Step 5: Choose Evaluation Design and Methods

Tool: *Design and Methods Example Tables*

Introduction:

At this stage, you have reviewed the Training Evaluation Framework (Step 1), thought about your situational factors (Step 2), completed the Evaluation Considerations Tool (Step 3), refined your evaluation questions, and begun to identify indicators (Step 4). You now have a good idea of the outcomes you intend to evaluate and the factors that may influence your evaluation.

This step will help you choose an evaluation **design** that fits the goals and objectives of your evaluation. It will also help you identify and choose between the many quantitative and qualitative **methods** available.

Note that the purpose of this section is to briefly introduce you to a few key evaluation designs and methods specifically used in training evaluation. It is not a comprehensive review of research design and methods. Regardless of your level of

> **Design and Methods**
>
> **Design** refers to the overall structure of the evaluation: how indicators measured for the intervention (training) and non-intervention (no training) conditions will be examined. Examples include:
> - Experimental design
> - Quasi-experimental design
> - Non-experimental design
>
> **Methods** refer to the strategies that are used to collect the indicator data.
> Examples include:
> - Pre- and post-knowledge testing
> - Survey
> - Observation

expertise, as you plan your evaluation it may be useful to involve another evaluator with advanced training in evaluation and research design and methods. Whether you are a highly

experienced evaluator or new to the field, input from another expert can be useful as you finalize the training evaluation design and methods you've chosen. Your colleague can also assist you as you plan, develop tools, and implement the evaluation, and may help you plan for final analysis and report writing.

Choosing your evaluation design and methods is a very important part of evaluation planning. Implementing strong design and methods well will allow you to collect high quality and relevant data to determine the effectiveness of your training program. Without good data, it's impossible to infer a link between training and outcomes.

Your choice of design and methods will be influenced by the situational factors and considerations you addressed in Steps 2 and 3, and may help to mitigate the problems or potential confounders that you identified.

Choosing Your Evaluation Methods:

After spending some time on the Evaluation Considerations Tool and thinking about your program's evaluation resources and sources of readily available data, begin to think about the possible methods that you might use to collect data on the indicators you've chosen.

The table below outlines some methods you might use, selected because they are among the most common in the literature regarding training outcomes.

Note again that although the TEFT is presented as a series of steps, in reality, the process of thinking about evaluation design and methods may not follow a straight line. It may instead be an iterative process, in which you come up with an idea and need to revisit previous steps, to see if it makes sense. For example, you might prefer to think about design before thinking about methods; this would work just as well.

Table 1. Possible Methods and Data Sources for Outcome Evaluation

Methods, Tools	Data Sources	Type of Data Collected
Written/oral responses: surveys questionnaires interviews journals	**Training participants ("trainees")**	Content knowledge
		Attitudes, feelings
		Self-reports of trainees' behavior
		Reports of perceived outcomes
	Patients	Reports of provider behavior
		Self-report of patient understanding
		Self-report of patient feelings/behavior
		Self-report of patient health status
	Trainees' co-workers, supervisors, and other contacts	Reports of trainees' behavior
		Reports of perceived outcomes
Group feedback: focus groups	**Trainees, co-workers, supervisors, patients, stakeholders**	Perceptions regarding programs
		Perceptions regarding needs
		Perceptions regarding outcomes, changes
notes, checklists	**Trained observers**	Observations of trainee behavior in the classroom or a simulated work setting
		Observations of trainee performance on the job
Document review, data extraction Institutional/clinical records	**Records, documents**	Patient health data
		Provider, organization, or population-level performance data

Choosing Your Evaluation Design:

You will need to choose a design that balances feasibility with an ability to help you infer a causal link between the training and the outcome. In order to infer causality, you will need to compare data for the intervention (training) with the non-intervention (no training).

