Review of the published literature for the impact and cost-effectiveness of six nutrition related emergency interventions

Report prepared by the ENN

December 2004

Authors:

Arabella Duffield
Garth Reid
Damian Walker
Jeremy Shoham

Executive Summary

The ENN received funding from CIDA to conduct a review of the published evidence for the impact and cost-effectiveness of six key humanitarian interventions commonly implemented in emergencies (general rations, supplementary feeding, therapeutic feeding, measles vaccination, vitamin A supplementation and bednet distributions). The overall aims of the review were to identify gaps in the literature and develop methodologies and institutional mechanisms for filling these gaps.

The recent interest in evidence-based public health practices has emerged from the more general movement towards evidence based medicine. Clinical decisions are based on the best available scientific data rather than on customary practice. Set against this background, the review has focused on a narrow definition of impact – it is measured as a change in population nutrition prevalence or mortality rates due to an intervention.

A hierarchy of study types is generally recognised in clinical medicine. The randomised control trial is considered the 'gold standard' method and provides the highest level of evidence while studies with controls also provide a high level of certainty in most cases. Observational studies and case-series data provide the weakest level of evidence. Given the practical difficulties of conducting an RCT in an emergency situation, this report uses a framework (adapted from Habicht et al, 1999[1]) with which to assess the strength of the different studies reviewed. All types of economic evaluation information were included in the review.

Standard methods to collate and appraise the literature for a systematic review were employed. The method involved (i) a search of the 5 most relevant databases, (ii) a secondary reference search, (iii) a hand search of the main journals, and (iv) expert advice on the literature. The quality of each study was assessed using standard critical appraisal techniques.

The most important finding of the review is that very few studies assessing the impact of any of the interventions in an emergency context have been published. There is virtually no publicly available information on the cost-effectiveness of different nutrition-related interventions commonly implemented in emergencies.

The number of published impact and economic evaluation studies undertaken in emergencies located by the search

Type of intervention	Impact assessment	Economic evaluation
General food/ration distribution (GFD)	9	0
Supplementary feeding programme (SFP)	15	1
Therapeutic feeding programme (TFP)	16	1
Vitamin A supplementation	0	0
Bednets programmes	0	1[2]
Measles immunisation programme	0	0

No studies assessing the impact of vitamin A supplementation, bednets programmes or measles immunisation programmes in emergencies were found. There were, however, quite a number of high level studies which reported on the positive impact of these programmes in

[1] Habicht, JP, Victora, CG and Vaughan, JP (1999) *Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact.* International Journal of Epidemiology 28: 10-18.

[2] The economic evaluation of bednets was a cost-effectiveness study which included a measure of impact so this study could be classified in either column.

non-emergency settings. How far these studies are transferable to emergency programmes will depend on the context (security situation, etc) of the programme area. A limited number of studies assessed the impact of GFD and SFP in emergencies, however the majority of these were observational and do not provide very plausible evidence of impact. The evidence base for TFPs is somewhat stronger.

The lack of published impact and cost effectiveness information - particularly in relation to emergency feeding and food security support programmes - is of enormous concern. There are key areas of uncertainty regarding both the utility of certain types of intervention, e.g. SFP or GFD versus cash transfer, and over issues of design within programme types, e.g. community versus administrative targeting in general ration programmes. There are also rapidly emerging new types of programming at the interface of HIV and nutrition for which impact and cost information is urgently needed. This lack of impact and cost-effectiveness information militates against cross-sectoral comparison of interventions in relation to nutrition and mortality impact.

There are a number of understandable reasons for the dearth of published information, e.g. the ethical difficulties of undertaking research in emergencies, and the fact that there are far fewer epidemiologists involved in emergency feeding than in more medically oriented interventions like measles vaccination. However, one over-arching key factor is the absence of an agency with responsibility for taking an overview of the effectiveness of different types of intervention. This lack of corporate accountability has allowed the institutional status quo to prevail. Thus agencies, which have built up expertise and mandates around certain types of intervention (or intervention design), will continue to practice these interventions in emergencies without serious examination or challenge.

This review argues that one way to address the gap in information on impact and cost-effectiveness is to make greater use of the so called ‘grey literature’ (unpublished information held mainly by implementing agencies which may be in a variety of forms, e.g. project reports, annual audits, monitoring forms, etc). Greater standardisation of agency reporting will enhance capacity to use this type of information. However, it is also probable that much of this grey literature could be used retrospectively to answer a number of questions. The review discusses how to increase access to, and use of, the grey literature. In conjunction with this, specialised impact studies could also be commissioned to address key questions. The review examines how these studies may be carried out for each of the six interventions by identifying the most ethically feasible and methodologically robust approach.

The review also explores the gap in information on costs of interventions and methodologies for obtaining such information. It is recognised that this is not a straightforward discipline, methodologies need to be developed and reporting standardised.

Given the multiplicity of stakeholders (and vested interests) in this sector the review argues the case for creation of an independent body/institutional mechanism with responsibility for increasing information on impact and cost-effectiveness in this sector. Without establishing such a body, it is likely that little will change. This body would take responsibility for identifying key gaps in knowledge regarding impact and cost-effectiveness. It will develop and co-ordinate mechanisms for making greater use of the grey literature and promoting impact studies. The agency would also have an advocacy role where emerging evidence indicates a need for change in implementation practice.

In the event that there is insufficient support for establishing such a body a more piecemeal and potentially realisable alternative may be for donors (individually or as a group) to take more responsibility. This could entail funding/helping to establish research/implementing agency partnerships, which aim to address specific questions in particular programme areas where impact and cost-effectiveness information are urgently needed.

Acronyms

AAH	Action against Hunger
ACC	Agency Committee on Coordination
ACF	Action contre la Faim
AIDS	Acquired Immune Deficency Syndrome
ALNAP	Active learning network for accountability and performance in humanitarian action
ARV	Anti retroviral
CAB	Commonwealth Agricultural Bureau
CASP	Critical appraisal skills programme
CBA	Cost-benefit analysis
CBR	Case based reasoning
CDC	Centre for disease control
CEA	Cost-effectiveness analysis
CIDA	Canadian International Development Agency
CINAHL	Cumulative Index of the Nursing and Allied Health Literature
CMR	Crude mortality rate
CRD	Centre for reviews and dissemination
C-SAFE	Consortium for the southern African food emergency
CSAS	Centric systematic area sampling
CSB	Corn soya blend
CTC	Community therapeutic care
CUA	Cost-utility analysis
DALY	Disability-adjusted life years
DEC	Disasters emergency committee
DOT	Direct Observation Treatment
DRC	Democratic Republic of Congo
DYLG	Discounted year of life gained
EMOP	Emergency operation
ENN	Emergency nutrition network
EPI	Expanded programme on immunisation
ERHS	Ethiopian rural household survey
FD	Food distribution
FFW	Food for work
GFD	General food/ration distribution
HAZ	Height for age z-score
HEED	Health Economists Evaluation Database
HH	Household
HIS	Health Information System
HIV	Human Immunodeficiency Virus
ICRC	International Committee of the Red Cross
IDP	Internally displaced person
ITN	Insecticide treated bednets
MEP	Minimum evaluation procedure
MOU	Memorandum of Understanding
MSF	Medecins sans Frontiers
MUAC	Mid upper arm circumference
NGO	Non-governmental agency
NHS	National health service
NICS	Nutrition Information in Crisis
NVASP	National Vitamin A Supplementation Programme (NVASP)
OECD	Organisation for economic cooperation and development

OTP	Outpatient therapeutic programme
PE	Protective efficacy
PHC	Primary health care
PLWHA	People living with HIV/AIDs
PMTCT	Prevention of mother to child transfer
PVO	Private voluntary organisation
QALY	Quality-adjusted life years
RCT	Randomised controlled trial
RNIS	Refugee nutrition information system
RUTF	Ready to use therapeutic food
SC	Stabilisation centre
SCN	Standing committee on nutrition
SC UK	Save the Children UK
SFP	Supplementary feeding programme
TFC	Therapeutic feeding centre
TFP	Therapeutic feeding programme
UK	United Kingdom
UNHCR	United Nations High Commission for Refugees
US	United States
WAZ	Weight for age z-score
WB	World Bank
WFP	World Food Programme
WFH	Weight for height
WHM	Weight for height median
WHO	World Health Organisation
WHZ	Weight for height z-score

1.0 Introduction

The objectives of this review were as follows;

i) to conduct a robust search of the published literature for evidence of impact and cost-effectiveness of six key types of humanitarian interventions (general rations, supplementary and therapeutic feeding, vitamin A supplementation, bednet programmes and measles vaccinations)
ii) to assess and summarise the published evidence and identify gaps
iii) to identify methodological constraints to assessing impact and cost-effectiveness of these types of intervention in emergencies
iv) to identify methods and analytical frameworks which would strengthen knowledge of impact and cost-effectiveness
v) to identify mechanisms, processes and strategies which will lead to an increased body of evidence on impact and cost-effectiveness for these types of intervention.

In this introductory section of the review, we provide a brief description of why it is important to measure impact and cost-effectiveness. We then explain the definition of impact that we have chosen to use. We also summarise some of the latest thinking on the appropriateness of different models of evaluation, which might be useful when examining the impact of nutrition-related interventions in emergency-affected populations. A description of the interventions which are assessed in this report - general food/ration distribution (GFDs), supplementary feeding programmes (SFPs), therapeutic feeding programmes (TFPs), vitamin A supplementation, bednets programmes and measles vaccinations – is provided in Annex 1.

1.1 Why is it important to measure impact and cost-effectiveness?

Although questioning the impact of humanitarian assistance is not new, it has moved up the humanitarian agenda in recent years (Hofmann et al, 2004). Over the last decade spending by aid agencies on emergencies has quadrupled to over US$6 billion (Griekspoor et al, 1999). Such dramatic increases in the level of expenditure on humanitarian aid operations are inevitably leading to increased scrutiny of the way emergency assistance is provided (Hallam 1996). Also, reforms within the West's public sector have led to the introduction of new management systems focusing on results (Macrae et al, 2002)

The greater interest in impact analysis of humanitarian interventions over the past few years has manifested itself in a number of ways, for example the advent of results based management into routine workings of organisations like WFP. As a result, new monitoring systems have emerged, for example the community household surveillance system in southern African EMOP countries (WFP/C-Safe).

Despite the increase in spending on emergencies, public health actions are often faced with severe financial constraints. Under these circumstances, information on the relative effectiveness and costs of different interventions in different contexts would be extremely useful to decision makers. However, to date, cost-effectiveness has seldom been considered in the prioritisation and evaluation of emergency interventions. This may be because many aid workers consider it impossible, and in many cases unethical, to consider the cost and cost-effectiveness of emergency aid (Hallam, 1996).

Relative information about costs may be crucially important where there are unresolved issues over comparative efficacy of different types of intervention or over optimal programme design. Such issues are particularly relevant to emergency feeding programmes. There are, for example, many unanswered questions over costs and effectiveness of emergency GFD programmes, i.e. in relation to market support, cash for work or cash transfers, all of which

could theoretically meet the same objectives at lower cost. There are design issues, e.g. the efficacy of community based targeting compared to allocation of ration cards to household heads. There are also major unresolved issues regarding the role and impact of emergency supplementary feeding, i.e. what is the overall evidence for impact of this type of programme, what situations predispose to SFPs reaching Sphere targets and can equivalent impact be achieved at lower cost by increasing the general rations? Impact studies to demonstrate optimal design of TFPs are also critically needed, i.e. a comparison of home versus centre based treatment.

Furthermore, as the nutritional context for certain global regions change due to the HIV/AIDS crisis, the nutritional objectives and modus operandi of emergency feeding programmes also become altered. Modified objectives need to be tested through impact assessment and cost-efficiency study before new and innovative programming is rolled out on a large scale.

1.2 Definitions of impact

There are several different definitions of impact each with their own limitations. The most commonly used definition within the development sector is provided by the OECD/DAC (2002):

'The positive and negative, primary and secondary, long-term effects produced by a development intervention, directly or indirectly, intended or unintended.'

The OECD/DAC definition refers explicitly to development interventions, and applies only imperfectly to humanitarian assistance in emergencies. In contrast to the emphasis on the long term in this definition, humanitarian interventions tend to have a short-term focus, and this is not captured here (Hofmann et al, 2004). Moreover, if impact is defined as concerned only with lasting change, then the idea of 'short-term impact' becomes a contradiction in terms. Oxfam accordingly defines impact as lasting or significant change in people's lives, in recognition of the fact that, in humanitarian response, saving someone's life is significant, even if the effect is not *lasting*, and that the individual is again subject to life-threatening risk at some later point (Roche, 1999).

This report uses a much more precise (and in some ways limited) definition of impact than that provided by OECD/DAC:

'Impact is measured as the change in population nutrition prevalence or mortality rates due to an intervention.'

This working definition employs a scientific approach to measuring impact as opposed to either the deductive/inductive approach which is more anthropological and socio-economic, or participatory approaches which depend on obtaining views of those participating from a programme (Hallam, 1998). By defining impact so narrowly this review will, of course, miss many of the impacts which an intervention might have on an emergency-affected population. For example, a GFD programme might prevent people having to migrate to find food – this is a positive impact that our analysis will not pick up on. Alternatively, a centre-based TFP may prevent a child from dying in the short term but could have negative impacts on the household's long-term food security, if the carers' were unable to plant for the following season because they were looking after the child in the centre during the crucial planting season.

This report does not attempt to assess the wider impact of the programmes described because it was beyond the scope of the terms of reference. However, the experience generated from completing this review makes it seems unlikely that there is much published information about the wider impact of these interventions.

A variety of terms such as ‘outcome’, ‘results’ and ‘effect’ are also related to the term ‘impact’. The distinctions between all of these terms are not always clear, and they are sometimes used inter-changeably. There is particular confusion between the terms ‘outcome’ and ‘impact’ (Hofmann, 2004). The concept of an ‘impact chain’ or ‘results chain’ (Roche, 1999) is often used to show causality between an action and its ultimate impact. Reduced to its simplest form, the impact chain looks like this:

inputs ► activities ► outputs ► outcomes ► impact

Although the terminology varies, the literature generally distinguishes between two types of indicators along this chain: those that relate to the implementation of a programme (inputs, activities, and output indicators) and those concerned with the effects of programmes (outcomes and impact indicators) (Hofmann, 2004). An example is provided in table 1.1.

Table 1.1 Types of indicators: example of measles immunisation programmes

Implementation of the programme			Effect of the programme	
Input indicator	Activity indicator	Output indicator	Outcome indicator	Impact indicator
No. of vaccines administered	No. of people trained	Percentage vaccinated	Measles cases decreases	Mortality decreases

Care is needed when applying this idea of a chain to emergency humanitarian aid. First, there is an important distinction between long and short impact chains; the fewer the links in the chain, the easier it is to assess whether a given input achieves an impact (Roche, 1999). For example, if a child is in a centre-based TFP and he/she receives treatment from the TFP and does not receive food or any other input from other sources and the child recovers, then we can be fairly certain that the TFP has had an impact on his/her nutritional status. In such a situation, the impact chain is short. By contrast, there is a longer chain of causality between a GFD programme and any possible impact on nutritional status, due to the variety of factors which can impact nutritional status that are unrelated to the intervention. This may make impact harder to demonstrate (Hofmann, 2004).

Another way to think of the connection between an intervention and its impact in emergency programmes is the idea of “causal nets”. Often there is no clear chain linking the intervention to an impact, but rather there is a net of different factors that contribute to the impact. This 'causal net' can be broken down into small sections and investigated with appropriate study designs for each section.

It is important to distinguish between the terms ‘effectiveness’ and ‘efficacy’ when assessing the impact of emergency programmes. The efficacy of an intervention is defined as its effect under “ideal” circumstances. The effectiveness of an intervention is defined as its effect under normal conditions in field settings. In emergency programmes, where many things can go wrong, effectiveness is a more useful measure. This report tries to distinguish, as far as possible, when programmes are ineffective whether or not the ineffectiveness is due to a problem of coverage or whether the programme has failed to have an impact because it is not making a difference (in terms of mortality or nutritional risk) to the beneficiaries enrolled on the programme.

1.3 Evaluation methods

Over the past several decades, a strong move towards evidence-based medicine has emerged. In the context of evidence-based medicine, clinical decisions are based on the best available

scientific data, rather than on customary practices or the personal belief of the healthcare provider. There is now a parallel movement towards evidence-based public health practices. The movement is intended to utilise the best available scientific knowledge as the foundation for public health-related decision making (des Jarlais et al, 2004).

When we are conducting, or planning to conduct, research (or a review of research) in emergencies, we need to decide what type of study we should assess or conduct. This is not always straightforward but in a seminal piece of work for UNICEF, Habicht et al (1999) described a framework for the evaluation of health and nutrition projects. This framework is very useful when thinking about what evaluations donors need in order to assess whether or not the nutrition and health programmes routinely undertaken in emergencies are actually effective.

Section 1.3 of this report will attempt to summarise Habicht et al's framework and identify the most pertinent points of the research undertaken for UNICEF. The discussion below draws heavily from four important papers.[3] This introduction section will also consider some recent discussions about the usefulness of RCTs in public health research (Victora et al, 2004; des Jarlais et al, 2004).

The main thrust of Habicht et al's argument is that different types of evaluations must be used to influence different types of decisions in a project cycle. In other words, the complexity and precision of an evaluation must depend on who the decision maker is and on what types of decisions will be taken as a consequence of the findings. For example, the following sequence of basic issues to be addressed is of particular interest to different audiences:

- Is the intervention performing as expected? (Programme managers, administrators, and funders)
- Is the intervention worth continuing? (Administrators and funders)
- Should the intervention be extended? (Administrators and funders)
- Is the intervention causally linked to improved nutrition? (Researchers, scientists, and others concerned with basic mechanisms of cause and effect)

1.3.1 What kind of information should be sought?

The first step in planning an evaluation must be to decide what to evaluate. This report employs a model of evaluation based on the following five basic indicators: (i) provision, (ii) utilisation, (iii) coverage, (iv) outcome and, (v) impact. Note that there are many other models of evaluation for example, the ALNAP model or the DEC model (Hofmann et al, 2004).

---

[3] Habicht, JP, Victora, CG and Vaughan, JP (1999) *Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact.* International Journal of Epidemiology 28: 10-18.

Habicht, JP, Mason, JB and Tabatabai (1984) Basic concepts for the design of evaluation during programme inmplementation. In *Methods for the evaluation of the impact of food and nutrition programmes.* Ed Sahn, DE, Lockwood, R and Scrimshaw, N. United Nations University Press.

Mason, JB and Habicht, JP (1984) Stages in the evaluation of ongoing programmes. In *Methods for the evaluation of the impact of food and nutrition programmes.* Ed Sahn, DE, Lockwood, R and Scrimshaw, N. United Nations University Press.

Carletto, C and Morris, S (2001) Designing methods for monitoring and evaluating food security and nutrition interventions. In *Methods for rural development projects.* Ed Hoddinott, J. International Food Policy Research Institute: Washington DC.

Provision means that services are available and accessible to the target population and are of adequate quality. Utilisation implies the measurement of the rate of use of these services. The issue of coverage asks whether the target population is being reached. The provision of a service by a project, if extended to and properly utilized by a sufficiently large number of beneficiaries, should have an impact on certain variables of interest among the beneficiary population. A number of relations and assumptions link the provision of the service to its impact. A thorough understanding of the existence and strength of these linkages will have a major effect on the form of interventions proposed by the project and, ultimately, on the design of the evaluation system. Table 1.2 presents the outcomes of interest in a logical order leading from provision to impact.

Table 1.2 Outcomes of interest for evaluations of emergency nutrition programmes

Indicator	**Question**
Provision	Are the services available? Are they accessible? Is their quality adequate?
Utilisation	Are the services being used?
Coverage	Is the target population being reached?
Outcome	Have the beneficiaries' nutritional status improved?
Impact	Were there population level improvements in patterns of food insecurity or malnutrition?

As stated above, the first three indicators (provision, utilisation and coverage) are also known as performance indicators. An assessment of these indicators is known as a performance evaluation. This is in contrast to an impact evaluation, which assesses outcome and impact.

It is important to note here that there is no point in undertaking an impact assessment unless an assessment of the performance indicators suggest that it is likely that an impact may have been achieved. WHO's minimum evaluation procedure (MEP) for water and sanitation programmes explains this point elegantly:

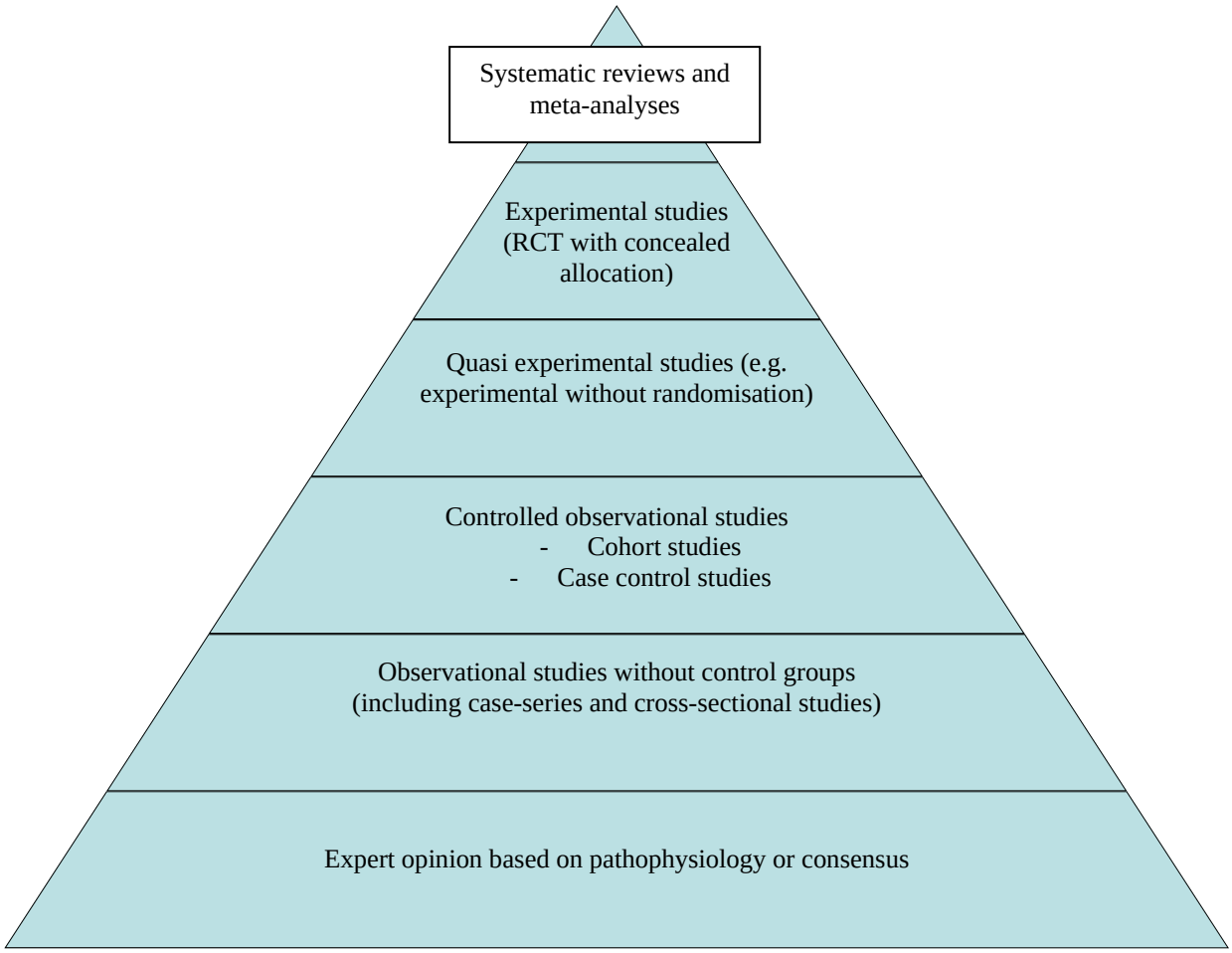
'The ultimate objectives of allocating resources for water supply and sanitation investments are to improve the health, welfare and economic status of the users of the facilities constructed. These objectives cannot be fully achieved unless the facilities are firstly, functioning in the correct way and, secondly, utilised by the community. Thus the MEP is designed to evaluate functioning and utilisation.'

Hence, the first objective of an evaluation exercise is usually to assess service provision. Once this is done, it may be important to evaluate the level of utilisation of such services by the intended beneficiaries and their coverage (take-up) by the project's target groups. It is only when the correct service is provided in a timely manner and properly utilised by a sufficiently large number of beneficiaries that one can plausibly expect an impact on the indicator of interest. Only in these cases is an impact evaluation required or justified.

1.3.2 What level of accuracy does the evaluation need?

The diagram below shows a pyramid which ranks the different strengths of different types of studies commonly employed in medical research. The studies at the top of the pyramid are considered to provide the strongest evidence (and hence are the most relevant and important studies) and those beneath are considered less strong.

Figure 1.1 Commonly employed evaluation methodologies ranked according to their strength



Hierarchical ranking of different types of studies have been described by many authors and organisations working in clinical medicine. (for example, the National Health Service (NHS)). In addition to this hierarchy of evidence, individual studies can be ranked according to a standard scoring system to enable authors to make a judgement about whether or not a study is of good quality. For example, an RCT can be scored as “good”, “medium” or “poor” quality. The scoring system employed in this report is described further in the methods section.

Habicht et al (1999) distinguish between three types of evaluations which provide varying levels of confidence in the results: adequacy, plausibility and probability. These distinctions are similar, but not identical, to those found in the pyramid above.

Adequacy assessments: did the expected changes occur?
An adequacy assessment simply determines whether some outcome actually occurred as expected, for example, did food security/nutritional status improve? Inferences about the adequacy of programme outcomes depend on the comparison of the performance, or impact, of the project with previously established criteria. These criteria may be fixed (or absolute) for example, 80% of the population receives food aid. Or they may change over time, for example, a 10% decline in the prevalence of malnutrition.

Adequacy assessments do not require control groups if results are set against fixed criteria. However, for assessing the adequacy of change over time (for example, change in the prevalence of malnutrition over time), then at least two measurements will be required.

This type of assessment may be particularly relevant when evaluating performance indicators such as the provision, utilisation, or coverage of a particular project activity (for example, the distribution of improved seed varieties). It is less useful for impact evaluation, since it is unable to isolate the effects of the project from those of other concurrent processes, such as improvement in malnutrition due to a harvest, general socio-economic improvements, and the presence of other projects in the area, etc. In many cases, an adequacy evaluation can only state that 'there has been an improvement in X…but whether or not this is entirely, or only partially, due to the project is not known.'

Adequacy assessments are frequently used by programme managers who need to find out whether or not they are reaching their performance indicators. Because these types of assessments are often unable to provide substantial evidence of impact, they are less useful for donors who need to make decisions about project effectiveness, although they may suggest the direction in which the project is likely to have an impact.

In terms of the pyramid, Habicht's adequacy studies are usually observational studies. The studies are often cross-sectional (showing an association between one variable and another) or use case-series data (a collation of reports on the treatment and outcomes of individuals).

Plausibility assessments: did the programme seem to have an effect over and above other external circumstances?
Some decision makers need a greater degree of confidence that any observed changes were, in fact, due to the programme. Plausibility assessments go beyond adequacy assessments by trying to rule out confounding factors, which might have caused the effect. A statement is said to be plausible if it is 'apparently true, or reasonable, winning assent, a plausible explanation.'

Plausibility assessments attempt to control for the influence of confounding factors by choosing control groups before an intervention is undertaken, or afterwards during data analysis. There are several different types of control groups including historical, internal and external groups (see annex 2 for definitions of these terms). The use of any control group described makes the results of an evaluation more plausible than if no controls are used at all. A combination of control groups is useful.

Plausibility assessments encompass a continuum, ranging from weak to strong statements. At the lower end of the plausibility scale are the simple comparisons with a control group, with an attempt to rule out possible confounding. At the higher end of the scale, one may have several comparisons and mathematical simulations. To reach the highest end of plausibility, one must formally discard all other likely explanations for the observed improvements.

In terms of the pyramid, Habicht et al's plausibility studies are usually controlled observational studies, such as cohort or case-control studies (level 4). A quasi-experimental study (e.g. an experimental study without randomisation) may also fall under the plausibility category (level 3).

Probability assessments: did the programme have an effect (P<x%)?
Probability evaluations aim at ensuring that there is only a small known probability that the difference between programme and control areas were due to confounding, or bias, or chance. These evaluations require randomisation of treatment and control activities to the comparison groups. RCTs would fit into Habicht et al's probability evaluations categories.

According to the thinking described in figure 1.1, the gold standard of evaluations are meta-analyses of RCTs which involve randomly selecting some individuals for treatment and others for a placebo. RCTs have been the benchmark and champion of the era of evidence-based medicine over the last 20 years and RCTs are still the gold standard for specific interventions in precise, non-complex contexts for evidence-based medicine (Lawlor et al, 2004).

Although probability assessments are usually considered the gold standard of academic efficacy research they may not be so useful in emergency settings. There are a number of reasons why this may be, including:

- The need to overcome political and/or ethical problems in randomisation of the intervention. In emergencies this is often unfeasible. In some situations it may be possible to overcome these problems by using the 'stepped wedge design' in which the intervention is deployed in a randomised sequence, but eventually extended to all eligible communities or individuals. However, in emergencies this is often unfeasible.
- The evaluator needs to be present at the very early stage of planning to ensure randomisation. Under normal emergency circumstances this is simply not the case. Expert evaluators are normally not part of national staff teams and often, even at headquarters, there may only be one or two people who are able to plan and implement a satisfactory evaluation.
- The costs of full-scale evaluations may be very high.
- In some situations, the stringent conditions of probability trials may result in situations that are artificially different from the reality to which the results must be extrapolated. In this situation, the assessment may lack external validity which means that the result may not be generalisable and the donor will not know whether or not the programme can be expanded to other areas.

(Reed et al, 2002; Banatvala and Zwi, 2000; Victora et al, 2004; des Jarlais et al, 2004; Black, 1996).

1.3.3 Choosing the evaluation design

Table 1.3 (adapted from Habicht et al) shows some areas of evaluation which may typically concern different decision makers in the field of health and nutrition. Remember that complex evaluations (for example, those with a plausibility approach) should not be carried out before ensuring, through less costly interventions, that the process is moving in the expected direction.

Table 1.3 Possible areas of concern of different decision makers

Type of evaluation	**Provision**	**Utilisation**	**Coverage**	**Outcome/Impact**
Adequacy	Health centre manager International agencies		District health manager International agencies	
Plausibility	International agencies			Donor agencies, scientists
Probability	Donor agencies, scientists			

From table 1.3 we can see that Habicht et al ideally recommend that donors need, at a minimum, plausibility evaluations of impact in order to make their decisions. However, the authors do make an exception to this rule for programmes for which the efficacy of the intervention is already well known. In a perfect world, interventions would only be widely

applied after their clinical and public health efficacy has been proven. However, efficacy is often not demonstrated before interventions are initiated. The known efficacy of an intervention, therefore, is another important factor affecting the choices of evaluation design.

An example of this could be a measles immunisation programme. The efficacy of a measles immunisation is well proven. If adequacy evaluations show that the cold chain is operational and that the coverage is high, there is little need to evaluate the impact of immunisation on disease rates. The case is rather different, however, when assessing the use of home gardens to promote vitamin A status. Their efficacy has not yet been established. Demonstration of increased ingestion may be insufficient to persuade donors of the utility of this approach without measures of vitamin A status and at least a strong plausibility design.

1.3.4 Combining different evaluation designs

The realisation that RCTs are impractical for research in most emergency situations has led more researchers to move towards using a range of research methodologies for evidence based public health (Waters and Doyle 2002; Black 1996; Concato, 2004; Des Jarlais *et al*, 2004; Kirkwood, 1997; Lawlor *et al*, 2004). Petticrew & Roberts (2003) argue against the rigid hierarchy of evidence and emphasise the need to match research questions to specific types of research (a 'horses for courses' approach).

The causal chain between context, intervention and outcome is very complex in emergencies. Thus in all evaluations of emergency nutrition programmes, even a probability assessment, it is necessary to include some information about process indicators to strengthen confidence that the impact is attributable to the programme. For example, a reader would be more convinced of the effectiveness of a GFD programme if they were provided with information on food basket monitoring rather than just information on nutritional status.

Clearly, the more evidence of the congruency of different impact evaluation types the better. We will return to this point in the final section of this report.

1. 4 Overview of economic evaluation

Economic evaluation attempts to identify ways in which scarce resources can be efficiently employed. The basic task of any economic evaluation is to identify, measure, value and compare, the costs and consequences of the alternatives being considered (Drummond et al, 1997). More specifically, economic evaluation techniques include cost-minimisation analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis (see annex 2 for definitions of the different types of economic evaluations). In such applications, health programmes are compared for their benefits and costs, where costs refer to the value of opportunities foregone from not employing resources elsewhere. Benefits are gauged by the consequences of a health programme on people's well-being or health status. The various evaluation techniques estimate costs in a similar fashion, but differ in the measurement of health outcomes.

One point to consider about cost-effectiveness analyses is that decision makers must have criteria against which to decide whether or not an intervention is actually cost-effective or not. In cost minimisation analyses, two or more interventions that have identical outcomes (e.g. number of cases treated) are assessed to see which provides the cheapest way of delivering the same outcome. These evaluations can stand alone. However, cost-effectiveness analyses are somewhat different because they work out how much it costs to obtain an outcome, for example how much it costs to cure a child from severe malnutrition.

Often researchers simply state that a programme is effective and conclude that the intervention is cost-effective, but never state what their criteria is. When no comparators are

stated, what defines 'cost-effective'? A quick comparison with the results of other studies, which reported their results using the same outcome measure, can help place results in context (and within an implicit or explicit cost-effectiveness league table). Alternatively, a willingness-to-pay threshold of the cost per marginal unit of effect a decision-maker is prepared to pay can be used.

