of 1.16 (95% confidence interval: 1.11 to 1.20) suggests that occupational educational programs with an additional physical activity component lead to improved smoking cessation; the confidence interval shows little uncertainty surrounding this effect.

In interpreting the results from meta-analyses, it is important to consider whether there was substantial heterogeneity across included studies. Both a visual inspection of the graph (checking the extent to which all confidence intervals overlap) and the I^2 value (here 15%, possible range 0-100%) suggest that only little heterogeneity is present.

Apply the findings in your context

At this point you have formulated a clear question, searched for relevant evidence, identified a relevant systematic review, determined that the systematic review was well-conducted and located and interpreted the results. Based on the systematic review, Choate et al. 2017, you determine that workplace educational interventions on smoking cessation that include the promotion of physical activity could be effective for reducing smoking rates in Rwandan adults.

When applying evidence, however, it is important to remember that the research evidence on its own is not sufficient to make a decision. A useful tool which highlights other aspects which may be important in the decision-making process is the GRADE Evidence to Decision (EtD) framework.^{7,8} These aspects include:

- **Problem:** Is the problem a priority?
- **Desirable effects:** How substantial are the desirable anticipated effects?
- **Undesirable effects:** How substantial are the undesirable anticipated effects?
- **Certainty of evidence:** What is the overall certainty of the evidence of effects?
- **Values:** Is there important uncertainty about or variability in how much people value the main outcomes?
- **Balance of effects:** Does the balance between desirable and undesirable effects favor the intervention or the comparison?
- **Resource requirements:** How large are the resource requirements? What is the certainty of the evidence of resource requirements? Are the net benefits with the incremental cost?

- **Feasibility:** Is the intervention feasible to implement?
- **Equity:** What is the impact on health equity?
- **Acceptability:** Is the intervention acceptable to key stakeholders?

ANOTHER TOOL WHICH HIGHLIGHTS OTHER ASPECTS WHICH MAY BE IMPORTANT IN THE DECISION-MAKING PROCESS:

- **The WHO-INTEGRATE framework⁹**
https://gh.bmj.com/content/4/Suppl_1/e000844

In a meeting at the Ministry of Health, you are provided with the following pieces of information:

- The available budget for implementing the plan to reduce the prevalence of smoking has been reduced by 50%
- The plan should be implemented beginning in 3 months
- Several of the largest employers in the country have recently expressed interest in implementing physical activity programmes at the workplace

Now in deciding whether or not you will recommend that the Ministry of Health should focus on implementing such interventions, you need to think of not only whether it will effectively lead to reduced smoking rates among adults, but also whether implementing such a programme will be affordable, feasible and acceptable to various stakeholders. In this case, you feel that the limited budget and short timeline may be problematic. At the same time you feel that the expressed willingness of several large employers to participate and the likelihood of a positive benefit with very little harm outweigh these concerns. Given all these considerations, your recommendation to the Ministry of Health is a strong recommendation for implementing workplace educational interventions on smoking cessation that include the promotion of physical activity.

TYPES OF RECOMMENDATION:

- Strong/Conditional recommendation against ...
- Conditional recommendation for either ... or ...
- Strong/Conditional recommendation for ...

If necessary provide a justification, subgroup considerations, implementation considerations, steps for monitoring and evaluation, research priorities.

STEP 05

Evaluate the process

After we have made our recommendation, it is important for us to reflect over this process. After all, this type of intervention may or may not be successful in Rwanda – some adaptation may be necessary in the future, or another intervention may be more appropriate. We may also continue to collaborate with the Ministry of Health or others in the future in helping to inform public health decisions, and it is important that we learn from what went well and what could be improved upon.

Evaluating the EBPH process involves critical reflection on the following:

- Did we ask the right question?
- Did we search efficiently?
- Did we find the best evidence?
- Did we identify the strengths and weaknesses of the evidence?
- Did we interpret the results correctly?
- Did we consider all other important factors when applying the evidence?
- What can we do better next time?

