HUMANITARIAN ACADEMY · HUMANITAIRES JOBS

Mini MBA Programmes and MEAL: design, manage and prove impact

Complete manual

A programme of exactly 3 months (12 weeks, 144 hours) to design, manage and evaluate humanitarian programmes: needs assessment and severity, results chain and indicators, targeting and cash transfers, planning and budget, MEAL plan and data quality, sampling, analysis, accountability, evaluation, impact measurement, value for money and adaptive management. Corrected numerical exercises, three monthly exams with case studies, and a verifiable diploma from the Humanitarian Academy.

https://humanitairejobs.org/en/academy/mini-mba-programmes-meal/

Free document from the Humanitarian Academy. It may be printed and shared freely for training purposes, without modification.

Contents

1. Week 1 : Assessing needs: data, severity and priorities
2. Week 2 : Designing a programme: results chain, indicators and targets
3. Week 3 : Targeting and modalities: who to help, how, and with how much
4. Week 4 : Planning and budgeting a programme: from work plan to burn rate
5. Week 5 : The monitoring system: MEAL plan, data quality and remote monitoring
6. Week 6 : Surveys and sampling: getting the numbers right
7. Week 7 : Analysing data: useful statistics, pitfalls and visualisation
8. Week 8 : Accountability to affected people: participation, complaints and feedback
9. Week 9 : Evaluating: criteria, terms of reference, matrix and use
10. Week 10 : Measuring impact: counterfactual, methods and contribution
11. Week 11 : Value for money: cost-efficiency, cost-effectiveness and trade-offs
12. Week 12 : Adaptive management, learning and reporting: leading with evidence

Week 1 : Assessing needs: data, severity and priorities

Objectives

- Place needs assessment within the humanitarian programme cycle and choose the right type of assessment for the phase of the crisis.
- Combine secondary and primary data, spot biases, and classify the severity of an area with an explicit rule.
- Move from findings to decisions: response analysis, feasibility, added value and a reasoned choice of priorities.

Needs assessment in the humanitarian programme cycle

The **humanitarian programme cycle** (HPC) links five steps: needs assessment and analysis, strategic planning, resource mobilisation, implementation and monitoring, review and evaluation. The first collective product is the **Humanitarian Needs Overview** (HNO), which feeds the response plan (since 2024, most countries publish a single document, the HNRP). An NGO that designs a programme without reading its country's HNO deprives itself of the best available synthesis and risks proposing what others already cover.

The type of assessment depends on the phase of the crisis:

Phase	Typical tool	Timeframe	Decision question
Sudden shock	Multi-sector initial rapid assessment (MIRA or equivalent)	72 hours to 2 weeks	Where are the life-saving needs, and who is working where?
Protracted crisis	Multi-sector needs assessment (MSNA), representative household survey	Annual	What severity by area and by group, how many people in need?
Ongoing programme	In-depth sector assessment (SMART nutrition, market, protection)	As needed	Should the modality, targeting or transfer value be adjusted?

Three notions must be distinguished. A **need** is the gap between the observed situation and a standard (for example 15 litres of water per person per day according to Sphere). **Vulnerability** is the likelihood of suffering harm from the shock, which depends on exposure and capacities. **Capacities** are the resources of people, communities and institutions. A good diagnosis describes all three; a list of gaps without capacities leads to responses that replace what was already working.

> Golden rule: every question in an assessment must serve a decision identified in advance. An "interesting" question that changes no decision costs respondents time and dilutes the

analysis.

Secondary data: start with what exists

The **secondary data review** (SDR) gathers what is already known before collecting anything: HNO, cluster reports, demographic and health surveys (DHS), IPC analyses, other NGOs' assessments, data from government technical services, displacement tracking (IOM DTM), market prices, protection reports. A well-done SDR halves the size of a primary data collection, because it lets you ask only the questions that remain unanswered.

Each source is rated on two dimensions. **Source reliability**: known method, described sample, author without a conflict of interest. **Information credibility**: confirmed by other sources, consistent with what is known, recent. Information from a reliable source but two years old in an area of massive displacement is weak; information from an average source confirmed by three others is solid.

The SDR ends with a **gap analysis**: for each decision question, what is known, with what level of confidence, and what remains to be collected? It is this table, not a list of topics, that justifies the questionnaire.

- **Source triangulation**: the same information confirmed by independent sources.
- **Method triangulation**: a household survey and focus groups that converge.
- **Investigator triangulation**: different teams obtaining the same results.