The ENN was unable to find any published information (for example, guidelines) on criteria for cost-effectiveness in emergencies. In a personal communication (email), CIDA Programme Against Hunger, Malnutrition and Disease uses 350 Canadian $ (equivalent to US $252) per death averted as an informal yardstick, although this does not apply to the use of food aid. The World Bank (WB) classifies intervention which have a cost-effectiveness per DALY averted < $50 as ‘very attractive’ and < $150 as ‘attractive’ (World Bank, 1993), but these cut-offs are, of course, arbitrary.

In this review we have tried to assess all the different types of published economic evaluations of the six specified nutrition-related interventions in emergency situations.

2.0 Methods

2.1 Methods to review the published literature

The methods used to collate and appraise the published literature for this systematic review are based on standard systematic review methodology developed by the National Health Service Centre for Reviews and Dissemination (NHS CRD, 2001)[4]. This method involves four stages: database search, secondary reference search (references obtained from papers found at the first stage), hand search of the main journals, and expert advice on the literature.

For this review, we followed the search strategy used by Wyness (2003) who undertook a review of the published literature available for the impact of SFPs and TFPs in complex emergencies as part of her MSc. Wyness (2003) identified a list of keywords, determined which bibliographic databases were most appropriate for the search, and developed a method of categorising appraised papers into poor medium and good categories. The authors of the current work are indebted to Wyness for allowing us to use and adapt her methods.

The details of the search strategy are outlined below.

2.1.1 Search strategy for impact studies

A list of keywords relating to complex emergencies, malnutrition, food distribution, measles vaccination, insecticide treated bednets (ITNs) and vitamin A supplementation were generated by scanning available papers and background material related to severe malnutrition in complex emergencies, as well as consultation with experts in the field. The list of keywords compiled is presented in Section 1 of Appendix III. Note that because the reporting of impact of the feeding programmes is not well standardised, choosing the terms to use for the searches was not straightforward. It is likely that we have missed some relevant terms and, hence, may have missed some papers.

A search strategy was initially developed for MEDLINE. This was based on a thesaurus search combined with a MeSH term search, for the majority of keywords and information from experts. This created a sensitive search strategy, which was considered necessary due to the anticipated paucity of published, peer-reviewed literature. The search strategy was piloted on several databases and the references were scanned to identify the proportion of relevant articles that the search strategy identified. After further consultation with systematic review specialists at the University of Aberdeen, minor amendments were made to the search strategy.

2.1.2 Electronic database searching

Due to the relatively short timeframe of the review, only the five electronic databases listed below were searched. Secondary searching of the references identified was also carried out. A longer time frame would have allowed greater hand searching and expert opinion to be gathered, which are important parts of systematic reviews. Each search strategy was developed according to the tools available within each bibliographic database.

- MEDLINE (1980 – July 2004)
  National Library of Medicine, USA: the electronic version of *Index Medicus* accessed by using the search software Ovid.

[4] The CRD was set up in 1994 to provide the NHS with information on the effectiveness of health care treatments and services.

- EMBASE (1980 – July 2004)
  Elsevier Science Publishers BV, The Netherlands: the electronic version of *Excerpta Medica* accessed by using the search software Ovid.

- CINAHL (1980 – July 2004)
  CINAHL Information Systems, USA: Cumulative index of the Nursing and Allied Health Literature accessed by using the search software Ovid.

- The Cochrane Library (1980-Feb 2003)
  The Cochrane Collaboration, UK: accessed by using Update Software, 2003; Issue 1, (Internet version)

- CAB Nutrition Abstracts & Reviews (1973 – 2003)
  Commonwealth Agricultural Bureau International Publishing, UK, accessed using the search software Ovid.

A very large number of articles were initially picked up during the database search for all the different intervention types, however very few of the studies were eventually included in the review. For example, 1,504 articles were picked up for vitamin A, but only 11 were eventually included. This was probably because we employed such a large number of search terms (see annex 3.1) in order to make the search as sensitive as possible.

2.1.3 Search strategy for cost effectiveness studies

Papers were identified for review using the following databases for the years 1980-2004: PubMed (National Library of Medicine), HEED (Health Economists Evaluations Database) and PopLine (a reproductive health database). We used economic evaluation terms in combination with the terms used for the interventions (see annex 3.1). This approach was supplemented by iterative reviews of reference lists attached to papers. We also contacted selected experts within this field including authors of the more recent papers identified for inclusion.

2.1.4 Hand-searching for studies

The *Field* Exchange publication, produced by ENN, is not strictly part of the 'published' literature because it does not have a peer-review process. However, Field Exchange probably has more information on nutrition related interventions in emergencies than any other source. Hence, we hand-searched all the 21 back issues of Field Exchange for this review. We also hand-searched issues of the Disasters journal back until 1980 (Disasters is peer-reviewed), because we expected that it also would include a large number of relevant studies.

Even though a very wide search strategy was employed to pick up as many articles as possible in the database, a surprisingly high proportion of the feeding programme studies were only picked up during the hand-searching of the journals.

2.1.5 Discussion with selected experts

Discussion with selected experts also provided a relatively high proportion of the studies on feeding programmes. This has implications for the replication of this review (see section 2.5).

2.2 Study inclusion criteria

The search strategy was divided into distinct study areas:

1. Feeding programme (food distribution, SFP and TFP) literature
2. Measles immunisation, vitamin A supplementation and bednet efficacy literature
3. Cost-effectiveness literature for all the programmes described above.

Because these study areas are so diverse, unique inclusion criteria had to be used for each one in order that the key articles in each area could be identified. Having different inclusion criteria for the different study areas does not create any methodological weakness for our review, because no comparison was made between areas. Table 2.1 describes the inclusion criteria used for each search strategy.

Table 2.1 Inclusion criteria used for each study area of the literature review

Inclusion criteria	**GFD**	**TFP**	**SFP**	**Measles**	**Vitamin A**	**Bednets**	**Cost-effectiveness**
English language	√	√	√	√	√	√	√
Published from 1980- July 2004	√	√	√	√	√	√	√
Studies on Children	√	√	√	√	√	√	**X**
Complex emergencies	√	√	√	**X**	**X**	**X**	√
Studies with control groups	**X**	**X**	**X**	√	√	√	**X**
Studies which measure impact on nutritional status and/or mortality	√	√	√	√	√	√	**X**

Inclusion criteria for studies looking at the impact of the feeding programmes were studies assessing the growth or mortality of children aged 6-59 months in a complex emergency setting, published in English since 1980. Studies of all types were included in this section of the review (i.e. not just studies with control groups) because we knew that very few high-level studies have been undertaken. Tables in annex 4 A4.1-4.3 summarise the feeding programme studies reviewed.

The inclusion criteria for studies looking at the impact of bednets were studies assessing the growth of children aged 0-59 months, which had control groups and were published since 1980. Table A4.4 in annex 4 summarises the three studies reviewed in this area. The inclusion criteria for studies looking at the impact of vitamin A supplementation were studies assessing the growth of children aged 0-59 months, which had control groups and were published since 1980. Table A4.5 in annex 4 summarises the eleven studies reviewed in this area. The inclusion criteria for studies looking at the impact of vitamin A supplementation were studies assessing the growth of children aged 0-10 years, which had control groups and were

published since 1980. Table A4.6 in annex 4 summarises the three studies reviewed in this area.

Studies with control groups were identified for the bednets, measles and vitamin A programmes because they provide a higher level of evidence than studies without control groups. Initially the inclusion criteria for these three topics focused on complex emergencies, however there were insufficient studies assessing children's growth following measles immunisation, vitamin A supplementation or bednet utilisation to carry out the review. As a result, studies with control groups that were carried out in non-emergency settings were identified and reviewed. The generalisability of studies carried out in non-emergency settings to complex emergencies is dealt with in sections 3 and 4 of this report. Seminal articles on the association between measles immunisation, vitamin A supplementation, bednet utilisation and reduced child mortality were also identified to give a background to the review.

The inclusion criteria for the cost effectiveness literature were English language papers and cost or economic evaluations. Some articles were gathered in the search process, but were excluded where the reference to cost-effectiveness was qualitative rather than quantitative, e.g. when an author states, "the intervention is cost-effective", but there was no evidence to support the statement. A summary of each study identified is provided in table A4.7.

2.3 Critical appraisal of included studies

Each of the studies was defined according to its level of evidence (as described in section 1.3.2). We recognised the following types of studies: meta-analyses of RCTs, RCT, quasi-experimental studies, cohort study, case control study, observational study without control group (including case series and cross sectional studies).

The quality of each individual study was also assessed. Critical appraisal forms for each type of study design adapted from standard validated forms from the NHS CRD or the Critical Appraisal Skills Programme (CASP) (NHS CRD, 1996; Public Health Resource Unit, 2003) (see annex 3.2). In these separate critical appraisal forms, studies were rated as being poor, medium or good quality. The scoring itself was a simple summation of binary allocations of 0/1 for each question. Scores were subsequently classified according to whether the paper fulfilled <33% (poor), 34-67% (medium) and 68% (high).

Table 2.2 Quality assessment score developed from the CASP score

Quality of the individual study	**Percentage of criteria scored as medium or good**
Good	>=67%
Medium	>=33 and 67<%
Poor	<33%

2.4 Methodological limitations

We have tried to conduct a robust search of the published literature for evidence of impact and cost-effectiveness of six key types of humanitarian interventions in a relatively short period of time. The limited time available meant that we were not able to obtain all the references which looked like they might have relevant studies, thoroughly search all the references of the studies reviewed, and find out about the existence of other studies from all the experts in this field. This may mean that we have missed some relevant studies.

We cannot be certain that the review presented here is completely replicable. This is because a relatively high proportion of the studies were identified during discussions with experts in the field. This is particularly true of the studies assessing the impact of the feeding programmes. Clearly this method may not produce the same results twice because different experts may be consulted and even the same experts may give different answers.

3.0 Search findings

This section of the review summarises the results of the published literature search for each of the six interventions. For each intervention, we first review the strength of the evidence available, i.e.the evaluation methodologies and the studies' CASP scores, and then summarise the results of the studies. The section concludes with a review of the published economic evaluation studies for the six interventions. A summary of all the studies reviewed in this section can be found in annex 4.

3.1 What published impact evaluations are available for general feeding programmes?

There are only a very limited number of impact assessments of general ration distributions in emergencies in the published literature. Much of what has been written about the effectiveness of food aid in terms of food security is based on ex-post research and explores whether the food aid has reached its intended beneficiaries (Barret, 2002). These evaluations can take place at several different levels and can assess whether or not the food aid reachedthe most needy country, the most needy region within the country, and the most needy beneficiaries within the region. Assessments of this type tell us about the coverage of the GFD. This information allows us to calculate how much food was received by what percentage of households in an area. However, it does not give any indication of the effect of the food on the household's food security situation, or its impact on the nutritional status or mortality rates of the household's members.

3.1.1 Strength of evidence for the GFD studies

Table A4.1 (annex 4) presents a summary of the published papers which have assessed the impact of general food distributions on beneficiaries' nutritional status or mortality rates. The first point to note about table A4.1 is how short it is – the search only found nine published studies assessing the impact of GFD programmes. Some general points to note about the studies are:

- Only two of the studies presented are cohort studies (Quisumbing 2003; Yamano et al, in press). One of these (Yamano et al, in press) is actually an ecological cohort study which is weaker than a classic cohort study.
- The remainder of the studies are all observational. Three of the observational studies assessed the relationship between ration size/adequacy and nutrition / mortality rates:
    - Two studies assessed the change in population's nutritional status over time as ration size changed (Toole et al, 1988; Toole and Bhatia, 1992).
    - One study looked at the association between ration size and mortality rates in a number of different camps (ACC/SCN, 1994).
- Two studies assessed micro-nutrient deficiencies in ration-dependent populations:
    - One study assessed the association between anaemia and adequacy of the ration (Kemmer, 2004).
    - One study assessed the prevalence of vitamin A deficiency in a ration-dependent population (Wolde-Gebriel et al, 1993).
- A further two observational studies assessed the change in nutritional status at the beginning and end of a GFD programme (Sadler, 2001), while the final observational study assessed the relationship between length since last ration and child's nutritional status (Warrack-Goldman et al, 1985).
- Both of the cohort studies were rated as having a good CASP score.
- All the observational studies were rated as either having a good/medium CASP score.

The observational studies presented by Sadler and Warrack-Goldman are interesting but do not fit the adequacy criteria presented by Habicht et al. In both studies, the general food distribution was accompanied by other intervention. Also, both studies were conducted in rural, free-living populations, which had access to other ways to generate income or produce food. It is, therefore, not possible to be sure that the food distribution was responsible for any change in the populations' nutritional status.

Intuitively, one would expect the most convincing evidence for impact of general ration programmes to come from research within refugee camps. A refugee or IDP camp, where there may be considerable control over resources and services within the camp, is a situation where there is likely to be greatest knowledge of all food security related factors. Hence the intuitive link between interventions like food aid deliveries, food security and nutritional status may be realistic (Clay and Stokke, 2000).

The most comprehensive database of nutritional status of emergency affected camp-based populations is currently the Nutrition Information in Crisis (NICS) of the UN Standing Committee on Nutrition (SCN), based in Geneva -formerly the Refugee Nutrition Information System (RNIS). An analysis of the reports from the RNIS between 1992 and 1994 was presented by the ACC/SCN in 1994. The two other studies of general food distributions and mortality presented in this review are those by Toole et al (1988) and Toole and Bhatia (1992).

However, the results of the RNIS analysis have to be treated with care. First, they do not allow for correlation of actual shortfalls in consumption with mortality as rations may be deliberately low because beneficiaries are partly self-sufficient, and secondly, other non-intervention factors may affect levels of wasting and the resulting mortality, e.g. market stimulations/prices, freedom of movement and income earning opportunities (Clay and Stokke, 2000), morbidity patterns and health. The Toole studies (1988, 1992) assess the change in malnutrition where refugees had only very limited access to food sources other than a GFD, however, SFPs and health programmes were implemented in conjunction with the increase in GFD in both studies.

Thus, none of the observational studies described fit the stringent plausibility criteria described by Habicht et al because in order to reach the highest end of plausibility, one must formally discard all other likely explanations for the observed improvements and this was not done. Recently, two new studies assessing the impact of food distribution programmes in Ethiopia have been published. These studies, which have employed controls, go some way to improving the dearth of impact evaluations for these types of programmes and are described in detail below.

Yamano et al (in press) used data from three national surveys conducted in Ethiopia between 1995-96 as part of the Rural Integrated Household Survey Programme to look at the impact of shocks (proportion of family plot damaged) and food-for-work (FFW) and free distribution of food (FD) on child growth. As expected, the study found that the linear growth of children aged 6-24 months was vulnerable to shocks. To assess the impact of FFW and FD, child growth was plotted over a six-month interval against child age at the first measurement, both for children in food aid receiving communities and non-food aid receiving communities. The regression controlled for individual child, household and community characteristics[5].

[5] Individual child characteristics controlled for included initial height, gender, age. Household characteristics included: mother's age, educational information on household members, gender of the household head, composition of the household, household assets (land, plough, animals, radio), and the source of drinking water. Community characteristics included community level expenditure (net of food aid) as a proxy to control for chronic poverty/income at the community level, dummy variables

Quisumbing also evaluated the impact of FFW and FD distributions on the change in anthropometric status in children aged 0-59 months in Ethiopia (Quisumbing, 2003). The assessment used data from all four rounds of the Ethiopian Rural Household Survey (ERHS). The first three data rounds were collected in 1994/95 and the final round in 1997. The ERHS covered approximately 1,500 households in 15 villages across Ethiopia. This assessment compared the change in anthropometry in individual children whose families did or did not take part in FFW or receive FD. Again, the analysis controlled for individual child, household and community characteristics[6] and two separate analyses were presented – one for high-asset households and another for low-asset households.

The Yamano and Quisumbing studies go some way to providing evidence that food aid does have an impact on children's nutritional status, in part because they use longitudinal data. However, even these studies have some weaknesses:

- The Yamano et al study is an ecological study - the study assessed the differences in growth between children living in an area which received food aid and children living in an area which did not receive food aid. So the intervention was measured at the community level but the impact was measured at the individual level. This reduces the strength of the study because it is not possible to account for variation between individuals in receiving the food aid in the intervention area.
- The Quisumbing study compares children whose families take part in FFW or receive FD to children whose households do not live in the same village. It is not clear how the author controls for some of the differences between these two groups. It is likely that the households taking part in the programme have different characteristics to those which do not, even within the same wealth group. Indeed, the analysis of the data shows that household size, for example, is significantly different for households which receive FD compared to those that do not. Some of these differences may also be associated with nutritional status and hence the analysis may be biased.
- The Yamano et al study compares children whose communities take part in FFW or receive FD to children living in communities that do not. Again, it is likely that these communities will have different characteristics. Although the authors have attempted to control for many exogenous variables, including programme placement, it is not possible to control for all of them (for example, some communities may have more political power than others and, hence, be more likely to receive food aid and health care interventions). One would need to know more about why one community was chosen for food aid distribution and not another.

Some of the problems of both these studies arise from the fact that they were not planned at the beginning of the interventions. In fact, the studies were not planned to evaluate the interventions at all but the authors realised that the datasets could provide valuable information on the interventions later. This is one of the main problems in trying to evaluate the impact of emergency nutrition interventions. The result is that information on the impact of general food distributions is not very strong and probably does not meet the most stringent conditions of plausibility.

3.1.2 Summary of the results of the GFD studies

---

for peri-urban areas, availability of tarmac road in the zone as well as elevation (a proxy for malaria infestation).

[6] Individual child level characteristics included: gender, age, sex. Household level characteristics included: maternal height, asset holdings, net expenditure. Community level characteristics: rainfall, livestock, disease, shocks.

It is not possible to generalise from the studies described above, as they are all assessing the impact of a GFD on populations with different levels of malnutrition and different levels of access to other food sources. Moreover, the GFDs were conducted differently in each of the populations (in terms of targeting mechanisms, ration size and quality, etc) against a background of a variety of other interventions (SFP present/not present, varying quality of health services etc). All of these variables make it virtually impossible to produce a summary of the results. Thus we will only reflect very broadly on the findings of each study.

- Impact on population level mortality
    - Toole et al's studies of camps in Eastern Sudan, Thailand and Ethiopia suggest that population rates of mortality only decreased when the general ration size was adequate (Toole et al, 1988; Toole and Bhatia, 1992).
    - An analysis of the reports received by the RNIS from 1992-1994 showed an association between mortality rates and ration size. Benchmark CMR levels below 1/10,000/day were associated with rations providing over 2,000 kcals per person per day, while less than 1,500 kcals per day were associated with mortality rates ranging from 2 –10 times higher (ACC/SCN, 1994).
- Impact on population's nutritional status
    - The Yamano study concluded that children aged 6-24 months in communities which were receiving food aid grew, on average, 1.6 cm faster than if no food aid had been available. The impact of FFW was larger than the impact of FD, perhaps because FFW programmes are self-targeted. The results for children aged 25-60 months were similar, but with smaller margins. The authors did not discuss the impact of the food aid on weight-for-height measurements (acute malnutrition).
    - Toole et al's studies of camps in Eastern Sudan, Thailand and Ethiopia suggest that population rates of malnutrition only decreased when the general ration size was adequate (Toole et al, 1988; Toole and Bhatia, 1992).
    - Wolde-Gebriel reported a very high prevalence of vitamin A deficiency in a free-living population, which had depended on a ration to compensate for poor food production for 6 years (Wolde-Gebriel, 1993).
- Impact on the beneficiaries' nutritional status
    - The study by Warrack-Goldman et al (1985) did not find an association between individuals' nutritional status and the length of time since the household last received a ration[7].
    - Quisumbing concluded that both FFW and FD significantly improve the weight-for-height z-scores of children aged 0-5 years in children from either wealth group, however no absolute figure (for example amount of change in z-scores) was given. The programme had no impact on children's height-for-age.
    - Kemmer et al found that children living in households who reported that their ration would run out before the next distribution were more likely to be anaemic than children living in households where the ration would last until the next distribution (Kemmer et al, 2004).

All the studies, except that reported by Warrack-Goldman et al report a positive association between the distribution of food aid and either decreased mortality rates, or decreased rates of malnutrition. Although these results provide some evidence of the positive impact of GFD programmes, important information gaps for at least two other nutritional impacts of GFD programmes remain - impact on micro-nutrient status and impact on people aged more than 5 years old.

[7] No information on ration size or other sources of food is presented in the study. It is difficult to measure the impact of a food distribution programme based only on information from households about how long ago the last distribution was.

Only two published studies reported on the impact on micronutrient status of GFD programmes (Kemmer et al, 2004; Wolde-Gebriel, 1993). Both of these studies found very high rates of micronutrient deficiencies in ration-dependent populations. The authors attributed the deficiencies to the poor quality of the rations. These results imply that the GFD were not effective in preventing micro-nutrient deficiencies in these populations.

This literature review has shown that impact of the general ration on micro-nutrient deficiencies is also rarely considered in emergencies. This is mainly due to (i) the lack of field friendly method to assess micro-nutrient status (Seal, 1998), (ii) the relative rarity of clinical micronutrient problems in emergencies, especially since the introduction of fortified blended foods (e.g. corn soya blend (CSB)) into general rations in the early 1990s, (iii) the fact that that low levels of wasting can mask poor micronutrient status (Assefa, 2001), and (iv) some forms of deficiency disease, e.g. pellagra and scurvy, appear to affect older age groups more predominantly and would, therefore, be missed in a standard nutritional survey which measures and weighs children under five years of age (Duce et al, 2003).

GFD programmes are meant to improve the nutritional status of the whole population, but this search of the published evaluations has only focused on studies measuring impact among children under five years of age. Nutritional surveys traditionally measure among this age group. However, there are a growing number of reported cases where the most nutritionally vulnerable are another demographic group. It has been observed amongst populations affected by food crisis that food consumption of children may be protected by adults forgoing meals. As a consequence, the nutritional status of children may be the last to suffer. Also, in some contexts other groups may be more nutritionally vulnerable. For example, in Bosnia, a number of nutrition and health surveys were conducted on the 'at-risk' populations of the besieged enclaves in 1992 and early 1993. These surveys collected anthropometric data on under-fives and found no signs of acute malnutrition. However, surveillance systems set up to collect data in the same enclaves at the end of 1993 collected nutritional, health and socio-economic data on all household members and found that while the nutritional status of children remained normal, the elderly (over 60 years of age) showed elevated signs of wasting while adults experienced substantial weight loss. These findings imply that adults and the elderly should be included in studies looking at the impact of general rations in certain situations (Watson, 1995). The HIV epidemic in Southern Africa is another example of when children may not be the most vulnerable group and there is a need to measure the impact of GFD programmes on other age groups.

3.2 What published impact evaluations are available for SFPs?

Table A4.2 presents a summary of the results of the 15 studies found in the literature search involving both targeted and blanket SFPs in emergencies. The information for each study has been separated into three different levels:

- Impact of the programme on children enrolled
- Impact of the programme on the population's nutritional status
- Coverage of the programme.

This information has been separated out because although it is useful to know that an SFP has had an effect on individual children enrolled, the programme will not have had a population level impact unless coverage is relatively high. In fact, the information on impact on individual enrolled children may more properly be described as outcome data. The coverage data provides information on the process of the programme. The data about change in a population's nutritional status provides us with an overview of the impact of the SFP.

3.2.1 Strength of evidence of the SFP studies

The first point to note about table A4.2 is just how short it is, given the very large number of SFPs being undertaken in emergency affected populations throughout the world (this is discussed further in section 5.1). By examining table A4.2, several conclusions about the strength of the published evidence for the impact of SFPs can be drawn.

- The search produced only one cohort study assessing the impact of an SFP in an emergency (Stefanak and Jajoura, 1989). All of the other studies are observational. This means that the evidence base for the effectiveness of these programmes is weak.
- All of the evidence for the impact of the programme on individual children enrolled comes from case-series data, where children are measured repeatedly throughout the programme to assess the programmes' success. Only two of these studies are rated as 'good'. All the others are rated as either of medium or poor strength by the CASP scoring system, which again implies a weak evidence base. In addition, it is likely that the reliability of the measurements is low in many cases, as many different project staff have to deal with measuring large numbers of children at the SFP every week/fortnight/month – this will probably result in relatively large intra- and inter-observer measurement error.
- The quality of the information on programme coverage is also weak. Of the eight surveys which present coverage information, three used cross-sectional survey techniques (asking children if they were registered on the programme during the survey) and two compared the figures of children enrolled in the SFP to expected figures of malnutrition calculated during an earlier nutrition survey. A further three assessments reported coverage without explaining the data collection technique. The majority of the coverage surveys were scored as 'poor' by the CASP scoring method[8]. In fact, measuring the coverage of feeding programmes is not straightforward for several reasons. this is discussed in detail in Annex 5. One point to note is that if children are registered in a SFP more than once (due to cheating or re-admission after becoming malnourished again), then it is likely that coverage appears higher than it actually was. This is likely to be the case in some of the longer duration studies for example, the studies by Vautier (1999) and Vasquez-Garcia (1999).
- Only six of the studies presented information on the impact of the programme on the populations' nutritional status. Four of these studies involved using repeated cross-sectional surveys and the others used camp monitoring data. Again, the majority of these studies were ranked as poor quality by the CASP system.

Six of the studies provide information on both coverage and the impact of the SFP on enrolled children. These studies provide adequacy level evidence that the SFP is working. In other words, it is possible to say that x% of the children were enrolled and that of this x%, y% recovered. However, whether or not the recovery was due to the SFP or some other factor is impossible to tell from the data presented above. The studies using population level nutrition data as evidence of overall impact of the programme suffer from the drawbacks described above under general food distributions.

None of the studies fit the plausibility criteria described by Habicht et al because there is no way of knowing whether or not the change seen was due to the SFP or something else. By

[8] One problem with some of the coverage study reports (particularly those from the ENN) is that very limited description of the coverage surveys was given. This results in many of these surveys being classified as 'poor' quality. This is probably for two reasons. Firstly, the coverage survey is not normally the focal point of the article but is included in it for completeness and hence detailed information is not provided. Secondly, the authors probably assume that Field Exchange readers know that an MSF nutritional survey usually consists of a standard two-stage cluster survey assessing nutritional status of children aged 6-59 months. However, unless all this information is given in the paper, the study must be scored as 'poor' according to the CASP scoring system.

definition, an SFP should be run in conjunction with other services – it is added on to the general ration, or serves as a supplement to households' own food and income sources – and hence to assess its impact without a control group is very difficult.

The conclusion to draw from these observations is that the evidence base for the effectiveness of SFPs is weak.

3.2.2 Summary of the results of the SFP studies

It is extremely difficult to generalise from the results of the studies above for three reasons: (i) there are too few of them, (ii) they have been undertaken in different settings – some in refugee camps, others in rural areas, some where an adequate GFD programme is in place, others where an inadequate GFD programme is in place, some where infrastructure is poor, others where infrastructure is better, etc, and (iii) the interventions themselves differ in terms of admission/exit criteria, protocols, etc. Given that there are such a small number of studies and that these are so variable, over-analysis of their results is probably ill advised hence the analysis presented below:

- According to the Sphere standards, 75% of children who exit from an SFP should have 'recovered':
    - From the six studies which presented their results in such a way as to be able to see if they had met the criterion, 50% reported that this indicator of programme success had been achieved.
    - The majority (75%) of the remaining studies (n=4) assessing weight gain in enrolled children showed that most children did gain weight when enrolled in an SFP.
- According to Sphere guidelines, coverage rates of SFP programmes should be more than 50% in rural areas, more than 70% in urban areas and more than 90% in camp settings:
    - Of the six studies that reported programme coverage, 86% fulfilled the Sphere criteria.
    - It was noted that high coverage of an SFP is generally more easily obtained in a refugee camp setting than in a rural area.

It is noteworthy that although many authors present 'positive' results in their assessment of an SFP, several of them question the usefulness of such a programme after the immediate emergency has passed. For example, Vasquez-Garcia (1999) and Taylor (1983) both suggest that an enlarged general ration or poverty reduction programmes might be more useful mechanisms of helping the population than SFPs. Roesel (1988) demonstrates that wet feeding is inappropriate even in refugee camp settings. Unfortunately, without data on the costs of these programmes it is difficult to judge whether or not these ideas are correct, but it could easily be argued that providing an income transfer through an SFP is not going to be very cost-effective (see below).

The study by Descenclos et al (1989) deserves a special mention because it was assessing the prevalence of scurvy rather than weight-for-height (WFH) types of malnutrition. The results of this study show that the prevalence of scurvy among children enrolled in the SFP in Somali refugee camps ranged from 16-28%. Given that this is only one study, it is, again, impossible to extrapolate the findings to other SFPs, however it is clear that the SFP and the general ration provided in these camps failed to achieve their objective of either preventing or reducing malnutrition.

Some SFPs have multiple objectives which will include preventing/reducing malnutrition and mortality, nutrition education, income transfer and 'getting food out' to beneficiaries in insecure areas (annex 1). These objectives have not been assessed in this review of the impact

of SFPs and it is possible that more positive results may have been recorded had these types of objective been examined.

A further limitation of this review has been that, except for one study, the focus of the review has been on change in WFH among children aged 6-59 months. This means several potential impacts may have been missed:

- Impact on nutritional status other than improvement in WFH. There is some evidence from SFPs implemented in non-emergency conditions that, depending on their initial nutritional status, some children will be more likely to gain in weight by growing taller (i.e. gain height) and others more likely to gain weight but will not increase their height (Rao and Joshi, 1992). If this is also true in emergency situations, then it is possible that some of the positive impact of the programme has not been recorded.
- Nutritional impact of interventions will not only occur through growth/weight gain, but will also be expressed in terms of improving/increasing tissue integrity/immunity/physical activity levels. These impacts will not be measured in traditional nutritional surveys. A seminal study by Beaton and Ghassemi (1982) examined the effectiveness (and impact) of a large number of supplementary feeding programmes in stable situations. The authors concluded that there was no demonstrable impact on nutritional status using anthropometric measures. While this finding was partly explained by the poor operational performance of some of the SFPs, another aspect was the positive impact these programmes may have had on variables such as physical activity levels and immunocompetence, i.e. there may have been a significant nutritional impact which was not measured in the study. This phenomenon is rarely considered in emergency contexts.
- We have not reviewed the evidence of impact of SFPs on pregnant or lactating womens' nutritional status.

3.3. What published impact evaluations are available for therapeutic feeding programmes?

Table A4.3 presents a summary of the results of the TFP impact studies found in the literature search[9]. Where possible, the information for each study has been separated into two levels:

- Impact of the programme on children enrolled
- Coverage of the programme.

None of the studies assessed the impact of the programme on the population's nutritional status, so this information is not given.

3.3.1 Strength of evidence of the TFP studies

It is possible to draw several conclusions about the strength of the published evidence for the impact of TFPs by looking at table A4.3:

- Only 16 published studies looking at the impact of TFPs in emergencies were found during the search process. The bulk of the studies described in the table (10/16) were published by the ENN as a record of a meeting which discussed advances in home-based TFPs[10].

[9] Note that new material about the impact of TFPs, particularly home-based TFPs, is being received by the ENN on a fairly regular basis at the moment. The information presented here is what is available at the end of July 2004.

[10] Unfortunately CASP scores could not be calculated for the ENN reports as the original studies were not available to the author of this report.

- Eleven of the sixteen studies were observational and relied on programme records. Of the remaining five studies, two were randomised trials assessing the efficacy of different type of phase II treatment protocols, one was a trial assessing the outcome of children with kwashiorkor undergoing different treatment routines, one was a retrospective cohort study assessing different protocols for home based therapeutic feeding, and one was a semi-randomised cohort study assessing the risk of mortality of children who were or were not treated for severe malnutrition.
- The impact of the programme on individual children enrolled was measured in all 16 studies. CASP scores were only available for six of these studies. The quality of the six studies was satisfactory - all of the studies were scored as either 'medium' or 'good'.
- Programme coverage was estimated in eight of the 16 studies. The methods for measuring coverage varied significantly but all of the studies used observational techniques. The most commonly used methods to obtain coverage data was to combine data from programme attendance figures with expected numbers of severely malnourished children based on earlier nutrition surveys (three studies). Two surveys estimated coverage using the classic approach described by MSF (1995)[11]. Two of the studies used the active-case finding method and a further study based coverage information on screening records. Unfortunately CASP scores were unavailable for all but one of the coverage surveys.

Overall, the published evidence for the impact of TFPs is stronger than that for either SFPs or food distributions. One study (Perra and Costello, 1995) actually fits the plausibility criteria described by Habicht because it compared mortality and weight gain in groups of severely malnourished children which had and had not received treatment for severe malnutrition in rehabilitation centres. The children were randomised to a control/treatment group naturally (because there were insufficient spaces available in the treatment centres) and hence, there was a 'real control'.

Three of the other non-observational were all prospective trials comparing the outcome of individual children on different treatment regimes. These studies clearly provide some evidence of the impact of different regimes but tell us nothing about the overall efficacy of an intervention (compared to no intervention) because it is not known what the children's mortality rates would have been without the programmes. The study by Collins and Sadler (2002) compared outcomes of the same treatment for children with different types of severe malnutrition, but again no information is available on the outcomes of children who were not treated at all.

Eight of the studies presented provide information on both coverage and the impact of the TFP on beneficiary children. These studies provide adequacy level evidence that the TFP is working. In other words, it is possible to say that x% of the children were enrolled and that of this x%, y% recovered. Again, without control data, it is not possible to know how many of the children would have recovered without the programme.