Table 2: Possible Designs for Outcome Evaluation			
Design Type	**Examples**	**Strengths**	**Challenges**
Experimental: Compares intervention with non-intervention Uses controls that are randomly assigned	Randomized controlled trial (RCT) Pre-post design with a randomized control group is one example of an RCT	Can infer causality with highest degree of confidence	• Most resource-intensive • Requires ensuring minimal extraneous factors • Sometimes challenging to generalize to "real world"
Quasi-Experimental: Compares intervention with non-intervention Uses controls or comparison groups that are not randomly assigned	Pre-post design with a non-randomized comparison group	Can be used when you are unable to randomize a control group, but you will still be able to compare across groups and/or across time points	• Differences between comparison groups may confound • Group selection critical • Moderate confidence in inferring causality

Table 2: Possible Designs for Outcome Evaluation			
Design Type	**Examples**	**Strengths**	**Challenges**
Non-Experimental: Does not use comparison or control group	Case control (post-intervention only): Retrospectively compares data between intervention and non-intervention groups Pre-post with no control: Data from one group are compared before and after the training intervention	Simple design, used when baseline data and/or comparison groups are not available and for descriptive study. May require least resources to conduct evaluation.	• Minimal ability to infer causality

As you think about these options, it is very useful to consult with colleagues to get feedback on your proposed evaluation level, questions, indicators, design, and methods. It will also be useful to "walk though" the evaluation in your mind, reviewing the process of collecting the data and the resources that will be required to do this. Think about whether your plan will answer your questions, is feasible, and makes good use of your resources.

Evaluation Designs and Methods— Examples from Published Literature:

Literature already published on training outcome evaluation provides some valuable insight into evaluation design and methodology. The examples below explore evaluation designs and methods that have been reported at the three levels of the Training Evaluation Framework: individual, organizational, and systems/population (Figure 1). Presented here are examples of experimental, quasi-experimental, and non-experimental designs and a range of methods, including interview, focus group, observation, written surveys and

Keep an open mind

Keep in mind that there is no single best design or best method; all designs and all methods are possibilities for each outcome level and category. Your choices will depend upon the outcomes you evaluate, the situational factors you identified in Step 2, and the considerations you addressed in Step 3.

examinations, and clinical data review. For your reference, the corresponding literature is cited in the last page of this document. A full list of articles reviewed for the development of this framework is also available on the TEFT website.

Evaluating Outcomes at the Individual Level:

The examples in Table 3[*] present methods and designs from published articles. These articles reported on the effect of training on individuals. As indicated in the Training Evaluation Framework, changes that could be evaluated include new knowledge, skills, and/or attitudes among the training participants, changes in their on-the-job performance (as a result of this new knowledge), and changes in patient health (as a result of improved health care worker performance). The majority of training evaluations which have been reported were conducted at this point in the causal pathway toward better health outcomes.