Primary data: choosing methods and controlling bias

Four methods complement each other. The **key informant interview** (KII) questions people who have an overview: village chief, nurse, teacher, trader, head of a women's association. Quick and cheap, it gives trends but not proportions. The **focus group discussion** (FGD), 6 to 10 homogeneous people (same sex and age group), explores perceptions, norms and preferences. **Direct observation** (latrines, markets, queues) verifies what is reported. The **household survey** on a representative sample is the only method that produces generalisable proportions, provided that sampling is rigorous (week 6).

The most frequent biases:

- **Selection bias**: interviewing only villages accessible by road, or households present during the day (missing daily workers).
- **Courtesy and aid-expectation bias**: people exaggerate or minimise according to what they think the enumerator wants to hear or the aid they hope for. It is reduced by clearly explaining that no aid is promised and by favouring observable indicators.
- **Enumerator bias**: leading wording, a man questioning women about violence. Train, pre-test the questionnaire and form mixed teams.
- **Survivorship bias**: in an area of high mortality or departure, the most affected households are no longer there to answer.

> Minimum ethics: informed consent, the possibility to refuse without consequence, no sensitive question without a referral protocol (gender-based violence, child protection), and personal data reduced to the strict minimum.

Measuring severity: explicit rules and people in need

Comparing areas requires a common **severity scale**. The Joint Intersectoral Analysis Framework (JIAF) and the Integrated Food Security Phase Classification (IPC) both use five phases, from 1 (minimal) to 5 (catastrophic). The IPC classifies an area according to a **20% rule**: the area is placed in the most severe phase that, together with the higher phases, concerns at least 20% of its population.

Example. Population of 100,000 people: phase 1: 35%; phase 2: 30%; phase 3: 22%; phase 4: 10%; phase 5: 3%. Phases 4 and 5 combined: 13%, below the threshold. Phases 3, 4 and 5 combined: 35%, above 20%. The area is therefore classified in **phase 3**. The number of people in need of urgent action (phase 3 or above) is 35,000.

The number of **people in need** (PIN) of a sector is not the total population of the area: it is the sum of people whose situation reaches the defined severity threshold (often phase 3 or above). Adding up the PIN of several sectors counts the same person several times. For the intersectoral figure, the maximum per area and per group is therefore used, or an analysis at household level.

Severity must be **disaggregated** by sex, age and disability. For disability, the Washington Group Short Set asks six questions (seeing, hearing, walking, remembering, self-care, communicating); a person is counted as living with a disability if they answer "a lot of difficulty" or "cannot do at all" to at least one question.

Protection, gender and market analysis

An assessment that looks only at material gaps misses the risks. **Protection analysis** identifies threats (who does what to whom), people's vulnerabilities to these threats, and their capacities: risk = threat × vulnerability / capacity. It identifies groups excluded from aid (people without documents, minorities, isolated older people) and the risks that aid itself can create: distributions that expose people to theft, dangerous journeys, tensions between displaced people and the host population.

Gender analysis describes who decides, who controls resources, who has access to services, and how the crisis changes these roles. For example, it avoids paying cash assistance to a household member who does not control its use, or setting distribution times incompatible with women's tasks.

Market analysis determines whether cash transfers are feasible: are markets functioning, can traders respond to an increase in demand without inflation, is physical and financial access safe? The reference tools (rapid market assessment, emergency market mapping and analysis) answer three questions: availability of goods, prices and their trends, restocking capacity.

> Do no harm: for each response option, write at least one risk it creates and the measure that reduces it.

From analysis to decision: response analysis

Response analysis links what is known about needs to what the organisation will do. It compares options (in-kind, cash, vouchers, services, advocacy) on explicit criteria: relevance to priority needs, people's preferences, feasibility (access, markets, security, capacities), cost-effectiveness, risks, and what other actors are already doing (3W or 4W matrix: who does what, where and when).

A simple method is the **weighted matrix**: each criterion receives a weight set before scoring the options, to avoid choosing first and justifying afterwards. For example, relevance 30%, preferences 20%, feasibility 20%, cost-effectiveness 20%, risks 10%. An option scoring 4/5 for relevance, 5 for preferences, 3 for feasibility, 4 for cost-effectiveness and 3 for risks totals $0.3 \times 4 + 0.2 \times 5 + 0.2 \times 3 + 0.2 \times 4 + 0.1 \times 3 = 3.9$.

The final decision must also take into account the organisation's **added value**: presence, acceptance, expertise, local partners. If an actor already adequately covers a need in an area, the best decision is often to go elsewhere or to complement, not to duplicate.