Nine of the eleven observational studies presented were assessing the outcome of home-based TFPs. The evidence from these types of studies is somewhat weaker than those presented from centre-based programmes because it is possible (although unlikely) that some of the improvement seen in the beneficiaries was due to non-programme based interventions. For example, it is possible that mothers gave their children other food in addition to the programmes' food to speed up their recovery.

In conclusion, the published evidence for the impact of TFPs is stronger than that available for either food distributions or SFPs.

[11] In this method, the agency conducts a standard cross-sectional anthropometric survey of children aged 6-59 months. During the survey, the teams assess whether or not the severely malnourished children are registered on the TFP. See annex 5 for more details on how to measure coverage.

3.3.2 Summary of the results of the TFP studies

Although there is more impact evidence from TFP studies, it is still difficult to generalise about the results. The reasons for this are similar to those for SFPs: (i) there are too few studies, (ii) the studies have been undertaken on very different groups of children – with or without oedema, with or without having completed phase I treatment, with or without a high level of HIV infection, etc, and (iii) the interventions themselves differ tremendously in terms of admission/exit criteria, protocols and setting (in hospitals of various levels, beneficiaries' homes in rural areas, etc). All of these variations make it very difficult, and probably inadvisable, to generalise about the impact of TFPs in emergencies – note that this observation was also made by the editors of the ENN conference report (ENN, 2003a).

- The Sphere guidelines indicate that more than 75% of children exiting from a TFP should have 'recovered'. Of the 14 studies that measure impact of their programme in this way, only eight (57%) meet this criterion.
- The Sphere guidelines also indicate that coverage for a TFP should be more than 50% in a rural area. Of the eight studies that measured the impact of their programme, five (63%) fulfilled this criterion.

Some readers may be surprised that only 57% of the published studies on TFPs have achieved the Sphere standards for recovery rates. The relatively low figure is probably because many (10/16) of the studies reviewed are reporting on the relatively new home-based treatment protocols. Although home based treatment programmes can achieve high recovery levels (see for example, Collins and Sadler, 2002; SC-UK/Valid for ENN, 2003a, many of the programmes reviewed were undertaken in very difficult circumstances (for example, in Afghanistan) or by agencies and/or staff which were relatively new to the home based treatment regime (for example, in South Sudan) (ENN, 2003a). Higher rates of recovery were generally reported for children based in therapeutic centres, for example, Prudhon et al (1997). There are at least three important points to note about these findings:

- Historically, improvements in the recovery rates of therapeutic feeding in centres have taken some time to improve. Often the improvement in rates of recovery is due to improved staff training - for example, SC UK noted a significant improvement in mortality rates in TFPs in both Kinshasa, Democratic Republic of Congo (DRC) and Bangladesh after staff received training from a well-known physician (SC UK personal communication). It is probable that as agencies gain more experience in home-based treatment, the recovery rates will start to increase.
- Three of the studies of home-based care were carried out in Malawi, which is known to have very high HIV rates. Children living with HIV/AIDs have a poorer prospect of recovery from severe malnutrition than those without.
- The coverage of centre-based TFPs is usually much lower than that of home based TFPs. This point is critical. If a centre-based TFP is only curing a small proportion of the severely malnourished children, its overall impact will be limited compared to a home based TFP, which may cure a large proportion of children with higher coverage.

The third point, about coverage, is absolutely critical. When assessing the impact of TFPs in the future, measures of coverage should always be included. There are, however, difficulties in measuring coverage and a standard technique has not yet been agreed (see annex 5).

Given the huge amount of variation in the differences in the environment in which the TFPs are undertaken, it is difficult to draw out robust conclusions about which methods work best under what circumstances from the evidence presented. As reported in section 5, many NGOs regularly implement TFPs without publishing the results of their programmes.

Prudhon et al (1996, 1997) have developed a tool to assist programme managers in predicting the expected number of deaths during treatment of severe malnutrition in centre-based TFPs based on the beneficiaries' admission weights and heights. The idea behind this tool.the Prudhon index', is that a child's chances of recovery will inevitably depend on the degree of his/her malnutrition, thus in order to assess how one treatment centre is performing compared to another it is necessary to control for the severity of malnutrition at admission. This tool could be usefully employed to compare the results of different TFPs in the future (or during a grey literature review). The index, however, will not solve all the problems of comparability because factors such as underlying disease (e.g. HIV) will still vary from one TFP to another.

3.4 What published impact evaluations are available for bednet programmes?

In 2004, the Cochrane group carried out a systematic literature review on the efficacy of insecticide-treated nets (ITNs) in preventing mortality and morbidity in children[12]. They reviewed five trials carried out in sub-Saharan Africa (Burkina Faso, Gambia, Ghana and two in Kenya) in non-emergency settings. The report found that ITNs are highly effective in reducing childhood mortality and morbidity[13] (Lengler, 2004).

Given the results of the Cochrane review, this report has not reviewed any studies of the efficacy of ITNs in preventing mortality or morbidity. However, a search assessing the impact of ITNs on malnutrition was conducted. The initial searches found no studies of bednet programmes conducted in emergencies[14] so the search was widened to include non-emergency settings. Only studies with a control group were included in the search (see section 2).

During the search, three relevant references were identified. Table A4.4 shows the summary results of the studies.

3.4.1 Strength of the evidence of the bednet programme studies

From table A4.4 it can be seen that:

- All three of the studies reviewed were RCTs.
- Two of the three studies were scored as medium using the CASP system and the third was scored as poor.

We chose to review only trials for the bednets section of this literature review in view of the likelihood that there would be too many published studies without controls to review in a short timeframe. The fact that the reviewed studies have a control makes their results more plausible than otherwise. Because the results of the poorly ranked study agree with those from the other studies, we are inclined to give this study more weight than might normally be the case[15].

[12] The age of the children was not specified.
[13] ITNs provide a 17% protective efficacy (PE) compared to no nets (relative rate 0.83, 95% confidence interval (CI) 0.76 to 0.90) and 23% PE compared to untreated nets (relative rate 0.77, 95% CI 0.63 to 0.95).
[14] Note that one cost-effectiveness study for the use of bednets in emergencies was found during the search (see section 3.7).
[15] However, we feel that the authors of the report from Kilifi (Snow, 1997) may have over-stated the significance of their findings as although the overall impact of the programme was significant, no change in nutritional status between control and treatment group was seen for seven of 11 age groups.

The three published studies were undertaken in rural populations of Kenya and The Gambia. This reduces the strength of the evidence base because it may not be appropriate to generalise the results outside of these countries. The generalisability of the studies will depend on whether the malaria transmission patterns are the same in other places. For example, in most places, malaria is transmitted mainly at night which is why bednets are useful, but if there was an area where a higher proportion of malaria was transmitted during the day, then it may be that the bednets are less effective.

A serious problem with the evidence base for the impact of bednets programmes on mortality or nutritional status is that the search could only locate studies which assessed these programmes in non-emergency conditions. Although it seems logical to assume that an ITN programme may work anywhere, there could be several important reasons why they may not work as effectively in emergencies. For example, displaced populations may not have access to shelter in which case it would be difficult for them to hang their nets. It may also be harder to provide the education which is needed with the nets when people are displaced. If people are moving around with their possessions, they may not prioritise taking their bednets with them, or bednets might get torn during the move, in which case agencies may have to distribute more than one bednet. Spraying of bednets may also be more difficult in a camp situation because of over-crowding. Alternatively, sanitation facilities in camps may be very different to those in 'normal' rural settings, resulting in more mosquitoes so that the ITN programme has a bigger impact.

3.4.2 Summary of the results of the bednet programme studies

This review could only find three published trials looking at the impact of ITN's on nutritional status. This means that care should be taken in making generalisations about the impact of ITNs on nutrition.

- All three of the studies reviewed found a significant positive impact of ITN's on WAZ.
- Both of the studies which measured MUAC also found that MUAC was significantly higher in the intervention group than the control group.
- Only one of the studies reported on HAZ. The study found no significant impact on HAZ.
- None of the studies reported on the impact of WHZ.

From the very limited number of studies presented it can be concluded that ITNs not only decrease malaria–related morbidity but also increase children's growth. This may be due to reductions in the probability of *p.falciparum* infection per se, or through a reduction in the occurrence of acute clinical attacks. These effects could only be disentangled through detailed prospective studies of morbidity and weight changes, which will be complicated by the ethical necessity to intervene when children are diagnosed either with malnutrition or malaria (Snow et al, 1997).

Two studies, which were excluded because they assessed the impact of ITNs on older age groups, noted that the impact of the programme on weight gain was reduced in older age groups (Leenstra et al, 2003; Friedman et al, 2003). This may be either because older children have a reduced burden of malaria and/or because they have reduced compliance and retire to bed later reducing the protection conferred by ITNs.

A weakness of these studies (for our purposes) is that none of the authors disaggregated their results by children's initial nutritional status. It would have been useful to know whether malnourished children gained more or less weight during the programme than children who were not malnourished. This weakness also applies to the Cochrane review – it is not known whether malnourished children who sleep under ITNs are likely to have an even more reduced risk of mortality than normally nourished children.

3.5 What published impact evaluations are available for vitamin A supplementation programmes?

It is widely accepted that vitamin A supplementation has a positive impact on the mortality of children aged 6 months to 5 years. A recent review of the role of vitamin A in reducing child mortality and morbidity (Ramakrishnan and Martorell, 1998) concluded that supplementation reduces mortality by 23% except in very young infants (<6 months) by reducing the severity of diarrhoea and measles and post-measles complications.

Given the results of these reviews and several well-known meta-analyses[16], this report has not reviewed any studies of the efficacy of vitamin A in preventing mortality or morbidity in children. However, we did conduct a review of the impact of vitamin A supplementation on the nutritional status of young children. As for bednets, the initial searches found no studies of vitamin A supplementation programmes conducted in emergencies[17] so the search was widened to include non-emergency settings. Only studies with a control group were included in the search (see section 2).

Eleven relevant studies were identified. A summary of these studies is given in table A4.5.

3.5.1 Strength of the evidence for the vitamin A distribution programme studies

From table A4.5 it can be seen that:

- All 11 of the studies reviewed were trials, nine of the 11 were randomised trials.
- Six of the studies were scored as medium using the CASP system and five were scored as high.

As for the bednets review, we chose to review only trials that assessed the impact of vitamin A supplementation, thus ensuring a relatively high level of evidence. Because all the studies have control groups they fit Habicht et al's plausibility criteria. All of the studies were of an acceptable quality.

None of the authors of these trials described the context as a complex emergency setting, however the situation in the South Kivu study (Donnen, 1998) was probably not dissimilar to emergencies in other rural settings.

It is not clear how transferable the results of these studies are to emergency settings. Theoretically, a vitamin A supplementation could be more or less effective in an emergency setting compared to a non-emergency. The impact of vitamin A distribution programmes

[16] There have been four independent meta-analyses assessing the impact of vitamin A on mortality. They identified average reductions in mortality of 23% (Beaton et al, 1993), 30% (Fawzi et al, 1993, Glaxiou and Mackerras 1993), and 34% (Tonascia 1993).

[17] Note that one cost-effectiveness study for the use of bednets in emergencies was found during the search (see section 3.7).

probably depends on how deficient the children are in vitamin A to begin with, and also how high the population incidence of measles and/or diarrhoea is. When a population is more deficient and the incidence of measles is higher, then it might be expected that supplementation has a more important impact. Given that in most emergencies the risk of vitamin A deficiency and measles is even higher than in a non-emergency situation (because of poor ration quality and over-crowding), this implies that the impact of a supplementation programme would be greater in an emergency context. However, in emergencies, children may be consuming less fat in their diet, which would make it harder for them to absorb the vitamin A. This may reduce the effectiveness of these programmes in emergencies.

The coverage of a vitamin A distribution programme may also change in emergencies when the usual mode of delivery of the supplement, e.g.:through the Ministry of Health (MoH), may be disrupted. Coverage will vary according to whether or not the population is in camps (easy to reach, should be a higher coverage) or scattered (harder to reach, probably lower coverage) and also on security risks incurred by attending a supplementation programme. In addition, uptake of the programme may not be prioritised in an emergency when adult members of the population are busy collecting food, water and other essential items.

3.5.2 Summary of results of the vitamin A supplementation studies

From the 11 trials reviewed here it is difficult to generalise about the findings because (i) they were conducted in different settings where access to healthcare etc varied, (ii) the children had different initial nutritional status, and (iii) interventions followed different protocols. Thus the summary below can only provide a very general overview of the results:

- Of the 10 studies assessing the impact of vitamin A supplementation alone, 50% of the studies found that some measure of growth was significantly improved in the intervention compared to the study group. Of the five studies which showed an impact:
    - One study showed an impact of supplementation on weight and MUAC (but not height) only in children who were vitamin A deficient at the baseline
    - Two studies showed an impact of vitamin A on height (Hadi, 2000; Muhilal, 1988)
    - Onestudyshowed an impact of vitamin A on weight and MUAC but not height (West, 1988)
    - Two studies showed an impact of vitamin A on height (Hadi, 2000; Muhilal, 1988)
    - One study showed an impact of vitamin A on MUAC but not weight or height (West, 1987)
- The only study which assessed the impact of vitamin A and zinc, given in combination or separately, found no impact of supplementation on growth (Rahman et al, 2002).

From the mixed results above, it is difficult to draw firm conclusions about whether or not vitamin A supplementation has an impact on the growth of children. A review by Ramakrishnan and Martorell (1998) concluded that based on the findings currently available supplementation is unlikely to improve the growth of young children who are only mildly to moderately vitamin A deficient.

However, the study by Donnen et al (1998) suggests that supplementation may have an impact on the growth of children who are more severely vitamin A deficient. Interestingly, the majority of the studies reviewed excluded children with clinical signs of vitamin A deficiency after treatment with a high dose of vitamin A. Furthermore, the effect of vitamin A supplementation was not compared among children with different serum retinol

concentrations in most of the studies. It is possible that the impact on growth would have been higher in those with lower concentrations, as found by Donnen et al (1998).

Several of the authors, who reported the results of trials which failed to show an impact of vitamin A supplementation on growth, noted that it was likely that vitamin A deficiency was probably not the only deficiency that the study children were suffering from. The researchers suggested that multiple micronutrients were probably responsible for slowing the growth of children and, therefore, that other nutrients (in addition to vitamin A) would probably need to be supplied to children in order to see a change in the rate of growth.

3.6 What published impact evaluations are available for measles vaccination programmes?

Measles vaccination has been taking place worldwide since 1964[18]. In 1989, the World Health Assembly resolved to reduce measles morbidity by 90% and measles mortality by 95%. By 1996, the incidence and death rates for measles world wide were reduced by 78% and 88% respectively (WHO, 1996). This is one of the success stories of global public health in the last 20 years.

A recent structured review by Cooper et al (2003) identified nine articles that assessed the relationship between measles vaccination and childhood mortality. Only two of these articles had a control group. The very small number of articles is probably due to the fact that given that we know how effective immunisation is in preventing mortality from measles, it would be both unnecessary and unethical to conduct further trials.

In this review we report on a search for studies which assessed the impact on measles immunisation on malnutrition. The initial search revealed that there were no studies in emergency conditions so the search was widened to include non-emergency settings. During the search, three suitable references were identified. Table A4.6 summarises the three studies.

3.6.1 Strength of the evidence for the measles immunisation studies

Each of the studies summarised in table A4.6 investigates the relationship between measles immunisation and malnutrition from a different perspective. One investigates the effectiveness of measles immunization on malnutrition related mortality, the second on the immunological response to measles vaccination, and the third the metabolic effects of acute measles in malnourished children.

- All of the studies reviewed were case control studies.
- Two of the studies were scored as medium using the CASP system and the third study was scored poor.

Because case control studies are viewed as being less robust than RCTs in the hierarchy of evidence, the evidence of impact on measles immunisation on nutritional status is less convincing than that for bednets or vitamin A supplementation.

None of the measles immunisation studies were conducted in emergency-affected populations - one study was carried out in Nigeria, one in India and the third does not specify, but is presumed to be India since it was published in the Indian Journal of Pediatrics (Kapoor, 1991). It is not easy to know how far one can generalise the results of studies in non-emergencies to those in emergencies.

[18] See weblink http://www.who.int/vaccines-diseases/history/history.shtml

In emergencies, the cold chain required to keep the vaccine effective is less likely to be in place. Because of over-crowding in emergencies, one would expect more measles, therefore the impact of programme is likely to be greater (as more to prevent). As for vitamin A, levels of coverage will depend on where the emergency-population is living and how secure the area is. Uptake of services may be less of a priority for carers who are busy collecting food, water etc in an emergency.

3.6.2 Summary of the results of the measles immunisation programme studies

- One study (Kapoor, 1991) found that immunisation with measles vaccine was associated with a significant reduction in mortality due to malnutrition.
- One study (Phillips, 2004) identified that measles immunisation was associated with significantly higher WAZ and WHZ in malnourished children. However, measles immunisation was not associated with significantly higher HAZ.
- One study (Bhaskaram, 1986) found no association between levels of malnutrition and the size of the immune response to the vaccine.

From these studies it appears that measles immunisation is associated with higher WAZ and WHZ in children. Furthermore, the immune response to the vaccination does not appear to be affected by malnutrition. Therefore, the measles vaccine should be as effective in combating measles in malnourished children as non-malnourished children. In fact, because malnourished children are more likely to die from measles than well nourished children, severely malnourished children should be prioritised to receive measles vaccinations before other children – this is common practice in emergencies for most agencies which vaccinate children enrolled in SFPs and TFPs.

3.7 What published economic-evaluations are available?

The search for cost-effectiveness studies of the six interventions found just three studies conducted in emergency conditions. These are summarised in table A4.7. No cost or cost-effectiveness studies were identified for GFD[19], vitamin A supplementation or measles immunisation programmes from an emergency setting.

Given that there are so few studies, they are described in detail below:

- Bednets evaluation

Rowland et al (1999) evaluated the cost-effectiveness of using a *chaddar* to prevent malaria transmission among Afghan women whom, in common with many Muslim peoples of Asia, wear such a veil or wrap to cover the head and upper body. The cloth also doubles as a sheet at night, when they are used by both sexes. A randomised controlled trial was undertaken, in which 10% of the families of an Afghan refugee camp (population 3,950) in north-western Pakistan had their *chaddars* and top-sheets treated with permethrin insecticide, at a dosage of 1 g/m^2 , while a further 10% had their *chaddars* treated with placebo formulation. Malaria episodes were recorded by passive case detection at the camp's health centre. From August to November, the odds of having a falciparum or vivax malaria episode were reduced by 64% in children aged 0-10 years and by 38% in refugees aged < 20 years in the group using permethrin-treated *chaddars* and top-sheets. Incidence in refugees over 20 years of age was not significantly reduced.

[19] The following papers were identified but not included in the review because they did not fit the inclusion criteria: (i) Walker DJ. Improving the efficiency and cost-effectiveness of food aid grain delivery. Disasters 1996; 20(2): 133-143, (ii) a cost analysis of a food logistics operation, inHallam (1996); and (iii) an example of a 'back-of-the-envelope' cost-effectiveness analysis of aid logistics, in Anon (2000).

The cost of the permethrin treatment per person protected ($0.17) and per case prevented ($1.07) was similar to that for treating bednets (and cost only 10-20% of the price of a new bednet).

- SFP

Young et al (1988) evaluated the cost of 29 samples of food-aid biscuits collected during a survey of feeding programmes in Ethiopia and Sudan. All but two brands were produced in Europe. Only 20 brands were recommended for use as emergency rations or as nutritious supplements; eight brands were similar to traditional baked biscuits and four were infant rusks. The biscuits were compared on the basis of ingredients, energy density, unit size, protein content, fortification with vitamins and minerals and costs. A wide range in all characteristics was found. Biscuits high in protein tended to have a low energy density. Energy density and unit size influenced the volume eaten and ease of transport, storage and also distribution to recipients. The cost of energy (per 500 kcal) ranged between 14p for baked biscuits and 47.5p for the unbaked compressed products.

- TFP

Caldwell and Hallam (ENN, 2004) presented the cost per beneficiary of community therapeutic care (CTC) and discussed aspects of these costs, underlying assumptions and other factors and issues affecting cost. The costs discussed were all based on CTC Pprogrammes operated by Concern Worldwide and supported by Valid International. The cost per beneficiary for the CTC programme as a whole (including SFP, OTP and SC elements) varies from Euros 114 in south Sudan to Euros 62 in Ethiopia. South Sudan is recognised as a difficult and expensive country in which to operate. High costs occurred for the transportation of food, medicines and staff much of which has to be flown in from Kenya. In addition, this programme was only operational for 4 months at the time of the analysis and therefore, the full cost of set up/capital expenditure has been borne by relatively few beneficiaries. One would expect the cost per beneficiary at the end of the programme to be substantially less. Two major factors affected Malawi's cost; high overhead costs allocated to the CTC programme and the high cost of vehicles. Ethiopia is a well-established Concern Worldwide field operation, with many programmes which are relatively inexpensive to run. Hence the cost per beneficiary in this case is lower than in the two other examples. One of the most important factors affecting the cost per beneficiary is the number and the density of beneficiaries. The cost per beneficiary is based on the number of beneficiaries admitted to the programme as opposed to the number recovered.

3.7.1 Strength of the evidence of the economic-evaluation studies

From table A4.7 it can be seen that:

- Two of the studies were cost-analysis studies (Young et al, 1988; Caldwell and Hallam, in press) and one was a cost-effectiveness study (Rowland et al, 1999).
- One of the studies was ranked as a 'poor' quality study and two as 'medium', according to the economic evaluation criteria.

These findings show that to date there are very few good quality studies on the cost-effectiveness of the six emergency interventions with which this report is concerned. In fact, when the search was widened to include cost-effectiveness studies of other health interventions in emergencies, only a few more studies were found[20].

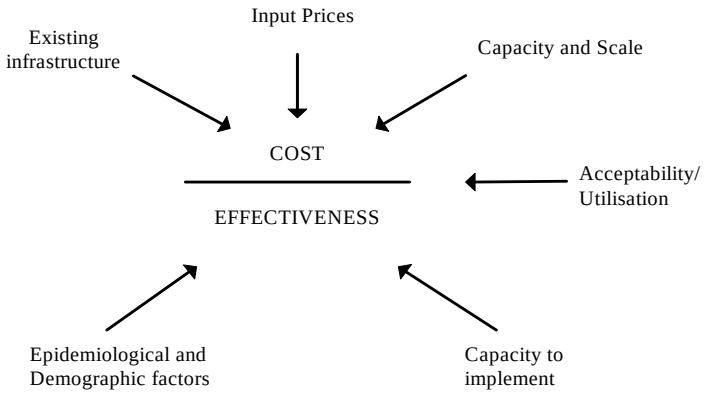
[20] Several papers have examined the cost and cost-effectiveness of other health care interventions in emergency settings. These include: visceral leishmaniasis (Griekspoor et al, 1999); tuberculosis (Biot et al, 2003); primary health care (Kasis et al, 2001; an example cited in Hofman et al, 2004); cholera vaccine (Naficy et al, 1998; Murray et al, 1998); HIV prevention (an example cited in Hofman et al,2004).

3.7.2 Summary of the results of the economic-evaluation studies

Given that the search only found three cost-effective studies and each of these assesses a different intervention, it is not possible to produce a summary of the economic evaluation studies.

If economic, epidemiological and behavioural factors were the same everywhere there would be no need to consider the generalisability of economic evaluation data, one could simply apply the same findings to different settings. Unfortunately, differences in these factors exist, both within and between settings (see figure 3.1).

Figure 3.1 Key factors affecting cost-effectiveness ratios



Major differences in existing infrastructure, capacity and scale make it is very difficult to extrapolate at all from non-emergency to emergency settings, as services are almost always provided in a very different manner. It is probably only the cost of the vitamin A capsule, bednet, measles vaccine, biscuit, etc. which is the same in emergency and non-emergency programmes. Indeed, although many relief operations involve undertaking essentially the same activities, such as the provision of food aid, the setting up of airlift capacity and the emergency provision of water and health services, each relief operation is unique to the circumstances in which it takes place. Cost-effectiveness analysis of one particular relief activity or operation will, therefore, tend to produce information that is highly context-specific (Hallam 1996).

Despite there being major problems with using economic evaluation data derived from development settings in emergency settings, because of the paucity of published studies assessing the cost-effectiveness of the six interventions in emergencies, the search was extended to assess these interventions in non-emergencies. The findings of this search are summarised in annex 6. A total of 22 economic-evaluation studies were located. A very brief over-view of the results is presented below:

- One economic evaluation of different modes of interventions for a TFP in Bangladesh estimated that per recovery (child attained a certain WFH) average costs ranged between

US$29-156 depending on whether the child was cared for at home, as an in-patient or through domiciliary care (Ashworth and Khanum, 1997).

- The economic evaluation studies of vitamin A estimate that the cost per death averted ranges from US $67 – 276 in development settings (Ching et al, 2000; Loevinsohn et al, 1997).
- The economic evaluation studies of bednets estimate that the cost per death averted ranges from US $219 to US $2,958 per death averted in development settings (Hanson et al, 2004).

The economic evaluation studies of measles immunisation programmes range from US $82-970 per death averted in development settings (Shepard et al, 1986; Robertson et al, 1985; Williams, 1989; Ponnighaus et al, 1980; Walker et al, 2000). None of the figures above are really meaningful unless they are compared to a cut-off figure of what donor agencies are prepared to spend to prevent a death. The ENN was unable to find any published information (for example, guidelines) on this. The informal cut-off forCIDAs Programme Against Hunger, Malnutrition and Disease, of around 350 Canadian $ (equivalent to US $252) per death averted, has been noted earlier. In this situation, all four of the interventions described above could potentially be considered ‘cost-effective’ in development settings.'

4.0 Summary of results

This section aims to summarise and review the strength of the published literature described in section 3. A discussion about why so little evidence is available is also provided. The results of the studies are summarised and information gaps are highlighted.

4.1 How much published impact and cost-effectiveness literature is available?

The most important thing to note about this review of the published studies assessing nutrition related interventions in emergencies is that there are very few studies available. Given the very large amounts of funds which are being spent on these types of interventions in emergencies, it is astonishing that there is so little information in the public domain about their impact and cost-effectiveness (Table 4.1).

Table 4.1 The number of published impact and economic evaluation studies undertaken in emergencies located by the search

Type of intervention	**Impact assessment**	**Economic evaluation**
General ration distribution	9	0
Supplementary feeding programme	15	1
Therapeutic feeding programme	16	1
Vitamin A supplementation	0	0
Bednets programmes	0	1[21]
Measles immunisation programme	0	0

Given the limited success of the search, the ENN sent a short questionnaire to some of the largest NGOs and UN agencies working in emergency nutrition. The headquarters' nutritionists of these organisations were asked to fill in table 4.2 below. The nutritionists were asked to focus only on supplementary and therapeutic feeding programmes between January 2003 and 2004 in the Horn of Africa (North and South Sudan, Ethiopia, Kenya, Eritrea and Somalia).

Table 4.2 NGO emergency programmes from January 2003-2004 in the horn of Africa[22]

	Number of programmes	Number of evaluations	Number of published evaluations
Supplementary feeding programme	20	7	1
Therapeutic feeding or CTC/OTP	14	4	1

UNHCR also provided the following table. Note that WFP is currently only implementing supplementary feeding programmes in Southern Africa and hence their results are not shown.

[21] The economic evaluation of bednets was a cost-effectiveness study which included a measure of impact, so this study could be classified in either column of table 4.1

[22] NGOs who responded and countries where they worked: ACF/AAH – Ethiopia, Somalia and North and South Sudan, Merlin – Ethiopia, Tearfund – Ethiopia and Eritrea, Concern – South Sudan, Eritrea and Ethiopia, Oxfam – North Sudan, SC UK – Ethiopia, North Sudan

Table 4.3 UNCHR emergency nutrition programmes from January 2003-2004 in the horn of Africa

	Number of programmes	Number of evaluations	Number of published evaluations
Supplementary feeding Programmes	40	Nil[23]	0
Therapeutic feeding or CTC/OTP	25	Nil	0

- Of the 65 SFPs implemented by these agencies between 2002 and 2003:
    - One evaluation report was published (in the ENN) – less than 2%.
    - Seven formal evaluations were conducted – less than 11%.
- Of the 39 TFPs implemented by these agencies between 2002 and 2003:
    - One evaluation report was published (in the ENN) – less than 3%
    - Fourformal evaluations were conducted – less than 11%.

Unfortunately, from the information provided above, it is not possible to know whether the formal evaluations conducted measured impact at the individual level, the population level or included coverage. Moreover, it is also not clear what type of evaluation was undertaken and hence, what the strength of the evidence would be. However, if these types of figures are representative of NGO practice (and they probably are of large European NGOs, less is known about US PVOs or the smaller NGO/PVOs), then the results of the search of the published literature are not surprising.

It is interesting to note how few organisations have undertaken evaluations, when most of the NGOs who responded to this questionnaire probably collect information on the impact of the programme – at least at the beneficiary level, through routine monitoring activities. Currently, we do not know what the NGOs do with their programme monitoring information, but it is likely that at least some of these data are thrown away when a programme ends. If this is the case, future researchers interested in collating grey literature information will have to specifically ask NGOs and UN agencies to hold on to these data (see section 5.1).

4.2 Strength of the available evidence

In general, there is less higher level evidence available for the impact of feeding programmes (SFP, TFP and GFD) than the other types of intervention (vitamin A distribution, measles vaccination and bednets). Virtually no studies employing a control group have been conducted for feeding programmes in emergencies. There are many reasons for this, including both practical and ethical difficulties, as well as issues which concern the generalisability of the results of a highly controlled study. There is more high level evidence available for the other types of interventions, however all the RCTs have been conducted in non-emergency settings, hence the applicability of these findings in emergencies needs to be considered.

The evidence provided for the impact of SFPs and GFD programmes is particularly weak. Although, the GFD studies score relatively well according to the CASP system (six out of

[23]Note: All UNHCR/implementing partners' selective feeding programmes performances are reported on monthly basis using the UNHCR HIS guidelines. These reports are submitted to UNHCR. Feeding programmes are also examined during the annual joint UNHCR/WFP missions. Recommendations for the programmes (e.g. to phase out, integrate, continue, increase coverage, etc) are made during this process. Feeding programmes are also partially assessed during nutrition surveys, and the results and recommendations from these surveys usually published in the RNIS.

eight of the studies were scored as 'good'), none of them fit Habicht et al's (1999) stringent criteria for plausibility and only two of the studies employ a control group. None of the SFP studies have a control[24] and the majority of the studies are ranked as 'medium' or "poor" by the CASP system.

The published evidence for the impact of TFPs is more convincing than that available for the other feeding programmes because there are more high level studies assessing the impact of TFPs. This finding was predictable, at least for health centre-based TFPs, where beneficiaries are entirely dependent on the programme inputs because the impact of a closed programme is more plausible than that of a programme which is run in tandem with other inputs (as is the case for an SFP) (Hofmann et al, 2004).

The recent development of home-based treatment for severe malnutrition is probably partially responsible for the relatively large amount of information available on TFPs in the past two years. Several agencies are beginning to assess when home rather than centre based programmes are more/less appropriate and this has resulted in a number of studies being conducted. Two RCTs comparing different treatment protocols for severely malnourished children were found in the literature review (ENN, 2003a and 2003b). Although the results of these trials provide important information, it is not clear how transferable their results would be to a non-trial setting. The information provided from the real-life programmes probably provides more realistic estimates of impact.

High quality meta-analyses of the impact of vitamin A and bednets on the mortality of children exist and therefore, no search was conducted for studies assessing these relationships. Similarly, measles immunisation is known to prevent mortality from measles so no studies looking at this relationship were assessed. Instead, this report reviewed studies assessing the impact of vitamin A supplementation, bednets and measles immunisation programmes on the nutritional status of children. The search only included studies with controls, which meant that the studies were of a relatively high evidence level. The studies also generally reported acceptable CASP scores.

Relatively high level studies assessing the impact of vitamin A supplementation, bednets and measles immunisation programmes on malnutrition in development settings were available. This is probably due to the fact that these programmes are undertaken in routine non-emergency settings, often in middle income countries, where there is more time (and money) to allow for rigorous evaluations to take place. However, no such studies were found in emergency settings. This creates interpretive difficulties as it cannot be assumed that impact will be similar in emergency circumstances. Indeed, the effectiveness of such programmes may change in emergency contexts, particularly where there is population displacement. There are several different ways effectiveness could be affected in an emergency, including:

- Programme utilisation rates (uptake of services) may change if peoples priorities no longer include going to the clinic (they may be too busy looking for food, for example)
- Coverage rates may change according to the security situation, whether a population is camp-based and easily accessible or scattered and harder to access
- If the incidence of a disease is higher, e.g.: measles because of over-crowding, then the overall impact might be greater.

One difficulty is that it is not possible to predict in what way the programme's impact will be affected prior to the emergency.

[24] One study (Stefanak and Jarjoura, 1989) compared wet and dry SFPs in the same area, but none of the studies compare SFP to no programme at all.

The methods applied to estimate costs in the economic evaluation studies reviewed give rise to questions of reliability, validity and transparency[25]. First, studies failed to systematically report the methods used to compute costs, and/or do not provide all the necessary data inputs such that re-estimation of the results would be possible (e.g. by providing menus of disaggregated resource use and cost data). A complete identification of costs was often missing from these studies - detailed information on the measurement of costs, e.g. how staff time and overheads were allocated was omitted; the valuation of costs was at times opaque, as evidenced by the absence of a base year relevant to the value of the chosen currency or the appropriate exchange rate and details concerning adjustments made for differential timing of costs; and a comprehensive description of the competing alternatives was rarely provided. These types of omissions make it difficult for researchers and programme managers to assess the reliability of the cost data.

