Introduction : Benchmarks and benchmarking systems

Benchmarking may be considered as a process of setting standards for various criteria and periodically confirming that these standards are being met. A statement about a standard and a criteria that might be used to benchmark diabetic care in a primary healthcare setting might state:

> *Eighty percent of diabetic patients should have had an eye examination in the previous twelve months.*

Eighty percent is the *standard* and the rest is the *criteria*.

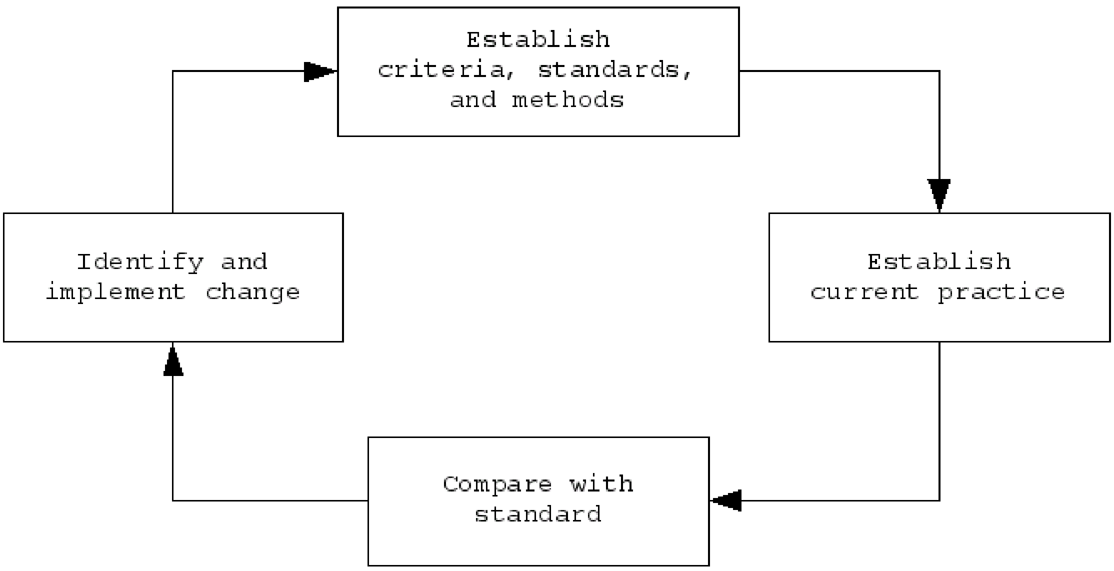
Benchmarking of services often requires more than one standard / criteria statement:

> *Eighty percent of diabetic patients should have had an eye examination in the previous twelve months.*
>
> *Eighty percent of diabetic patients should have had their blood pressure checked and recorded in the previous twelve months.*
>
> *Ninety-five percent of patients with both hypertension and diabetes should have had an eye examination in the previous twelve months.*

Standard / criteria statements may be accompanied by a statement of how the statement will be confirmed or denied:

> *Eighty percent of diabetic patients should have had an eye examination in the previous twelve months* ***based on data extracted from a retrospective review of a representative sample of the medical records of diabetic patients****.*

Standard / criteria / method statements (*benchmarks*) are usually nested within an operational framework. In healthcare settings, a framework known as the *clinical audit cycle* is frequently employed:



The clinical audit cycle makes use of benchmarks in an operational framework to help deliver improved patient care.

The benchmarks require an operational framework in order to be of use. The combination of benchmarks and an operational framework form a *benchmarking system*.

This document outlines a benchmarking system for humanitarian responses to calamity and war.

Outline of a benchmarking system for humanitarian responses to calamity and war

The proposed benchmarking system treats a humanitarian response as having four distinct stages (the *four-stage model*). The four stages are:

> **Stage 1 : Detection of crisis** – Any response system requires a means of deciding if a response is required. This is termed *detection*. Early detection facilitates a timely response. Instead of waiting for a crisis to become apparent, humanitarian actors need a system that will identify crises in their early stages. Early detection of crises facilitates planning of responses so that they are both timely and appropriate.
>
> **Stage 2: Confirmation and needs assessment** – Once a potential crisis has been identified actions must be taken to confirm the existence of the crisis. If a crisis is confirmed then actions to determine the scope and nature of the crisis are required. These actions define the scope and nature of the required response. Confirmation and needs assessment ensures that interventions are appropriate, adequately targeted, properly planned, and properly funded.
>
> **Stage 3 : Delivery** – Once a crisis has been detected, confirmed, and appropriate responses have been identified, individual programs (e.g. general rations, food-for-work programs) are implemented. These programs need to deliver interventions of sufficient quality covering a high proportion of eligible beneficiaries.
>
> **Stage 4 : Closure** – Program delivery needs to be sustained until the crisis has passed. Only then can individual programs be closed and the response terminated. Premature closure risks a re-emergence of the crisis. Late closure risks damaging markets, society structures, and development efforts.

Benchmarking in the proposed system is about evidence.