The final product is an **analysis note** of a few pages: priority needs by area and by group, severity, coverage gaps, compared options, recommendation and level of confidence. This is what management will read, not the database.

Exercise: classify two areas and choose where to intervene

Situation. Two districts affected by drought. District A (60,000 inhabitants): phase 1: 30%; phase 2: 32%; phase 3: 25%; phase 4: 11%; phase 5: 2%. District B (80,000 inhabitants): phase 1: 45%; phase 2: 37%; phase 3: 12%; phase 4: 5%; phase 5: 1%. Another NGO is already distributing food to 15,000 people in phase 3 or above in district A. District B's markets are functioning; district A's are severely disrupted.

Task. 1) Classify each district according to the 20% rule. 2) Calculate the number of people in phase 3 or above in each district and the number not covered. 3) Recommend an area and a modality, and justify in three sentences.

Answer. District A: phases 4 and 5 = 13%, phases 3 to 5 = 38%: phase 3. People in phase 3 or above: 0.38 × 60,000 = 22,800; not covered: 22,800 − 15,000 = 7,800. District B: phases 3 to 5 = 18%, below the threshold; phases 2 to 5 = 55%: phase 2. People in phase 3 or above: 0.18 × 80,000 = 14,400, all not covered. District B has more people not covered (14,400 against 7,800) despite a less severe classification, and its markets allow cash transfers. A defensible response: cash transfers targeted at households in phase 3 or above in district B, and in-kind support in district A in coordination with the other NGO. An area's classification is not enough to decide: absolute numbers and coverage must be considered.

Key takeaways

- Every assessment question must serve a decision known in advance.
- Secondary data review first, with reliability and credibility ratings and a gap analysis.
- Only a representative survey produces generalisable proportions; KIIs and FGDs give trends and explanations.
- 20% rule: an area takes the most severe phase reached, together with the higher phases, by at least 20% of its population.
- Response analysis compares options on weighted criteria set in advance, taking existing coverage into account.

Quiz

1. Distribution of an area: phase 1: 40%; phase 2: 35%; phase 3: 15%; phase 4: 8%; phase 5: 2%. According to the 20% rule, the area is in:

   A. Phase 2

   B. Phase 3

   C. Phase 4

   D. Phase 1

2. Why can't the people in need of each sector be added up to obtain an intersectoral total?
   A. Because sectors use different currencies
   B. Because the same person may be in need in several sectors and would be counted several times
   C. Because the total would be too low
   D. Because donors forbid it

3. Which method can estimate a representative proportion of households without access to safe water?
   A. Three interviews with village chiefs
   B. Two focus group discussions
   C. A household survey on a correctly sized random sample
   D. Observation of a water point

4. According to the Washington Group Short Set, a person is counted as living with a disability if:
   A. They have a medical certificate
   B. They answer "some difficulty" to one question
   C. They answer "a lot of difficulty" or "cannot do at all" to at least one of the six questions
   D. They are over 60

5. In a weighted matrix, why set the weights before scoring the options?
   A. To go faster
   B. To avoid choosing an option first and then adjusting the criteria to justify it
   C. Because the donor requires it
   D. To obtain higher scores

Quiz answers : 1. B · 2. B · 3. C · 4. C · 5. B

Week 2 : Designing a programme: results chain, indicators and targets

Objectives

- Build a coherent results chain, from activity to impact, and a theory of change whose assumptions are testable.
- Choose valid, measurable indicators, with a definition, baseline, target and source, and use the sector's standard indicators.
- Set realistic targets from baseline data and real capacity, and integrate gender, age and disability from the design stage.

The results chain: what is yours and what is not

A **results chain** links five levels: **inputs** (money, staff, equipment), **activities** (train, distribute, build), **outputs** (direct products: 400 latrines built, 1,200 households trained), **outcomes** (changes in behaviour or situation: 80% of households use the latrines), **impact** (lasting long-term change: fewer cases of diarrhoea among children under five).

The decisive boundary lies between output and outcome. The organisation **controls** its outputs: if the latrines are not built, that is its responsibility. It **influences** outcomes, which also depend on people and context. It **contributes** to impact, along with many other actors. A report that presents only outputs says what was done, not what changed; a project that promises impact without a measurable outcome promises what it cannot demonstrate.

> The "so what?" test: for each output, ask "so what?" until you reach a change that matters to people. If the answer stops at "people received a kit", the programme has no defined outcome.

Each step from one level to the next rests on **assumptions**. Building latrines leads to their use only if they are culturally acceptable, safe at night for women, maintained and accompanied by awareness-raising. These assumptions are exactly what monitoring must check.