---

[*] Notice that in the table, all outcomes are in color. The color of the text corresponds to its place in the framework. So, for example, the outcome "skills and attitudes of health care workers" is purple, because on the framework, outcomes of this type are in the purple arrow.
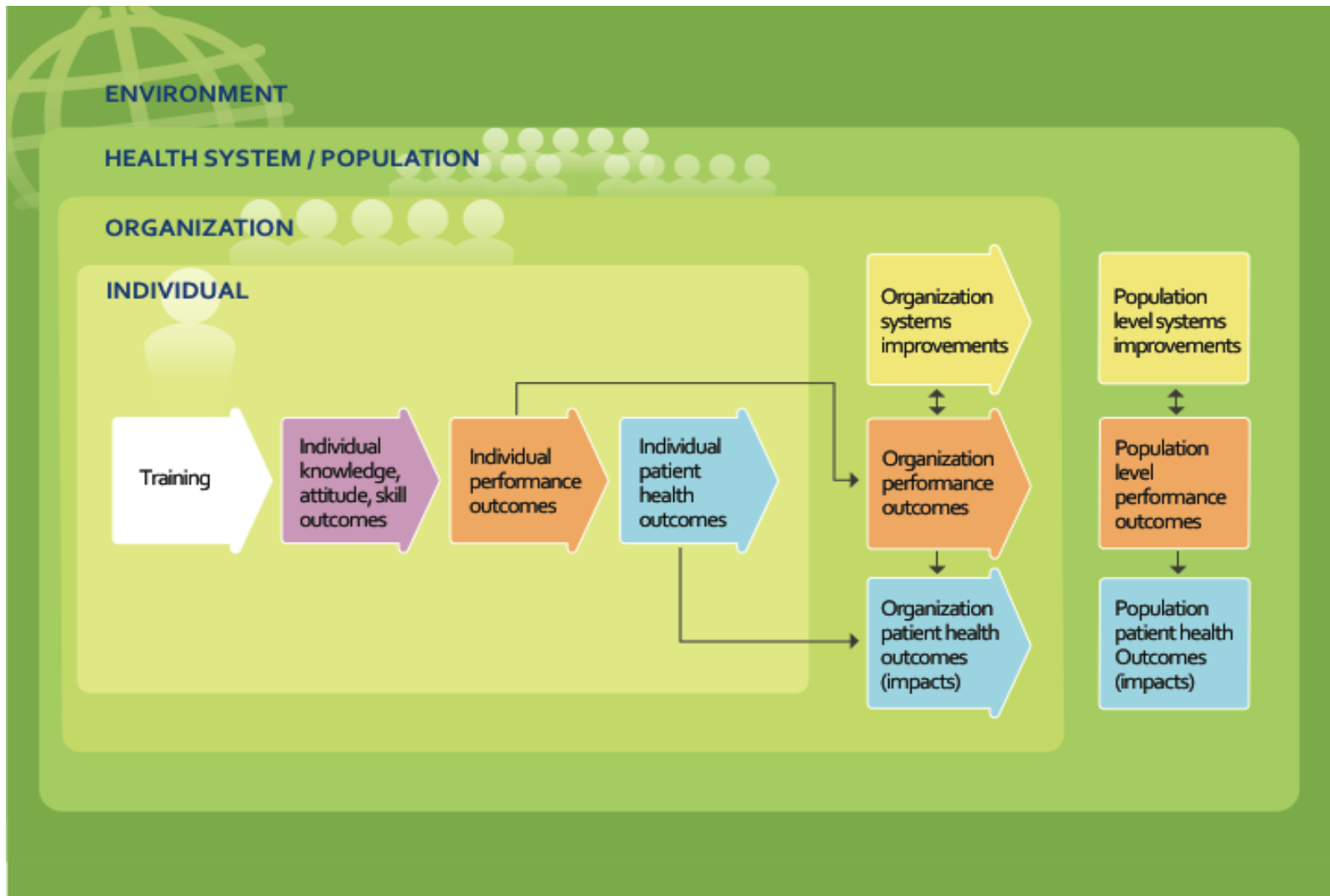
Figure 1: Training Evaluation Framework

Table 3. Evaluation Designs and Methods for Measuring Changes at the Individual Level

1. Randomized Controlled Trial: An experimental design in which the individuals being studied (e.g., training participants) are randomly assigned to either an intervention condition or a control condition.

Examples	Design and Methods
Armstrong et al (2012)[1]: An evaluation of the outcomes of a 15-hour training intervention aimed at providing social workers with cognitive behavioral strategies for work in the primary mental health care setting.	**Outcomes evaluated:** **Skills** and **attitudes** **How the evaluation was designed:** Forty social workers were randomly assigned to either an intervention condition (20 received the training), or a control condition (20 did not receive the training). **How the data were collected:** **Observation** using expert observers "blind" to whether the trainees received or did not receive the training intervention. • Observers scored videotaped sessions of simulated counseling sessions conducted pre-training and 1 week post-training. • Scoring was based on a validated measure of competencies related to the training. **Self-report** of confidence was measured using a questionnaire. • Participants rated their confidence in applying the trained methods on 5-point scales.

2. Pre- and Post-Intervention with Control: *A quasi-experimental design in which trainees are assigned to either an intervention or control condition, but they are not randomly assigned. Measures are taken before and after the intervention in both the training group and control group (non-randomized) to compare effects of the intervention.*

Examples	Designs and Methods
Kumar et al (2009)[2]: A conventional 8-day	**Outcomes evaluated:** **Knowledge and skills**

integrated management of neonatal and childhood illness training package was compared to a 5-day training package that was developed and administered in an interrupted mode of 3 days and 2 days duration with a break of 4 days in between, in a district of Haryana state in northern India.	**How the evaluation was designed:** Primary health care workers were assigned to either an intervention condition (conventional 8-day training), or a control condition (interrupted 5-day training). **How the data were collected:** **Pre-post questionnaire:** A multiple choice questionnaire was administered before and after the training to assess participant knowledge. **Case studies:** Video case demonstrations pre- and post-training were used to assess clinical decision making skills. **Observation:** Observers rated trainees on performance in role plays and conducting simulated client examinations using skills observation checklists.