> Evidence that a crisis exists (Stages 1 and 2)
>
> Evidence of the scope and nature of the crisis (Stage 2)
>
> Evidence that high quality / high coverage interventions are delivered (Stage 3)
>
> Evidence that the response can be scaled down or closed altogether (Stage 4)

The proposed system is not an evaluation tool but an integrated planning and evaluation tool. It is intended to be generally applicable at global, regional, national, district, and individual program levels.

The scope of this proposal

This proposal outlines a benchmarking system for humanitarian responses to calamity and war. It is intended to provide an operational framework for using, whenever possible, existing benchmarks such as those detailed in the SPHERE Project's "Humanitarian Charter and Minimum Standards in Disaster Response". It is also intended that, whenever possible, benchmarks will be confirmed or denied using standard methods such as those developed in UNICEF's Standardised Monitoring and Assessment of Relief and Transitions (SMART) initiative. New benchmarks and methods will only be proposed if corresponding benchmarks and appropriate methods are not available. The outlined benchmarking system is **not** intended to compete with or replace alternative systems. It is intended to compliment other benchmarking efforts by providing a framework in which they can be best used for the planning, implementation, and evaluation of humanitarian responses. Any new benchmarks and methods in this document are preliminary proposals and are intended as starting points for discussion.

The status of the document

This document is a draft consultation document. The views expressed are solely the responsibility of the authors and do not represent the policy or aims of [redacted], [redacted], [redacted], or the [redacted].

Requirements for consultancy

This document includes a set of required / desired outputs for the consultancy process. These are introduced as each stage is described. One overall output of the consultancy should be to collate views on the sense and suitability of the four-stage model as a planning and evaluation tool for humanitarian responses to calamity and war.

Stage 1 : Detection of Crisis

Any response system requires a means of deciding if a response is required. This is termed *detection*. Early detection facilitates a timely response. Instead of waiting for a crisis to become apparent, humanitarian actors need a system that will identify crises in their early stages. Early detection of crises facilitates planning of responses so that they are both timely and appropriate.

At present, crisis detection is carried out in a rather ad-hoc manner through a combination of reports such as those from established early warning and surveillance systems and reports and requests from officials and NGO workers in affected areas. Such reports are limited in coverage and may lack sensitivity. Crises may be detected but are likely to be detected late. For example, reports from NGOs are likely to be restricted to areas in which NGOs are already active and the responsible NGO may not consider that a crisis is occurring until extremely high levels of key indicators such as the prevalence of acute malnutrition and the under five years mortality rate are observed. In such cases the response will usually be too late to prevent widespread losses of lives and livelihoods.

It is proposed, therefore, that crisis detection should be strengthened through the establishment of a system for routine collection, collation, and analysis of key indicators. Examples of key indicators are:

Economic :

Market prices of key food stuffs
Market prices of agricultural inputs
Terms of trade (livestock : cereal price ratio)
Daily labour rates

Agricultural :

Reports of livestock disease
Reports of plant infestation / disease
Productivity / crop assessments and predictions
Grazing land quality assessments

Meteorological / Ecological :

Rainfall
Vegetation cover
Land usage

Security / Civil Society :

Reports of displacement / population movement
Reports of insecurity / violence
Reports of widespread human rights violations

Public Health :

Vital (mortality) statistics
Outbreak reports
Other public health surveillance data

Other data sources :

Situation reports from local actors (including NGOs)

Comprehensive spatio-temporal coverage is required. For example, terms of trade (livestock : cereal price ratio) data should be collected on a routine basis, possibly informed by local agricultural calendars, in all agro-ecological / food-economy zones. Where data is not available, efforts should be made to improve coverage. Such efforts should concentrate on the collection of key indicators at low cost rather than on creating large and expensive systems that are likely to prove difficult to sustain. In some cases (e.g. crop yield assessments) it will be possible to substitute local data with remotely sensed data.

Limited contextual information (e.g. agricultural calendar and disease calendars) will also be required in order to usefully analyse indicator data.

Data analysis

Data should be analysed using a systematic / algorithmic approach and yield either a two-class (i.e. alert / not alert) or three-class (alert / watch-list / not alert) classification. In both two-class and three-class systems, an alert classification would automatically prompt a Stage 2 confirmation and needs assessment action. In a three-class system a watch-list classification may trigger (e.g.) more intensive data collection.