From theory of change to logical framework

The **theory of change** explains why and how the expected change will happen, with its causal assumptions, the evidence supporting them and the role of other actors. The **logical framework** (logframe) is its operational translation: a matrix of four columns (intervention logic, indicators, means of verification, assumptions) and four rows (overall objective, specific objective, results, activities).

The **vertical logic** reads from bottom to top: if the activities are carried out and the assumptions at that level hold, then the results are achieved; if the results are achieved and the assumptions hold, then the specific objective is achieved. The **horizontal logic** reads by

row: each level has measurable indicators and a source that actually allows them to be measured.

Frequent errors in proposal reviews:

- A specific objective that restates the results ("households received transfers") instead of a change ("targeted households improve their food consumption").
- Several specific objectives in a small project, which scatter measurement.
- Assumptions that are in fact risks the organisation should manage ("staff are recruited on time"). An assumption is outside the project's control; if it is critical and highly unlikely, it is a killer assumption and the project must be redesigned.
- Means of verification that nobody will produce ("national statistics" that do not exist at district level).

Indicators that really measure

A good indicator is **valid** (it measures what it claims to measure), **reliable** (two enumerators obtain the same result), **sensitive** (it moves if the situation changes within the project's duration), **simple** to collect and **useful** for a decision. The SMART criterion (specific, measurable, achievable, relevant, time-bound) applies mainly to the associated target.

Each key indicator must have a **reference sheet** (PIRS): precise definition, numerator and denominator, unit, disaggregation, collection method and tool, frequency, person responsible, baseline and target. Without a sheet, two offices calculate the same indicator in two ways and the figures do not add up.

Example of a rigorous definition. "Percentage of targeted households with an acceptable food consumption score (above 35)". Numerator: households in the sample with a score above 35. Denominator: targeted households in the sample. Disaggregation: sex of the head of household, presence of a person with a disability, status (displaced, host).

The sector has **standard indicators** that are better used than reinvented, because they allow comparison and aggregation: food consumption score (FCS: thresholds of 21 and 35 in most contexts, 28 and 42 where oil and sugar consumption is high), reduced coping strategies index (rCSI), household hunger scale (HHS), mid-upper arm circumference (MUAC: below 115 mm for severe acute malnutrition in children aged 6 to 59 months), indicators of the WASH Joint Monitoring Programme (JMP), and the clusters' indicator registries.

Baselines and realistic targets

A target without a **baseline** is meaningless: "70% of households with acceptable consumption" is huge progress if you start from 30%, and a decline if you start from 80%. The baseline is measured before activities start, with the same method as the final measurement. When that is impossible (sudden emergency), recent secondary data are used and measurement takes place in the first weeks, with documentation.

A realistic target is calculated, not guessed. It combines: the baseline, the project's real **coverage** (share of the population concerned that is reached), the expected **effectiveness** of

the intervention according to the available evidence, and external factors (season, prices).

Example. Baseline: 40% of households with acceptable consumption. The project reaches 80% of targeted households. Comparable projects moved 50% of the households reached from insufficient to acceptable consumption. Among the 60% of households below the threshold, 0.8 × 0.5 = 40% move above it: 0.6 × 0.4 = 24 points. Target: 40 + 24 = 64%, which will be prudently rounded to 60% if the lean season is harder than expected.

> An overambitious target does not motivate: it forces you to explain a failure that is not one, and pushes people to embellish the data.

Gender, age, disability and inclusion from the design stage

Inclusion is not added at the end: it is designed. The IASC **Gender with Age Marker** (GAM) assesses projects on concrete elements: analysis of differentiated needs, adapted activities, participation of different groups in design and decision-making, and accessible feedback mechanisms. A project that does not collect disaggregated data cannot show that women, older people or people with disabilities benefit from it.

In practice: disaggregate indicators by sex, age (at least by brackets 0-5, 6-17, 18-59, 60 and over) and disability; adapt sites (ramps, schedules, distance, priority queues); include the cost of accessibility in the budget (often 3 to 5% of the costs of the activities concerned); consult women and men, young and old, separately.

The Humanitarian inclusion standards for older people and people with disabilities and the IASC Guidelines on the inclusion of persons with disabilities (2019) set out minimum actions: **identify** (disaggregated data), **remove barriers** (physical, communication, attitudinal), **empower** (participation in decisions) and **build the capacity** of organisations of persons with disabilities.