Second, even among those studies that document their methods, these vary widely, making different studies, even within the same country and programme setting, largely incomparable. In addition, the treatment of shared resources (which are used jointly by one or more programmes) and capital (defined as goods that last for more than one year), costs varies among studies[26]. Furthermore, either due to poor practice or intent, the studies do not include a constant or exhaustive list of inputs, which may lead to under-estimating the cost of an intervention, and hence an over-estimation of cost-effectiveness. While some researchers are thorough in their inclusion of capital, administrative, overhead, depreciation, and opportunity costs, others only performed an incremental analysis focusing on the recurrent costs that represent the cost difference to the institution, adding further to the lack of comparability between studies. Unfortunately, it is difficult to know the impact the variable methods had on estimates of cost and cost-effectiveness. However, it should be noted that similar weaknesses have been documented for studies applied to the non-emergency health care field (Walker and Fox-Rushby, 2000).

4.3 Why is there so little information and research available?

There are a number of reasons why there is so little published evidence available:

1. The ethical difficulties of mounting RCT research in emergencies (Habicht et al, 1999; Victora, 2004).

2. There is no single body or agency responsible for assessing the overall effectiveness of any of these programmes in emergencies, or comparing effectiveness of different types of programme design. Thus, while agencies may be responsible for demonstrating that programmes meet certain standards, e.g. SPHERE, there is no overview of what proportion of programmes meet these standards and/or the cost of programmes in meeting these standards in specific contexts.

   Thus, there is no collation of overall evidence about the impact of programmes and hence no entity which can publish findings and advocate for change. As a result, in some cases, vested interests and inappropriate 'technical' mandates may prevail. There are, of course, some notable exceptions and ACF and Valid International, in particular, have been working hard at publishing the results of their community based

[25] These comments are based on a review of all the emergency papers identified.

[26] Hallam (1996) states that there is currently no standard way of dealing with the issue of overheads among humanitarian agencies. Furthermore, he states that "NGOs, when collecting from the public, generally wish to downplay the proportion of funding that goes on overheads, yet when seeking funds from official donors like to raise overheads as high as possible to increase the level of 'own' resources."

therapeutic feeding interventions. There is, however , no equivalent drive to assess overall impact of SFPs or GFD programmes.

3. In order to undertake and publish good quality studies, researchers need training in epidemiology. Many nutritionists (who usually run SFPs) or logisticians (who usually run GFDs) do not have this training. This may explain the bias towards better reporting for TFPs, measles, vitamin A and bednets programmes, i.e. there are greater numbers of medically trained staff working in these areas who are more likely to understand the principles of epidemiology and the need to publish. The capacity to perform economic evaluations is also inadequate.

4. Impact assessments are expensive and time-consuming. In an emergency situation they may not be seen as the first priority. This may also account for the greater amount of impact information on measles, vitamin A and bednets interventions, as these are carried out on a routine basis in stable situations where there is more time (and money) for evaluation.

5. Impact assessments need careful planning. The intrinsic unexpected nature of most emergencies militates against planning (certainly natural disasters). There may therefore, be insufficient time, or personnel available, to conduct a proper impact assessment under these circumstances.

6. Incentive structures in agencies promote defensive behaviour and a culture of blame so that transparent impact assessment may be threatening (ALNAP 2002, 2003a)

7. Short term funding mechanisms and high staff turnover rates militate against a learning environment (ALNAP 2002, 2003b)
8.
9. There is often a lack of clarity as to the objectives and desired outcomes of interventions (ALNAP 2002,2003a)

4.4 Summary of the results of the published impact and economic evaluation studies

It is difficult to draw any meaningful conclusions about the results' of the studies described above because (i) there are very few studies, (ii) the studies assess impact on populations living in different contexts (in terms of infrastructure, access to food and health care, security etc), (iii) the protocols of the interventions are different in many of the studies, (iv) the studies' evaluation techniques (for example, measuring coverage) vary. All of these factors argue against attempting generalisation and summary statements of the results. Hence, the discussion below presents some very general points about the results and suggests where gaps in the literature exist.

Note that it is not clear how far the results presented are due to over-reporting of positive results. This phenomenon is well known in other areas of research (Easterbrook et al, 1991) and there is no reason to suspect that biased reporting would be any different for this field. The ENN's Field Exchange, which was set up to help field based practitioners improve practice by recording field experience and thereby strengthening institutional memory, does encourage the reporting of both positive and negative results. However, reports of agencies not wanting to show their "dirty linen" in public have been recorded even for Field Exchange.

General Ration Distribution Programmes
All the available studies show a positive impact of GFD programmes on mortality and malnutrition. However, there are very large gaps in the information for these programmes. In

particular, we have no comparative impact information for different modes of GFD (e.g.:community based targeting versus administrative targeting, or cash versus food distributions).

With the current HIV crisis, food aid is increasingly being called upon to have a number of diverse roles and objectives. These range from increasing incentives to attend programmes, like Prevention of Mother to Child Transfer of HIV (PMTCT) and Tuberculosis treatment (DOT), increasing food security of PLWHA in home based care programmes, enhancing the efficacy of ARV treatment, protecting the vulnerable in neighbourhood care, and orphans and vulnerable children programmes. While this study has focussed upon evidence for a nutritional and mortality impact of feeding programmes, there will be an increasing need to measure the impact of GFD programmes in relation to the expanded set of objectives that food aid is being asked to fulfil in an emergency and/or protracted emergency context. Currently there is little, if any, information on the effectiveness of these newer types of feeding initiative.

There is also a gap in the literature which assesses different types of ration and their impact on micro-nutrient status. As described in section 3.1.2, this is partially because micronutrient deficiencies are difficult to measure in emergency situations, however there is still a need to obtain information on this subject.

There is no economic information available for any of the GFD programme types. This means that we cannot compare the cost-effectiveness of any of the different ways to provide food to households in emergencies.

Supplementary feeding programmes
The evidence for the effectiveness of SFPs is equivocal in this review. Of the 15 studies reviewed, only six presented their results in such a way as to be able to measure whether or not they met the sphere standards for recovery (i.e. improve the beneficiaries' nutritional status) and coverage. Only 50% of these studies did achieve the Sphere recovery standards although most (75%) of the other studies assessing weight gain in children did record a positive impact of SFPs. A further 83% of the studies achieved the coverage rates suggested by Sphere (see section 3.2.3 for more detail).

Only one economic evaluation study assessed SFPs. This study compared the costs of different types of biscuits commonly available for feeding programmes in emergencies. Given that the biscuits are probably only a small fraction of the costs of a whole SFP, this information is insufficient to make conclusions about whether or not SFPs are cost-effective.

There is an urgent need to assess whether targeted SFPs have a greater impact and are more cost-effective than other types of feeding programmes – for example, an enlarged general ration.

There is also no published information about the impact or cost-effectiveness of SFPs on age groups other than children aged 6-59 months, despite large numbers of programmes for pregnant and lactating women. There is a dearth of information about the impact of these programmes on micronutrient deficiencies. There is no information about the impact of blanket SFPs at all. Finally, this search found no published studies assessing the non-nutritional impact of SFPs (or GFDs) in emergencies. It is likely that some of the benefits of these types of programmes (for example, increase in physical activity due to higher energy levels, etc) are therefore being missed.

Therapeutic feeding programmes
The review of the TFP studies has shown a surprisingly low attainment for the Sphere recovery indicators (57%). This may be because many of the studies assessed home-based

rather than centre-based treatment and home based treatment protocols are still being developed and refined. However, a higher than expected proportion (63%) of studies achieved the Sphere-recommended coverage rates (see section 3.3.2 for more detail).

The relative effectiveness, and also cost-effectiveness, of the two types of TFP in different contexts is really the most important question about TFPs that donors and other agencies need to answer currently. This review reported on one economic evaluation of a home-based implementation of TFP undertaken in three different emergency settings. More of this type of work is required.

Bednet programmes
There is good quality, convincing evidence that bednets reduce mortality risks for children in non-emergency settings. There is also some, limited, evidence that these programmes have a positive impact on children's nutritional status. Economic evaluations are also available for development settings. However, gaps remain in the bednet literature for emergencies. One study showed that bednets could be economically viable in a refugee emergency in Afghanistan. How far this study is generalisable is unknown (see section 4.2).

Vitamin A supplementation programmes
There is robust, convincing evidence that vitamin A supplementation reduces the mortality risk of children in non-emergency settings. The evidence for the impact of supplementation on nutrition, however, is equivocal (see section 3.4.2 for more detail). No published information on the costs of these programmes in emergencies is available. As described above, it is not clear how either the impact or costs of these programmes would change in emergency settings (see section 4.2).

Measles immunisation programmes
Measles immunisation programmes reduce children's mortality risk in development settings. There is also a limited amount of evidence that suggests that immunisation may reduce population level malnutrition rates. Comparable cost data for measles programmes conducted using different strategies (campaigns rather than as part of the standard Expanded Programme on Immunisation (EPI) programme) is also available from studies in development settings, but not emergencies. Again, it is not clear how the impact or cost-effectiveness of these programmes would change in emergency settings (section 4.2).

4.5 How generalisable are the findings from this report?

One final point to note about the generalisability of the results described in the studies reviewed in this report is that the literature on emergencies is probably biased towards the most severely affected (and, in some locations, more accessible or more highly publicised) populations of operational interest, and towards situations where it is more practical to collect data and where there is a more compelling operational reason to do so (Seaman, 1993). It is not known how these biases will affect the conclusions about the effectiveness of the interventions.

5.0 What steps can the humanitarian community take?

The section begins with a general discussion of how future researchers may use the grey literature (i.e. unpublished studies) to further assess the impact and cost-effectiveness of the six nutrition-related interventions. This is followed by a more detailed description of the different types of analysis and assessments that should be undertaken in order to fill in the gaps in our current knowledge, described above. Each of the different interventions is considered separately as different models of analysis and assessment are required. A summary of the description of the gaps in information, or the most pressing data needs, is presented first. This is followed by suggestions about how to analyse the existing grey literature found in historical project reports. There is also guidance regarding the type of impact assessments that may be appropriate for specific interventions, with a view to providing valuable information for future planning.

Finally, this section explores institutional mechanisms for improving the information base on impact and cost-effectiveness.

5.1 The grey literature

The so called 'grey literature' held by agencies that implement humanitarian programmes is a tantalising 'pot of gold' to most researchers. This unpublished literature (mostly residing on the office shelves of project/country officers working for agencies) falls into numerous categories, i.e. monthly project reports, project mid and end term reports, baseline, mid-project and end of project surveys, internal agency project evaluations, external agency evaluations, emails between project staff and headquarters nutritionists/specialists, audited reports to donors, country accountant reports, country programme manager reports to head office, unpublished research held by academics and academic institutions, etc. It is conceivable, if not likely, that the information and data held in this enormous volume of literature holds certain answers to questions around operational practice and optimal resource allocation. It is also likely that a degree of expert guidance on how these types of reporting may be modified in the future could lead to their significantly increased utility for programming purposes.

As described in section 1, there is currently a move to wards using a range of research methodologies for evidence based public health. Petticrew & Roberts (2003) argue that a greater emphasis should be put on methodological appropriateness rather than study design. Waters and Doyle (2003) raise the issue of a lack of reviews in childhood malnutrition and stress the importance of locating 'difficult to locate' intervention studies which may be in the grey or unpublished literature. Previous work by Wyness, Mandava and Knight (2004, unpub) showed that there are two fundamental problems; a lack of standardised reporting in the published literature, and authors within this field are very mobile between posts and are very difficult to track down for further information - even then the information may not be sufficient. Note that the project reports in the grey literature may actually be more standardised in terms of reporting Sphere outcomes, than the published literature and hence that this information will be easier to assess than the results of the published literature.

Given the limited evidence that is published in this field, it is likely that systematic reviews which do not include the appropriate grey literature may well be biased, as well as limited. We, therefore, suggest that the grey literature should be included in future reviews but that it should be systematically collated and the quality assessed in a standardised way. The standard methodology for a systematic critical literature review of published studies needs to be extended for the grey literature in this field, because the basic issues of reducing subjectivity in a review of the evidence, and also for ensuring reproducibility, remain.

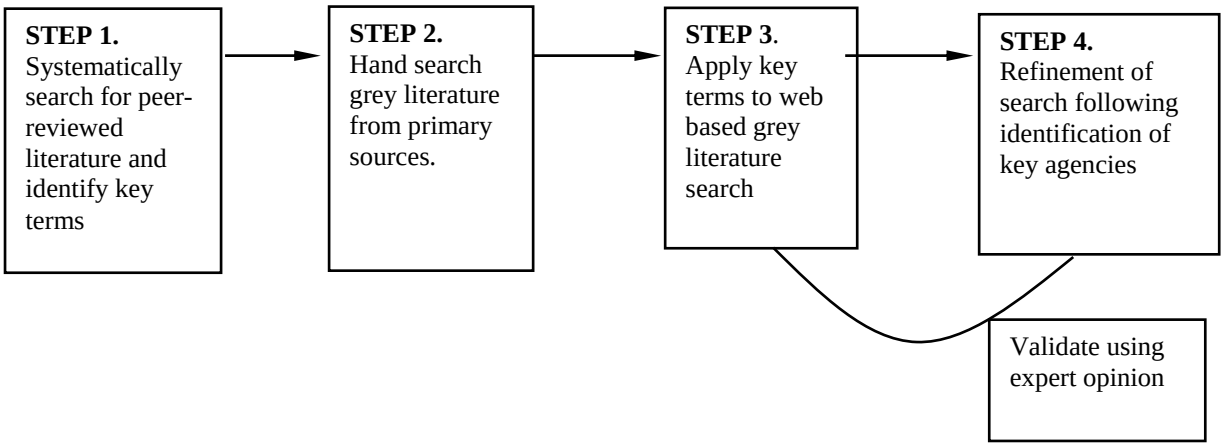
5.1.1 How to find the grey literature

The challenge is how to create an explicit, objective, reproducible method of systematically collating and appraising grey literature when it is, by definition, difficult to define, difficult to find, randomly stored and of varying strength. A standard systematic critical literature review of published literature consists of four stages: database search, secondary reference search (references obtained from papers found at the first stage), hand search of the main journals, expert advice on the literature field.

One focus for a grey literature review would be to assess what is published on the World Wide Web. The World Wide Web is divided into two parts, the surface web and the deep web. The surface web contains websites that can be identified using search engines, where as the content of the deep web resides in searchable databases, the results from which can only be discovered by direct query.

A 4-step process has been developed to identify grey literature. The first step is to carry out a systematic literature review of the study area through databases. This enables the key articles and authors to be identified. It also identifies the terminology commonly used in the studies. The second step is to hand search the grey literature from primary sources (main organisations involved). This provides additional information on the terminology used in the field and important articles, but may be unsystematic. The third step is to apply the key terms identified in steps 1 and 2 to the web based grey literature search. This identifies articles in the surface web. The fourth step is a refinement of the search to focus on the websites of the key organisations and search within their surface websites for grey literature that exist in their databases. Steps 3 and 4 can be repeated as the search becomes more focused as part of an iterative loop. The final articles are validated using expert opinion (figure 5.1).

Figure 5.1 Process to identify the grey literature



These foursteps include elements of the 4 stages of a standard systematic critical literature review described above. Hand searching is required of all relevant files and documents identified by experts in the main organisations involved. The only way to obtain project reports, for example, may be by actually asking key individuals working within the agencies which implement the programmes. This may not be so reproducible because it will depend largely on personal contacts. To be more systematic in accessing grey literature and so reduce the subjectivity, a strict search pro-forma will be required which is used in the same way for each organisation. It is anticipated from previous research in this field (Wyness, Knight and Mandava, 2004) that most literature will be found through expert advice and hand searching.

Note that “grey” cost data may be harder to obtain than impact data because agencies may consider this type of information as confidential. This suggests the importance of establishing a mechanism whereby cost data is made more available (see section 5.3).

5.1.2 Assessing the quality of the grey literature

Currently, there is no standard method of assessing the quality of grey literature. Researchers will probably have to adapt the CASP score methodology. A further issue that needs consideration is how to combine the grey literature with other (peer reviewed) literature to build up evidence for causal associations between inputs and outcomes. The different quality scales for differing types of studies (case control, cohort or cross sectional survey, standard critical appraisal tools) have to be brought together with common criteria measured across the studies or a sensitivity analysis conducted using a consensus of the assessed potential for bias in each study.

This review has focused on the quantitative studies with controls, but there is the question of whether the grey literature including the descriptive observational reports, may increase the generalisability of findings from quantitative studies, and also answer relevant questions on the performance indicators of a programme. For this, a quality scale or tool needs to be developed from quality assessment or evaluation theories, appropriate to this field, that, as above, again can be brought together with the current quality scales. Appropriate quality assessment could include a) the organisational management approach realising a network of stakeholders involved and including local participation and ownership and also b) the evaluation approach (beyond the one focused on in this current project looking at studies with control groups), combining quantitative and qualitative techniques to produce the evidence decision-makers need , which of course includes economic evaluations.

Finally, if the grey literature is collated, it needs to be both accessible and useful for decision makers. For example, it would be useful to be able to use the evidence from past programmes to predict what will happen in new situations. Thus, there needs to be pattern recognition of the main factors that affect impact. Case Based Reasoning (CBR) work is currently being piloted in the field of childhood malnutrition in complex emergencies. CBR may include a wide range of studies. CBR takes weighted data from previous ‘similar’ situations, and synthesises it to predict what the outcome will be for the current situation. CBR can be used in situations where one would intuitively think it to be unique, but ‘unique’ complex situations can have common performance indicators and impact factors. CBR is also a methodology to allow quick access to relevant evidence for a situation.

5.2 Types of analysis needed to fill current knowledge gaps

5.2.1 General ration distribution programmes

Currently there is virtually no scientifically rigorous published information about the impact of general food distributions on population’s mortality or nutritional status. But is this type of information really needed? If the initial needs assessment was correct and a population is malnourished because it needs food (rather than health or watsan interventions) then it can be assumed that providing well targeted food at household level will positively impact on nutritional status and, hence, mortality. In this situation, the knowledge most required is “what is the most effective and efficient way to improve household food security?”

There is currently a lack of information needed about comparative cost and effectiveness of the different types of GFDs or food security support measures (cash for work/food for work, income transfer, etc) in different contexts. For example, it would be useful to know whether or not a general ration targeted by the community or by the administration/implementing

agency works best in different contexts. This is an especially critical question given the current groundswell of opinion that community-based targeting is the most effective route for GFD and the seemingly unchallenged rolling out of this type of design in the recent southern Africa emergency. Other key questions relate to the comparative costs and effectiveness of GFDs versus market support in urban areas, cash for work/food for work, cash transfers.

What can we get from the grey literature?
As a first step towards obtaining cost and effectiveness data, a meta-analysis of GFD programme cost and process indicators should be undertaken. This should include information on programme coverage (including inclusion and exclusion measures) in different contexts. At a minimum, the information shown in the table below should be included.

Table 4.1 Minimum information needs for a database assessing the cost and effectiveness of different types of GFDs

Type of GFD	Targeting objectives	Actual targeting figures			Cost per beneficiary	Context
		Proportion of beneficiaries included in the distribution who should have been included	Proportion of beneficiaries who should have been included in the distribution but were not (exclusion measure)	Proportion of beneficiaries who were included in the distribution but should not have been (inclusion measure)		

Data collated in the way described above could then be analysed to answer these kinds of questions:

i. In context A, what types of GFD was most likely to achieve the highest coverage of the population?
ii. In context B, what is the most effective way to reduce inclusion errors?
iii. In context C, what is the cheapest way to deliver food to households?

The kind of information needed to construct this type of database should be routinely collected by most programmes during food basket monitoring (Jaspers and Young, 1995; WFP, 2003). Much of this information should already be held by WFP and ICRC and/or implementing partners. Data on costs may be more difficult to obtain at the moment, but an effort should be made to collect this information in the future.

Another role for the grey literature is to obtain more information about the impact of a GFD on population micro-nutrient status. The SCN receives reports about population level micro-nutrient deficiency disease outbreaks every year - mainly from refugee/IDPs food aid dependent populations. These outbreaks imply that the current assumption that food aid dependent populations can obtain micronutrient requirements outside of the food provided to them in their general ration is not always correct. However, it is not currently possible to estimate the size of this problem because no analysis of the frequency of the outbreaks has been made. Furthermore, the reliance on fortified blended foods (usually shipped from overseas) to ensure dietary micronutrient adequacy for food aid dependent populations has not always proven to be effective (mainly due to logistical and resource issues), yet there is no overview of the frequency with which this strategy fails.

There is a need to collate all reports of micronutrient deficiency disease outbreaks and estimate the prevalence of the outbreaks against the number of GFD programmes. For the

moment, this may only be possible for refugee and IDP camps, as agencies do not currently record micro-nutrient deficiency outbreaks in a systematic way. A combination of the RNIS database and UNHCR records should have sufficient information to allow for a historical analysis of this type.

In order to obtain more information on the impact of a GFD programmes on micronutrient deficiencies in the future, it will be necessary for the agencies which implement GFD programmes on a large-scale (such as WFP and ICRC and their implementing partners) to start recording information about outbreaks in a systematic way.

What types of impact evaluations should we be undertaking in the future?
Following Habicht's example, the different types of evaluations that could be employed to assess a GFD programme can be found in table 4.2 below.

Table 4.2 Examples of possible evaluations for general ration programmes

Type of evaluation	**Provision**	**Utilisation**	**Coverage**	**Outcome/Impact**	
				Food security	**Nutritional status**
Adequacy	Availability of food at distribution points	Amount of food distributed at the distribution points and use of food by poor HH *- food basket monitoring* *- post-distribution monitoring*	Measurement of % of poor HH which received the general ration, through: *- coverage surveys (looking at inclusion errors and exclusion errors)*	Measurement of trends in food security in intervention area, through: *- food security assessment before and after the intervention*	Measurement of trends in nutritional status in intervention area, through: *- anthropometric survey before and after the intervention*
Plausibility	As above, but comparing interventions with control areas	As above, but comparing interventions with control areas	Comparison of coverage of general ration between intervention and control areas	Comparison of trends in food security between intervention and control areas (or dose-response)	Comparison of trends in nutritional status between intervention and control areas (or dose-response)
Probability	As above, but intervention and control areas would be randomised	As above, but intervention and control areas would be randomised	As above, with randomisation	As above, with randomisation	As above, with randomisation

A probability assessment, or RCT, of a GFD programme is not feasible in an emergency setting. An adequacy evaluation which measures the impact of a GFD programme on household food security or nutritional status would need to include a food security assessment and an anthropometric survey, before and after the intervention. However, these types of evaluations will not enable the donor to know whether or not the improvement seen is due to

the intervention, or any one of many other factors e.g.: a new harvest, a watsan intervention, etc.

Given the problems with probability and adequacy evaluations, we can only attribute improvement to a GFD intervention by comparing a change in nutrition/food security in the intervention population with a change in a control population.

The problem is how to decide which population to use as a control? Ideally, the control group will exhibit identical characteristics (on aggregate) to the beneficiary group, except for participation in the project. Potential controls and the comparative advantages and disadvantages of the different types of control groups are described below:

(a) A population living in very similar circumstances (same livelihood group, access to markets, etc) as the intervention population, however, the control population is not given the food in order for the evaluation to be conducted.

(b) A population living in very similar circumstances (same livelihood group, access to markets, etc) as the intervention population, however, the control population was not given the food due to a programmatic problem for example, insecurity or a lack of resources.

(c) A population living in very similar circumstances (same livelihood group, access to markets etc) as the intervention population, however, the control population was given the food sometime after the intervention population as the programme was being phased in.

(d) The same population at the end of the intervention when the general ration distributions are decreased or stopped altogether.

Ethically, situation (a) is impossible. In an emergency, it is not possible to withhold food from one group for the sake of an evaluation. Situation (b) is better in terms of ethics, but one would need to carefully consider why one group was chosen to receive the intervention and not the other. If one area is more insecure than another, then it is possible that the population's food security is worse anyway and hence any difference seen is due to the security, not the programme. If a lack of resources was the main reason, then who made the decision about which group was going to receive food and which one wouldn't? Was the choice political? Do the politically favoured group have a better/worse food security situation anyway?

Situation (c) may be used to assess the impact of development programmes. One example of this was the mid-term evaluation of the impact of the Rural Development Plan of the Western Region (PLANDERO) in Honduras on the nutrition and food security of the population (Carletto and Morris, 2001). The use of phasing in controls overcomes the ethical problems of situation (a) and some of the problems about the groups not being identical described for situation (b), although again one needs to ask how the decision to start with one group and not the other was made.

The main problem with situation (c) is the timeframe. Development programmes usually have a much longer timeframe over which to measure the programmes impact so phasing in is possible. For example, the PLANDERO programme was phased in over several years. But emergency programmes, particularly general ration distributions, do not usually last several years (except in a camp situation, but then there is the problem of ethics again – withholding food from one section of the camp population). If the timeframe is less than one year, the problem of seasonality may arise. If the first part of the general ration distribution happens before a harvest and the second part after the harvest, it will be difficult to attribute the impact

to the intervention alone. The PLANDERO programme was phased in at the same time each year in order to control for the impact of seasonality.

A further problem with situation (c) is that it is necessary to know which population groups will be phased in at what time before the programme starts. The evaluators assessing the PLANDERO programme were fortunate because they knew which communities were earmarked for inclusion before the programme started. This is unlikely to be the case in most emergency situations.

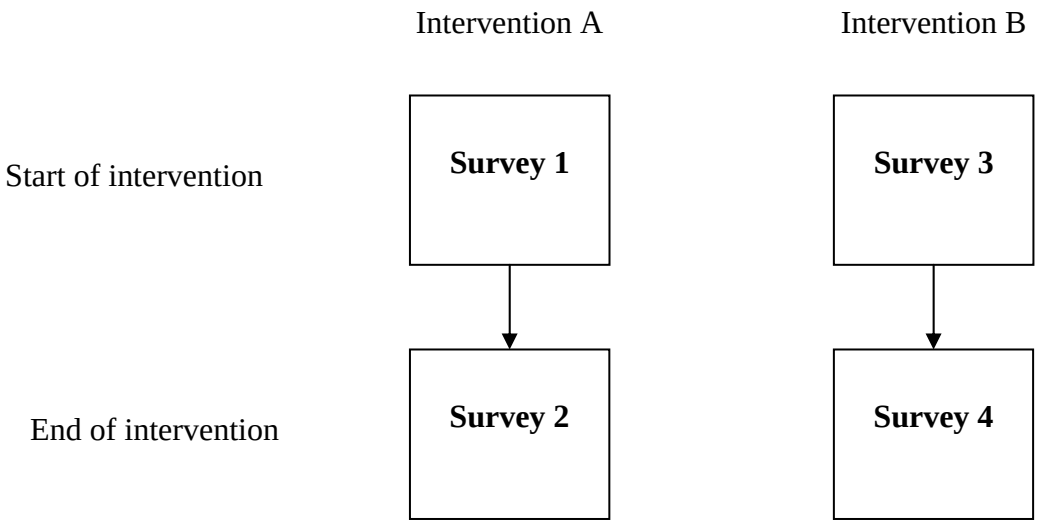
Some authors have suggested that situation (d) is a useful method to demonstrate impact – this involves monitoring the nutrition/food security situation when the intervention is withdrawn or phased out. An example of this comes from Angola. In 1995, access to land in Malange had improved but was still limited by sporadic fighting and the widespread laying of landmines. Local traders were operating aircraft and had established rudimentary markets. As a result of the improvement, the international humanitarian community was attempting to decrease the amount of food aid provided. It was quickly noted that in order to determine the optimal level of food aid requirements, reductions in the general ration should be carried out in conjunction with surveillance activities. The surveillance showed that a gradual reduction in the general ration did not necessarily have a negative impact on nutritional status. However, the near total withdrawal of the ration in December 1995 resulted in an increase in the level of acute malnutrition among children under-five. This suggested that certain population groups within Malange were still, at least partially, dependent on external assistance. More in-depth qualitative studies revealed that a proportion of the population, especially those who were displaced from rural areas and had no access to land within the peri-urban secure boundaries of the city, were particularly vulnerable if the ration was withdrawn (Borrel and Salama, 1999). Information on speed of change in population level nutritional status in relation to reduction in general ration programme will also strengthen confidence in this type of assessment.

Situation (d) overcomes the ethical problems of situation (a) – provided the intervention is not withdrawn just for the sake of evaluation, and also overcomes the problems described in (b). However, although the Malange example given above does imply that the withdrawal of food aid did have an impact on the population's nutritional status again, without a control group it would not be possible to be absolutely certain that the effect was due to the withdrawal of the food aid rather than another factor such as disease, for example there may have been an increase in diarrhoeal disease amongst the under-fives in December 1995. In order to be certain that the intervention caused the impact it would be necessary to control for these confounding factors.

Situations (c) and (d) only occur rarely in emergencies. Furthermore, if situation (d) did occur, it is likely that most agencies' first priority would be to start the intervention as soon as possible, rather than undertake an assessment. So, although interesting cost and impact information can be obtained from scenarios (c) and (d), it is unlikely that these scenarios could be used to generate much information.

It will probably be easier (especially from an ethical standpoint) and more useful to undertake impact studies which compare the efficacy of various different ration or resource distribution models. In such a situation (shown schematically in figure 5.1), one would collect cost, process and impact information on two different interventions running in parallel in similar populations.

Figure 5.1 Schematic representation of an evaluation to measure impact of a feeding programme



The evaluation model shown in figure 5.1 will allow the following analyses:

Analysis	**Comparison**
Outcome	2 vs 4
Change	(2-1) vs (4-3)
Effectiveness	(2-1)/1 vs (4-3)/3

A combination of process indicators and nutrition or food security impact monitoring indicators should be used to compare similar populations with different intervention models at the same time. Key process indicators would be food basket monitoring (for quality and quantity), and targeting (inclusion and exclusion errors). Population impact monitoring could be conducted as before/after food security assessments and/or anthropometric surveys in the different intervention areas.

This type of evaluation would be useful, for example, when comparing a cash-for-work and food-for-work programme. In Ethiopia, SC UK is currently implementing two emergency programmes in neighbouring administrative areas with similar agro-ecological and health environments. One programme is providing food for the beneficiaries and the other is providing a cash-equivalent of the food. It would be very useful to compare the outcomes and costs of these two programmes.

The cohort study described in section 3.1 (Quisumbing, 2003) is similar to the method proposed above. However, this study used retrospective data. In the future, it would be more useful to have prospective studies, so that all the necessary information about confounding factors can be collected.

Various conditions need to be in place before a prospective study, like the one described above, can be undertaken. These conditions are described in more detail in section 5.3.2.

Summary Points and Recommendations

- There are a number of key unanswered questions regarding the comparative impact and cost-efficiency of different types of GFD and emergency food security support measures during food crises, e.g. GFD versus cash transfers, community based targeting versus administratively controlled targeting, etc.
- The grey literature should be fully mined to obtain more comprehensive information on coverage and targeting variables (from food basket monitoring data) and on incidence of micronutrient deficiency outbreaks relative to number of overall GFDs (sources would include RNIS).
- It can be assumed that GFDs will have a positive nutritional/mortality impact if the initial needs assessment is correct and coverage and targeting are good
- Although certain emergency scenarios, e.g. when gradually phasing out a GFD, may provide opportunities for controlled impact assessment, such opportunities are rare and usually involve methodological weaknesses.
- It is therefore more appropriate from an ethical, methodological and utility perspective to compare costs and effectiveness of different ration or distribution models (in similar agro-ecological-health areas) in order to address key unresolved issues around optimal emergency GFD/food security support approaches.

5.2. 2 Supplementary Feeding Programmes

Of all the interventions reviewed in this report, SFPs have the weakest rationale. Most SFP's objectives include the prevention and/or treatment of mild to moderate malnutrition. While there are few major controversies about the need for TFPs and GFDs, emergency SFPs still provoke vigorous debate about whether and how such programmes should be implemented (Shoham, 1994; Mason, 2002; Mourey, 2004; Curdy, 1994). There are several reasons for this. Targeted SFPs are predicated on the basis of general ration adequacy yet this cannot be guaranteed in many emergency contexts, especially where the agency implementing the GFD is not the same as the one implementing the SFP. Where general rations are inadequate (this frequently occurs for a variety of reasons), the basis for a targeted SFP is seriously undermined and impact is likely to be diminished. There are also questions over the cost efficiency of setting up a distribution system in parallel to the GFD, when SFPs usually only provide up to 10% of household caloric needs. ICRC policy is to simply provide a large general ration (2,400 kcals), which would allow weight recovery of the mild and moderately malnourished. There are also significant practical difficulties around attending SFPs and opportunity cost (especially for on site feeding in areas of difficult terrain or insecurity).

There are also issues around blanket feeding of under-fives and other vulnerable groups. These programmes are predicated on the basis that GFDs/food security will improve in a short period of time yet implementing agencies often have little control of GFDs While it is true that most SFPs are implemented either to allow recovery of moderately and mildly malnourished children or to prevent nutritional deterioration (blanket), some SFPs are implemented for other reasons (for example, if security is very poor and this is the only way to get food out into a community or for nutrition education purposes etc). Attainment of these additional objectives is almost never tested by evaluation or reviews.

Despite these theoretical and conceptual problems and the relatively large amount of critical reviews written by various experts, SFPs continue to be implemented in a large number of emergencies. This is partially due to donor policy, but also because many of the smaller agencies are unable to implement large-scale GFDs but still want to " do something" in a crisis. Also, SFPs require less technical expertise/capacity than a TFP (medical expertise) or GFD (logistical capacity).

Given the theoretical and conceptual problems surrounding SFPs, this review argues for the need to urgently review the effectiveness of these programmes especially as the evidence base of SFPs is weaker than that for any of the interventions examined (section 4.3).