The data analysis algorithm should be both *sensitive* and *specific*. It is envisaged that the algorithms used in Stage 1 will favour sensitivity over specificity with specificity favoured by the Stage 2 confirmatory action. It is important that the data-analysis algorithm is capable of clearly identifying the indicators contributing to an alert or watch-list status. The algorithm should be clearly stated and readily understandable by both data-providers and data-users. Such an algorithm could be expressed as a series of "if ... then" *production-rules* which will typically resolve a *quantitative* indicator, or a combination of quantitative indicators, to a *qualitative* classification. An example of a production-rule based on a single quantitative indicator is:

RULE 1 : IF the livestock : cereal price ratio has declined by X for two reporting periods
THEN [ALERT BY RULE 1]

An example of a production-rule based on a combination of quantitative indicator is:

RULE 2 : IF the livestock : cereal price ratio has declined by X for Y reporting periods
AND daily labour rates have declined for Z recording periods
THEN [ALERT BY RULE 2]

Some production-rules may provide intermediate results that may be used by other production-rules. This could also form the basis of arriving at a watch-list classification in a three-class system. For example:

RULE 3 : IF rainfall is X% below the historical average for previous reporting period
AND reduced rainfall in previous reporting period will reduce cereal yields
AND vegetation cover is lower than expected in this reporting period
THEN [WATCH-LIST BY RULE 3]

RULE 4 : IF daily labour rates have declined in this reporting period
AND [WATCH-LIST BY RULE 3]
THEN [ALERT BY RULE 4]

RULE 5 : IF report of major cereal crop infestation
AND [WATCH-LIST BY RULE 3]
THEN [ALERT BY RULE 5]

The advantage of using a system of production-rules are:

The ease with which statements about what to do in predetermined situations can be encoded and read by others.

Production-rule systems are easy to computerise allowing large sets of production-rules (a *rule-base*) to be validated with relative ease against expert judgements on the same data.

The system is modular and individual rules can be added, deleted, or modified independently. This allows large systems of rules to be developed and checked quite rapidly.

The reasoning and data behind an alert or watch-list classification is explicit. For example, if a system using the example rules 3 and 4 (above) returns [ALERT BY RULE 4] then the reason for the alert is the recent decline in daily labour rates in the context of both poor recent rainfall and low vegetation cover at this point in the local agricultural calendar. This will be essential in both justifying further action and in providing a means by which the sensitivity of the Stage 1 system may be improved (i.e. by informed modification of the rule-base).

Automated data-analysis systems can be easily created and modified.

The advantage of using an algorithmic approach is that an alert or watch-list classification is made solely on technical grounds rather than a tangle of technical and political grounds. This will allow political actors to make decisions informed by technical data that has no political bias (this is not, of course, to suggest that political factors are irrelevant in humanitarian actions).

Requirements for consultancy

The output of an initial consultancy should identify:

The requirement for either a two-class or three-class system.

Identification and selection of a **minimal** set of key indicators.

Identification and selection of data-collection methods for key indicators.

Identification and selection of required **minimum** contextual information.

Identification of responsible data collectors and data sources.

Identification of responsible data analysts.

Identification of responsible data users.

Identification of means for dissemination of alert and watch-list classifications.

Preliminary versions of production-rules.

These outputs will be collated and subjected to a second round consultation concentrating on:

Finalisation of the **minimum** dataset and production-rule system.

Issues surrounding implementation of the Stage 1 reporting and classification system.

Stage 2: Confirmation and needs assessment

Once a potential crisis has been identified (i.e. by Stage 1 in the proposed framework) actions must be taken to confirm the existence of the crisis (confirmatory action). If a crisis is confirmed then actions to determine the scope and nature of the crisis are required (needs assessment action). These actions define to the scope and nature of required response. Confirmation and needs assessment ensures that interventions are appropriate, adequately targeted, properly planned, and properly funded.

Confirmatory action:

Confirm Stage 1 indicator values that triggered the alert classification(s) over all probable affected areas.

Confirm Stage 1 indicators collected by remote sensing and field investigations over all probable affected areas.

Collect additional rapid-assessment food-security indicators over all probable affected areas (probably using a semi-quantitative food security assessment method such as that proposed by SMART or the HEA method developed by Save the Children).

Analyse collected data using a production-rule system limited to a two-class (i.e. alert / not alert) outcome. This analysis is performed separately for all probable affected areas. An alert classification at this stage triggers a needs assessment action for the affected areas.

Needs assessment action:

Quantitative assessment by survey of:

Prevalence of acute malnutrition (wasting).

Recent period mortality.

The SMART survey methods for the quantitative assessment of the prevalence of acute malnutrition and recent period mortality could be used for needs assessment. There are, however, problems with the methods proposed by SMART in this context:

Prevalence of acute malnutrition – The most commonly implemented SMART method suitable for use with free living rural populations is a two-stage cluster-sampled method based on the EPI coverage survey method. This method is resource intensive (particularly in terms of suitably qualified staff and vehicles) and is limited to providing a single estimate over the entire survey area. SMART uses weight-for-height (W/H) as an indicator of nutritional status. This indicator is not a particularly good predictor of mortality and is both expensive and complicated to measure. In addition, programs following the CTC model of delivery tend not to use W/H for screening and admission. This means that estimates based on a W/H indicator may not be useful for planning CTC-type interventions.

Recent period mortality – The SMART method for recent periods mortality estimates the crude death rate (CDR) for a population. In the early stages of a crisis without widespread violent attacks on adults, CDR is likely to be insensitive to changes in mortality. This is because mortality, particularly due to food availability, is likely to increase in young children (i.e. under-fives) before increasing in the general population.