Checking the design before submitting it

Before submitting a proposal, a short internal review prevents most rejections and bad surprises during the project. Five questions are enough:

- **Coherence**: does each activity contribute to a result, and each result to the specific objective? An activity without a result should be removed; a result without activities is unrealistic.
- **Measurability**: does each indicator have a definition, baseline, target, source, budget and person responsible?
- **Proportionality**: does the monitoring system cost between 3 and 10% of the budget (the usual range), in line with what the donor and management require?
- **Assumptions**: are critical assumptions tracked by a context indicator?
- **People**: have communities been consulted on the modality, targeting and complaints mechanisms?

The logframe is not set in stone. It is revised when the context changes, provided the reason is documented, the donor is consulted on any change of objective or target, and versions are tracked. A logframe that has never changed in a volatile context is often a logframe that nobody uses.

Exercise: repair a logframe and set a target

Situation. A 12-month WASH project proposes: specific objective "400 latrines built"; result 1 "households adopt good hygiene practices"; specific objective indicator "number of hygiene kits distributed"; assumption "staff will be recruited on time". Baseline: 35% of households wash their hands with soap at critical times. The project will reach 70% of households in the area; in comparable projects, 40% of the households reached that did not have the practice adopt it.

Task. 1) Identify the four errors. 2) Propose a correct specific objective and indicator. 3) Calculate a realistic target for handwashing.

Answer. Errors: the specific objective is an output; result 1 is an outcome placed at results level; the indicator measures a distribution, not a change; the assumption is an internal risk that the organisation must manage. A possible specific objective: "reduce the risk of water- and sanitation-related diseases in targeted households", indicator: "percentage of targeted households that wash their hands with soap at critical times (observation and self-report)". Target: 65% of households do not have the practice; 0.7 × 0.4 = 28% of them adopt it; 0.65 × 0.28 = 18.2 points. Target: 35 + 18 = about 53%.

Key takeaways

- The organisation controls its outputs, influences outcomes and contributes to impact.
- The logframe reads from bottom to top (if… then…) and by row (indicator, source, assumption).
- A key indicator has a reference sheet: definition, numerator, denominator, disaggregation, method, frequency, person responsible.
- Target = baseline + coverage × expected effectiveness applied to the share of the population concerned.
- Disaggregate by sex, age and disability from the design stage, and budget for accessibility.

Quiz

1. "1,500 households received a hygiene kit" is:
   A. An impact
   B. An outcome
   C. An output
   D. An assumption

2. A critical and highly unlikely assumption in a logframe is:
   A. Unimportant
   B. A killer assumption: the project must be redesigned
   C. An indicator
   D. A risk to ignore

3. Baseline: 50% of households with acceptable consumption. Coverage: 60%; effectiveness: 50% of the households reached below the threshold move above it. Calculated target:
   A. 65%
   B. 80%
   C. 55%
   D. 60%

4. Using the most common thresholds, a food consumption score of 30 is:
   A. Acceptable
   B. Borderline
   C. Poor
   D. Impossible

5. Which element is most often missing from a poorly defined indicator?
   A. A title
   B. A precise numerator and denominator
   C. A colour
   D. A translation

Quiz answers : 1. C · 2. B · 3. A · 4. B · 5. B

Week 3 : Targeting and modalities: who to help, how, and with how much

Objectives

- Choose a targeting method suited to the context and measure its inclusion and exclusion errors.
- Compare in-kind aid, vouchers and cash transfers on explicit criteria, including risks.
- Calculate a transfer value from a minimum expenditure basket and adjust it for inflation.

Why target, and what poor targeting costs

With resources below needs, **targeting** decides who receives aid. It produces two errors. The **inclusion error**: households that do not meet the criteria receive aid (waste, loss of trust). The **exclusion error**: eligible households do not receive it (direct harm, often to the most marginalised). Reducing one generally increases the other: very strict criteria exclude vulnerable households; broad criteria include households that do not need it.

Calculation. Of 1,000 households assessed, 400 are truly eligible. The project selects 450, of which 330 are eligible. Inclusion error: 120 / 450 = 26.7% of the households assisted are not eligible. Exclusion error: (400 − 330) / 400 = 17.5% of eligible households are left out.

> In humanitarian work, the exclusion error is the more serious one when survival is at stake. A targeting policy must state explicitly which of the two errors it is more willing to accept, and why.

Targeting also has a **cost**: surveys, verification, complaints, social tensions. When more than 70 to 80% of an area's population is in need, a general distribution (geographic targeting only) is often fairer and cheaper than household-by-household targeting.