3. Pre- and Post-Intervention with No Control: *A non-experimental design in which data are collected before and after a training to see whether changes occurred. Because of possible confounders, this design limits the ability to conclude that the intervention caused any changes seen, but findings can be suggestive of causality.*

Examples	Designs and Methods
Akweongo et al (2011)[3]: In 5 African countries (Burkina Faso, Ethiopia, Ghana, Kenya, and Malawi) a malaria prevention training program was implemented among community medicine distributors (CMDs). CMDs were trained to educate caregivers, and to diagnose and treat malaria cases in <5-year olds.	**Outcomes evaluated: Provider performance** and **health outcomes** (patient knowledge is included in health outcomes category) **How the evaluation was designed:** Data were collected before and 1 year after the CMDs were trained, in order to assess the effect of the training upon treatment practices and patient attitudes regarding care provision by CMDs. **How the data were collected:** **Records review:** Registers kept by CMDs were reviewed and data extracted. **Household surveys:** Surveys were administered to households to assess treatment provided. **Focus group discussions and in-depth interviews:** Patients (parents of pediatric patients) and community stakeholders participated in focus group discussions to determine perceptions related to malaria care and treatment.

4. Case-Control: *A non-experimental design in which data are analyzed retrospectively for 2 groups of subjects: a group which had the training intervention, compared with a group that did not have the intervention.*

Examples	Designs and Methods
Idigbe et al (2006)[4]: In Nigeria, a study was conducted to determine the effectiveness of a national antiretroviral training program.	**Outcomes evaluated:** Knowledge , skills and performance **How the evaluation was designed:** This was a study in which data for a sample of the health care workers who had received the national training (and who were randomly selected after a pool of potential evaluation subjects was compiled) were compared with that of a group of untrained health personnel from institutions where no personnel had received the trainings. The data were collected at one time point, between 6 and 9 months after the trainees had received the training. **How the data were collected:** **Pre-post knowledge questionnaires:** Knowledge questionnaires were completed by trainees and subjects in the comparison groups to assess content knowledge related to the national ART training program. **Patient exit surveys:** Clients were surveyed as they exited their appointments to assess their perceptions regarding the quality of care that they received. **In-depth interviews:** Heads of ART units were interviewed with regard to their units' management, strengths and weaknesses, and how these might affect the performance of the trained health workers. **Focus group discussions:** Trainees participated in focus groups to explore their perceptions of the training and its effects on their skills and the quality of services in the antiretroviral treatment clinics. **Observation checklist:** Observers used a checklist to inventory facility resources and systems, including available drugs, laboratory equipment, human resources, and supplies.

Evaluating Outcomes at the Organizational Level:

There are also many examples of research designs and methods used for evaluating the effect of training at the organizational level. As illustrated in the Training Evaluation Framework (Figure 1), these changes include improvements in health systems at the organizational level, changes in organizational performance, and changes in patient health outcomes at an organization. Table 4 presents some research designs that have been used to measure training effect on organizations, along with the corresponding literature cited for each type of outcome.

Table 4. Evaluation Designs and Methods for Evaluating Changes at the Organization Level
1. Pre-Post Time Series: *A non-experimental design in which the indicators of change were tracked over multiple time points before, during, and after the intervention.*

Examples	**Design and Methods**
Kayongo et al (2006)[5]: In a large hospital improvement project in Rwanda, CARE and partner organizations implemented a multi-component intervention to improve emergency obstetric care. In addition to training of staff, the hospital was renovated, essential equipment was provided and quality improvement projects were introduced.	**Outcomes evaluated:** Systems changes and facility-level provider performance **How the evaluation was designed:** The evaluation was incorporated into a comprehensive improvement project which tracked project progress weekly as well as analyzing indicator data yearly over 4 years. **How the data were collected:** **Patient interviews:** Patients were interviewed regarding their experiences with clinical services during a baseline assessment conducted to develop intervention plans. **Progress notes:** Progress of improvement projects were tracked with weekly meetings. **Records review:** Registers kept by facilities were reviewed and data extracted for several indicators related to emergency obstetric care. These were analyzed and tabulated yearly over a period of 4 years.