The prevalence of acute malnutrition may be estimated using mid-upper-arm-circumference (MUAC) which is the preferred admission criteria for the newer community-based service delivery models (e.g. CTC). Experiences with MUAC and MUAC-for-height (MUAC/H), measured using a device known as the QUAC stick, suggest that these indicators are simple and cheap to collect and will allow more versatile and rapid survey methods capable of providing information on the spatial distribution of prevalence and need to be implemented. Suitable survey designs are available. One particular design (e.g. centric-systematic-area-sampling) has been field-tested for prevalence and coverage estimation in emergency contexts. Survey methods that provide information regarding the spatial distributions of prevalence will provide information useful for program planning and implementation such as allowing priority to be given to areas most in need of intervention during the initial stages of an intervention and for informing the location of (e.g.) distribution points and treatment centres. This sort of information cannot be provided by currently available SMART methods.

In situations where there are patches of high prevalence of acute malnutrition there is a risk that the SMART method for prevalence estimation will return false-negative results (i.e. indicate that prevalence is low overall when it is, in fact, high in specific areas). This is because the survey method is restricted to returning an average prevalence for a wide area and because the sampling method concentrates data-collection in the most populous communities. Areas of low population density are often not sampled. This may cause surveys to evaluate prevalence as being low even when prevalence is high in areas outside of urban centres.

Recent period mortality may be better estimated, in this context, by the under fives years mortality rate (U5MR). A method for doing this using the previous birth history (PBH) technique is already available and has been field-tested.

The currently available SMART indicators appear to be unsuited to the requirements of a Stage 2 needs assessment action. The SMART project could, however, take responsibility for the development, testing, and documentation of more suitable indicators and methods, such as those outlined above.

Other indicators (e.g. water utilisation in litres / person / day) may be added. In many cases suitable validated instruments will be available.

Data analysis

The analysis of data from cross-sectional quantitative surveys can be standardised so that the results of each survey can be summarised into a small set of estimates (e.g. U5MR as the number of deaths per 10,000 person-days, prevalence of acute malnutrition (global, moderate, and severe) as percentages) which may be used in a production-rule system to define the size and scope of the overall intervention. Such an analysis will also require summary information from the Stage 2 confirmatory action. An example of a rule-base for a Stage 2 needs assessment is:

RULE 1 : IF prevalence of global acute malnutrition > X%
THEN [GENERAL RATION]

RULE 2 : IF prevalence of global acute malnutrition > X%
THEN [SUPPLEMENTARY RATION]

RULE 3 : IF prevalence of severe acute malnutrition > Y%
THEN [COMMUNITY-BASED THERAPEUTIC CARE]

RULE 4 : IF U5MR > 1.5 / 10,000 / day
THEN [COMMUNITY-BASED THERAPEUTIC CARE]

RULE 5 : IF daily labour rates have declined below subsistence levels
THEN [FOOD FOR WORK OR CASH FOR WORK]

RULE 6 : IF crop yields < Z% of normal
THEN [SEED DISTRIBUTION]

Rules for defining the size of an indicated intervention (e.g. to arrive at planned beneficiary numbers) are also required.

It should be noted that these production-rules are intended as heuristics (*rules of thumb*) intended to inform rather than dictate the scope and nature of the required response. Indicated responses will require expert evaluation for practicality and utility in the specific context by experienced field personnel and local officials.

Requirements for consultancy

The output of an initial consultancy should identify:

Identification and selection of a **minimal** set of food-security indicators.

Identification and selection of data-collection methods for food security indicators.

Preliminary versions of production-rules for analysis of confirmatory data.

Identification of additional indicators for needs assessment surveys.

Views on indicators / methods for needs assessment data collection.