In reflecting on the process, for example, you may consider the following aspects highly relevant for improving the EBPH process:

- You find two additional relevant systematic reviews by looking at PDQ-Evidence, and decide that in the future you should search more comprehensively for evidence
- You wonder whether you thought enough about the implications for equity with regard to the intervention; those living in poverty will perhaps not be protected by interventions at the workplace, and an expansion of or addition to the intervention may be appropriate

⁷ Alonso-Coello P, Schunemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *Bmj* 2016;353:i2016. doi: 10.1136/bmj.i2016 [published Online First: 2016/06/30] <https://www.bmj.com/content/bmj/353/bmj.i2016.full.pdf>

⁸ Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst* 2018;16(1):45. doi: 10.1186/s12961-018-0320-2 [published Online First: 2018/05/31] <https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-018-0320-2>

⁹ Rehfuss EA, Stratil JM, Scheel IB, et al. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. *BMJ Global Health* 2019;4:e000844.

GLOSSARY:

Absolute effect: Difference between the baseline risk of an outcome and the risk of outcome after the intervention is applied. Absolute effect is based on the relative magnitude of an effect and baseline risk.

Bias: A systematic error or deviation in results from the truth. The main types of bias arise from systematic differences in the groups that are compared (selection bias), the care that is provided, exposure to other factors apart from the intervention of interest (performance bias), withdrawals or exclusions of people entered into a study (attrition bias) or how outcomes are assessed (detection bias). Reviews of studies may also be particularly affected by reporting bias, where a biased subset of all the relevant data is available.

Boolean operators: In searching, terms which form the basis of database logic; they connect search words to either narrow or broaden the results. The basic Boolean operators include 'AND', 'OR' and 'NOT'.

Certainty of evidence: The extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation. The certainty of evidence reflects the extent to which we are confident that an estimate of the effect is correct.

Cluster-randomised controlled trial: A trial in which clusters of individuals (e.g. clinics, families, geographical areas), rather than individuals themselves, are randomly allocated to one of two or more interventions. In such studies, care should be taken to avoid unit of analysis errors.

Confidence interval: A measure of the uncertainty around the main finding of a statistical analysis. Estimates comparing an experimental intervention with a control, are usually presented as a point estimate and a 95% confidence interval. This means that if someone were to keep repeating a study in other samples from the same population, 95% of the confidence intervals from those studies would contain the true value of the unknown quantity. Alternatives to 95%, such as 90% and 99% confidence intervals, are sometimes used. Wider intervals indicate lower precision; narrow intervals, greater precision.

Confounder: A factor that is associated with both the intervention (or exposure) and the outcome of interest.

Effectiveness: The extent to which a specific intervention, when used under ordinary circumstances, does what it is intended to do.

Estimate of effect/overall estimate of effect (effect estimate): The observed relationship between an intervention and an outcome expressed as, for example, a number needed to treat to benefit, odds ratio, risk difference, risk ratio, standardised mean difference, or weighted mean difference.

External validity: The extent to which results provide a correct basis for generalisations to other circumstances. For instance, a meta-analysis of trials of elderly patients may not be generalisable to children.

Forest plot: A graphical representation of the individual results of each study included in a meta-analysis together with the combined meta-analysis result.

GRADE (Grading of Recommendations, Assessment, Development and Evaluation): An approach for rating the certainty of a body of evidence and grading recommendations in health care. The GRADE system classifies the certainty of evidence in one of four grades: High, moderate, low, very low.

Health inequity: Differences in health that are avoidable and also considered unfair or unjust.

Heterogeneity: The variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies, or the variation in internal validity of those studies. OR: statistical heterogeneity: degree of variation in the effect estimates from a set of studies.

I²: A measure used to quantify heterogeneity. It describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). A value greater than 50% may be considered to represent substantial heterogeneity, although this threshold is arbitrary in nature.