2. Cluster Randomized Controlled Trial: *A design in which clusters of health facilities are randomly assigned to either an intervention condition (facilities where providers received the training) or a control condition (facilities where providers did not receive the training).*

Examples	Designs and Methods
Gilroy et al (2004)[6]: In Mali, the impact of IMCI (Integrated Management of Childhood Illness) training on the quality of counseling provided to parents of sick children was evaluated.	**Outcomes evaluated:** Skills and Performance **How the evaluation was designed:** **Cluster randomized controlled trial:** Ten community health centers in southern Mali were randomly assigned to either the training group or the comparison group. **How the data were collected:** **Observation:** Health providers' consultations with pediatric patients' caregivers were observed by trained observers using a 10-point scale related to counseling behaviors. **Interviews:** Patients' caregivers were interviewed using structured questionnaires to collect socio-demographic, literacy and language data.

3. Three-Arm Cluster Randomized Controlled Trial: *A design in which clusters of health facilities were randomly assigned to one of three conditions: an intervention condition (facilities where providers received one version of the training), a second intervention condition (facilities where providers received a different version of the training), or a control condition (facilities where providers did not receive the training).*

Examples	Designs and Methods
Ngasala et al (2008)[7]: In Tanzania, the impact of training in microscopy-based diagnosis of childhood malaria was assessed at primary health	**Outcomes evaluated:** Performance and health outcomes **How the evaluation was designed**: 16 health facilities in southern Tanzania were randomly assigned to one of three training groups: training in clinical algorithm plus microscopy (Arm-I), or clinical algorithm only (Arm-II), or to no training (Arm-III).

care facilities.	**How the data were collected:**
	Records review: Demographic and patient health data were collected for patients. The primary outcome indicator was the proportion of eligible patients receiving prescriptions of antimalarial drugs. Secondary outcomes included prescriptions of antibiotics, cost of drugs and the proportion of patients reporting symptoms following treatment. The cost of drugs was calculated using price list from Medical Store Department.
	Interview: At the end of the study, interviews were conducted with health care workers in Arm-I to assess whether and under what circumstances they prescribed antimalarial treatment to patients with negative blood smear results.

Evaluating Outcomes at the Health Systems/ Population Level:

In the published literature reporting outcomes of health care worker training programs, the proportion of examples at the health systems/population level is the smallest of the 3 levels. As illustrated in the Training Evaluation Framework (Figure 1), changes at this level include improvements in national and regional health systems, and changes in health care worker performance and patient health outcomes at the population and community levels. Table 5 presents some research designs that have been used to measure training effect on health systems and populations, along with the corresponding literature cited for each type of outcome.

Table 5. Evaluation Designs and Methods for Evaluating Changes at the Health System/Population Level

1. Pre- and Post- Intervention with Comparison Population
In this quasi-experimental design, populations are assigned to either an intervention or control condition, but they are not randomly assigned. Measures are taken before and after the intervention in both the intervention group and the comparison group (non-randomized) to compare effects of the intervention.

Examples	Design and Methods
Marsh et al (2004)[8]: A study conducted in rural Kenya was designed to evaluate the effect of a training program for drug retailers to sell early effective treatment for malaria.	**Outcomes evaluated:** Performance **How the evaluation was designed:** The implementation of the training intervention was implemented in a phased manner. Communities which received the training in the first phase (early-implementation areas) were compared to communities which received the training in the later phase (late-implementation areas). This evaluation design also provided before-training and after-training data for both the early-implementation and the late-implementation areas. **How the data were collected:** **Household surveys:** Households within each area were surveyed regarding their over-the-counter malaria drug use. **Observation:** Simulated client visits to trainees were conducted, using a standardized scenario and simple assessment checklist of retailers' skills in responding to requests for over-the-counter medication to treat a child with a fever.