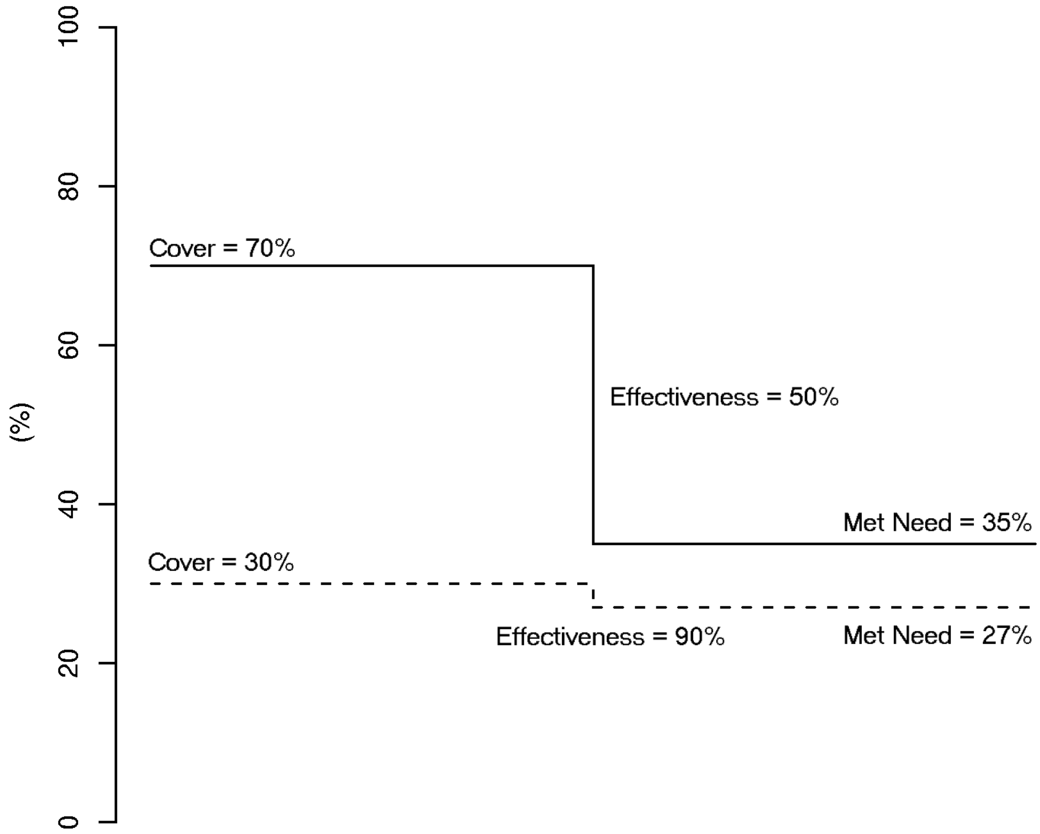
Preliminary versions of production-rules for analysis of needs assessment data.

These outputs will be collated and subjected to a second round consultation concentrating on:

Finalisation of the **minimum** datasets and production-rule systems for confirmatory and needs assessment analysis.

Stage 3 : Delivery

Once a crisis has been detected, confirmed, and appropriate responses have been identified, individual programs (e.g. general ration distributions, food-for-work programs) are implemented. These programs need to deliver interventions of sufficient quality covering a high proportion of eligible beneficiaries. *Met need* is a product of quality (assuming that quality equates with effectiveness) and coverage. Low coverage means low met need regardless of program quality:



Stage 3 delivery assessments should, therefore, measure both the quality and the coverage of delivered goods and services. Measuring one without the other is meaningless in terms of assessing whether programs are meetings needs.

Measuring intervention quality

Indicators / benchmarks of delivery will use, where available, those defined by the SPHERE project to assess the quality of individual programs and groups of similar programs (e.g. all therapeutic feeding programs). Many of these indicators will already be collected routinely by implementing agencies. Failure to collect, analyse, and report these indicators can be taken, by itself, as indicative of poor program quality.

Many interventions will have multiple quality-of-delivery indicators. For example, a general food distribution may have indicators relating to the ration diversity, ration sufficiency, completeness of delivered rations, timeliness of delivery, and the size of catchment for each distribution point. In such cases, it may be desirable to create a composite benchmark with a binary (i.e. pass / fail) outcome.

Measuring intervention coverage

Coverage has both spatial and temporal components. Good coverage is high coverage at all places at all times. This means that coverage needs to be measured routinely over the entire program area. Wherever possible methods of estimating program should be capable of:

Being integrated into routine program activities.

Acting to improve program coverage:

Directly by (e.g.) identifying uncovered eligible individuals for referral into programs.

Indirectly by identifying areas of low coverage within the wider program area where remedial action is required.

Providing an estimate of current need (since the Stage 2 needs assessment will only provide approximate estimates of need).

Providing estimates of both overall coverage and of the spatial distribution of coverage.

At present there is no SMART indicator or method that meets these requirements. A suitable method using a spatial sampling technique known as centric-systematic-area-sampling (CSAS) has been developed for nutritional interventions and this may prove suitable for assessing the coverage of other interventions. The SMART project could, however, take responsibility for the development, testing, and documentation of suitable indicators and methods.