Overall measure of efficacy for feeding programmes
There is a need to compare the overall efficiency of SFPs in different contexts. Currently, most agencies record coverage and individual children's recovery rates. These two variables could usefully be combined to obtain a proportion which reflects number of children recovered out of the total number of malnourished children.

$$\text{Proportion of met needs} = \frac{\text{No. of malnourished children who recovered on the programme}}{\text{No. of malnourished children}}$$

This measure would reflect both the coverage and the efficacy of the programme – i.e. it would represent the effectiveness of the programme. The calculation is very straightforward provided the term "recovered" is well defined. A potential definition of 'recovered' could be 'children whose WFH median reached more than 85% for two distributions in a row'. In other words, recovered means planned exits. Defaulters and deaths should not be counted as recovered and nor should children who remain in the programme at the end of its lifespan.

In order to calculate this proportion for two different programmes, the information in table 4.3. is required.

Table 4.3 Information needed to calculate the proportion of needs met from feeding programme performance indicator records

Programme	Total number of children aged 6-59 months in programme area	Prevalence of acute moderate malnutrition in children aged 6-59 months	Number of children treated in the SFP	Number of children who recovered
Programme A				
Programme B				

A worked out example is shown in table 4.4

Table 4.4 Worked example of calculations of met need for two feeding programmes

Programme	Total number of children aged 6-59 months in programme area	Prevalence of acute moderate malnutrition in children aged 6-59 months	Number of children treated in the SFP	Number of children who recovered	Programme coverage	Programme recovery rate	Proportion of needs met
Programme A	30,000	15%	3,000	2,345	= 3,000 / (0.15* 30,000) = 3,000 / 4,500 =67%	= 2,345 / 3,000 = 78%	= 2,345 / (0.15* 30,000) = 2,345 / 4,500 =52%
Programme B	30,000	15%	2,000	1,250	= 2,000 / (0.15* 30,000) = 2,000 / 4,500 =44%	= 1,250 / 2,000 = 62%	= 1,250 / (0.15* 30,000) = 1,250 / 4,500 =28%

From table 4.4 the following conclusions can be drawn:

i. Programme A was more effective because it met the needs of 52% of the malnourished children compared to 28% in programme B.
ii. The reason that programme A was more effective than programme B was because it had a higher coverage rate (67% vs 44%) and a higher rate of recovery within the programme (78% vs 62%).

This kind of information could be extremely useful for donors. If it was matched to cost data (i.e.: information about how much each recovery costs), then there would be potential to begin comparing the effectiveness of different targeted SFPs in different contexts, as well as obtaining an overview of the general effectiveness of these programmes.

There is also a pressing need for more information on efficacy of emergency blanket SFPs. These programmes are normally predicated on the basis that all children should be targeted as general rations are inadequate. The rationale is that the blanket feeding acts as a holding operation and provides an opportunity for the implementing agency to lobby for the introduction of 'better' general rations. However, as stated above, the assumption that better general rations will follow is not always valid so that these programmes become open-ended and immensely expensive – sometimes duplicating a GFD infrastructure. An analysis of the proportion of programmes which are successfully superseded by improved GFD would be valuable in determining the utility of this type of intervention.

What can we get from the grey literature?
As a first step towards obtaining cost and effectiveness data, a meta-analysis of targeted SFP cost and process indicators should be undertaken. This should include information on programme coverage and recovery rates within the programmes. Calculations (such as the ones shown above) could also be made if information on population figures and the initial prevalence of malnutrition were available. Furthermore, by collating information on the

programme's context – for example data on population density, infrastructure/terrain, culture, security, general ration/access to food, health environment, it should be feasible to begin determining contexts in which targeted SFPs are more likely to achieve desired impacts.

The kind of information needed to construct this type of database should be routinely collected by SFP staff as part of programme management and reporting and will, therefore, be available at country, and possibly headquarters, level (Shoham, 1994; Clay and Stokke, 2000***)***. Data on costs may be more difficult to obtain at the moment, but efforts should be made to collect this information in the future.

The types of questions a meta-analysis could answer are:

i. In context A, what proportion of met needs can we expect?
ii. Is it reasonable to expect an SFP to reach 80% of a population in an urban area?
iii. In a situation of high HIV prevalence, what are the average recovery rates of children enrolled on an SFP?

There are many more questions which could be framed and answers found.

Agency programme reports should also be able to provide answers to questions about degree of success of blanket SFPs with regard to acting as a holding operation and leading to improved general ration adequacy.

What types of impact evaluations should we be undertaking in the future?
The information from the grey literature should help us to understand how many programmes meet the Sphere standards in what contexts and also, which factors predispose SFPs to be successful. However, given all the uncertainty about SFPs, there is a real need for more rigorous information about their overall impact. Table 4.5 presents some ideas for evaluations of SFPs based on Habicht et al's framework.

Table 4.5 Examples of possible evaluations of a targeted supplementary feeding programme

Type of evaluation	**Provision**	**Utilisation**	**Coverage**	**Outcome/Impact**
Adequacy	Availability of food at distribution points	Amount of food distributed at the distribution points to moderately malnourished children, through: *- food basket monitoring* *- post-distribution monitoring*	Measurement of % of moderately malnourished children who received the supplementary ration, through: *- coverage surveys*	Measurement of trends in individual recovery rates Measurement of trends in numbers of moderately and severely malnourished in intervention area, through: *- anthropometric survey before and after the intervention (focusing on rates of severe and moderate malnutrition)* *- surveillance[27] of incidence of severe acute malnutrition* *- mortality survey before and after the intervention* *- surveillance of mortality*
Plausibility	As above, but comparing interventions with control areas	As above, but comparing interventions with control areas	Comparison of coverage of SFP between intervention and control areas	Comparison of trends in numbers of severely malnourished and/or mortality between intervention and control areas
Probability	As above, but intervention and control areas would be randomised	As above, but intervention and control areas would be randomised	As above, with randomisation	As above, with randomisation

As for GFDs, an adequacy evaluation of an SFP will not give sufficient information about the programme's impact because the possibility of other factors being responsible for the change in nutritional status is too high. Probability evaluations, such as an RCT, would be unethical and impractical. This leaves us with the potential for undertaking various different types of plausibility evaluations, which are outlined below.

[27] Surveillance involves regular and ongoing data collection activities, analysis and timely dissemination to stakeholders (periodicity varies according to the individual system) whereas surveys usually denote a one off assessment which may be repeated to measure trends and/or impact. In general, surveillance systems may be difficult to set-up in emergencies except in a camp situation. Surveillance data is usually the preferred method to collect information on mortality because information on mortality collected in surveys is often biased for more information).

1. Cohort study. This involves comparing the prevalence of malnutrition in two identical agro-ecological areas – one where an SFP is up and running and one where no SFP is taking place. It would be necessary to ensure that other interventions (e.g.: GFD and medical interventions) were identical in both areas and to continue measuring SFP coverage and individual child impact in the SFP area. Ideally, prevalence data on malnutrition would be available at both the beginning and end of interventions in both areas, although a compromise may be to utilise data on both areas at the end of the survey. There would also be a need for cost/population data for both areas.
2. Case-control study. If there was an SFP operating in one area and not in another then at the end of the programme it would be feasible to match children of the same age/sex/maternal education/SEC/ration, etc, in SFP and non-SFP areas. A comparison of the matched pairs' nutritional status could then be made. Again, SFP and non-SFP areas need to be as similar as possible in terms of agro-ecology and other interventions.
3. Screening study. At the beginning of a programme a large and detailed screening would be necessary. During this time, all the children's names, ages and measurements are taken. The programme starts and some of the malnourished children go into the programme and others don't. Three months later a new screening to improve coverage is undertaken. It would then be possible to compare the progress of those who went into the programme to those who did not.
4. The time-in programme method (Mason and Habicht, 1984). This involves comparing outcome information for children who have been in a programme for a substantial period with others just entering, on a cross-sectional basis. An improvement in those who have been 'treated' for a certain time compared to those just entering can then be used to determine whether or not a programme has achieved its objective of improving moderate malnutrition. Comparisons should be made of children of the same age to exclude the effects of ageing. It would also be necessary to ensure that there has been no overall deterioration or amelioration of the population's nutritional status during the study time frame when using this method (Mason and Habicht, 1984).
5. The correlation study design (Mason and Habicht, 1984). This design model relies on a large enough cross-section of the participant population being exposed to the programme at varying levels of intensity. Data on outcome indicators (like nutritional status), programme indicators and confounding factors are needed. If the number of observations is large enough and the observations on the variables are significantly diverse, then it may be possible to use statistical techniques of control to allow for the confounding factors and variables chosen. An example of a correlation study would be to collect data on a cross-section of the population measuring a range of factors which could possibly affect outcome, including data on programme delivery. The analysis would then examine the degree of correlation of outcome with programme delivery, taking into account other factors. In its usual form, this type of analysis involves multiple regression techniques requiring extensive computing. Alternatively, multiple group comparisons can be made – these are more easily calculated (Mason and Habicht, 1984).

Method (1) has serious ethical issues – if the areas are so similar then it would probably be necessary to implement an SFP in both populations. Methods (2) and (3) are both less problematic in terms of ethical objections. However, for method (2) it would be necessary to look very carefully at the differences in possible confounding factors that might account for any differences in nutritional status between the matched pairs. For example, was it really only the SFP that distinguished the two areas, or were there other issues, like access, which also differed? If access was different, then access to markets could be different, in which case household food security could also have been affected, etc.

Similarly, although method (3) appears relatively straightforward, the researcher would need to be very careful to look at the difference between the household characteristics of these children – why do some mothers send their children to the SFP and not others? Are the more educated mothers sending their children? Are the children living furthest away from the distribution point attending the programme? Also, method (3) requires substantial detailed information to be collected at the start of a programme during a screening. Screenings are normally fairly busy periods so it may be difficult to add on an extra work task.

The "time in programme" method (4) presents similar problems to method (3). One obvious disadvantage of the method is self-selection (Mason and Habicht, 1984). Children entering an SFP at the start of the programme may have different characteristics to those entering the programme a few months later. For example, those entering at the start of the programme may live closer to the distribution site, etc. These differences could have an impact on the children's nutritional status. There is therefore a need to check to see if the mean nutritional status of children at entry into the programme differs over the programme's lifespan.

Probably the most reliable method is the correlation study type (5). However, this type of assessment has very large data needs and would be expensive and time-consuming.

The methods described above only measure the impact of an SFP compared to no SFP. A key question for practitioners is whether or not an SFP is a more cost-effective measure of improving children's nutritional status than a bigger general ration. In order to answer this question, an evaluation similar to the one described for GFDs (shown in figure 5.1) will be needed. In this scenario, it would be necessary to compare the impact and costs of an enlarged general ration to a GFD + SFP in two very similar populations. This could be planned by encouraging WFP and ICRC to work in close proximity in an emergency[28]. Such a situation occurred in certain prefectures in Rwanda following the 1994 genocide and resulting displacement.

Summary and Recommendation:

- There is substantial controversy over the efficacy of emergency SFPs and a large number of questions over the cost-effectiveness of this type of intervention.
- There is extremely little high-level published evidence of impact of targeted and blanket SFPs in emergencies.
- There is a need to begin collating and analysing information on recovery in SFPs and programme coverage from the grey literature (agency programme reports), in an attempt to gauge overall efficacy of targeted SFPs and begin understanding the context specific factors that predispose SFP's to success.
- There is a need to review the grey literature on blanket SFPs in order to develop an over view of the relative success of these programmes in acting as a holding operation (preventing nutritional decline) and in terms of being superseded by implementation of adequate general rations.
- Given the level of uncertainty over SFP efficacy, there is a need for more rigorous analysis of overall impact than can be undertaken through use of the grey literature.
- Although RCT studies are unethical and impractical for emergency SFPs, certain types of plausibility study may be possible.
- Most plausibility studies would involve a degree of methodological weaknesses
- Correlation studies would be most methodologically robust, although these require a substantial amount of data and are expensive undertakings.

[28] ICRC policy is to implement GFDs without SFPs providing 2400 kcals/capita, while WFP policy is to implement a GFD providing 2100 kcals/capita in conjunction with a targeted SFP.

- An additional and key area for study would be a number of comparisons between the impact and cost-effectiveness of an expanded GFD programme with a normal GFD programme (2,100 kcals per capita) in conjunction with a targeted SFP. Such studies, which should be conducted in a number of contexts, would not raise ethical issues and could be made relatively robust methodologically.

5.2.3 Therapeutic Feeding Programmes

Currently, the major debates about TFPs revolve around which model of implementation – home or centre based - should be undertaken in different contexts (ENN, 2003a). These discussions are related to issues of quality (proportion of recoveries) and coverage. Debates about the protocols for home-based programmes also continue. There are also a number of highly technical debates, e.g. optimal antibiotic regimes and malaria treatments.

In order to contribute to the debate over community versus centre based treatment there is a need for agencies to begin calculating the proportion of met needs for all their TFPs in the future (as described in the SFP section above), i.e. arrive at an overview and begin defining the importance of context specific factors in determining impact.

What can we learn from the grey literature?
An analysis of the grey literature on TFPs could provide the answers to a number of important questions about the cost-effectiveness of different types of TFPs in different contexts. The database would need to contain the same type of information that is shown for SFPs in section 4.2.2.

A deeper analysis of the grey literature information which is generated from TFPs could be refined using the Prudhon index. This index allows a researcher to adjust for the beneficiaries' initial level of malnutrition at admission. It is important to adjust for this when comparing the efficacy of treatment protocols, because children who are more severely malnourished have a higher risk of mortality whatever the protocol (Prudhon et al, 1997). However, this type of analysis requires information on individual children enrolled in the various programmes. It is not known how many agencies keep this detailed programme information after the programmes have closed down.

What types of impact evaluations should we be undertaking in the future?
The types of evaluations described for SFPs would also be suitable for TFPs. As yet,unpublished cohort studies comparing the effectiveness of home-based and centre-based programmes have already been undertaken by some agencies (e.g. ACF in Sierra Leone and MSF and Concern in Malawi). However, several agencies have also undertaken RCTs of TFPs (ACF, and University of Dijop). These types of programmes lend themselves more easily to RCT studies because the care is focused at the individual level, rather than the community. We encourage researchers to continue with these types of studies because they provide good evidence of impact and, in the long term, should lead to better programming.

Information on the relative costs of TFPs is still extremely limited. We reported above on the one published study that examines the costs of centre-based TFPs in different contexts (Hallam et al, in press). More cost-effective evaluations are required in the future.

Summary and recommendations

- The current main issue over TFP impact relates to comparison of community versus centre based care of the severely malnourished.
- There are also certain highly technical areas of uncertainty, e.g. different antibiotic regimes/malaria treatments.

- The grey literature (agency programme reports) should be utilised to generate more information about the cost-effectiveness of TFPs overall, and the importance of context specific factors in determining impact.
- The Proudhon index should be employed where possible to deepen this analysis of the grey literature.
- A number of RCTs have already been undertaken of TFPs in emergency settings. These should continue to be implemented in different contexts.
- There is an urgent need to fill the information gap regarding overall costs and cost components of TFPs.

5.2.4 Vitamin A supplementation programmes

The evidence base for the efficacy of vitamin A in decreasing mortality is strong and several meta-analyses of RCTs are available. There is currently insufficient information to fully understand the impact of vitamin A supplementation on the growth of children. However, given that we know that vitamin A supplementation reduces mortality, it may be unethical to continue further trials to assess the impact on growth. It might, however, be useful to undertake a meta-analysis of the all the available data on vitamin A supplementation and growth. Ideally, this review would categorise children into different levels of serum deficiency at baseline. There are several studies looking at the cost-effectiveness of vitamin A distributions in routine settings.

As far as we know, there is no reason to believe that the efficacy of vitamin A would be reduced in an emergency. In fact, as most emergency affected populations have a poor nutritional status and may also be at heightened risk of measles (particularly if the population is displaced), vitamin A supplementation may have an even more beneficial effect. However, cost-effectiveness figures on supplementation in emergencies will probably be very different from those in a development setting where the health-infrastructure is already in place.

What types of impact evaluations should we be undertaking in the future?
We suggest that vitamin A impact monitoring is not necessary. Instead, information on process indicators (basically coverage) and costs in different settings are required.

5.2.5 Bednet programmes

There is good quality information available on the impact and cost-effectiveness of ITNs on mortality in children in development settings. There has also been one study assessing the cost-effectiveness of such a treatment in an emergency (Afghanistan). The information about the impact of bednets on the nutritional status of malnourished children is less convincing. However, we believe that an RCT assessing the impact of ITNs on the nutritional status of malnourished children would be unethical, as ITNs have already been shown to decrease mortality.

It is not clear how transferable the findings of the studies in development settings are to emergency settings. In particular, displaced populations may not have access to shelter, in which case, it would be difficult for them to hang their nets. It may also be harder to provide the education which is needed with the nets when people are displaced. Further research on these topics is required.

The costs of the bednet programmes would, again, be different in emergencies than in a development setting. For example, if people are highly mobile then they may not prioritise carrying their bednets with them. In this situation, an agency might have to supply more than one net per person in a relatively short time period.

What types of impact evaluations should we be undertaking in the future?

Given the information above, we suggest that bednet impact monitoring would be useful in different emergency settings. In particular, cost-effective studies like the one in Afghanistan (Rowland et al, 1999) would be very helpful.

5.2.6 Measles immunisation programmes

The efficacy of measles immunisations in preventing mortality is well known. The relationship between measles immunisation and malnutrition is less clear-cut but it seems obvious that by reducing measles morbidity and the diarrhoea which accompanies measles, the vaccine will prevent weight loss. Cost-effectiveness data for measles, both by campaign and routine EPI programme, are available in development settings. None of this information is available for emergency settings.

What types of impact evaluations should we be undertaking in the future?
It would be unethical and unnecessary to conduct any kind of trial assessing the efficacy of measles at this point. However, information on costs and coverage of programmes using different implementation models (campaign/house to house vaccination, etc) in different contexts could be useful.

Summary and recommendations

- There is no need to conduct efficacy assessments of either vitamin A or measles immunisation programmes.
- Information on the cost and coverage of different types of vitamin A and measles immunisation programnmes in a variety of contexts would be useful for future planning.
- Information on the effectiveness of bednets in different emergencies should be collected and collated in such a way as to help decision makers make informed choices about when and where such programmes would be useful.

5.2.7 Economic evaluations

There is virtually no published information on the relative costs of the different interventions in emergencies. Cost information can serve as a critical input into the processes of setting priorities and the efficient allocation of resources. Ideally, planners would have access to cost-effectiveness information comparable across a range of strategies. Eventually, we need models in which local rapid-assessment data can be 'plugged in' (Griekspoor et al, 1999). In the meantime, any cost information needs to be clearly specified and presented so that adjustments could be made for different economic, epidemiological and programme settings.

What can we get from the grey literature?
Information about the costs of the interventions described in this report almost certainly exists in project reports held by agencies. Efforts should be made to unearth and collate all available data. A collection of the existing cost information, specific to local epidemiological and economic conditions and health system capabilities and constraints, would inform national and regional policy-makers interested in making best use of constrained resources. In turn, donors' responsibilities will be clearer when local decision-makers understand the inevitable resource allocation choices to be made. Ideally, a databank of this information would be developed (as suggested by Griekspoor et al, 1999) and made available to researchers, and local and international policy-makers.

What types of economic evaluations should we be undertaking in the future?
The methods applied to estimate costs in the studies reviewed in section 3 give rise to questions of reliability, validity and transparency[29]. One of the problems of reviewing

[29] These comments are based on a review of all the emergency papers identified.

historical cost records held by agencies is that the information may not be organised in a way that is convenient for cost analysis. For example, Hallam (1996) notes that:

- Many agencies do not record their expenditure by activity or beneficiary group, particularly those working on multi-sectoral relief programmes. Their accounts are primarily produced for auditing requirements, and not to allow for detailed cost-effectiveness analysis of projects.
- Where expenditure is classified by activity, no standard classifications are used, so the same project can be recorded in different ways by the NGO, UN agency and donor involved.
- Complex sub-contracting takes place between and among bilateral agencies, UN organisations and northern and southern NGO partners, involving the transfer of human and financial resources, as well as aid-in-kind. It becomes difficult to trace the flow of funds, let alone work out end-use of resources.

Given these kinds of issues, it may be that we will only be able to get good quality economic information from prospective studies. Perhaps the minimum we can recommend is the development of a common reporting format, which is explicit and transparent, e.g. disaggregated reporting of price and quantity data. This process may need to be driven by donor agencies, i.e. donors request that their implementing partners start to report on costs in a standardised manner.

It is important to recognise that costing studies are driven by local information needs and circumstances, and thus may not be designed to measure the same information. The widely varying purposes that underlie costing exercises can explain, in part, the variability in methodologies used and inputs costed. For example, programme managers/decision makers may require cost information to analyse on-going costs to identify potential cost savings and to improve the efficiency of the service, to obtain an accurate estimate of the budget necessary to maintain it, toprovide information on the total costs of the intervention with a view to replication, and to to examine the relative cost-effectiveness of alternative ways of delivering a particular intervention or relative to each other.

Useful cost information could probably be obtained by increasing the use of modelling as a predictive tool (Walker et al, 2000). Economic evaluations can be expensive, and a potentially cost-effective way of proceeding is to develop models to predict the impact of health care interventions in developing countries (Goodman et al. 1999). Models can be used to identify meaningful gaps in data and hence guide future research and development.

In the early stages of an emergency operation, the heavy demands upon relief personnel and organisational structures preclude adequate data collection. Furthermore, the lack of counterfactuals means that it is impossible to know how many lives may have been saved by the relief operation and thus it is impossible to generate even highly approximate estimates of the cost per life saved. These situations present themselves as ideal opportunities to apply modelling techniques. They can also be performed in an *ex ante* fashion, to estimate the cost-effectiveness of preparedness measures (Hallam, 1996). In fact, models have already been used to predict the cost-effectiveness of different TB and cholera interventions in emergency settings (Biot et al, 1999; Naficy et al, 1998). However, increased validation of models is imperative, and careful thought needs to be given to the processes and judgements involved in this.

When undertaking economic evaluations in the future, there is a need to consider how to take account of wider costs such as opportunity costs to beneficiaries. If we only cost out the providers' costs, we will encourage shifting of the expense of programmes on to some of the world's poorest people. Humanitarian aid managers are often guilty of only including the direct costs of their programmes in their analysis of cost effectiveness, and ignoring other costs borne by the communities being helped. Hallam (1996) provides the example of

Tanzania and Zaire, where the lack of provision of cooking fuel to refugees, along with the supply of slow-cooking beans and whole grain maize instead of flour, led to intense deforestation around the camps. This cost, which is a direct result of inappropriate assistance and inadequate account being taken of the needs of the refugee population, was being borne almost solely by the host communities in Tanzania and Zaire. Caldwell and Hallam (in press) provide the example of traditional TFC programmes, which incur significant costs for the families and communities of programme beneficiaries, that needs to be included in any comparative analysis of CTC and the TFC approach. In the latter, mothers are removed from their families for up to a month, in order to stay with the child in the TFC. Siblings of the malnourished child are deprived of maternal care for this period. Furthermore, the mother is unavailable to work in the fields or participate in other income-generating activities during this time. All of this imposes a significant opportunity cost on the family and community – a cost that is largely avoided in the CTC model.

Finally, even with consistent identification, measurement and valuation of costs, the results of the economic evaluations described above will generally not be comparable because of the lack of a common outcome measure. For example, although it may be possible to work out whether a home or centre based programme is a more cost-effective intervention to cure severely malnourished children, there is still no method to help decision makers know whether or not a TFP or a measles immunisation programme would save more lives per dollar. There are 'uncomfortable' ethical questions around this type of consideration. To aid decision-making at national and international levels, measuring the effectiveness of interventions necessitates the inclusion of final outcome measures, e.g. disability-adjusted life-years (see annex 2 for more information on DALYs).

Summary and recommendations

- There is virtually no cost information available.
- Researchers should consider trying to create a database of cost-impact information from the grey literature available.
- In the future, donors should consider requesting all agencies to report costs of programmes in a transparent, standardised manner. This information could then be used to create a database.
- Researchers should consider using modelling to predict cost-effectiveness of different interventions in various contexts.
- Researchers need to think about how best to include beneficiaries' costs when undertaking economic evaluations in the future.
- If donors want to be able to compare the cost-effectiveness of different interventions, there is a need to develop final outcome measures, such as DALYs, for these programmes.

5.3 Institutional mechanisms for moving forward

5.3.1 Institutional accountability

As noted above in section 4.2, one of the reasons why there is so little published evidence about the impact and cost-effectiveness of these programmes is that there is no agency or group that is responsible (and therefore accountable) for ensuring the effectiveness of these types of programmes in emergencies.

Debates about accountability flourished in the 1990s, as humanitarian agencies faced growing levels of scrutiny and criticism, and the automatic assumption that humanitarian aid was a good thing began to be questioned. There was a realisation that aid alone, even when well

delivered, might have only negligible impacts in situations where other political, economic or social factors were far more important in determining humanitarian outcomes.

Political economy approaches highlighted a series of difficult dilemmas around the delivery of aid in conflict, such as the risks of aid being diverted to warring parties (Cliffe and Luckham, 2000; Le Billon, 2000). This focused attention on the possible negative and unintended consequences of humanitarian aid. Linked to this were a number of initiatives aimed at increasing accountability, such as the Code of Conduct, Sphere, ALNAP, the Ombudsman project, the Humanitarian Accountability Project (now the HumanitarianAccountability Partnership International), and People in Aid. These all have in common a concern for the quality, performance and accountability of humanitarian aid. Two elements are particularly relevant here. First, greater efforts and attention are being put into humanitarian evaluation (ALNAP, 2003b). Second, there has been a focus on the development of standards and indicators through the Sphere process (Sphere, 2004). Potentially, these provide benchmarks against which humanitarian impact can be measured, though critics argue that Sphere is overly focused on the technical aspects of aid delivery (Dufour et al, 2004). Several international NGOs have introduced impact assessment systems that aim to improve accountability at the organisational level. Action Aid's Accountability, Learning and Planning System and Save the Children UK's Global Impact Monitoring, stress both upwards and downwards accountability (British Agencies Aid Group, 2002; SC UK, 2003; Starling, 2003). UN agencies have also taken initiatives, for example WFP's introduction of rights based monitoring into project appraisal mechanisms.

5.3.2 Whose responsibility is it to analyse intervention effectiveness and cost?

The dearth of published information on impact and costs of emergency GFDs and SFPs is unacceptable. At a conceptual level, there are many reasons why emergency SFPs may routinely fail to have a nutritional impact. Indeed the 'infamous' review by Beaton and Ghassemmi in 1982) of SFPs in stable situations concluded that these had minimal nutritional impact. A similar overview and understanding of effectiveness of emergency SFPs is long overdue. While there is less controversy over GFDs, there are important efficacy and cost issues to resolve over programme design, e.g. community based targeting versus administrative targeting utilising ration cards, the effectiveness of CSB provision in preventing micronutrient deficiency outbreaks, GFDs versus cash transfers, etc. Furthermore, as GFDs take on increasingly complex roles in relation to the HIV pandemic, the need to fine tune programme design and assess whether multiple objectives are being met will grow. There are also design issues for TFPs. The questions are (i) how can we establish an institutional mechanism whereby the overall cost effectiveness of different types of intervention design can be assessed/monitored, and (ii) how we can feed this information back to donors and implementing agencies to effect change in the name of effective resource allocation and economic efficiency. The idea that we would accept an equivalent lack of scrutiny of impact and efficiency of resource allocation in the health sector in the developed world is unthinkable.

At a pragmatic level, it may have to be assumed that donors and implementing agencies have institutional and political vested interests and sensitivities around information on cost-effectiveness of interventions – especially where certain types of programme have been implemented and supported for decades and/or where agencies have mandates and technical expertise/capacity built around certain types of intervention. As a result, it is important to think through institutional mechanisms which are perceived as neutral and non-threatening to donors and implementing partners.

Establishing an impact and cost-effectiveness monitoring agency

One potential model for enhancing information and analysis on cost-effectiveness may be the establishment of a non-affiliated agency with a mandate to monitor programme effectiveness and cost. Such an agency could operate like the RNIS (although would, ideally, not be located within the UN system because the agencies may be too involved). The agency would thus receive regular information from implementing partners on programme outcomes and costs. Much of this information would not be in the published literature ('grey literature'), although as with the RNIS, agencies would need to be encouraged to standardise reporting as much as possible.

This 'non-affiliated' agency would also have responsibility for commissioning controlled or cohort studies to address particular questions. Because of the ethics of using control groups in humanitarian crises, these studies would inevitably mainly address issues of optimal design where 'we don't know the answer', although there may be opportunities (sometimes unplanned) for other types of controlled study (see section 5.2) to add to the body of knowledge regarding overall efficacy of a particular type of intervention. The agency could carry out these studies in conjunction with 'expert institutions' such as CDC, London School of Hygiene and Tropical Medicine, Aberdeen University, etc. Donors could be encouraged to commit a set budget for these types of study based on an agreed framework and code of ethics so that when opportunities come up, the agency can instigate the study as speedily as needed. Note that this type of prospective analysis of impact and cost effectiveness will require NGOs or other agencies to be fully prepared ahead of the 'next' emergency.

The agency charged with coordinating impact and cost-effectiveness work would take on responsibility for publishing study findings where appropriate, as well as producing regular updates on grey literature findings. Areas requiring further study would also be highlighted regularly. Most importantly, the agency would take on advocacy responsibility where necessary.

Another role for the agency would be to work closely with donors in order to standardise implementing agency reporting on costs. As with defining and analysing impact, this is a highly technical area which will require technical support from those with expertise in health economics and experience of humanitarian contexts.

Given the potential sensitivity that may exist around certain types of information, there will be a need to establish Memoranda of Understanding (MOUs) with participating agencies. These would govern the use of information passed onto the agency, the degree of anonymity around information dissemination, and inter-agency peer review of material published and disseminated by the agency.

The agency would function mainly as a co-ordination and advocacy body. The questions that it would address would undoubtedly change over time. It may be that establishing and gaining sufficient financial and political support and buy-in for such an agency would require that it initially focuses upon one or two areas, e.g. information on SFPs and community based targeting of GFD versus traditional household registration approach. Assuming that the agency is able to 'prove it's worth', it may be able to address an increasing number of programmatic questions within the humanitarian field, e.g. comparisons of interventions impact on nutrition across a range of sectors.

Donor support to university/research bodies for focussing on specific areas of programming

Another mechanism for filling the information gap, requiring less start up investment cost,may be for donors to tender for contracts in the research sector for working on impact and cost-effectiveness of specific types of programming and design. Each research body

would then establish an inter-agency steering group made up of those agencies that implement specific types of programme, e.g. GFDs, SFPs, etc. These agencies would then provide access to their grey literature, as well as be involved in actual impact studies.

Summary and Recommendations:

- The lack of an agency with responsibility for overseeing impact and cost-effectiveness of certain key humanitarian activities has resulted in insufficient impact and cost-effectiveness information to inform decisions about optimal programme choice and design.
- There is a need to establish an agency responsible for coordinating activities aimed at filling the information gap on impact and cost-effectiveness of humanitarian interventions.
- Such an agency should focus activities on key programming areas where substantial unresolved questions exist, e.g. Expanded GFD versus GFD plus SFP
- The agency would coordinate use of the grey literature and implementation of ethically acceptable impact studies.
- The agency would also establish frameworks with donors and collaborating implementing agencies regarding research protocols, anonymity of case findings and dissemination
- The agency would also undertake an advocacy role where appropriate.

Annex 1: Description of the different types of feeding programmes

A1.1 General feeding/ration distribution programmes (GRDs)

There are many different types of food security programmes in emergencies. For the purposes of this paper, we will examine a general ration programme which is targeted at the poorest sections of the population. Assume the objectives of this general ration programme will be to:

- Improve the population's food security
- Improve the population's nutritional status

A1.2 Supplementary feeding programmes (SFPs)

Supplementary feeding programmes typically have one or both of the following objectives:

- to prevent mortality amongst mild and moderately malnourished individuals
- to prevent increasing levels of malnutrition, or maintain nutritional status of vulnerable groups.

These different objectives correspond to the distinction between 'blanket supplementary feeding' directed at all vulnerable groups (children under five and pregnant and lactating women), and 'targeted supplementary feeding' directed at the moderately malnourished (those children whose nutritional status is between 70 and 80% WFH). Rations may be either dry ('take-home') or wet (to be eaten 'on-the-spot').

When assessing the impact of SFPs, it is important to distinguish between targeted and blanket programmes because their objectives are different. The targeted programmes aim to address moderate acute malnutrition in individuals which means there will be a decrease in the number of cases of moderate malnutrition (because they are cured). This does not necessarily mean that all the moderate malnutrition will disappear since the programme is only meant to cure people who become malnourished, not prevent new cases from developing. However, an effective targeted SFP should prevent any new cases of severe malnutrition from developing and should decrease mortality from moderate malnutrition.

The blanket programmes aim to maintain the nutritional health of vulnerable population-groups and thus has an objective to reduce the overall prevalence of moderate (and severe) malnutrition in the community.

In addition to the objectives described above, SFPs may also have other aims. Additional objectives might include the promotion of health and increased EPI coverage (Curdy, 1994). The impact of health objectives may be relatively easily assessed. However, in some situations, objectives may be less explicitly stated. For example, an SFP may aim to ensure food access in situations of conflict where general ration distributions may be targeted by combatants, or enhance household food security of refugee impacted households or of households supporting prisoners (Borrel, 1997). These types of objective rarely appear in agency emergency nutrition guidelines but may be stated in programme proposals submitted to donors. Such objectives are often country and population specific. If the objectives are not stated quantitatively, it is difficult to measure whether or not they have been achieved. In the examples given by Borrel above, some form of food security monitoring would need to be introduced to test whether programmes have achieved desired impact.

A1.3 Therapeutic Feeding Programmes (TFP)

A Therapeutic Feeding Programme's usual objective is:

- to reduce the mortality rate of severely malnourished individuals within the entire affected population, including children, adolescents, elderly and adults.

There are currently two types of TFP programme which are commonly implemented. The traditional centre-based programme and the newer home-based approach.