Internal validity: The extent to which the design and conduct of a study are likely to have prevented bias. Variation in quality can explain variation in the results of studies included in a systematic review. More rigorously designed (better quality) trials are more likely to yield results that are closer to the truth.

Intervention: The process of intervening on people, groups, entities or objects.

Mean difference: A method used to combine measures on continuous scales (such as weight), where the mean, standard deviation and sample size in each group are known. The weight given to the difference in means from each study (e.g. how much influence each study has on the overall results of the meta-analysis) is determined by the precision of its estimate of effect. This method assumes that all of the trials have measured the outcome on the same scale.

MEDLINE: An electronic database produced by the United States National Library of Medicine (NLM). It indexes millions of articles in selected journals, available through most medical libraries, and can be accessed on the Internet.

Meta-analysis: The use of statistical techniques to pool the results of included studies.

MeSH headings (Medical Subject Headings): Terms used by the United States National Library of Medicine to index articles in Index Medicus and MEDLINE. The MeSH system has a tree structure in which broad subject terms branch into a series of progressively narrower subject terms.

Natural experiment study: A natural experiment is an observational study that takes advantage of a naturally occurring event or situation (e.g. policy interventions) that cannot be controlled or manipulated by the investigator and can be exploited by investigators to answer a particular question.

Non-randomised intervention study: A comparative study of an intervention against some control intervention (or no intervention) that is not a randomised controlled trial. There are many possible types of non-randomised intervention study, including quasi-randomised or

non-randomised controlled trials and quasi-experimental studies.

Observational studies: A study in which the investigators do not seek to intervene, and simply observe the course of events. Changes or differences in one characteristic (e.g. whether or not people were exposed) are studied in relation to changes or differences in other characteristic(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. Study designs typically include retrospective and prospective cohort studies and case-control studies.

Odds ratio: The ratio of the odds of an event in one group to the odds of an event in another group. In studies of treatment effect, the odds in the treatment group are usually divided by the odds in the control group. An odds ratio of '1' indicates no difference between comparison groups. For undesirable outcomes an OR that is less than '1' indicates that the intervention was effective in reducing the risk of that outcome.

Outcome: A component of a participant's clinical and functional status after an intervention has been applied, that is used to assess the effectiveness of an intervention.

PRISMA (Preferred Reporting Items for a Systematic Review and Meta-analysis): The PRISMA Statement consists of a 27-item checklist and a four-phase flow diagram. The aim of the PRISMA Statement is to help authors improve the reporting of systematic reviews and meta-analyses.

Power: The probability of rejecting the null hypothesis when a specific alternative hypothesis is true. The power of a hypothesis test is one minus the probability of Type II error. In clinical trials, power is the probability that a trial will detect, as statistically significant, an intervention effect of a specified size. If a clinical trial had a power of 0.80 (or 80%), and assuming that the pre-specified treatment effect truly existed, then if the trial was repeated 100 times, it would find a statistically significant treatment effect in 80 of them.

Publication bias: Occurs when the likelihood of a study being published varies with the results it finds. Usually, this occurs when studies that find a significant effect are more likely to be published

than studies that do not find a significant effect, thereby making it appear from surveys of the published literature that treatments are more effective than is truly the case. Can occur through both preference for significant (positive) results by journals and selective releasing of results by interested parties. Often publication bias takes the form of slower or less prominent publication of trials with less interesting results.

PubMed: A free access Internet version of MEDLINE.

p-value: The probability (ranging from zero to one) that the results observed in a study (or results more extreme) could have occurred by chance if in reality the null hypothesis was true.

Quasi-experimental study: Study design in which the researcher does not control intervention allocation, instead using design aspects, such as geographical or temporal variation, to assess intervention effectiveness. Common examples include the controlled before-after study (CBA) (also referred to as a difference-in-differences study), the interrupted time-series (ITS) study.