Targeting methods

Methods are often combined in several stages:

- **Geographic**: choose the most severe areas (week 1). Simple and hard to dispute, but imprecise within an area.
- **Categorical**: observable criteria (household headed by a single woman, person with a disability, child under two, recently displaced). Transparent, but these categories only partly match real vulnerability.
- **Community-based**: a representative committee identifies households according to publicly explained criteria. Uses local knowledge, but risks elite capture and the exclusion of minorities. It is controlled through random verification and an independent complaints mechanism.

- **Vulnerability score** (proxy means test): a household questionnaire produces a score calculated from variables correlated with poverty (assets, housing, composition). More objective, but costly, opaque to people and sensitive to false declarations.
- **Self-targeting**: the aid attracts only those who need it (cash for work at a below-market wage, less valued products). Cheap, but it can exclude people who cannot work.

The choice depends on four factors: the proportion of people in need, the availability of reliable data, social cohesion (risk of tensions if targeting appears unfair), and the time available. In all cases, the **criteria must be public** and understood by the community, and a complaints mechanism must allow an exclusion to be challenged.

In-kind, vouchers or cash transfers

Cash transfers now account for about one fifth of international humanitarian assistance. Comparative data show that, for equivalent results, they often cost less to deliver than in-kind aid, let people choose according to their priorities and support local markets. They are not always the right answer.

Criterion	In-kind	Vouchers	Unrestricted cash
Non-functioning markets	Suitable	Not very suitable	Unsuitable
Need for a specific product (therapeutic food, certified seeds)	Suitable	Suitable	Not very suitable
Needs varying between households	Not very suitable	Moderate	Suitable
Logistics cost	High	Medium	Low
Dignity and choice	Low	Medium	High

The **risks** of cash can be managed: local inflation (price monitoring, transfer caps), insecurity (electronic payment, split distributions), diversion by a household member (choice of recipient based on gender analysis), sanctions and counter-terrorism (vetting of financial service providers). The idea that cash would be diverted to alcohol or tobacco is contradicted by many studies: spending on "temptation goods" does not increase significantly.

Multipurpose cash (MPCA) covers several basic needs at once; it is coordinated within the cash working group, which sets a common transfer value so that organisations do not compete with each other.

Calculating a transfer value

A transfer value starts from the **minimum expenditure basket** (MEB): the monthly cost of the goods and services a household needs to cover its basic needs (food, water, hygiene, rent,

energy, transport, communication). It is calculated for a typical household size, from prices recorded in markets.

The transfer value is not necessarily the whole MEB: what the household already covers itself (income, other aid) is deducted, and the programme's objective is taken into account. **Basic formula: gap = MEB − average resources of the target household.**

Example. Monthly MEB for a household of 6: 210 USD, of which 130 for food. Average monthly income of targeted households: 75 USD. Gap: 135 USD. The cash working group has set a common value of 80% of the gap to avoid creating dependency and to preserve own income: 0.8 × 135 = 108 USD per month.

Inflation. If the price of the basket increases by 12% in three months, the real value of a fixed transfer of 108 USD falls to 108 / 1.12 = 96 USD in initial purchasing power. Programmes therefore include an adjustment rule: for example, revision as soon as the MEB changes by more than 10%, based on monthly price monitoring.

> Household size: a fixed amount per household disadvantages large households. Programmes often use brackets (1-3, 4-6, 7 people and more) with a cap, to remain simple to explain.

Payment mechanisms and data protection

The **delivery mechanism** is chosen after the modality: hand-delivered cash, payment agents, mobile money transfer, prepaid card, bank account. The criteria: network coverage, access of women and older people to phones, distance to cash-out points, fees, agent liquidity, and identification requirements ("know your customer").

Electronic payment reduces the risk of theft and improves traceability, but creates **data risks**: beneficiary lists shared with an operator, authorities or other actors, which can expose people at risk (refugees, minorities). The ICRC Handbook on data protection in humanitarian action recommends: data minimisation, a defined purpose, written sharing agreements, a data protection impact assessment before any new system, and a limited retention period.

Reconciliation compares three lists every month: registered households, payments sent, payments actually withdrawn. Discrepancies (uncollected payments, duplicates, phone numbers shared by several beneficiaries) are signs of error or fraud to be dealt with quickly.

Post-distribution monitoring

Post-distribution monitoring (PDM) checks, two to six weeks after a distribution or transfer, that the aid arrived, that it is being used and that it is producing the expected effects. It covers a random sample of beneficiaries and has four blocks:

- **Process**: amount or quantities received, waiting time, distance, travel cost, safety, respectful treatment.
- **Use**: expenses made with the transfer, resale of in-kind goods, sharing.
- **Outcomes**: food consumption score, coping strategies, debt.