In a centre based programme ,treatment is conducted in a Therapeutic Feeding Centre (TFC). The beneficiaries receive specialised diets and medical treatment, as well as close individual follow-up. The treatment is divided into two distinct Phases. Phase I of a TFC usually comprises a 24-hour intensive care unit where medical complications are treated and where nutritional treatment (controlled energy and protein content) is started. This comprises of eight meals per day, over 24 hours, with feeding day and night. When the patient has passed the critical phase, they will be transferred into the Phase II section which is preferably a day-care unit, in service eight to nine hours a day, or a 24-hour unit. During Phase II, the patient will receive four to six meals per day of high energy content with nutritional and medical follow-up. Protocols for centre-based treatment programmes have been published by WHO (1999).

Currently, there is no internationally agreed standard protocol for a home based feeding programme. The programmes usually include a stabilisation phase and an outpatient phase. The stabilisation phase is for the treatment of severe malnutrition for children with complications: life-threatening problems are identified and treated, specific deficiencies are corrected, metabolic abnormalities are reversed and feeding with Ready to Use Therapeutic fFoods (RUTF) is begun. This is an inpatient phase and takes place in a stabilisation centre (SC) which may be located in a hospital or clinic. The outpatient therapeutic programme (OTP) provides specialised ready to use foods and simple medical protocols through existing health infrastructure to severely malnourished children in their own homes. The ENN has recently produced a report of a meeting about home-based treatment programmes, which contains more details about the different strategies employed by different agencies (ENN, 2003a).

Annex 2: Definition and explanation of terms commonly used in evaluation methodology

A2.1 Different types of control groups

Historical control group: the same target population. This approach entails a comparison of the change from before to after the programme, with an attempt to rule out external factors.

Internal control group: geographical areas (or individuals or institutions) that should have received the full intervention but didn't, either because they refused it, or could not be reached by the programme. Often the reception of a programme is variable. This means that some people will receive the intervention for a longer period of time, or will receive the programme more intensively. In this case, a dose-response relation may be observed. These approaches require the collection of cross-sectional data from different groups at the end of the programme. The case-control method is another example of use of an internal control group.

External control group: geographical areas (or individuals or institutions) areas without the programme. Comparison can be cross-sectional (intervention versus control at the end of the programme) or longitudinal-control (comparing intervention and control at the beginning and the end of the cycle).

A2.2 Economic evaluations

Brief overview of the different types of economic evaluations
Economic evaluation techniques include cost-minimisation analysis, cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis:

- Cost Minimisation Analysis (CMA): two or more interventions that have identical outcomes (e.g. number of cases treated) are assessed to see which provides the cheapest way of delivering the same outcome.
- Cost Effectiveness Analysis (CEA): measures the outcome of approaches in terms of 'natural units', e.g. for emergency interventions, this could be the number of cases of measles averted.
- Cost Utility Analysis (CUA): these evaluations use a measure of utility (reflecting people's preferences). The outcomes are then expressed in terms of measures such as quality-adjusted life-years (QALYs) or disability-adjusted life-years (DALYs).
- Cost Benefit Analysis (CBA): expresses outcomes (e.g. the number of lives saved) in terms of monetary units.

In such applications, health programmes are compared for their benefits and costs, where costs refer to the value of opportunities foregone from not employing resources elsewhere. Benefits are gauged by the consequences of a health programme on people's well-being or health status. The various evaluation techniques estimate costs in a similar fashion, but differ in the measurement of health outcomes.

The different ways of measuring benefits leads to a trade-off between the scope for potential use and the practicality of various evaluation techniques. Furthermore, the type(s) of outcome measures, and hence type of economic evaluation, used infers the purpose of the evaluation. First, there is technical efficiency, which is a narrow definition as it concentrates on maximising the achievement of a given objective within a given budget, e.g. what is the cheapest strategy for vaccinating children – fixed, outreach or mobile teams? In short,

technical efficiency is concerned with 'doing things right'. For this reason it is sometimes dubbed 'low level' efficiency. CEA, based on 'natural units', primarily addresses this type of efficiency, in which analysts compare strategies to meet a given goal.

Second, there is 'high level' efficiency defined as allocative efficiency. While all the different types of economic evaluations can be used to assess technical efficiency, this is not true for assessments of allocative efficiency. This describes the search for the optimal allocation of resources across a mix of programmes that cannot all be fully funded, to produce the greatest gain to society, i.e. allocative efficiency is concerned with 'doing the right things' (in the 'right' way it should be stressed). In this broader definition of efficiency, different health care interventions with different objectives and outcomes are compared, and comparisons across sectors can be achieved, e.g. malaria control versus immunisation, health versus education. For this reason, CUA, which uses more complex measures of outcomes, can be used to assess allocative efficiency within the health sector, in which comparisons between competing programmes are made, i.e. how should a Ministry of Health budget be distributed between different programmes? However, this approach is still restricted to comparisons of programmes within the health sector, i.e. quasi-allocative assessments.

In theory, CBA has the widest scope of all types of analyses because outcomes are monetised enabling inter-sectoral comparisons, i.e. how should a government budget be distributed between different ministries. In practice, the valuation of health benefits is difficult and preference for CEA over other types of evaluation for evaluating health care programmes has been emerging since the late 1970s, in both developed and developing countries (Warner and Hutton 1980; Walker and Fox-Rushby, 2000).

A brief explanation of Disability Adjusted Life Years - DALYs
The DALY is an indicator of the time lived with a disability and the time lost due to premature mortality. The duration of time lost due to premature mortality is calculated using standard expected years of life lost with model life-tables. The reduction in physical capacity due to morbidity is measured using disability weights. The value of time lived at different ages is calculated using an exponential function which reflects the dependence of the young and the elderly on the adults. Streams of time are discounted at 3 percent.

There are five components of disability-adjusted life years (DALYs):

1. Duration of time lost due to a death at each age – measured based on the potential limit for life, which has been set at 82.5 years for women and 80 years for men.

2. Disability weights – the degree of incapacity associated with various health conditions. Values range from 0 (perfect health) to 1 (death) with four other prescribed points along the interval in between, representing a set of accepted disability classes.

3. Age-weighting function = $Cxe^{-\beta x}$
Where:
- C is a constant = 0.16243
- B is a constant = 0.04
- x = age
- e is a constant = 2.71

Indicates the relative importance of healthy life at different ages.

4. Discounting function = $e^{-r(x-a)}$
Where:
- r is the discount rate, fixed at 0.03
- a = onset of disease year

- x = age
- e is a constant = 2.71

Indicates the value of health gains today compared to the value of health gains in the future.

5. Health is added across individuals – two people, each losing 10 DALYs, are treated as the same loss as one person losing 20 years.

Annex 3 Search terms and critical appraisal forms

A3.1 Search terms

A3.1.1 Complex emergency terms

1. context factors.mp.
2. complex emergenc$.mp.
3. drought.mp.
4. civil war.mp.
5. conflict.mp.
6. hostile environment.mp.
7. disaster relief.mp.
8. exp Relief Work/
9. disaster situation.mp.
10. exp DISASTERS/
11. humanitarian.mp.
12. food distribution.mp.
13. food security.mp.
14. food aid.mp.
15. food donation.mp.
16. food crisis.mp.
17. food availability.mp.
18. exp REFUGEES/
19. refugee camp.mp.
20. aid worker.mp.
21. human displacement.mp.
22. internal displacement.mp.
23. geographic isolation.mp.
24. nutrition emergenc$.mp.
25. humanitarian relief.mp.
26. political instability.mp.
27. economic instability.mp.
28. hunger season.mp.
29. exp Emergency Medical Services/

A3.1.2 Malnutrition terms

1. exp MALNUTRITION/
2. exp Nutrition Assessment/
3. exp Nutrition Policy/
4. Nutrition Disorders/
5. exp Child Nutrition Disorders/
6. nutritional support.tw.
7. nutritional status.tw.
8. malnutrition$.mp.
9. nutritional assessment$.mp.
10. protein deficienc$.mp.
11. nutritional disorder$.mp.
12. exp Infant Nutrition Disorders/
13. chronic energy deficiency.mp.
14. undernourished.mp.
15. undernutrition.mp.
16. nutritional management.mp.
17. nutritional treatment.mp.
18. nutritional treatment.mp.

19. exp Child Nutrition/
20. childhood nutrition.mp.
21. nutritional risk.mp.
22. famine.mp.
23. starving.mp.
24. malnourished.mp.
25. child nutrition.mp.

A3.1.3 Outcome terms

1. outcome indicator$.mp.
2. target weight.mp.
3. exp Infant Mortality/
4. exp MORTALITY/
5. mortality.mp.
6. mortality rate.mp.
7. fatality rate.mp.
8. exp MORBIDITY/
9. morbidity.mp.
10. length of stay.mp.
11. recorvery.mp.
12. discharge.mp.
13. readmission rate.mp.
14. exp Patient Transfer/
15. discharge criteria.mp.
16. defaulter.mp.
17. good appetite.mp.
18. impact indicator.mp.
19. exp "OUTCOME AND PROCESS ASSESSMENT (HEALTH CARE)"/
20. process indicator.mp.
21. program$ effectiveness.mp.
22. exp Program Evaluation/
23. program evaluation.mp.
24. wieght gain.mp.
25. weight.mp.

A3.1.4 GFD terms

1. Relief Work/ec, mt, st, sn [Economics, Methods, Standards, Statistics & Numerical Data]
2. relief food.mp.
3. food aid.mp.
4. general ration distribution.mp.
5. Starvation/mo, di, pc, ep, et, th [Mortality, Diagnosis, Prevention & Control, Epidemiology, Etiology, Therapy]
6. famine.mp.
7. food donation.mp.
8. feeding programme.mp.
9. Hunger/
10. food security.mp.

A3.1.5 Supplementary and therapeutic feeding terms

1. supplementary feeding program
2. selective feeding program
3. therapeutic feeding program
4. therapeutic feeding unit

5. therapeutic feeding center
6. community based rehabilitation
7. hospital based rehabilitation
8. feeding station
9. emergency feeding program
10. general food distribution
11. nutriset
12. oral rehydration salt
13. oral rehydration solution
14. oral rehydration formulation
15. rehydration solution
16. supplement
17. fortified milk
18. initial phase
19. rehabilitation phase
20. intravenous fluid
21. meal ready to eat
22. therapeutic food
23. mineral mix, vitamin mix
24. formula diet
25. nutritional intervention
26. high energy milk
27. antifungal and malnutrition
28. antibiotics and malnutrition
29. amoxycillin, ampicillin, benzylpenicillin cotrimoxazole, chloramphenicol, gentamicin

A3.1.6 Vitamin A terms

1. exp Vitamin A/
2. exp VITAMIN A DEFICIENCY/
3. vitamin a.mp.
4. vitamin a deficienc$.mp.
5. exp Retinol/
6. retinol.mp.

A3.1.7 Bed net terms

1. exp Malaria/
2. malaria.mp.
3. bed nets.mp.
4. exp Mosquito Control/
5. mosquito control.mp.
6. mosquito nets.mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
7. exp Insecticides/ or Insecticides, Botanical/ or exp Pyrethrins/ or exp Insect Vectors/
8. insecticides.mp.
9. pyrethrins.mp.
10. insect vectors.mp.
11. exp Malaria, Falciparum/ or exp Culicidae/ or exp Anopheles/
12. falciparum.mp.
13. culicidae.mp.
14. anopeles.mp.
15. exp "Bedding and Linens"/
16. (bedding and linen$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading]
17. Permethrin.mp.

A3.1.8 Measles terms

1. exp MEASLES/
2. exp IMMUNIZATION/
3. exp Measles Vaccine/
4. exp Immunization Programs/
5. immunization programs.mp.
6. Measles.mp.
7. measles treatment.mp.
8. measles immunisation.mp.
9. measles vaccine.mp.
10. morbilli.mp.
11. exp Measles virus/
12. measles virus.mp.

A.3.1.9 Cost effectiveness terms

1. cost*
2. cost* and benefit*,
3. cost* and effect*
4. cost* and utiliti*

A3.2 Critical Appraisal forms[30]

Trials (Experimental Studies)

Criteria for assessment	**Poor**	**Fair**	**Good**
Are the aims clearly stated?			
Was the study randomised? *(were there adequate approaches to sequence generation – computer-generated random nos or random number tables OR inadequate approaches to sequence generation – use of alternation, case record nos, birth dates or week days)*			
Was the treatment allocation concealed? *Adequate approaches: centralised/pharmacy-controlled randomisation, serially-numbered identical containers, onsite computer based system with randomisation sequence that is not readable until allocation, other approaches to prevent foreknowledge of the allocation sequence to clinicians/patients.* *Inadequate: use of alternation, case record numbers, birth dates or week days, open random numbers lists, serially numbered envelopes (even opaque envelopes)*			
Were groups similar at baseline in terms of prognostic factors?			
Were the eligibility criteria specified?			
Were outcome assessors blinded to the treatment allocation?			
Was the care provider blinded?			
Was the patient blinded?			
Aside from the intervention(s) being evaluated, were the participants treated identically in the different groups?			
Were all participants who entered the study accounted for at the end?			
Did the analyses include an intention to treat analysis?			
Were the results precise? *(Confidence Intervals or p-values reported)*			
Was clinical as well as statistical significance considered?			

[30] These are all adapted either from Crombie I. Pocket Guide to critical appraisal and CASP Appraisal Tools http://www.phru.org.uk/~casp/appraisa.htm, or from NHS CRD Report 4, Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews.

Systematic Review

Criteria for assessment	**Poor**	**Fair**	**Good**
Is the purpose of the study clearly stated?			
Were the search methods used to find evidence on the primary question stated?			
Was the search for evidence reasonable comprehensive? *(bibliographic databases, ref lists, unpublished work, non-English publications)*			
Was the selection criteria for studies reported? *(study type, participants, inclusion criteria)*			
Was bias in the selection of articles avoided? *(explicit selection criteria used, independent screening by at least 2 reviewers)*			
Were the criteria used for assessing the validity of the studies that were reviewed reported? *(quality assessment forms)*			
Was the quality of the included studies appraised? *(validity criteria, more than 1 reviewer, independent screening)*			
If the results of the included studies have been combined, was it reasonable to do so? *(clinical heterogeneity among studies, reasons for variations discussed, results similar from study to study)*			
Were the main results of the review clearly reported? *(ORs, RR, numerical results, weighting of studies)*			
Were the results precise? *(Confidence Intervals)*			
Are the conclusions justified? *(conclusions consistent with results, they do not go beyond data, no evidence not interpreted as no effect, strength of recommendations for practice consistent with level of evidence, recommendations for research consistent with identified shortcomings)*			

:

Cohort study

Criteria for assessment	Poor	Medium	Fair
Clarity of study question and definition of outcome			
Is the purpose of the study clearly stated?			
Is there a clear definition of primary outcome(s)?			
Description of study sample			
Is the method of selection of the sample adequately described?			
Are the study exclusion and inclusion criteria stated?			
Is the baseline sample clearly described in terms of basic characteristics (age, sex, etc.)?			
Is the study sample sufficiently homogenous in terms of disease/diagnosis?			
Is the study sample sufficiently homogeneous in terms of co-morbidity?			
Control of bias in study design			
Are baseline values for groups compared?			
Has the study adequately controlled for confounding factors?			
Are the groups assembled at a similar point in their disease progression?			
Is the treatment/intervention reliably ascertained?			
Were the groups comparable on all important confounding factors?			
Was there adequate adjustment for the effects of these confounding variables?			
Was outcome assessment blind to exposure status?			
Was the proportion of the study sample followed-up adequate?			
Were drop-out rates and reasons for drop-out similar between study groups?			
Duration and completeness of follow-up			
Are reasons for loss of patients to follow-up stated?			
Are those lost to follow-up compared with the rest of the sample?			
Is there an appropriate length of follow-up?			
Statistical and analytical considerations			
Has the study sample size been justified?			
Are the data clearly presented?			
Was the data analyst masked to interventions?			
Has the type of statistical test and actual probability value been stated?			
Are statistical tests appropriate to the study?			
Have the data been analysed by intention to treat?			
Are the conclusions justified by evidence?			

Case-control study

Criteria for assessment	Poor	Medium	Good
Clarity of study question and definition of outcome			
Is the purpose of the study clearly stated?			
Is there a clear definition of primary outcome(s)?			
Description of study sample			
Is the method of selection of the sample adequately described?			
Are the study exclusion and inclusion criteria stated?			
Is the baseline sample clearly described in terms of basic characteristics (age, sex, etc.)?			
Is the study sample sufficiently homogenous in terms of disease/diagnosis?			
Is the study sample sufficiently homogeneous in terms of co-morbidity?			
Control of bias in study design			
Is the case definition explicit?			
Has the disease state of the cases been reliably assess and validated?			
Were the controls randomly selected from the source of population of the cases?			
Are the cases and controls comparable with respect to potential confounding factors?			
Were the interventions and other exposures assessed in the same way for cases and controls?			
Were the response rates and reasons for non-response the same in both groups?			
Is it possible that over-matching has occurred in that cases and controls were matched on factors related to exposure?			
Duration and completeness of follow-up			
Are reasons for loss of patients to follow-up stated?			
Are those lost to follow-up compared with the rest of the sample?			
Is there an appropriate length of follow-up?			
Statistical and analytical considerations			
Has the study sample size been justified?			
Are the data clearly presented?			
Was the data analyst masked to interventions?			
Has the type of statistical test and actual probability value been stated?			
Are statistical tests appropriate to the study?			
Have the data been analysed by intention to treat?			

Case-Series Study

Criteria for assessment	Poor	Fair	Good
Clarity of study question and definition of outcome			
Is the purpose of the study clearly stated?			
Is there a clear definition of primary outcome(s)?			
Description of study sample			
Is the method of selection of the sample adequately described?			
Are the study exclusion and inclusion criteria stated?			
Is the baseline sample clearly described in terms of basic characteristics (age, sex, etc..)?			
Is the study sample sufficiently homogenous in terms of disease/diagnosis?			
Is the study sample sufficiently homogeneous in terms of co-morbidity?			
Control of bias in study design			
Is the study based on a representative sample selected from a relevant population?			
Did all individuals enter the survey at a similar point in time?			
Were outcomes assessed using objective criteria or was blinding used?			
If comparisons of sub-series are being made, was there sufficient description of the series and the description of prognostic factors?			
Duration and completeness of follow-up			
Are reasons for loss of patients to follow-up stated?			
Are those lost to follow-up compared with the rest of the sample?			
Is there an appropriate length of follow-up?			
Statistical and analytical considerations			
Has the study sample size been justified?			
Are the data clearly presented?			
Was the data analyst masked to interventions?			
Has the type of statistical test and actual probability value been stated?			
Are statistical tests appropriate to the study?			
Have the data been analysed by intention to treat?			
Are the conclusions justified by evidence?			

Survey

Criteria for assessment	**Poor**	**Fair**	**Good**
Are the aims clearly stated? *Consider: what was goal of research, why important, its relevance*			
Is the design appropriate to the stated objectives? *Consider: if researcher has justified research design (have they discussed how they decided which method to use?)*			
Was recruitment strategy appropriate to the aims of the research? *Consider: has researcher explained how participants selected; any explanations why the participants they selected were the most appropriate to provide access to the type of knowledge sought by the study; any discussions around recruitment (e.g. why some people chose not to take part)*			
Was the sample size justified?			
Are the measurements likely to be valid and reliable?			
Are the statistical methods described?			
Is there a suggestion of haste?			
Were the basic data adequately described?			
Do the number add up?			
Was the statistical significance assessed?			
Were the findings unexpected?			
Are the conclusions justified?			

Economic evaluation studies review sheet

Background details

1. Author(s)
2. Country or area studied
3. Study year
4. Type of programme, e.g. vitamin A, measles, etc.
5. Strategies examined, e.g. fixed, outreach or mobile delivery
6. Type of analysis, e.g. CMA, CEA, CUA, CBA or cost analysis
7. Base year for costs
8. Results

Technical details

1. Is the perspective clear?
2. Are quantities of resources reported separately from their unit costs?
3. Is a generic outcome used?
4. Are the sources of data clear?
5. Are the time frame and analytic horizon clear?
6. Has discounting been performed when appropriate?
7. Has incremental analysis been used when appropriate?
8. Has sensitivity analysis been performed? If so yes, which technique?
9. Has affordability of the intervention been discussed? If yes, how?
10. Has the generalisability of the results been discussed? If yes, how?

Annex 4 Summary of the published studies

Table A4.1 Summary of the published studies assessing the impact of general ration distribution programmes

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Impact
Quisumbing (2003)	Assess impact of food for work and free food distributions on nutritional status	Rural Ethiopia (1994-1997)	• Retrospective cohort study (compared children living in households that received food aid to those living in households that did not) • Good CASP score • Statistics presented	• Children aged 0-5 yrs • N=1,181 children	Food for work: significant increase in WHZ for children in low asset households Free food: significant increase in WHZ in children in high asset households *Note: some gender differentials recorded in impact*
Yamano et al (in press)	Assess the impact of food aid distributions on nutritional status	Rural Ethiopia (1995-1996)	• Retrospective cohort study using ecological data (compared children living in areas that received food aid compared to those living in areas that did not) • Good CASP score • Statistics presented	• Children aged 6-59 months • N=2,089	Food aid has a significant impact on children's HAZ. No reported impact on WHZ or WAZ.
Toole et al (1988)	Assess association between ration size, prevalence of malnutrition and CMR	Refugee camp in Eastern Thailand (1979-1980) and refugee camp in Eastern Sudan (1984-1985)	• Observational study (series of cross-sectional surveys and screenings) • Good CASP score • No statistics presented	• Children aged 0-5 years measured for nutrition surveys • CMR measures whole population • Thailand: sample size varied from 41-1658 children	As ration size increases, prevalence of malnutrition and CMR decreases.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Impact
				• Sudan: sample size varied from 222-2941 children	
Toole and Bhatia (1992)	Assess association between ration size, prevalence of malnutrition and CMR	Refugee camp in Ethiopia (1988-1990)	• Observational study (series of cross-sectional surveys) • Good CASP score • Statistics presented	• Children aged 0-5 years measured for nutrition surveys • CMR measures whole population • N= approximately 1,350 children per survey	When the GFD was inadequate, the prevalence of malnutrition and CMR increased. However, as the ration size was increased, the prevalence of malnutrition and CMR decreased.
ACC/SCN (1994)	Assess association between ration size and CMR	Refugee and IDP camps in sub-Saharan Africa (1992-1994)	• Observational study (cross-sectional surveys in different sections) • Medium CASP score • No statistics presented	• CMR measures whole population • N=48 assessments in the camps, no sample sizes given	Mean CMR decreases as ration size increases
Sadler (2001)	Assess impact of food aid and other nutrition programmes (SFP and TFP) on nutritional status of children	Rural Ethiopia (2000)	• Observational study (repeated cross-sectional surveys) • Good CASP score • No statistical tests presented, but 95% confidence intervals shown.	• Children aged 6-59 months • N= approximately 900 for each survey	The prevalence of both severe and moderate acute malnutrition reduced significantly during the programme. Malnutrition seemed to decrease more than in a neighbouring area where no NGO provided assistance. .
Kemmer et al (2004)	Assess association between anaemia (Hb<110g/l) and iron deficiency (ZPP/H>80µmol/mol)	Burmese refugee population living in camps in Thailand	• Observational study (cross-sectional survey) • CASP score • Statistics presented	• Children aged 6-59 months • N= 857	Children living in households who reported that their ration would not last until the next distribution had a higher risk of being anaemic (but not iron

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Impact
	and adequacy of ration in a refugee population.				deficient) than children living in households which reported that the ration would last.
Wolde-Gebriel et al (1993)	Assess the prevalence of vitamin A deficiency in a village which had been partially dependent on food aid for 6 years.	Rural population in Ethiopia (1989)	• Observational study (cross-sectional survey) • Good CASP score • Statistics presented	• Children aged 0-150 months • N=240	Approximately 50% of children examined had one or more signs of vitamin A deficiency, 7% had Bitot's spots and 29% reported night blindness.
Warrack-Goldman et al (1986)	Assess association between the length of time since the last food distribution and nutritional status of children	Rural Mauritania (1983)	• Observational study (cross-sectional survey) • Medium CASP score • No statistics presented	• Children aged 6-59 months • N= 1,498	No association seen between children's nutritional status and the length of time since the household last received a food distribution

Table A4.2 Summary of the published studies assessing the impact of supplementary feeding programmes (SFPs)

Note that * is given next to the CASP score if there is a suspicion that the grading is lower than the actual quality of the study. This might occur when insufficient information is given about the study design but the authors of this report think that the assessment was probably well designed.

Author of study	**Objective of study**	**Setting**	**Study design, CASP score grade*, Statistics**	**Survey population (Age, size)**	**Impact / outcome**
Stefanak and Jarjoura (1989)	Assess weight gains in wet and dry supplementary feeding programmes	Resident population in Chad (1986)	*Impact on enrolled children* • Cohort study (different groups of children measured on a repeated basis throughout the study) • Good CASP score • Statistical tests presented	• Moderately acutely malnourished (>=70% WFH <80%) children 65-130 cm in height • N=774	Children enrolled on both the wet and dry feeding programmes improved their weights whilst enrolled on the programme[31].
Taylor (1983)	Assess impact of targeted supplementary feeding programmes	Refugees in camps in Somalia (1981)	*Impact on children enrolled* • Observational study (case-series data of children measured repeatedly over time) • Medium CASP score • No statistical tests presented *Coverage study* • Observational studies (cross-sectional cluster survey) • Medium CASP score • No statistical tests presented	• Moderately acutely malnourished children (<80% WFH >=70%) aged 6-59 months • N=622 • Children <=110cm • N=495	Substantial mean improvement in WFH% seen in children registered on SFP. Between 52-67% of eligible children were enrolled on the SFP.
Sadler (2001)	Assess impact of a targeted supplementary feeding programme on the nutritional status of	Rural Ethiopia (2000)	*Impact on children enrolled* • Observational study (case series data: children measured repeatedly over time) • Medium CASP score • No statistical tests presented *Impact on population's nutritional status*	• Moderately acutely malnourished children (assume <80% WFH >=70%) aged 6-59 months. • N= 5,407 children	More than 70% of all children who exited from the programme recovered (i.e. their WFH increased).

[31] The authors of this study also state that they compared the weight gains of 60 Chadian children 'not enrolled' in the programme. The children enrolled in the programme had a significantly higher weight gain than those not enrolled. However, the authors do describe any of the characteristics of the un-enrolled children (e.g. age, nutritional status, area of residence, etc) and hence it is not possible to know whether or not the two groups are comparable.

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
	children.		• Observational study (repeated cross-sectional two-stage cluster surveys) • Good CASP score • No statistical tests shown, but 95% confidence intervals presented.	• Children aged 6-59 months • N= approx. 900 for each survey	Significant decrease in population prevalence of malnutrition.
Vautier et al (1999a)	Assess outcome indicators of several targeted supplementary feeding programmes	Rural Liberia (1993-94) Returnee populations in Burundi (1994) Refugees in Goma, DRC (1994-95)	*Impact on children enrolled* • Observational studies (case series data: children measured repeatedly over time) • Medium CASP score • No statistical tests presented *Coverage* • Observational studies (figures from attendance at SFP combined with expected numbers of malnourished as predicted in earlier cross-sectional cluster surveys). • Poor CASP score* • No statistical tests presented	• Moderately acutely malnourished children (<80% WFH >=70%) with height <=130 cm (Liberia and DRC) and height <110cm (Burundi) • N=12,259 Liberia, N=9,197 Burundi, N=18,767 Goma • Moderately acutely children aged 6-59 months • Assume N=approx. 900 for each survey	Recovery rates of children who exited from the programme (i.e.: children's whose WFH increased): Liberia – 81% Burundi – 67% Goma – 79% Coverage for each area based on calculation from earlier survey result: Liberia – 70% Burundi – 30% Goma – 94%
Vautier (1999b)	Assess outcome indicators of a targeted supplementary feeding programme	Pastoralist population in Wadjir, Kenya (1998)	*Impact on children enrolled* • Observational study (case series data: children measured repeatedly over time) • Medium CASP score • No statistical tests presented *Impact on population's nutritional status* • Observational study (repeated cross-sectional two-stage cluster surveys) • Poor CASP score* • No statistical tests presented	• Moderately acutely malnourished children (assume <80% WFH >=70%), age not given (assume 6-59 months) • N=1,186 • Assume children aged 6-59 months • Assume N=approx. 900	79% of children who exited from the programme recovered (i.e.their WFH increased). Decrease in prevalence of malnutrition observed, but general ration distribution ongoing.

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
			Coverage • Observational study (figures from attendance at SFP combined with expected numbers of malnourished as predicted in an earlier cross-sectional cluster survey). • Poor CASP score* • No statistical tests presented	• Moderately acutely malnourished children, assume aged 6-59 months • Assume N=approx. 900	After the first three weeks coverage only 40%. Increased later during the programme (figures not given).
Vasquez-Garcia (1999)	Assess the outcome indicators of a targeted supplementary feeding programme	Pastoralist population in Mandera, Kenya (1996-98)	*Impact on children enrolled* • Observational study (case series data: children measured repeatedly over time) • Medium CASP score • No statistical tests presented *Coverage* • No information about methods • Poor CASP score* • No statistical tests presented	• Moderately acutely malnourished children (assume <80% WFH >=70%), less than 5 years old • N=11,250 • No information	Approximately 80% of the children who exited from the programme recovered (i.e. their WFH increased). Coverage estimated at 90%.
Toole and Bhatia (1992)	Assess association between prevalence of malnutrition and implementation of and SFP	Refugee camp in Ethiopia (1988-1990)	*Impact on population's nutritional status* • Observational study (series of cross-sectional surveys) • Good CASP score • Statistics presented *Coverage* • Observational study (series of cross-	• Children aged 0-5 years measured for nutrition surveys • N= approximately 1,350 children per survey • Children aged 0-5 years	The prevalence of malnutrition declined after the introduction of a comprehensive SFP. Coverage gradually increased

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
			sectional surveys) • Poor CASP score • No statistics presented	measured for nutrition surveys • N= approximately 1,350 children per survey	from 65 to 90%.
Munro (2002)	Assess a blanket supplementary feeding programme	Rural population in Zimbabwe (1992-3 and 1995-6)	*Coverage* • Observational study (series of cross-sectional nation-wide surveys) • Good CASP score • Statistics presented *Impact on population's nutritional status* • Observational study (series of cross-sectional nation-wide surveys) • Medium CASP score • Statistics presented	• Children aged less than 5 years old • N 3,758 • Children less than 5 years old • N not known	Programme did not reach the whole country. Average coverage of under-fives in programme areas in 1995: 50%. This figured varied by time and place. Prevalence of low MUAC similar in project and non-project areas. Unclear impact on nutritional status.
Roesel (1988)	Assess impact of a blanket supplementary feeding programme	Refugee camp population in Thailand (1983-1986)	*Impact on children enrolled* • Observational data (case-series data: children measured repeatedly over time) • Poor CASP score • Statistics presented *Impact on population's nutritional status* • Observational study (monthly monitoring of children's weight and height) • Poor CASP score* • No statistical tests presented	• All children aged 6-36 months old • Target population = 14,800 (actual figures not presented) • Children aged 6-36 months old • Camp population varied from 20-60,000. Assume children 15% of this?	By 1986, 74% of all children under three years who were registered in the programme gained weight each month. Prevalence of acute malnutrition decreased during the programme.
Barnabas et al (1982)	Assess weight gain of children in a	Refugees in Sudan (1981)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children)	• Malnourished children (definition of malnutrition	An average of 22% of children had gained weight since the

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
	supplementary feeding programme		• Poor CASP score • No statistics presented	and age range not presented) • N=149	previous distribution.
Desenclos et al (1989)	Assess prevalence of scurvy in children attending an SFP	Refugees in Somalia (1985-1986)	*Impact on children enrolled* • Observational study (cross-sectional survey of children) • Good CASP score • Statistics presented	• Children aged less than 5 years old • N=583	Children enrolled in the SFP were as likely to suffer from scurvy as those not enrolled.
Gibb (1985)	Assess outcome indicators of children enrolled in a supplementary feeding programme	Refugee camps in Sudan (1985)	*Impact on children enrolled* • Observational study (case-series data: repeated measurements of children over time) • Medium CASP score • No statistics presented *Coverage* • No description of how data collected • Poor CASP score • No statistics presented	• Malnourished children (definitions depended on camp, between <85% and >=70% WFH) aged less than 10 years • N=853 • No information	In 10-12 weeks, an average of 41% of children were discharged over 85% WFH. Attendance rates were reported to range from 75-94%.
Dzumhur et al (1995)	Assess growth of children enrolled in a supplementary feeding programme	Resident population in Sarajevo during the war (1995)	*Impact on children enrolled* • Observational study (case-series data: repeated measurements of children over time) • Medium CASP score • No statistics presented	• Malnourished children who were not sick (<10 percentile of WFA) • Children aged 1-14 years • N=633	58.3% of underweight children (who were not sick) were discharged above 10th percentile weight for age.
Henderson and Biellik (1980)	Assess impact of a supplementary feeding programme	Refugee camps in Somalia (1980)	*Impact on population's nutritional status* • Observational study (repeated cross-sectional surveys) • Medium CASP score* • No statistics presented	• Children aged <5 years • No sample size given	The prevalence of malnutrition amongst children and infants declined in all but one camp, in some cases decline was "dramatic".
Brown (1998)	Assess outcomes of a supplementary	Resident population in Kenya	*Impact on children enrolled* • Observational study (case series data: children measured repeatedly over time)	• Moderately acutely malnourished children	56% of children who exited from the programme recovered,

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
	feeding programme	(1997)	• Medium CASP score • No statistical tests presented	(<80% WFH >=70%), age not given (assume 6-59 months) • N=1,400 (?)	i.e.:their WFH increased. (High default rate)
			Impact on population's nutritional status • Observational study (repeated screenings of children) • Medium CASP score • No statistical tests presented	• Assume children aged 6-59 months • Sample size ranged from 898 to 2,596	Large decrease in prevalence of malnutrition as measured by MUAC.
			Coverage • No information on how coverage was calculated was provided. • Poor CASP score* • No statistical tests presented	• Moderately acutely malnourished children, aged 6-59 months	Coverage estimated at 76%.