Quasi-randomised controlled trial: An experiment in which trial participants are allocated to an intervention and control condition in using methods that are not random, but were intended to produce similar groups when used to allocate participants. Quasi-random methods include, for example, allocation by the person's date of birth, by the day of the week or month of the year, by a person's medical record number, or just allocating every alternate person. In practice, these methods of allocation are relatively easy to manipulate, introducing selection bias. A 'non-randomised controlled trial' uses other methods, such as convenience or self-selection to allocate participants into intervention and control conditions.

Randomisation: The process of randomly allocating participants into one of the arms of a controlled trial. There are two components to randomisation: the generation of a random sequence, and its implementation, ideally in a way so that those entering participants into a study are not aware of the sequence (concealment of allocation).

Randomised controlled trial: An experiment that compares participants that are randomly allocated to one of two or more

interventions, possibly including a control intervention or no intervention.

Reporting guidelines: Guidelines for the reporting of studies were originally designed to help scientific authors to report well on their work. Some reporting guidelines have been used by scientific journals as mandatory guidelines. Common examples include the CONSORT statement for RCTs, the STROBE statement for observational studies and the PRISMA statement for systematic reviews.

Relative effect: The relative effect for a dichotomous outcome from a single study or a meta-analysis will typically be a risk ratio (relative risk), odds ratio, or occasionally a hazard ratio.

Relative risk/risk ratio: The ratio of risks in two groups. In intervention studies, it is the ratio of the risk in the intervention group to the risk in the control group. A risk ratio of 1 indicates no difference between comparison groups.

Risk: The proportion of participants experiencing the event of interest. Thus, if out of 100 participants the event (e.g. a stroke) is observed in 32, the risk is 0.32.

Standard error: A measure of the variation in a sample statistic over all possible samples of the same size. The standard error decreases as the sample size increases.

Systematic Review: 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.

USEFUL RESOURCES:

NAME	WHY IS THIS USEFUL?	URL
Centre for Evidence-based Medicine	Contains talks, presentations and useful tools on evidence-based medicine	https://www.cebm.net/category/ebm-resources
National Collaborating Centre for Methods and Tools	Provides an overview of EBPH, resources, learning modules	https://www.nccmt.ca
Cochrane Training	Webinars and resources linked to conducting systematic reviews	https://training.cochrane.org/
Cochrane Effective Practice and Organisation of Care	Useful resources when conducting systematic reviews that include non-randomised studies	https://epoc.cochrane.org/resources
Cochrane Glossary	A comprehensive glossary of terms related to systematic reviews and evidence based practice	https://community.cochrane.org/glossary
Students 4 Best Evidence	A network of students that write blogs on topics related to evidence based practice, and provide reviews on relevant resources	https://www.students4bestevidence.net/
Open Epi calculator	Free software for epidemiologic statistics	https://www.openepi.com/Menu/OE_Menu.htm

NAME	WHY IS THIS USEFUL?	URL
GRADE Evidence to Decision (EtD) framework	A tool to facilitate the use of GRADE EtD frameworks	https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-018-0320-2
GRADEpro GDT	Software to help you GRADE your evidence and develop guidelines	https://gradepro.org
EQUATOR Library	Comprehensive searchable database of reporting guidelines	http://www.equator-network.org/
McMaster GRADE Centre	Online learning modules on GRADE. This is useful for authors of systematic reviews and guideline developers.	https://cebgrade.mcmaster.ca/

ACKNOWLEDGEMENTS:

This pocket book was developed by Jake Burns, Anke Rohwer, Solange Durão and Ingrid Töws, with input from the CEBHA+ Capacity Development working group. This work expands upon a tool developed at the LMU Munich for teaching evidence-based medicine to medical students.

FUNDERS



Federal Ministry
of Education
and Research



CEBHA+ is funded by the German Federal Ministry of Education and Research (BMBF) as part of the Research Networks for Health Innovation in Sub-Saharan Africa Funding Initiative.

**The funder did not have any role in the development of the workshop or the pocket guide.*

PARTNERS





www.cebha-plus.org

CEBHA+ is funded by the Federal Ministry of Education and Research (BMBF)