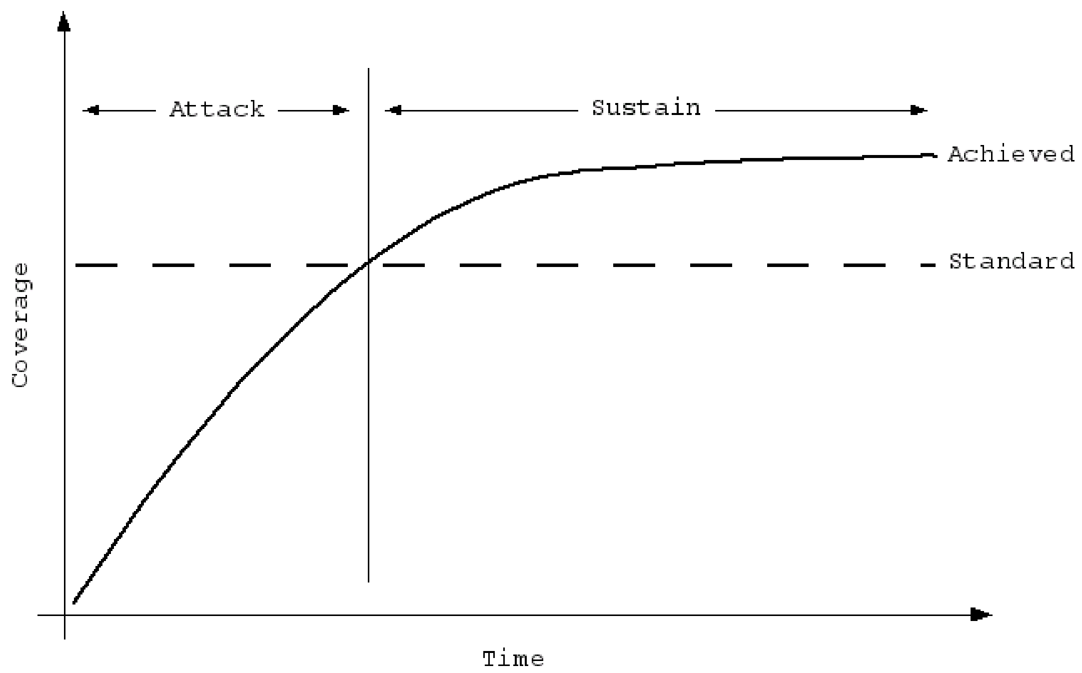
Current coverage benchmarks often ignore the division between the spatial and temporal component of coverage. Benchmark statements usually take the form:

A program of type T should have a coverage of X% in rural areas.

Might be better expressed as:

A program of type T should have a coverage of X% ***over Y% of the program area*** *in rural areas.*

Such a statement ignores the problem of achieving the desired level of coverage (the *standard*) in short periods of time and ignore the fact that coverage needs to be both achieved within a reasonable time (the *attack phase*) and, once achieved, sustained for the duration of the program (the *sustain phase*):



Coverage benchmarks might be better expressed as:

> *A program of type T should* ***achieve and sustain*** *a coverage of X% over Y% of the program area in rural areas* ***after Z months of operation****.*

Being able to indicate whether this benchmark has been met means that coverage needs to be measured frequently and routinely throughout the life of the program.

In some programs, such as provision of safe and protected water sources in WATSAN interventions, therewill only be an attack phase:

> ***All villages*** *in the intervention area will have a safe and protected water source* ***within X months****.*

Data analysis

Data analysis of single quality-of-delivery indicators require a single production-rule for each indicator. An example rule-base for a therapeutic feeding program is:

RULE 1 : IF mortality as (deaths / 100 admissions) > X%
THEN [FAIL BY RULE 1]

RULE 2 : IF defaults as (defaults / 100 admissions) > X%
THEN [FAIL BY RULE 2]

RULE 3 : IF cured as (cured / 100 admissions) < X%
THEN [FAIL BY RULE 3]

A simple compositing system could be used to indicate program failure if any component benchmark was not met:

RULE 4 : IF [FAIL BY RULE 3]
OR [FAIL BY RULE 3]
OR [FAIL BY RULE 3]
THEN [COMPOSITE FAIL BY RULE 4]

Analysis of coverage indicators during the attack phase is more difficult since it will require some assessment of trend in coverage. It may be sufficient to assess coverage for benchmarking (as opposed for internal programmatic reasons) at the end of the attack phase defined in the benchmark for a particular program type. For example:

> *A program of type T should* ***achieve*** *and sustain* ***a coverage of X%*** *over Y% of the program area in rural areas* ***after Z months of operation****.*

Such a procedure may also be used in interventions with an an attack phase but no sustain phase:

> ***All villages*** *in the intervention area will have a safe and protected water source* ***within X months****.*

Analysis of coverage indicators during the sustain phase will require the analysis of data from a sequence of coverage estimates with regard to the coverage standard for particular program types.

Analysis of coverage indicators may also be expressed as a set of production-rules:

RULE 1 : IF time since start = Z months AND coverage < X% in all areas
THEN [ATTACK PHASE COVERAGE FAILBY RULE 1]

RULE 2 : IF time since start > Z months AND coverage < X% in all areas for all months
THEN [SUSTAIN PHASE COVERAGE FAILBY RULE 2]

Missing data could also be interpreted as a program failure.

Requirements for consultancy

The output of an initial consultancy should:

Identify interventions where existing benchmarks are currently unavailable and suggest appropriate benchmarks.

Identify measurement methods for quality-of-delivery indicators where standardised methods do not already exist.

Identify methods for compositing benchmarks for interventions with multiple quality-of-delivery indicators.

Suggest appropriate coverage standards and attack phase durations for coverage indicators for different program types.

Suggest appropriate coverage standards for coverage indicators for program types without sustain phases.

These outputs will be collated and subjected to a second round consultation.

Stage 4 : Closure

Program delivery needs to be sustained until the crisis has passed. Only then can individual programs be closed and the response terminated. Premature closure risks a re-emergence of the crisis. Late closure risks damaging markets, society structures, and development efforts.