Table A4.3: Summary of published studies assessing the impact of therapeutic feeding programmes (TFPs)

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
ENN (2003b)[32] ACF	Assess individual children's outcome of phase II treatment at home and in a centre	Rural Sierra Leone (2003)	*Impact on children enrolled* • Randomised trial • Statistical tests presented	• Severely acutely malnourished children <70% WFH without oedema who had already passed through phase I treatment, aged 12-60 months • N= 95	90% of children in the TFC and 95% of children based at home recovered, i.e. reached >=85% WFH and no oedema.
ENN (2003b) Cheik Anta Diop University	Assess efficacy of phase II home based nutritional rehabilitation with imported or locally produced ready to use foods (RUTF)	Dakar, Senegal (2003)	*Impact on children enrolled* • Randomised trial • Statistical tests presented	• Severely acutely malnourished children (<70% WFH without oedema) who had already passed through phase I treatment, aged 6-59 months • N= 66	72% of children given local RUTF and 71% of children given imported RUTF recovered, i.e. reached >=85% WFH and no oedema.
Brewster et al (1997)	Assess outcomes of different types of centre-based therapeutic feeding protocols	Malawi – 2 central hospitals, 2 district hospitals, 3 rural clinics (1995)	*Impact on children enrolled* • Trial (prospectively followed the outcome of children undergoing different treatment routines) • Medium CASP score • Statistical tests presented	• Children with kwashiorkor (no age range given, mean age=28 months) • N=1,625	Between 64-77% of children gained weight during treatment, depending on the protocol.

[32] None of the ENN (2003) references have been assessed for CASP score because the original research papers are not available to the author of this report.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
Perra and Costello (1995)	Compare mortality and nutrition outcomes of severely malnourished children enrolled in a nutrition rehabilitation centre to those who are not.	Rural Guinea Bissau (1988-1991)	*Impact on children enrolled* • Cohort study (compared children who were rehabilitated to those who were not; semi-randomised to control/non-control group). • Good CASP score • Statistical tests presented *Coverage* • Observational study ?? (records from screenings) • **** CASP score • No statistical tests presented	• Severely malnourished children 6-47 months (<=60% WFA) • N =1,038 • Severely malnourished children 6-47 months (<=60% WFA) • N=1,038	Relative risk of mortality in the rehabilitated group 0.75 (equivalent to 25% reduction in mortality) after 36 months. Rehabilitated children had higher mean weight gain in first three months. Difference in weight gain significant for 18 months. Coverage estimated at 33%.
Collins and Sadler (2002)	Assess outcome indicators of an outpatient programme for severely malnourished children	Home based setting in rural Ethiopia (2000-2001)	*Impact on children enrolled* • Retrospective cohort study (compared recovery rates of groups of children with different types of severe malnutrition) • Good CASP score • Statistical tests presented	• Severely malnourished children (<70% WHM and/or oedema) • Aged 6-120 months • N= 170	85 % of children recovered (i.e.: increased their WHZ). Children with marasmic-kwashiorkor had higher mortality rates than children enrolled with marasmus or kwashiorkor.
Prudhon et al (1997)	Assess outcome indicators of an inpatient programme for severely malnourished children	Children in 18 feeding centres in 9 different African countries (1993-95)	*Impact on children enrolled* • Observational study (prospectively recorded children's progress in different centres) • Moderate CASP score • Statistical tests presented •	• Severely malnourished children (<70% WFM and/or oedema or MUAC <110mm) • Aged 6-59 months • N= 3,021	91% of the children recovered (i.e. reached 80% or 85% WHM - depending on the centre).
Prag and Helemersgaard (1982)	Assess effectiveness of a	Refugee camps in Thailand	*Impact on children enrolled* • Observational study (case	• Severely malnourished	80% of children enrolled on the

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
	hospital based intensive feeding programme.	(1979)	series: repeated measurements of the same children) • Good CASP score • No statistical tests presented	children (<70% WFH and/or oedema) • N= 370	programme recovered (i.e. gained sufficient weight and/or reduced oedema).
Sadler (2001)	Assess impact of a centre-based therapeutic feeding programme on nutritional status of children.	Rural Ethiopia (2000)	*Impact on children enrolled* • Observational study (case series: repeated measurements of the same children) • Medium CASP score • No statistical tests presented	• Severely acutely malnourished children (assume <70% WFH and/or oedema) aged 6-59 months. • N= 874 children	More than 90% of all children who exited from the programme recovered (i.e. their nutritional status improved).
ENN (2003a) MSF-Belgium	Assess outcome of home based therapeutic feeding programme	Rural population of Afghanistan (2002-2003)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time) *Coverage* • Observational data: (cross-sectional survey, but note that survey and programme areas not identical). •	• Severely acutely malnourished children (WFH <70% and/or oedema and/or MUAC <110 or WFH<80% and acute medical problem). No information on age range. • N = 756 • Assume 6-59 months • Assume N=900	63% of children recovered – i.e. they reached WFH >=80% and no oedema. Coverage estimated at 17%
ENN (2003; updated 2004)	Assess outcome of home based	Rural population of South Wollo,	*Impact on children enrolled* • Observational study (case	• Severely acutely	75% of children recovered – i.e.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
Concern/VALID	therapeutic feeding programme	Ethiopia (2003)	series data: repeated measurements of children over time) *Coverage* • Observational data: (active case finding method)	malnourished children (WFH <70% and/or oedema ++ and/or MUAC <110 or aged more than 6 months and <4 kg and no medical complications). No information on age range. • N = 794 • Assume 6-59 months	they reached WFH>85% and no medical complications. Coverage estimated at 78%
ENN (2003b) SCF/VALID	Assess outcome of home based therapeutic feeding programme	Rural population of North Sudan (2001)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time) *Coverage* • Observational study (figures from attendance at SFP combined with expected numbers of malnourished as predicted in an earlier cross-sectional cluster survey)	• Severely acutely malnourished children (WFH <70% with no complications), 6-59 months. • N = 836 • Assume 6-59 months • Assume N=900	81% of children recovered – i.e. they reached WFH >=75%. Coverage estimated at 32%.
ENN (2003; updated 2004))	Assess outcome of home based	Rural population of Southern Sudan	*Impact on children enrolled* • Observational study (case	• Severely acutely	73% of children recovered – i.e.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
Concern/VALID	therapeutic feeding programme		series data: repeated measurements of children over time)	malnourished children (WFH <70% and/or oedema+ and/or MUAC <110 or more than 6 months and <4 kg and no medical complications). No information on age range. • N = 610	they reached WFH >=85% and no medical complications.
ENN (2003a) MSF-France	Assess outcome of home based therapeutic feeding programme	Rural population of Niger (2001-2003)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time) *Coverage* • Observational data: (cross-sectional survey).	• Severely acutely malnourished children (WFH <- 3 z-scores and/or oedema and no medical complications). More than 12 months old. • N = 7,597 • Assume 6-59 months • Assume N=900	65% of children recovered – i.e. they reached WFH >-2 z-scores and no oedema or medical complications. Coverage estimated at 53%
ENN (2003a) Queen Elizabeth Hospital	Assess outcome of home based therapeutic feeding programme	Blantyre mixed urban/rural population in Malawi (2003)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time) *Coverage*	• Severely acutely malnourished children (WFH <70% and/or oedema and no medical complications) who have passed through phase I treatment in a centre. No information on age range. • N = 316	50% of children recovered – i.e. they reached WFH>85% and no medical complications.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
			• Observational study (figures from attendance at SFP combined with expected numbers of malnourished as predicted in an earlier cross-sectional cluster survey).	• Assume 6-59 months	Coverage estimated at 65%
ENN (2003a) College of Medicine, University of Malawi	Assess outcome of home based therapeutic feeding programme	Rural population of Southern Region of Malawi (2002-2003)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time) *Coverage* • Observational study (figures from attendance at SFP combined with expected numbers of malnourished as predicted in an earlier cross-sectional cluster survey)	• Severely acutely malnourished children (WFH <- 2 z-scores and/or oedema and no medical complication). No age information. • N = 458 • Assume 6-59 months • Assume N=900	74% of children recovered, i.e. they reached WFH >-2 z-scores and no oedema. Coverage estimated at 66%
ENN (2003a); Concern/VALID	Assess outcome of home based therapeutic feeding programme	Rural population in Dowa, Malawi (2003)	*Impact on children enrolled* • Observational study (case series data: repeated measurements of children over time)	• Severely acutely malnourished children (WFH <70% and/or oedema and/or MUAC <110mm with no medical complications). No information on age range. • N = 1,900	86% of children recovered, i.e. they reached WFH >85% and no oedema.

Author of study	Objective of study	Setting	Study design, CASP score grade, Statistics	Survey population (Age, size)	Outcome/ Impact
			Coverage • Observational study (active case finding method)	• Assume 6-59 months	Coverage estimated at 73%

Table A4.4: Summary of published studies assessing the impact of bednet programmes

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
Cochrane group (2004)	Assess the impact of insecticide-treated bed nets or curtains on mortality	Burkina Faso, The Gambia, Ghana and Kenya (2)	• Meta analysis of Randomised Controlled Trial (RCTs) • High CASP score • Statistical tests appropriate and well presented	• Children (no age group specified)	• ITNs significantly reduce the mortality and morbidity of children • ITNs provided a 17% Protective Efficacy (PE) compared to no nets • ITNs provided a 23% PE compared to untreated nets
D'Assandro et al (1995)	Assessing the impact of ITN on childhood mortality and morbidity	Rural population in The Gambia	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• N total = 115,895 • 18,911 children aged 1-4 years • 21,191 children aged 5-9 years	• WAZ and WHZ scores are significantly higher in children from treated than untreated children. • 25% reduction in all cause mortality in children aged 1-9 years
Ter Kuile et al (2003)	Impact of premethrin-treated bed nets on growth, nutritional status and body composition of young children	Rural population in Western Kenya	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• 1,890 children aged less than 3 years old	• Use of ITNs is associated with significantly higher mean WAZ score and MUAC**. • The use of ITNs was not associated with higher WHZ or HAZ * *= Children 3-35 months **=Children 6-35 months
Snow et al (1997)	Examine the effects of insecticide-treated bed nets to reduce malnutrition in infants	Rural population of Kilifi district (coastal) Kenya	• RCT • Poor CASP score • Statistical tests appropriate and well presented	• 1481 infants • 787 infants aged between 1 and 11 months slept under ITN • 692 control infants	• Overall, WAZ and MUAC measures were significantly higher among infants who used ITN compared with control infants. However, the difference in nutritional status between the two groups was insignificant for seven of the 11 age groups.

Table A4.5: Summary of published studies assessing the impact of vitamin A supplementation programmes

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
Fawzi, 1993	Assess whether vitamin A supplementation significantly reduces mortality in children	Multi country trials including South Africa, Tanzania, Nepal, Sudan, India and Indonesia	• Meta analysis • High CASP score • Statistical tests appropriate and well presented	• All children	• Vitamin A supplementation significantly reduces child mortality rates.
Donnen, P et al (1998)	Assess the efficacy of a 6~mo high dose vitamin A supplementation programme on the growth of moderately malnourished preschool children	Rural population living in South Kivu Province, Democratic Republic of Congo	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• 385 moderately malnourished (definition not clear, children had been discharged from nutrition hospital previously but most did not have low WHZ at baseline) children • Aged 0-72 months • Children who had received a vitamin A supplementation in previous 4 months were excluded • Follow up 1 year	• Vitamin A supplementation appeared to improve growth (weight and MUAC) of those children who were severely vitamin A deficient at baseline. There was no impact on height for girls, but a small positive impact for boys. • No significant impact of supplementation on WAZ, HAZ or MUAC in children without vitamin A deficiency at baseline.
Rahman et al (2002)	Identify whether simultaneous zinc and vitamin A supplementation increases the growth rate in malnourished children	Population living in urban slums of Dhaka city, Bangladesh	• RCT • High CASP score • Statistical tests appropriate and well presented	• 653 mildly to moderately under nourished children (not clearly defined) • Age 12- 35 months • Children who had received a vitamin A supplementation in previous 4 months were excluded • Severely malnourished (<60% WFA or clinical signs of micronutrient malnutrition including vitamin A deficiency) children who needed medical intervention were excluded	• Zinc supplementation and single dose vitamin A supplementation, either alone or combined did not improve the growth (WHZ, WAZ or HAZ) of these malnourished children.

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
Ramakrishnan et al (1995)	Assess the impact of high dose vitamin A supplementation on the growth of mild to moderately malnourished children	Rural population in South India	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• 592 children aged less than 3 years • Children with severe malnutrition (weight/age <60% of NCHS median or height/age <85% of NCHS) or ophthalmic signs of xerophthalmia, serum retinol of <0.35umol were excluded	• Vitamin A did not significantly increase in growth measured by WAZ or HAZ.
Kirkwood et al (1997)	Study the effect of prophylactic vitamin A supplementation on child growth	Northern Ghana	• 2 RCTs • High CASP score • Statistical tests appropriate and well presented	• Health study ~1500 children aged 6-59 months • Survival study~15,000 children aged 6-59 months • Children with clinical signs of xeropthalmia were treated and excluded from the study.	• In the health study, Vitamin A supplementation did not lead to any increased growth. • In the survival study, children aged more than 36 months in the vitamin A supplementation group had a significantly higher weight gain, but the authors noted that the gain was so small as to be functionally important in this age group. No other significant changes in growth were observed.
Hadi, H (2000)	Study the effect of vitamin A supplementation on the linear growth or preschool children	Central Java, Indonesia	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• 1,407 children aged 6-47 months • Only one child with clinical signs of vitamin A reported.	• Vitamin A supplementations lead to a significant increase in height but not in weight.
Lie (1993)	Assess the impact of vitamin A supplementation on growth, childhood diarrhoea and respiratory disease	China	• RCT • High CASP score • Statistical tests appropriate and well presented	• Children aged 6-36 months • N= 172 • All young children included in the trial irrespective of serum retinal levels. Very low levels of clinical vitamin A deficiency reported.	• Vitamin A supplementation was not associated with an increase in growth

Author of study	Objective of study	Setting	Study design, CASP score grade*, Statistics	Survey population (Age, size)	Impact / outcome
Rahmathullah, 1990	Assess whether weekly vitamin A supplementation has an impact on groowth	India	• RCT • Medium CASP score • Statistical tests appropriate and well	• N=15,419 • Children aged 0-71 months • Excluded all children with xeropthalamia	• No effect of vitamin A supplement on growth
West, 1988	Assess the impact of vitamin A supplementation on growth	Indonesia	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• N= 2,012 children • Aged 0-5 years • Excluded children with clinical signs of vitamin A deficiency	• Significant increase in weight and MUAC for boys aged greater than 2 years • No increase in the rest of the population
West, 1997	Assess the impact of vitamin A on supplementation on the growth of children	Nepal	• RCT • Medium CASP score • Statistical tests appropriate and well presented	• N= 3,377 children • Aged 12-60 months • Excluded children with clinical signs of vitamin A deficiency	• In the non-xeropthalmic group, Vitamin A supplementation had no significant impact on weight or height gain but an increase in MUAC and muscle area was recorded.
Muhilal, 1988	Assess the impact of improved vitamin A status on health growth and survival of children	Indonesia	• Controlled trial • High CASP score • Statistical tests appropriate and well presented	• N= 11,200 children • Aged 0-5 years • Excluded all children with clinical signs of vitamin A deficiency	• There is an significant increase in linear growth • There is an increase in weight gain, but it is not significant
Fawzi, 1998	Assess the impact of vitamin A supplementation on growth of preschool children	Sudan	• Controlled trial • High CASP score to follow • Statistical tests appropriate and well presented	• N= 21,251 children • Aged 6-72 months • Excluded all children with clinical signs of vitamin A deficiency	• Vitamin A supplementation did not have a significant impact on weight or height of children.

Table A4.6: Summary of published studies assessing the impact of measles immunisation programmes

*Note that * is given next to the CASP score if there is a suspicion that the grading is lower than the actual quality of the study. This might occur when insufficient information is given about the study design but the authors of this report think that the assessment was probably well designed.*

Author of study	**Objective of study**	**Setting**	**Study design, CASP score grade*, Statistics**	**Survey population (Age, size)**	**Impact / outcome**
Bhaskaram, 1986	Assess the effect of malnutrition on the development of protective levels of antibody titres	• Hyderabad city, India	• Case control study • Medium (63%) • Statistics are presented	• 190 children aged between 9 months and 3 years from a slum area of the city	• Immune response was unaffected by malnutrition
Kapoor, 1991	Assess the effectiveness of measles immunisation on malnutrition related mortality	• Unknown, presumably India since it is in the Indian Journal of Pediatrics	• Case control study • Poor (21%) • Statistics are presented but no information on what the tests used were	• Children aged 1-4 years	• Measles vaccination is associated with a significant reduction in mortality rate • The mortality rate of the study area is almost twice that of the control area, so no comparison can be made
Phillips, 2004	Identify whether measles severely compromises immune responsiveness in malnourished children	• Sokoto state, Nigeria	• Case control study. The control group was randomly allocated • Medium (66%)	• 130 children • 65 with measles • 65 without (control) • 45% of the child pop has malnutrition	• Measles immunisation was not associated with significantly higher age for height Z-scores • Measles immunisation was associated with significantly higher Weight for Age and Weight for Height Z-scores

Table A4.7: Summary of published studies assessing the cost-effectiveness of the 6 nutrition-related interventions

Reference	Area studied and study year	Interventions examined	Type of analysis	Base year for costs	Results (costs, cost-effectiveness)	Quality (good, medium, poor)	Notes
SFP							
Young et al (1988)	Ethiopia and Sudan (1985-86)	29 brands of food-aid biscuits	Cost analysis	Ns	The average cost per 500 kcal was 47.5p for compressed products and 14.0p for traditional biscuits, and the average cost per 10g of protein was 30.5p for compressed products and 11.0p for the traditional products	3 / 7 = 43% (medium)	There is a suspicion that the overall quality of this paper is lower than the score given.
TFP							
Caldwell & Hallam (in press)	Sudan, Malawi and Ethiopia (2003)	Community therapeutic feeding	Cost analysis	2003	Cost per beneficiary in euros = 114, 148 and 60 for Sudan, Malawi and Ethiopia respectively	5 / 9 = 56% (medium)	There is a suspicion that the overall quality of this paper is higher than the score given
Bednet programmes							
Rowland et al (1999)	Afghanistan (1996)	Permethrin-treated *chaddars* and top-sheets	Cost analysis / cost-effectiveness analysis	Ns	Cost per person protected = $0.17 Cost per case prevented = $1.07	2 / 8 = 25% (poor)	Not a conventional CEA as comparator not specified,however the implicit comparator is 'do-nothing'. Unclear whether any delivery costs have been included

Ns = not stated

Annex 5: Different methods to analyse coverage of emergency feeding nutrition programmes

Programme coverage is an important indicator for monitoring and evaluating selective feeding programmes. Coverage data is also useful to help project staff find out how well a project is doing and whether or not the project is meeting, or will meet, its objectives. In 2004, specific coverage indicators for selective feeding programs were included in the SPHERE project's humanitarian guidelines for the first time.

This annex outlines different approaches to estimate coverage and discusses their shortcomings. Recommendations about which method to use, and when, are given. Much of the material in this section is drawn from Myatt (in press).

A5.1 Standard approaches to assessing programme coverage

Currently standard approaches to assessing supplementary and therapeutic feeding programmes' coverage involve making use of anthropometric surveys either *directly* using the survey data, or *indirectly* using survey data, program enrolment data, and population estimates.

A5.1.1 Direct method to assess coverage

The direct method involves adding a question to the anthropometric questionnaire about whether or not each child is currently enrolled in a feeding programme. This is probably, currently, the most commonly used method to assess coverage. Coverage is then estimated using the following equation:

$$\text{Coverage} = 100 * \frac{\text{Number of eligible children found attending the programme during the survey}}{\text{Number of eligible children found during the survey}}$$

An eligible child is defined as a child who should be enrolled in the programme. For example, a moderately malnourished child (WHM <80% and WHM >= 70%) should be enrolled for an SFP and hence would be eligible for that type of programme. A severely malnourished child (WHM<70% and/or oedema) should be enrolled in a TFP and hence would be eligible for that type of programme.

A5.1. 2 The indirect method for assessing coverage

The indirect method involves comparing the number of malnourished children estimated to exist in a population through a nutrition survey to the actual number of children attending the programme. There is no need to add a question to the anthropometric questionnaire. This method is usually less accurate than the direct method because it requires relatively up-to-date information on population figures.

Coverage is estimated using the following equation:

$$\text{Coverage} = 100 * \frac{\text{Number of children attending the feeding programme}}{\text{Estimated prevalence of malnutrition} * \text{estimated number of children}}$$

A5.2 Disadvantages of the standard methods

The standard methods of estimating programme coverage have two major problems. The first is connected to sample size and the second to the assumption of homogeneity.

Sample size issues
The sample size calculated for an anthropometric survey allows the *prevalence* of acute malnutrition to be estimated with reasonable precision, but the sample size available to estimate *coverage* depends on the prevalence of acute malnutrition found by the survey. When the aim of the survey is to estimate the coverage of a feeding programme for *severe* acute malnutrition (a therapeutic feeding programme), the sample size will usually be too small to estimate coverage with reasonable precision. This means that you will have very wide confidence intervals around your estimate of coverage. This problem affects both the direct and indirect methods of estimating coverage. This problem will be less acute if you are estimating the coverage of a supplementary feeding programme (compared to a TFP) because there will be more moderately malnourished children in the population.

The indirect method suffers from a further drawback because the denominator used in the formula is subject to considerable uncertainty. The population estimate is usually derived from census data. In complex emergencies, certain factors may lead to census data not being accurate (e.g. political manipulation, the absence of a functioning civil society, population displacement, and poor security).

Homogeneity
Both the direct and indirect methods to measure coverage assume that the coverage of the feeding programmes is homogenous across the whole survey area[33]. This means that the methods assume that the programme coverage is the same across the whole catchment area for a survey. In a small geographical area, such as a refugee camp this assumption may be true. However, over a wider area the assumption is often unlikely to be true, especially for some centre-based programmes during start-up phase because coverage will be greater for areas close to centres. Also, in the start up phase some villages may not have information about the existence of centre-based facilities.

In general, anthropometric surveys which have been sampled using the two-stage cluster sample technique include more children from the most populous parts of the survey area (because a cluster is more likely to be sampled from this area) than from the outlying areas. The most populous areas are more likely to have a feeding centre and hence the coverage of the programme is likely to be higher in these areas. Thus, it is likely that the survey will over-estimate coverage.

The direct or indirect methods only produce one figure for the coverage of the whole survey area. If the homogeneity assumption is untrue and coverage is uneven, then it is useful to be able to identify where coverage is good and where it is bad so that you can improve your programme. The method described below is one way to do this.

A5.3 Centric Systematic Area Sampling

It has recently been suggested that the *centric systematic area sampling* (CSAS) method might be a useful way to assess feeding programme coverage in an area where the coverage is not homogenous and population figures are unsure (Myatt, 2003). The method has been

[33] Note that the standard sampling methods described in section 5 make the same assumption about the prevalence of malnutrition: it is assumed that the prevalence of malnutrition is the same throughout the survey area.

adapted from ecological studies where it is used to ascertain the spatial distribution of plant and animal species over wide areas.

The CSAS method abandons probability sampling and adopts active case-finding instead. The project area is split into 30 or so equal sized quadrats (squares of approximately equal area) and cases of severe malnutrition are sought. A simple count of cases enrolled in the programme compared to cases not enrolled in the programme is made. This figure can be compiled for all the quadrats to give an overall project coverage figure, or used separately to estimate coverage in each area. As more cases are seen using the CSAS methods, the confidence intervals are much narrower than when you use the standard method.

Detailed steps for conducting a CSAS are outlined in Myatt (in press).

A5.3.1 Advantages of the CSAS method

The main advantage of CSAS method is that it allows you to see the spatial distribution of the coverage of your programme. This is very useful for programmes because it means you can follow-up on the communities that aren't using the programme, and find out why they aren't using the programme.

Another major advantage of the CSAS method is that you can use it to refer children to the programme immediately. This, of course, is true of the direct method too – when you find a malnourished child who is not enrolled on a programme then you should always refer them if a programme is available. However, since the CSAS sees so many malnourished children (because of the active case finding method), you are able to refer more of them than by using the direct or indirect survey methods.

A5.3.2 Potential disadvantages of the CSAS method

The CSAS method assumes that the coverage of a programme is homogenous within any given quadrat. If the coverage is not homogenous within each quadrat, then the method will have the same problems as those described in above section A5.2. In fact, normally quadrats are relatively small areas – certainly much smaller areas than the whole programme area – so it is likely that the homogeneity assumption is true. If you are worried that the quadrats are not homgenous, then you could consider taking a greater number of smaller-sized quadrats for your survey.

A poor case finding method might systematically exclude some children. For example, children from minority groups or children living on the periphery of sampled communities may be excluded leading to bias. To avoid this you need to work really hard at your active case finding. Remind your key informants that you want to see all the sick/malnourished children in the community – not just the ones living in the centre of the village.

The method takes longer than the standard approaches because you spend one whole day in each quadrant, as opposed to just combining the coverage questions with a standard anthropometric survey. This is true but remember that active case-finding is central to successful programme implementation. The estimation of the coverage can be integrated into programme outreach work. This would allow continued estimation of coverage as part of routine programme activity.

Finally, one important drawback of the method is that although trials of CSAS methods indicate that the results obtained when estimating coverage of programmes designed to

correct severe acute malnutrition are very good[34], the results for supplementary feeding programmes (moderate malnutrition) are less accurate. This is because it is relatively straightforward for members of the community to identify severely malnourished children but harder for them to identify moderately malnourished children. This means that more moderately malnourished children will be missed by the active-case finding method and that the estimate of coverage for a supplementary feeding programme may be less accurate.

A5.4 What method of estimating coverage is most appropriate when?

All of the methods described above have some drawbacks. These are summarised in the table A5.1. The 'X's in the table show which method suffers from which drawback.

TablevA5.1 Summary of drawbacks faced by each of the coverage methods

Drawback	**Method**		
	Direct	**Indirect**	**CSAS**
Sample size	X	X	
Assumes homogeneity	X	X	
Cannot visualise distribution	X	X	
Need accurate population figures		X	
Less useful for active case finding	X	X	
Not useful for SFP			X
Time required			X

The most desirable method is the one which has the least drawbacks. Judging from this table it looks like the indirect method is the least useful. In particular, the need for accurate population data means that in most situations this method will probably return inaccurate results. The exception to this rule may be in a camp situation where you are pretty sure of the population figures and that the coverage of the programme is homogenous.

From table A5.1, it looks like the CSAS method is the most promising, at least when you are trying to estimate the coverage of a therapeutic feeding programme. The fact that the CSAS method can be incorporated into standard programme outreach activities gives it a big advantage over the other methods. The CSAS method is particularly relevant in a setting where you suspect that the rates of coverage will differ substantially within the programme area. So it may be very useful to use it near the beginning of a programme – it will help you see where you need to work on improving coverage.

When you are estimating the coverage of an SFP, it may be necessary to use the direct method. The problems with sample size should not be so serious when you are estimating the coverage rate of an SFP (compared to a TFP) as you would expect to find more moderately malnourished children. However, the assumption about heterogeneity may still be an issue.

[34] Using capture-recapture methods it has been found that the sensitivity of active case finding is around 100%. Specificity (including moderately malnourished children) is low at about 50%. This is not a concern for this application (coverage examination) where exhaustivity (i.e.: high sensitivity) is required to calculate the spatial pattern of coverage. As in many 'screening' contexts, specificity is sacrificed in order to achieve 100% sensitivity.

Alternatively, you could use the CSAS method but make extra efforts to look for as many moderately malnourished children as possible. You could even end up screening all the children in a village in each quadrant to assess the rate of coverage of an SFP – although this would be very time consuming.

Annex 6: Information from published economic-evaluation studies in development settings

All data on cost-effectiveness for vitamin A supplementation and measles immunisation comes from work from routine health care settings. These are discussed below after which there is a brief section about the extent to which they can be reasonably extrapolated to emergency settings. There is also a large cost-effectiveness evidence-base for bednets outside of emergency settings and this is also discussed below.

A6.1 Vitamin A interventions

Ching P et al (2000) used country-specific activity and coverage data to estimate the cost-effectiveness of adding vitamin A supplements into immunisation campaigns conducted in 1998 and 1999. The estimated incremental cost per death averted was US$72 (range: 36-142) in 1998 and US$64 (range: 32-126) in 1999. The estimated average total cost of providing supplementation per death averted was US$310 (range: 157-609) in 1998 and US$276 (range: 139-540) in 1999. Costs per death averted varied by campaign, depending on the number and proportion of the child population reached, number of doses received per child, and child mortality rates.

Individual studies from The Philippines, India and South Africa support their findings that vitamin A supplements are cost-effective. In 1993 in the Philippines, where vitamin A deficiency is a serious and widespread public health problem, the Philippines National Vitamin A Supplementation Programme (NVASP) was established. Two papers assessed the economics of the programme. Loevinsohn et al (1997) performed a cost-effectiveness analysis in the Philippines to examine whether vitamin A supplements should be given universally to all children 6-59 months, targeted broadly to children suffering from mild, moderate, or severe malnutrition, or targeted narrowly to pre-schoolers with moderate and severe malnutrition. The first year average cost of the universal approach was $67.21 per death averted, compared to $144.12 and $257.20 for the broad and narrow targeting approaches respectively. The authors, therefore, concluded that targeting vitamin A supplements to high-risk children is not an efficient use of resources. Furthermore, they note that decisions about targeting are complex because they depend on a number of factors including: the degree of clustering of preventable deaths, the cost of the intervention, the side-effects of the intervention, the cost of identifying the high risk group, and the accuracy of the 'diagnosis' of risk.

In the second paper, Fiedler et al (2000) presented a cost-effectiveness analysis of the NVASP, and of a hypothetical programme of vitamin A fortification of wheat flour that was conducted to inform policymakers as to how to modify the NVAS Programme. Employing a proxy effectiveness indicator of vitamin A deficiency (the intake of < 70% of the recommended daily allowance of vitamin A), in a series of simulations using individual child consumption data, the analysis found that fortification is more efficient in reducing inadequate vitamin A intake compared to the ongoing programme. However, due to the nature of food consumption patterns, fortification alone is not enough. Interestingly, an investigation of the cost and efficiency of geographically targeted supplementation programmes revealed that maintaining a universal supplementation programme in urban areas and, in rural areas, introducing a targeted programme to only the poorest municipalities (where the prevalence of vitamin A efficiency is the highest) will provide a more acceptable public health policy response than fortification alone.

Pandav CS et al (1998) published a paper which estimates the cost of providing iron and vitamin A supplementation through the primary health care system in India. The costs

included the proportionate cost of the building, workers' salary and the cost of the supplements. The cost of vitamin A supplementation to under-threes through the PHC system was estimated at Rs. 3.20 per beneficiary per year. Consequently, the authors concluded that vitamin A supplementation (and iron as well) through the PHC system was a low cost intervention. Unfortunately, the authors did not extend their analysis to assess the cost-effectiveness of such programmes.

Finally, Saitowitz et al (2001) sought to assist in the development and implementation of a national vitamin A supplementation programme at primary health care facilities for mothers and children. They estimated that the total annual, recurrent cost of a national programme would cost R16.4 million. The bulk of the costs would include personnel costs, comprising 68% of the total costs. Other costs included promotion (27%), vitamin A capsules (4%) and training (1%). The projections showed that the programme would be financially feasible and would reach the majority of children under 24 months of age.

A6.2 Bednet programmes

Hanson et al (2004) reviewed the existing cost-effectiveness evidence-base of malaria control interventions. For insecticide treatment of bednets, results ranged from $9-27, and for the provision and treatment of bednets, from $10-118. These results were produced from trial-based and modelling studies, and all came to the broad conclusion that bednets were a highly cost-effective use of resources.

Table A6.1: Cost-effectiveness results for bednets (US$1995) (sensitivity analysis results in brackets)[35]

Author (reference)	Area studied and study year	Interventions examined	Results (costs, cost-effectiveness)	
			Cost per death averted	Cost per DALY averted or DYLG[36]
Picard et al,1993	The Gambia	Insecticide treatment of bednets	$219 ($167-243)	$9 ($9-14)
Aikins et al.,1998	The Gambia	Insecticide treatment of bednets	Net costs $494 ($326-805)	Net costs $21 ($14-35)
Graves, 1998	The Gambia	Insecticide treatment of bednets	$829 ($447-2117)	-
Goodman et al, 1999	Sub-Saharan Africa	Insecticide treatment of bednets	-	$4-10
Binka et al, 1997	Ghana	Provision and insecticide treatment of bednets	$2112 ($992-2289)	$77 (37-84)
Some, 1999	Kenya	Provision and insecticide treatment of bednets	$2958 ($2838-3120)	-
Evans et al, 1997	Africa	Provision and insecticide treatment of bednets	-	$10-118
Goodman et al, 1999	Sub-Saharan Africa	Provision and insecticide treatment of bednets	-	$19-85

Source: adapted from Hanson et al. 2004

[35] To facilitate comparison between studies, the authors converted all costs to 1995 US$ using the US$ period average market exchange rate in the study year and the US Consumer Price Index.
[36] Discounted year of life gained

A6.3 Measles immunisation programmes

Measles immunisation is one of the most highly cost-effective preventive interventions available (World Bank, 1993). Table A6.2 presents some of the existing international evidence of the cost-effectiveness of measles vaccination. More recently, studies have examined the measles cost-effectiveness given through campaigns, either routine or emergency outbreak responses.