2. Cluster Randomized Controlled Trial (Population-level): *A design in which clusters of communities or populations are randomly assigned to either an intervention condition (communities where providers received the training) or a control condition (communities where providers did not receive the training).*

Examples	Designs and Methods
Bhandari et al (2002)[9]:	**Outcomes evaluated:** Health **How the evaluation was designed:**

A training evaluation was conducted to measure the effectiveness of an educational intervention to promote adequate complementary infant feeding practices. Health and nutrition workers in the intervention communities were trained to counsel on locally developed feeding recommendations.	**Cluster randomized controlled trial:** Eight communities in rural Haryana, India were matched based on socioeconomic indicators, child mortality and morbidity rates. The communities were then randomly assigned to either the training group or the comparison group. **How the data were collected:** **Home visits:** Newborns were enrolled in all of the communities and followed up with home visits every 3 months to the age of 18 months. Follow-up data collected in home visits included weight and length of the infants as well as dietary intake.

3. Pre- and post-Intervention with No Control: *A non-experimental design in which the indicators of change were tracked over 2time points: before and after the training intervention.*

Examples	Designs and Methods
Carlo et al (2010)[10]: A multi-country study, conducted in Argentina, Democratic Republic of the Congo, Guatemala, India, Pakistan, and Zambia, sought to evaluate a train-the-trainer model for training birth attendants living in rural areas in the WHO Essential Newborn Care (ENC) course.	**Outcomes evaluated: Performance** and **Health** **How the evaluation was designed:** This was a before-and-after intervention design conducted in 6 countries. Data collection for this study preceded a later study of further training, which was designed as a randomized, controlled trial. **How the data were collected:** **Medical records data extraction:** Reviews of hospital and clinic records were done, and data were extracted for comparison between pre- and post-training time points. Performance indicators included use of resuscitation techniques. Health indicators included Apgar scores, birth weight, and maternal and neonatal deaths.

Summary

Choosing your evaluation design and methods is a very important part of evaluation planning. Your choice of design and methods will be key factors in collecting high quality data which may, in turn, allow you to infer a link between training and outcomes. The examples presented in this document illustrate a number of the options you may choose from.

Remember that there is no single best design or best method, and that all designs and all methods are possibilities for each outcome level and category. Your choices will be influenced by the indicators you wish to track, the intended use of the evaluation findings, and the resources available for evaluation.

References

1. Armstrong G et al: An evaluation of the effect of an educational intervention for Australian social workers on competence in delivering brief cognitive behavioural strategies: A randomised controlled trial. BMC Health Services Research 2010, 10(1): 304.
2. Kumar D et al: The effect of interrupted 5-day training on Integrated Management of Neonatal and Childhood Illness on the knowledge and skills of primary health care workers. Health Policy and Planning 2009, 24:94-100.
3. Akweongo P et al: Feasibility and acceptability of ACT for the community case management of malaria in urban settings in five African sites. Malaria Journal 2011, 10:240.
4. Idigbe E et al. Evaluation of the Nigerian national antiretroviral (ARV) treatment training programme. Sahara J 2006, 3:488-502.
5. Kayongo M et al: Improving availability of EmOC services in Rwanda--CARE's experiences and lessons learned at Kabgayi Referral Hospital. Int J Gynaecol Obstet 2006, 92(3):291-298.
6. Gilroy P et al: Impact of IMCI training and language used by provider on quality of counseling provided to parents of sick children in Bougouni District, Mali. Patient Education and Counseling 2004, 54(1):35-44.
7. Ngasala B et al: Impact of training in clinical and microscopy diagnosis of childhood malaria on antimalarial drug prescription and health outcome at primary health care level in Tanzania: a randomized controlled trial. Malar J. 2008 7:199.
8. Marsh V et al: Improving malaria home treatment by training drug retailers in rural Kenya. Trop Med Int Health 2004, 9(4):451-460.
9. Bhandari N et al: An educational intervention to promote appropriate complementary feeding practices and physical growth in infants and young children in rural Haryana, India. India J Nutr 2004, 134(9):2342-8.
10. Carlo W et al. Newborn-Care Training and Perinatal Mortality in Developing Countries. N Engl J Med 2010, 362: 614-23.