Closure is often given the name “exit strategy”. This terminology belies the fact that many project closures are often anything but strategic with programs running for an allotted time before closure regardless of the levels of need in program areas. Such a “strategy” risks closing programs too early or too late in situations where the initial needs assessment is less than perfect. A better strategy might be to close programs when there is evidence that closure is indicated.

In programs with only attack phase coverage targets, such as a program to ensure that all villages in an intervention area have a safe and protected water source, the decision to close is made only when the program coverage target has been met.

In targeted programs, such as therapeutic feeding programs, closure may be indicated by a sustained drop in beneficiary numbers provided that sustain phase coverage standards have been met. This may be sufficient to make a decision to close. Such a decision may, however, require additional information such as a nutritional anthropometry assessment showing that the prevalence of acute malnutrition has dropped to an acceptable level and / or a favourable food-security assessment. It should be noted that abrupt closure is not the only option for such programs. A measured scaling-back of activities leading to closure may be a better option:

RULE 1 : IF sustain phase coverage targets met for X months
AND beneficiary numbers have dropped for Y months
THEN [SCALE-DOWN BY RULE 1]

RULE 2 : IF [SCALE DOWN BY RULE 1]
AND prevalence of global acute malnutrition < X%
AND food-security indicators favourable
THEN [CLOSE BY RULE 2]

In non-targeted programs, such as a general ration, it is appealing to base closure decisions on the same proximate indicators that were used to initiate the program. For example, if a general ration was initiated because the prevalence of acute malnutrition was above a threshold value then a decision to close may be made when the prevalence of acute malnutrition has fallen (and remained) below the threshold value. Such an approach is, however, likely to be too simplistic to be useful in many contexts since the indicators being used to make the decision to close a program will often be strongly influenced by current program activity. Continuing the example of a general ration, a large part of the observed reduction in the prevalence of acute malnutrition will, if a high quality and high coverage program is delivered, probably be due to the general ration. If the general ration is closed and (e.g.) daily labour remain below subsistence levels then closure may result in the prevalence of acute malnutrition rising to intervention levels. In this case, simplistic decision-making results in a “stop-go” program or, more likely, a premature closure with serious negative consequences for some part of the population in the intervention area. With such programs it may be better to base closure decisions on a wider range of indicators than those used to initiate the program. Such an analysis is likely to lead to the replacement of non-targeted programs with one or more targeted programs. In the above example (i.e. of an acceptable prevalence of acute malnutrition but low daily labour rates), a food-for-work (FFW) program could be initiated and the general ration closed only when the attack phase coverage target for the FFW program has been met.

Indicators, such as the prevalence of acute undernutrition, are commonly measured at the level of the entire program area. This assumes that recovery is even across the entire program area. This may not be the case. In situations where recovery is patchy it may be possible to close the program in some areas whilst concentrating program resources in areas that still require intervention. A process of targeted closure will, however, require the use of (e.g.) prevalence estimation techniques that are capable of providing information on the spatial distribution of prevalence. Currently available SMART methods do not provide such information. The SMART project could, however, take responsibility for the development of suitable methods.

In situations where there are patches of high prevalence of acute malnutrition there is a risk that the SMART method for prevalence estimation will return false-negative results (i.e. indicate that prevalence is low overall when it is, in fact, high in specific areas). This is because the survey method is restricted to returning an average prevalence for a wide area and because the sampling method concentrates data-collection in the most populous communities. Areas of low population density (i.e. those areas consisting of communities likely to be distant from health facilities, feeding centres, and distribution points) are often not sampled. This may cause surveys to evaluate prevalence as being low even when prevalence is high in areas outside of urban centres. Such an error may lead to premature program closure.

Data analysis

As with all other stages in this framework, production-rules may be of use in informing exit decisions.

Requirements for consultancy

The output of an initial consultancy should identify:

Identify indicators and production-rules for program scale-down, replacement / diversification, and closure.

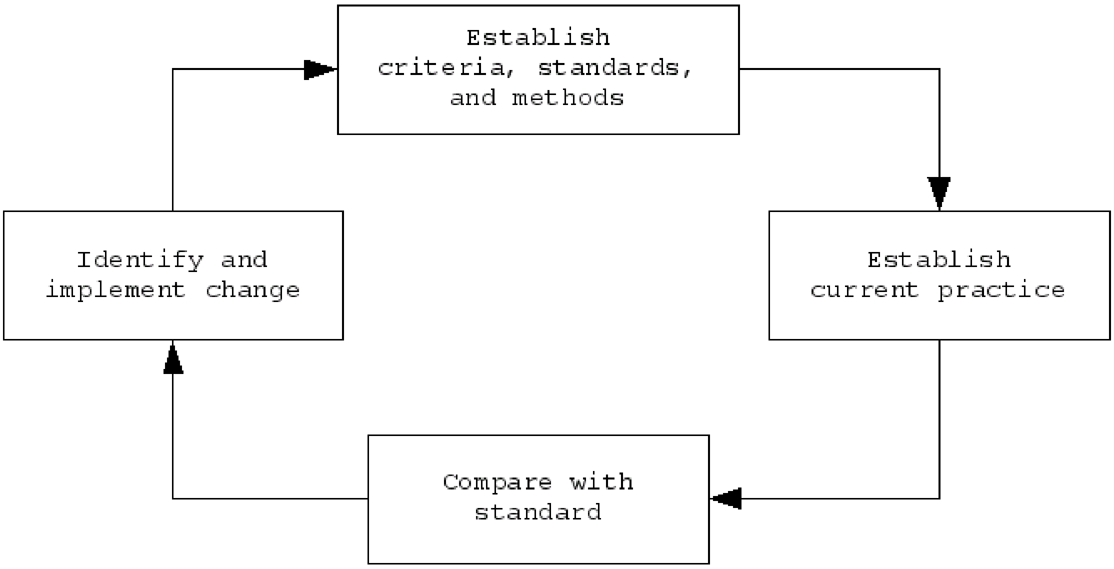
Views on the need for methods capable of providing information on the spatial distribution of prevalence.