Table A6.2: Cost-effectiveness results for routine measles immunisation (US$1999)[37]

Country (author)	Area studied and study year	Coverage	Results (costs, cost-effectiveness)	
			Cost per case averted	Cost per death averted
Shepard et al, 1986	Ivory Coast	61%	28.31	970.49
Robertson et al, 1985	The Gambia	71%	3.96	82.55
Williams, 1989	The Gambia	70%	4.71	94.64
Ponnighaus, 1980	Zambia	75%	-	787.79
Walker et al, 2000	Bangladesh	70%	5.01	288.15

In Zambia, the vaccination programme includes one dose of measles vaccine at 9 months of age. The objective of the paper published by Dayan et al (2004) was to compare the cost-effectiveness of the current one-dose measles vaccination programme with an immunisation schedule in which a second dose is provided either through routine health services or through supplemental immunisation activities. Given the parameters established for this analysis, the authors found that such a schedule would be cost-saving and the most cost-effective vaccination strategy for Zambia. This conclusion is in line with that found by Tulchinsky et al (1993) who published evidence from a number of countries suggesting that a two-dose measles vaccination programme, by improving individual protection and heard immunity, can make a major contribution to measles control and elimination of local circulation of the disease. Cost-benefit analyses cited in their paper also supported the two-dose schedule in terms of savings in health costs, and total costs to society.

Only Sniadack et al (1999) provided data on the impact of measles outbreak response immunisation from a developing country, which is probably of most relevance to an emergency setting. The authors found that lack of national campaigns or access to routine immunisation in a remote part of Peru caused the severe impact of the measles virus outbreak. The outbreak response immunisation campaign targeted non-measles case children aged 6 months to 15 years regardless of immunisation status, which was effective in terminating the measles outbreak, morbidity and mortality. This campaign cost approximately US$3000 and in 1998 saved 1155 person-days of work among 77 adults. It also prevented 87 diarrhoea and 46 pneumonia cases and averted 5 deaths, at a cost of $600 each.

[37] To facilitate comparison between studies, we converted all costs to 1995 US$ using the US$ period average market exchange rate in the study year and the US Consumer Price Index.

Annex 7: Bibliography

ACC/SCN (1994) *Update on the nutrition situation*. Agency Committee on Coordination/Sub-Committee on Nutrition: Geneva.

Aikins MK et al. The Gambian National Impregnated Bednet Programme: Costs, consequences and net cost-effectiveness. *Social Science and Medicine*, 1998, 46(2):181–191.

ALNAP (2002) *Humanitarian Action: Improving Performance through Improved Learning*. London: Overseas Development Institute.

ALNAP (2003a) *Annual Review 2003. Humanitarian Action: Improving Monitoring To Enhance Accountability and Learning*. London: Overseas Development Institute.

ALNAP (2003b) *Participation by Crisis Affected Populations in Humanitarian Action: A Handbook for Practitioners*. London: Overseas Development Institute.

Anonymous (1999). Use of health economics for decision-making in complex emergencies. Report of a WHO-FICOSSER meeting. Paris, 16-18 December, 1999. Cah Sociol Demogr Med. 2000 Apr-Jun;40 (2):159-88.

Ashworth, A and Khanum, S (1997) Cost-effective treatment for severely malnourished children: what is the best approach? *Health Policy and Planning.* 12(2):115-121.

Assefa, F (2001) 'Scurvy Outbreak and Erosion of Livelihoods Masked by Low Wasting Levels in Drought-Affected Northern Afghanistan', *Field Exchange*, vol. 12, pp. 14–16.

Banatvala N, Zwi AB (2000). Conflict and health. Public health and humanitarian interventions: developing the evidence base. *BMJ. 2000, 321(7253):101-5.*

Barnabas G, Lovel HJ, Morley DC (1982). Supplementary food for the few in a refugee camp. *Lancet*. 1982;2:48-48.

Barret, CB (2002) *Food aid effectiveness: "It's the targeting, stupid!"* First draft of paper prepared for WFP.
Beaton, G and Ghassemmi, H (1982) 'Supplementary Feeding Programmes for Young Children in Developing Countries', *American Journal of Clinical Nutrition*, vol. 35, pp. 864–913.

Beaton, GH, Martorel R, L'Abbe, KA et al (1982). Effectiveness of vitamin A supplementation in the control of young child morbidity and mortality in developing countries. Final report of CIDA. Toronto: University of Toronto, 1993

Bhaskaram, P, Madhusudan, J, Radhrakrishna, K.V, Raj, S (1986) Immunological response to measles vaccination in poor communities, *Human Journal of Clinical Nutrition,* 1986, 40, 4, 295-299.
Biellik, RJ, Henderson, P L (1980) Mortality, nutritional status and dietary conditions in a food deficit region: North Teso District, Uganda, December, 1980. *Ecology of Food & Nutrition.* 1981, 11, 3, 163-170

Binka FN, Mensah OA, Mills A (1987). The cost-effectiveness of permethrin impregnated bednets in preventing child mortality in Kassena-Nankana district of Northern Ghana. *Health Policy,* 1997, 41:229–239.

Biot M, Chandramohan D, Porter JD (2003). Tuberculosis treatment in complex emergencies: are risks outweighing benefits? *Trop Med Int Hyg*. 2003; 8(3): 211-218.

Black, N (1986) Why we need observational studies to evaluate the effectiveness of healthcare. *BMJ*, 1996, 312, 1215-1218

Borrel, A (1997). Selective feeding programmes for detainees in Rwanda. *Field Exchange.* 2: 4-6.

Borrel, A. and Salama, P (1999). Public Nutrition from an Approach to a Discipline: Concern's Nutrition Case Studies in Complex Emergencies, *Disasters*, vol. 23, no. 4, pp. 326–43.

Brewster DR, Manary MJ, Graham SM (1997). Case management of kwashiorkor: An intervention project at seven nutrition rehabilitation centres in Malawi. *European Journal of Clinical Nutrition*. 1997;51(3):139-147.

Briend A (1999). Ready-to-use therapeutic foods for treatment of marasmus. *The Lancet.* 353:1767-1768.

British Agencies Aid Group (2002). *Breaking New Ground? BOAG Agencies' Experience of Developing Systems for Assessing Organisational Performance*.

Brown V, Carbonell S (1998). Evaluation of the supplementary feeding programme carried out in Marsabit District, Kenya, April-November 1997. *Field Exchange* no. 5.

Brown LR,Yohannes Y, Webb P (1994). Rural labor-intensive public works: impacts of participation on preschooler nutrition: evidence from Niger American. *Journal of Agricultural Economics*, 1994, 76, 5, 1213-1218

Caldwell R, Hallam A. (in press). The cost of selective feeding. *Field Exchange*.

Carletto, C and Morris, S (2001). Designing methods for monitoring and evaluating food security and nutrition interventions. In *Methods for rural development projects.* Ed Hoddinott, J. International Food Policy Research Institute: Washington DC.

Ching P, Birmingham M, Goodman T, Sutter R, and Loevinsohn B (2000). Childhood mortality impact and costs of integrating vitamin A supplementation into immunization campaigns. *Am J Public Health*. 90(10):1526-9.

Clay EJ.(2003). Responding to change: WFP and the Global food aid system. *Development Policy review*. 21(5-6):697-709

Clay, EJ and Stokke, O (eds) (2000). *Food aid and Human Security.* EADI Book Series 24. London: Frank Cass.

Cliffe L and Luckham R (2000). 'What Happens to the State in Conflict?: Political Analysis as a Tool for Planning Humanitarian Assistance', *Disasters*, vol. 24, no. 4.

Collins S and Sadler K (2002). Outpatient care for severely malnourished children in emergency relief programmes: a retrospective cohort study. *Lancet.* 2002, Dec 7, 360, 9348 1824-1830

Concato J and Horwitz RI (2004). Beyond randomised versus observational studies. *Lancet.* 2004, 363, 1660-1661

Cooper WO, Boyce TG, Wright PF, Griffin MR (2003). Do childhood vaccines have non-specific effects on mortality? *Bulletin of the World Health Organization. 81(11):821-6, 2003.*

Curdy A (1994). *The Relevance of Supplementary Feeding Programmes for Refugees, Displaced or Otherwise Affected Populations.* Paper prepared for a workshop on the improvement of the nutrition of refugees and displaced people in Africa, Machakos, Kenya, 5–7 December 1994.

D'Alessandro U, Olaleye BO, McGuire W, Langerock P, Bennett S, Aikins MK, Thomson MC, Cham MK, Cham BA, Greenwood BM (1995). Mortality and morbidity from malaria in Gambian children after introduction of an impregnated bednet programme. *Lancet*, 1995, 345, 8948, 479-483.

Dayan GH, Cairns L, Sangrujee N, Mtonga A, Nguyen V, Strebel P (2004). Cost-effectiveness of three different vaccination strategies against measles in Zambian children. Vaccine. 2004;22 (3-4):475-84.

Deep Web Frequently Asked Questions. Accessed on 20/07/2004, Available at (http://www.brightplanet.com/deepcontent/deep_web_faq.asp).
Des Jarlais DC, Lyles C, Crepaz N and the TREND Group (2004). Improving the Reporting Quality of Non-randomised Evaluations of Behavioural and Public Health Interventions: The TREND Statement. *Am J Public Health.* 2004: 94: 361-366.
Desenclos, 1989

Desenclos JC, Berry AM, Padt R, Farah B, Segala C, Nabil AM (1989). Epidemiological patterns of scurvy among Ethiopian refugees. *Bulletin of the World Health Organisation*. 1989;67(3):309-316.

Donnen P, Brasseur D, Dramaix M, Vertongen F, Zihindula M, Muhamiriza M, Hennart P (1998). Vitamin A supplementation but not de-worming improves growth of malnourished preschool children in eastern Zaire. *Journal of Nutrition.* 1998, 128, 8, 1320-1327.

Drummond MF et al (1997). *Methods for the Economic Evaluation of Health Care Programmes*. New York: Oxford Medical Publications, Oxford University Press, 1997 (second edition).

Duce M. et al (2003) 'Suspected Thiamine Deficiency in Angola', *Field Exchange*, vol. 20, pp. 26–28.

Dufour C, de Geoffrey V, Maury H and Grunewald F (2004). Rights, Standards and Quality in a Complex Humanitarian Space: Is Sphere the Right Tool? *Disasters.* 28:2:124-141.

Dzumhur Z, Zec S, Buljina A, Terzic R (1995). Therapeutic feeding in Sarajevo during the war. *European Journal of Clinical Nutrition*. 1995;49(10):Suppl-2.

Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991). "Publication bias in clinical research," *Lancet.* 1991;337:867-872.

Emergency Nutrition Network (2003a). 'Community Based Approaches to Managing Severe Malnutrition'. Proceedings of an inter-agency workshop, Dublin, 8–10 October 2003.

Emergency Nutrition Network (2003b). 'Community Based Therapeutic Care: Summary of Published Research', *Field Exchange*, vol. 19.

Emergency Nutrition Network (2004). The Cost of Selective Feeding by Rose Caldwell and Alistair Hallam in Community Based Therapeutic Care. Special Supplement Series, no 2, pp 39-40.

Evans DB, Azene G, Kirigia J (1997). Should governments subsidize the use of insecticide-impregnated mosquito nets in Africa? Implications of a cost-effectiveness analysis. *Health Policy and Planning*, 1997, 12(2):107–114.

Fawzi WW, Herrera MG, Willett WC, Nestel P, El-Amin A, Mohamed KA (1998) The effect of vitamin A supplementation on the growth of preschool children in the Sudan. *American Journal of Public Health.* 1998. 87: 8, 1359-1362.

Fawzi WW, Thomas C, Charles MD Et al (1993). Vitamin A Supplementation and Child Mortality, A Meta-analysis. *JAMA:* February 17, 269/7, 898-903.

Fiedler JL, Dado DR, Maglalang H, Juban N, Capistrano M, Magpantay MV (2000). Cost analysis as a vitamin A program design and evaluation tool: a case study of the Philippines. *Soc Sci Med*. 2000;51(2):223-42.

Fox-Rushby JA, Hanson K (2001). Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health Policy Plan*. 2001 Sep;16(3):326-31.

Friedman JF, Phillips-Howard PA, Hawley WA, Terlouw DJ,Kolczak MS, Barber M,Okello N,Vulule JM,Duggan C, Nahlen BL, Ter Kuile, FO (2003). Impact of permethrin-treated bed nets on growth, nutritional status, and body composition of primary school children in western Kenya. *American Journal of Tropical Medicine & Hygiene*, 2003, 68, 4 SUPPL., 78-85,

Gibb C (1986). A review of feeding programmes in refugee reception centres in Eastern Sudan, October 1985. Disasters. 1986:17-24.

Glaxiou PP, Mackerras DEM (1993). Vitamin A supplementation and infectious disease: a meta-analysis. *British Medical Journal*. 1993, 306, 366-370.

Goodman CA, Coleman PG, Mills AJ (1989). Cost-effectiveness of malaria control in sub-Saharan Africa. *The Lancet*, 1999, 354:378–385.

Graves PM (1998). Comparison of the cost-effectiveness of vaccines and insecticide impregnation of mosquito nets for the prevention of malaria. *Annals of Tropical Medicine and Parasitology*, 1998, 92(4):399–410.

Griekspoor A, Sondorp E and Vos T (1999). 'Cost-effectiveness Analysis of Humanitarian Relief Interventions: Visceral Leishmaniasis Treatment in the Sudan', *Health Policy and Planning*, vol. 14.

Habicht, JP, Mason, JB and Tabatabai (1984). Basic concepts for the design of evaluation during programme implementation. In *Methods for the evaluation of the impact of food and nutrition programmes.* Ed Sahn, DE, Lockwood, R and Scrimshaw, N. United Nations University Press.

Habicht, JP, Victora, CG and Vaughan, JP (1999). Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact. *International Journal of Epidemiology* 28: 10-18.

Hadi H, Stoltzfus RJ, Dibley MJ, Moulton LH,West KP Jr, Kjolhede CL, Sadjimin T (2000). Vitamin A supplementation selectively improves the linear growth of Indonesian preschool children: results from a randomised controlled trial (2000) *American Journal of Clinical Nutrition* 71, 2, 507-513.

Hallam A (1998). Evaluating Humanitarian Assistance Programmes in Complex Emergencies. *Good Practice Review 7.* London: Overseas Development Institute.

Hallam A (1996). Cost-effectiveness analysis: a useful tool for the assessment and evaluation of relief operations. *Relief and Rehabilitation Network Paper 15.* Overseas Development Institute, London. 1996.

Hanson K, Goodman C, Lines J, Meek S, Bradley D, Mills A (2004). The economic of malaria control interventions. *Global Forum for Health Research: Geneva*. 2004. ISBN 2-940286-00-0

Henderson PL, Biellik RJ (1980). Health and nutrition service delivery to refugees in the Somali Democratic Republic, 1980. *Disasters*. 1981;5(2):104-112.

Hofmann C-A, Roberts L, Shoham J, Harvey P (2004). Measuring the impact of humanitarian aid. A review of current practice. *HPG Report 17*, June 2004. ISBN: 0-85003-718-2

Jaspars S and Young Y.(1995). General food distribution in emergencies: from nutritional needs to political priorities. *RRN: Good practice review no. 3*. ODI: London.

Kapoor SK, Reddaiah VP (1991). Effectiveness of measles immunization on diarrhea and malnutrition related mortality in 1-4 year olds. *Indian Journal of Paediatrics*, 1991, 58, 6, 821-823.

Kasis I, Lak L, Adler J, Choni R, Shazberg G, Fekede TD, Shoshani E, Miller D, Heyman S. (2001). Medical relief operation in rural northern Ethiopia: addressing an ongoing disaster.*Isr Med Assoc J*. Oct;3(10):772-7.
Kemmer TM, Bovill MEe, Kongsomboon W, Hansch SJ, Geisler KL, Cheney C, Shell-Duncan BK and Drewnowski A (2004). Iron deficiency is unacceptably high in refugee children from Burma. *Am. Soc. Nutr. Sciences.* 4143-4149.

Kirkwood BR, Cousens SN. Victora CG, de Zoysa I. (1997). Issues in the design and interpretation of studies to evaluate the impact of community-based interventions. *Trop Med & Int Health.* 1997 2(11), 1022-1029.

Kirkwood BR, Ross DA Arthur P, Morris SS, Dollimore N, Binka FN, Shier RP, Gyapong JO, Addy HA, Smith PG (1996). Effect of vitamin A supplementation on the growth of young children in northern Ghana. *American Journal of Clinical Nutrition*, 1996, 63, 5, 773-781.
Kwena AM, Terlouw DJ, de Vlas SJ, Phillips-Howard PA, Hawley WA, Friedman JF, Vulule JM, Nahlen BL, Sauerwein RW, ter Kuile FO (2003). Prevalence and severity of malnutrition in pre-school children in a rural area of western Kenya. *American Journal of Tropical Medicine and Hygiene*, 2003, 68, 4 Suppl, 94-99.

Lawlor DA, Smith GD, Kundu D, Bruckdorfer KR, Ebrahim S (2004). Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet*. 2004 363, 1724-1727

Le Billon P (2000). *The Political Economy of War:What Relief Agencies Need to Know*. Network Paper 33. London: Overseas Development Institute.

Leenstra T, Phillips-Howard PA, Kariuki SK, Hawley WA, Alaii JA, Rosen DH, Oloo AJ, Nahlen BL, Kager PA, ter Kuile FO (2003). Permethrin-treated bed nets in the prevention of malaria and anemia in adolescent schoolgirls in western Kenya *American Journal of Tropical Medicine and Hygiene*, 2003, 68, 4 Suppl, 86-93.

Lengeler C *2004) Insecticide – treated bed nets and curtains for preventing malaria (Cochrane Review). *The Cochrane Library*, 2004, issue 3, Chichester, UK: John Wiley & Sons, Ltd.

Lie C, Ying C, En-Lin W, Brun T, Geissler C (1993). Impact of large-dose vitamin A supplementation on childhood diarrhoea, respiratory disease and growth. *European Journal of Clinical Nutrition. 1993. 47: 2, 88-96.*

Loevinsohn BP, Sutter RW, Costales MO (1997). Using cost-effectiveness analysis to evaluate targeting strategies: the case of vitamin A supplementation. Health Policy Plan. 1997 12(1): 29-37.

Macrae J et al (2002). Uncertain Power: The Changing Role of Official Donors in Humanitarian Action. *HPG Report 15*. London: Overseas Development Institute.

Mason JB (2002). Lessons on nutrition of displaced people. *Journal of Nutrition.* 132:7:S2096-2103.

Mason, JB and Habicht, JP (1984). Stages in the evaluation of ongoing programmes. In *Methods for the evaluation of the impact of food and nutrition programmes.* Ed Sahn, DE, Lockwood, R and Scrimshaw, N. United Nations University Press.

Maxwell D (1999). Alternative Food Security Indicators: Revisiting the Frequency and Severity of Coping Strategies, *Food Policy*, vol. 24.

Maxwell, D (2001). The Coping Strategies Index: Monitoring Food Security Status in Emergencies, *Field Exchange*, no. 13.

Mourey A (2004). Nutrition intervention in crisis – some worrying practices. *Field Exchange*, no. 20: 28-29.

MSF (1995) *Nutrition Guidelines*. Paris: MSF.

Muhilal. Permeisih, D. Idjradinata, Y. R. Muherdiyantiningsih. Karyadi, D (1998). Vitamin A-fortified monosodium glutamate and health, growth, and survival of children: a controlled field trial. *American Journal of Clinical Nutrition. 1988. 48: 5, 1271-1276.*

Munro L (2002). Zimbabwe's Child Supplementary Feeding Programme: A Re-assessment Using Household Survey Data, *Disasters*, vol. 26, no. 3, pp. 242–61.

Murray J, McFarland DA, Waldman RJ (1998). Cost-effectiveness of oral cholera vaccine in a stable refugee population at risk for epidemic cholera and in a population with endemic cholera. *Bull World Health Organ*. 1998;76(4): 343-52.

Myatt, M (in press) A field trial of a new method for estimating coverage of emergency nutrition programmes. *WHO Bulletin*

Naficy A, Rao MR, Paquet C, Antona D, Sorkin A, Clemens JD (1998). Treatment and vaccination strategies to control cholera in sub-Saharan refugee settings: a cost-effectiveness analysis. *JAMA*. 1998 279(7): 521-5.

NHS CRD (1996). Undertaking Systematic Reviews of Research on Effectiveness: CRD guidelines for those carrying out or commissioning reviews. *CRD Report No.4.*

Orach CG (1999). Morbidity and mortality amongst southern Sudanese in Koboko refugee camps, Arua District, Uganda. *East African Medical Journal*. 1999;76(4):195-199.

Pandav CS, Anand K, Gupta S, Murthy GV (1998). Cost of vitamin A and iron supplementation to 'at risk' population. *Indian J Pediatr*. 1998;65(6):849-56.

Pawson R and Tilley N (1997). Realistic Evaluation. Whole book, Sage Publications Ltd, London, 1997.

Perra A, de L Costello AM (1995). Efficacy of outreach nutrition rehabilitation centres in reducing mortality and improving nutritional outcome of severely malnourished children in Guinea Bissau. *European Journal of Clinical Nutrition.* 1995;49:353-359.

Petticrew M, Roberts H (2003). Evidence, hierarchies and typologies: horses for courses. *Journal of Epidemiology and Community Health* 2003 57(7):527-9.

Phillips RS, Enwonwu CO, Okolo S, Hassan A (2004). Metabolic effects of acute measles in chronically malnourished Nigerian children *Journal of Nutritional Biochemistry,* 2004, 15, 5, 281-288.

Picard J et a (1993)l. A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia,West Africa. 8. Cost-effectiveness of bed net impregnation alone or combined with chemoprophylaxis in preventing mortality and morbidity from malaria in Gambian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 1993, 2:53–57.

Ponnighaus JM (1980). The cost/benefit of measles immunisation: a study from Southern Zambia. *J Trop Med Hyg*. 1980;83(4):141-9.

Prag J, Helmersgaard P (1979). Intensive feeding in Sa Kaeo, Thailand 1979. *Disasters*. 1982:343-348.

Prudhon C, Briend A Laurier D, Golden MHN, Mary JY (1996). Weight and height indices for assessing the risk of death in severely malnourished children. *Am. J. Epidemiol.* 144: 116-123.

Prudhon C, Golden MHN,Briend A, Mary JY (1997). A model to standardise mortality of severly malnourished children using nutritional status on admission to therapeutic feeding centre. *Eur. J. Clin. Nutr.* 51:771-777.

Public Health Resource Unit. Appraisal Tools, (2003). Available at: http://www.phru.org.uk/~casp/appraisa.htm. Accessed 04, Jul 2003.

Quisumbing AR(2003). Food aid and child nutrition in rural Ethiopia. World Development (Oxford), 13-14 November 2001, 31, 7, 1309-1324

Rahman MM, Tofail F, Wahed MA, Fuchs GJ, Baqui AH, Alvarez JO (2002). Short-term supplementation with zinc and vitamin A has no significant effect on the growth of undernourished Bangladeshi children. *American Journal of Clinical Nutrition*.2002, 75, 1, 87-91.

Rahmathullah L, Underwood B, Thulasiraj RD *et al* (1990). Reduced mortality among children in southern India receiving a small weekly dose of vitamin A. *New England Journal of Medicine*. 1990, 323, 929-935.

Ramakrishnan U and Martorell R (1998). The role of vitamin A in reducing child mortality and morbidity and improving growth. *Salud Pubica De Mexico.* 1998, 40/2, 189-197

Ramakrishnan U, Latham MC, Rajaratnam A (1995). Vitamin A supplementation does not improve growth of preschool children: a randomised double-blind field trial in South India. *Journal of Nutrition. 1995. 125: 2, 202-211*.

Reed BA, Habicht JP, Garza C (2002). Translating nutrition research into action in humanitarian emergencies. *Am. Soc. Nut. Sciences.* S2112-2116.

Robertson RL, Foster SO, Hull HF, Williams PJ (1985). Cost-effectiveness of immunisation in The Gambia. J Trop Med Hyg. 1985;88(5):343-51.

Roche C (1999). *Impact Assessment for Development Agencies: Learning to Value Change*. Oxford: Oxfam,

Roesel C (1998). From relief to development: supplementary feeding among Khmer refugees. *Health Policy and Planning*. 1988;3(3):227-236.

Rowland M, Durrani N, Hewitt S, Mohammed N, Bouma M, Carneiro I, Rozendaal J, and Schapira, A (1999). Permethrin-treated chaddars and top-sheets: appropriate technology for protection against malaria in Afghanistan and other complex emergencies. *Trans R Soc Trop Med Hyg.* 93(5): 465-72

Sadler K (2001). Rapid impact on malnutrition through a multi-faceted programme in Wolayita, Southern Ethiopia. *Field Exchange*, no.12.

Save the Children UK (2003). *Toolkits: A Practical Guide to Planning, Monitoring, Evaluation and Impact Assessment*. London: Save the Children.

Save the Children UK (2004). *Global Impact Monitoring Guidelines 2004*. London: Save the Children.

Saitowitz R, Hendricks M, Fiedler J, le Roux I, Hussey G, Makan B (2001). A proposed vitamin A supplementation programme for South Africa--design, coverage and cost. S Afr Med J. 2001;91(9):755-60.

Seal A(1998). The Measurement of Micronutrient Deficiencies – A Long Way To Go, *Field Exchange*, vol. 5, no. 13.

Seaman J (1993). Famine mortality in Africa *IDS bull.* Oct 24:4:27-32.

Shepard DS, Sanoh L, Coffi E (1986). Cost-effectiveness of the expanded programme on immunization in the Ivory Coast: a preliminary assessment. Soc Sci Med. 1986;22(3):369-77.

Shoham J.(1994) .Emergency Supplementary Feeding Programmes. *RRN: Good practice review 2*. ODI: London.

Shoham J.(2004). Review of the Role of Nutrition and Food Security Information in Assessing the Impact of Humanitarian Interventions. Background Paper. London: Overseas Development Institute.

Sniadack DH, Moscoso B, Aguilar R, Heath J, Bellini W, Chiu MC (1999). Measles epidemiology and outbreak response immunization in a rural community in Peru. *Bull World Health Organ*. 1999;77(7):545-52.

Snow RW, Molyneux CS, Njeru EK, Omumbo J, Nevill CG, Muniu E, Marsh K (1997). The effects of malaria control on nutritional status in infancy. *Acta Tropica.*, 1997, 65, 1, 1-10.

Some ES (1999). *Optimizing the community effectiveness of insecticide-impregnated bednets used for malaria control in coastal Kenya: Implications of perceptions, programme organization, compliance, and costs.* PhD Thesis, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, 1999.

Sphere Project (2004). *Humanitarian Charter and Minimum Standards in Disaster Response*, 2nd edition.

Starling S (2003). Balancing Measurement, Management and Accountability: Lessons Learned from SC UK's Impact assessment Framework.

Stefanak M A,Jarjoura D (1989). Weight gain in supervised and take-home feeding programmes in Chad. *Journal of Tropical Pediatrics*. 1989, Oct 35. 5, 214-217
Taylor, 1983

Talley L and Salama P (2003). Short report: assessing field vaccine efficacy for measles in famine-affected rural Ethiopia. *Am.J.Trop.Med.Hyg.*68:5:545-546

Taylor WR (1983). An evaluation of supplementary feeding in Somali refugee camps. *International Journal of Epidemiology*. 12(4):433-436.

Taylor A (2004). Outpatient therapeutic programme (OTP): an evaluation of a new SC UK venture in North Darfur, Sudan (2001). *Field Exchange*, no. 20: 26-27.

Ter Kuile FO, Terlouw DJ, Phillips-Howard PA, Hawley WA,Friedman JF, Kolczak MS, Kariuki SK, Shi YP, Kwena AM, Vulule JM, Nahlen BL (2003). Impact of permethrin-treated bed nets on malaria and all-cause morbidity in young children in an area of intense perennial malaria transmission in western Kenya: cross-sectional survey. *American Journal of Tropical Medicine & Hygiene.* 2003 April. 68 (4 Suppl): 100-107.

Tonascia JA (1993). Meta-analysis of published community trials: impact of vitamin A on mortality. Proceedings of the Bellagio Meeting on vitamin A deficiency and childhood mortality. New York: Helen Keller International, 1993.

Toole MJ, Nieburg P, Waldman RJ (1988). The association between inadequate rations, undernutrition prevalence, and mortality in refugee camps: case studies of refugee

populations in eastern Thailand, 1979-1980, and eastern Sudan, 1984-1985. *Journal of Tropical Pediatrics* Oct 34, 5, 218-224

Toole M J.and Bhatia R (1992). A case study of Somali refugees in Hartisheik A Camp, Eastern Ethiopia: health and nutrition profile, July 1988-June 1990. *Journal of Refugee Studies*. 5:3/4:313-326.

Tulchinsky TH, Ginsberg GM, Abed Y, Angeles MT, Akukwe C, Bonn J (1993). Measles control in developing and developed countries: the case for a two-dose policy. *Bull World Health Organ.* 1993;71(1):93-103.

Vautier F (1999). Selective feeding programmes in Wadjir: some reasons for low coverage and high defaulter rate. *Field Exchange*, no.6, 1999

Vautier F, Hildebrand K, Dedeurwaeder M, Baquet S, Herp van M (1999). Dry supplementary feeding programmes: an effective short-term strategy in food crisis situations. *Tropical Medicine and International Health*. 1999;4(12):875-879.

Vazquez-Garcia L (1999). Supplementary feeding in Mandera: the right intervention? *Field Exchange*, no.6, 1999.

Victora CG, Habicht JP, Bryce J (2004). Evidence-Based Public Health: Moving Beyond Randomised Trials. *Am J Public Health.* 94:400-405.

Walker D, Fox-Rushby J (2000). Economic evaluation of parasitic diseases: a critique of the internal and external validity of published studies. *International Health and Tropical Medicine* 5(4): 237-249

Walker D, Fox-Rushby J (2000). Economic evaluation of communicable disease interventions in developing countries: a critical review of the published literature. *Health Economics* 9(8): 681-698.

Walker D, Khan S, Akramuzzaman SM, Khan M, Fox-Rushby J, Cutts F (2000). *Cost-effectiveness analysis of measles control in Dhaka, Bangladesh*. Report prepared for WHO. HQ/98/454419 011638

Warner KE, Hutton RC (1980). Cost-benefit and cost-effectiveness analysis in health care. Growth and composition of the literature. *Med Care.* 18(11):1069-84.

Warrack-Goldman H, Brown BH. Binkin NJ (1986). Nutritional status of Mauritanian children during a drought emergency. *Ecology of Food & Nutrition*, 1986, 18, 3, 221-229

Waters E and Doyle J (2002). Evidence-based public health practice: improving the quality and quantity of the evidence. *Journal of Public Health Medicine*. 2002, 24/3, 227-229.

Watson F and Vespa J (1995). The impact of a reduced and uncertain food supply in three besieged cities of Bosnia-Hercegovina. *Disasters*. 19:3:216-234.

West KP Jr, Djunaedi E, Pandji A, Kusdiono. Tarwotjo I, Sommer A (1988). Vitamin A supplementation and growth: a randomised community trial. *American Journal of Clinical Nutrition. 1988. 48: 5, 1257-1264.*

West KPJr,. LeClerq SC, Shrestha SR, Wu LSF, Pradhan EK, Khatry SK, Katz J, Ramesh Adhikari. Sommer A (1997). Effects of vitamin A on growth of vitamin A-deficient children: Field studies in Nepal. *Journal of Nutrition. 1997. 127: 10, 1957-1965.*

WFP (2003). Report of the Technical Meeting on Key Issues in Emergency Needs Assessment. Rome, 28–30 October.

Williams PJ (1989). Effect of measles immunisation on child mortality in rural Gambia. *J Biosoc Sci Suppl*. 1989;10:95-104.

Wolde-Gebriel Z, Gebru H, Fisseha T and West CE (1993). Severe vitamin A deficiency in a rural village in the Hararge region of Ethiopia. *Eur.J.Clin.Nutr*. 47:2:104-114.

World Bank. *World Development Report: Investing in Health*. New York: Oxford University Press, 1993.

World Health Organisation (1996), Global Programme for Vaccines and Immunization. Progress toward 1995 measles control goals and implications for the future. *Geneva, Switzerland*: 1996; publication no. WHO/GPV/SAGE 96.04.

World Health Organisation. The History of Vaccination. A*ccessed on 15/07/2004*, Available at http://www.who.int/vaccines-diseases/history/history.shtml.

World Health Organisation (1999). Management of severe malnutrition: a manual for physicians and other senior health workers. *WHO*: Geneva.

Wyness, L (2003 unpub). Developing and evidence-based tool for redefining an approach to severe malnutrition in complex emergencies, *Unpublished data*, 2003

Wyness L, Mandava L, Knight J (2004 unpub). Systematic critical literature review effectiveness in malnutrition in complex emergencies. *Unpublished.* 2004.

Yamano T, Alderman H, Christiaensen, L (in press). Child Growth, Shocks, and Food Aid in Rural Ethiopia. *American J of Agricultural Economics.*

Young H (1987). Selective feeding programmes in Ethiopia and East Sudan - 1985/1986. *Disasters* 11:2:102-109.

Young H. and Jaspars S (1995). *Nutrition Matters. People, Food and Famine*. London: Intermediate Technology Publications.

Young H, Owen M, Clark J (1988). A comparison of biscuits used in emergency relief feeding programmes in Ethiopia and eastern Sudan--1985/86. *Eur J Clin Nutr*. 1988 Mar;42(3):261-71.

.