- **Accountability**: knowledge of the criteria and the complaints mechanism, satisfaction, problems reported (requests for payment, harassment).

A PDM is valuable only if it triggers decisions. For example: if more than 10% of beneficiaries report having paid something to be registered, an investigation is opened and the registration process is changed; if 40% resell the oil received, the composition of the basket is adjusted or the programme switches to cash.

Exercise: assess targeting and set a transfer

Situation. Of 2,000 households assessed, an independent verification shows that 900 are truly eligible. The community committee selected 1,000, of which 700 are eligible. Monthly MEB: 180 USD; average income of targeted households: 60 USD; common rule: 75% of the gap. Prices rise by 15% in the third month.

Task. 1) Calculate the inclusion and exclusion errors. 2) Calculate the transfer value. 3) Calculate its real value after the increase and propose a decision.

Answer. Inclusion: 300 / 1,000 = 30%. Exclusion: (900 − 700) / 900 = 22.2%. Both errors are high: the committee probably favoured certain households; random verification, an independent complaints mechanism and a review of the list are needed. Transfer: gap 120 USD; 0.75 × 120 = 90 USD. After the increase: 90 / 1.15 = 78.3 USD in initial purchasing power, a loss of 13%. The increase exceeds the usual 10% threshold: propose to the cash working group a revision to 90 × 1.15 = 103.5 USD, after first checking that the transfers themselves are not causing local inflation.

Key takeaways

- Inclusion error = ineligible households assisted / total assisted; exclusion error = eligible households not assisted / total eligible.
- Community-based targeting must be controlled by random verification and an independent complaints mechanism.
- The choice between in-kind, vouchers and cash depends first on how markets function and on the need.
- Transfer value = agreed share of (MEB − household resources), adjusted for inflation according to a written rule.
- Post-distribution monitoring must trigger decisions, not just produce a report.

Quiz

1. 500 households assisted, of which 400 are eligible; 600 eligible households in total. The exclusion error is:
   A. 20%
   B. 33.3%
   C. 16.7%
   D. 25%
2. In an area where 85% of the population is in need, the fairest and cheapest method is often:
   A. A household-by-household vulnerability score
   B. A general distribution with geographic targeting
   C. A lottery
   D. Targeting by the authorities
3. A transfer of 100 USD after a 25% price increase is worth, in initial purchasing power:
   A. 75 USD
   B. 80 USD
   C. 125 USD
   D. 100 USD
4. Which condition makes cash transfers unsuitable?
   A. Households with varied needs
   B. Markets that do not function or cannot respond to demand
   C. Widespread mobile phone use
   D. People who prefer cash
5. Monthly payment reconciliation compares:
   A. Headquarters' budget and expenditure
   B. Registered households, payments sent and payments withdrawn
   C. Prices in two markets
   D. Reports from two donors

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 4 : Planning and budgeting a programme: from work plan to burn rate

Objectives

- Build a realistic work plan: task breakdown, dependencies, critical path and float.
- Link the budget to activities through unit costs, and calculate cost per person and cost per result.
- Monitor implementation with the burn rate and detect underspending and overspending early.

From logframe to work plan

The logframe says what; the **work plan** says who does what, when and with what resources. Start from the logframe activities and break them down into tasks of 1 to 4 weeks (work breakdown structure): to "distribute transfers to 3,000 households", you must register, verify, contract a payment provider, train the teams, inform communities, pay, and carry out post-distribution monitoring.

Each task receives a **duration**, a single **person responsible** and its **dependencies**: most are "finish-to-start" (payment can only start after the contract is signed). The Gantt chart shows these tasks over time, but it is not enough: the critical path must be identified.

The work plan also integrates **context constraints**: rainy season cutting off roads, lean season, religious holidays, elections, agricultural calendar, and administrative delays (authorities' approval, customs clearance, recruitment). In many contexts, recruiting for a key post takes 8 to 12 weeks and importing goods 3 to 6 months; a plan that ignores this is wrong from day one.

> The first quarter of a project is almost always late. Planning expenditure in a linear way (one twelfth per month) guarantees a large variance by the third month.

The critical path and float

The **critical path** is the longest sequence of dependent tasks; its duration is the project's minimum duration. Any task on the critical path that is delayed delays the whole project. The other tasks have **float**: they can slip without consequence, within the limit of that float.