These outputs will be collated and subjected to a second round consultation.

Benchmarking and change

The four-stage framework concentrates on how indicators and benchmarks can, when expressed in production-rules, be used to detect a crisis, confirm a crisis, assess need, define the scope and nature of the required response, evaluate quality and coverage, and inform program closure.

The clinical audit cycle provides a framework in which benchmarks can be used to improve humanitarian practice:



Many humanitarian actors have a limited view of benchmarking that excludes change. For example, an agency may use SPHERE minimum standards (*Establish criteria, standards and methods*), review program performance (*Establish current practice*), evaluate program performance with regard to the SPHERE minimum standards (*Compare with standard)*) but do little when they find that their programs fail to meet SPHERE minimum standards preferring, instead, to continue with program designs, delivery models, and personnel policies that consistently fail to deliver programs of sufficient quality or of sufficient coverage. In other words, they fail to apply the *Identify and implement change* component of the clinical audit cycle. It is only by embracing change that benchmarks will be consistently met and, with time, exceeded. Benchmarking without a commitment to change is a futile exercise.

The clinical audit cycle framework can be applied to interventions at all levels. In the field at the program level, it may indicate the need to improve coverage causing the implementing agency to investigate barriers to accessing services implement appropriate changes to the program. At the agency level, it may indicate that a traditional model of service delivery (e.g. centralised therapeutic feeding programs) consistently fails to deliver adequate coverage and cause the agency to move towards more decentralised community-based model of service delivery. At higher levels, for example, it may indicate problems with mobilising resources at short notice and cause donors, NGOs, and UNOs to change decision-making and resource allocation procedures.

The cyclical nature of the clinical audit cycle also informs the choice of standards. Without such a framework standards may be set that can only be met in the longer term (aspirational standards) or can be easily met in the short term without changes in practice (normative standards) or, perhaps, set somewhere in between (compromise standards). The clinical audit cycle allows for standards to be set that can be met in the short to medium term and once consistently met are revised upwards. The idea is to create a “virtuous circle” of rising expectation and aspiration that is reinforced by a series of small steps achieved.

Requirements for consultancy

The output of an initial consultancy should identify:

> Identify ways in which identifying and implementing change to improve performance may be encouraged all levels.

These outputs will be collated and subjected to a second round consultation.

Benchmarks at multiple levels

Many of the examples given in this outline document consider benchmarks that can be applied to the delivery of individual programs by implementing agencies. It should be recognised, however, that barriers to good humanitarian practice exist at levels of the humanitarian system. For example, benchmarks, such as those proposed by the *Good Humanitarian Donorship* initiative, can be applied to donors.

A set of benchmarks for the four-stage model are also required. These will include stage-specific benchmarks. For example, benchmarks such as:

> *The minimal Stage 1 dataset will be collected from all agro-ecological zones*.
>
> *The minimal Stage 1 dataset will be collected on a quarterly basis.*
>
> *X% of Stage 1 analyses should be completed and reported within Y weeks of data becoming available.*

May apply to Stage 1 of the four-stage framework.

Benchmarks for the interface between stages and components of the four-stage model are also required. For example, benchmarks such as:

> *A Stage 2 confirmation action should be initiated within X days of a positive Stage 1 crisis detection result.*
>
> *A Stage 2 confirmation action should report within X days of initiation.*

May apply to the interface between Stages 1 and 2 of the four-stage framework.

Benchmarks such as:

> *A Stage 2 needs assessment action should be initiated within X days of a positive Stage 2 crisis confirmation.*
>
> *A Stage 2 needs assessment action should report within X days of initiation.*

May apply to the interface between the confirmation and needs assessment components of Stage 2 of the four-stage framework.

Requirements for consultancy

The output of an initial consultancy should identify:

> Benchmarks for levels other than at the level of implementing agencies and identify where they should be placed with the four-stage framework.
>
> Benchmarks applicable to the four-stage model.

These outputs will be collated and subjected to a second round consultation.