Example. A: assessment (3 weeks). B: tender for the payment provider (6 weeks, after A). C: household registration (4 weeks, after A). D: verification of lists (2 weeks, after C). E: first payment (1 week, after B and D).

- Path A-B-E: 3 + 6 + 1 = 10 weeks.
- Path A-C-D-E: 3 + 4 + 2 + 1 = 10 weeks.

Both paths last 10 weeks: both are critical and no task has any float. A one-week delay on registration or on the tender pushes back the first payment. If the tender took 4 weeks, path A-B-E would take 8 weeks and task B would have 2 weeks of float.

To **shorten** a project, act only on the critical path: speeding up a task that has float changes nothing. Two techniques: do in parallel what was in sequence (start registration during the end of the assessment, with a risk of rework) and add resources to a critical task (two registration teams instead of one).

The programme budget built on unit costs

A solid budget is built **from activities**, with verifiable unit costs: quantity × unit cost × duration or frequency. Each line has a justification (quotation, market price, salary scale, history). Donors mostly reject budgets where amounts are round, have no basis of calculation, or have no link with the logframe.

The budget usually distinguishes: **direct activity costs** (transfers, inputs, training), **staff** (with the share of time charged to the project), **office running costs** shared between projects according to a documented allocation key, **monitoring and evaluation**, and **indirect costs** (percentage set by the donor, often 7 to 10%).

Shared costs. An office costs 20,000 USD per month and serves three projects. The most defensible key is often staff time or budget share: if the project accounts for 40% of the office's direct expenditure, it takes 8,000 USD per month. The key must be written down, applied consistently and verifiable by an auditor.

> A budget is not a ceiling to be reached: it is a forecast. Spending at the end of the project to "use up" the budget is a management fault that shows up in audits.

Cost per person and cost per result

Two cost indicators inform decisions, provided they are well defined. The **cost per person reached** divides the total cost by the number of people assisted. It is simple but misleading: it favours programmes that give little to many people and says nothing about the effect.

The **cost per result** divides the cost by the effect obtained: cost per child cured of severe acute malnutrition, cost per household moving from insufficient to acceptable consumption, cost per water point still functional after one year. It allows approaches to be compared.

Example. Project A: 600,000 USD, 4,000 children admitted, cure rate 70%: cost per child cured = 600,000 / 2,800 = 214 USD. Project B: 450,000 USD, 3,500 children admitted, cure rate 55%: 450,000 / 1,925 = 234 USD. Project B seems cheaper per child admitted (129 against 150 USD), but it is more expensive per child cured.

The **share of costs that directly reaches people** (transfer ratio, or cost-transfer ratio for cash) measures delivery efficiency: for a cash programme, amount transferred / total cost. A ratio of 0.80 means that 80 cents of every dollar reach households. It is compared between similar programmes, never between very different modalities or contexts.

Monitoring implementation: the burn rate

The **burn rate** compares cumulative actual expenditure with the expenditure planned in the cash-flow plan, not with a straight line. It is monitored every month, line by line, by the finance manager and the project manager together.

Example. 12-month budget: 900,000 USD. End of month 6: cumulative planned expenditure 420,000 USD (the plan allowed for a slow start), actual expenditure 290,000 USD. Implementation rate: 290,000 / 420,000 = 69%. At the current average rate (290,000 / 6 = 48,333 USD per month), 610,000 USD would remain to be spent in 6 months, i.e. 101,667 USD per month: more than double the current rate. Action is needed now.

The causes of underspending are almost always operational: delays in access, recruitment or procurement, postponed activities. The response follows this order: understand the causes line by line; speed up what can be sped up without loss of quality; propose a budget revision or a no-cost extension to the donor **before** the last months (most donors require a request at least 30 to 60 days before the end); never make unnecessary end-of-project purchases.

Conversely, an overspend on one line is offset within the **flexibility** provided in the contract (often 10 to 20% between lines or headings without prior approval). Beyond that, the donor's written approval is mandatory, otherwise the expenditure risks being declared ineligible.

Programme human resources and organisation

The staffing plan follows from the work plan: how many enumerators are needed to register 3,000 households in 4 weeks? If one enumerator conducts 12 interviews per day over 20 working days per month, a team of 13 enumerators conducts 3,120 interviews in 4 weeks. The calculation avoids both teams that are too small (delays) and too large (costs, quality).

The **programme organisation chart** clarifies responsibility for each result and the segregation of duties: the person who selects beneficiaries must not be the one who approves payments or the one who handles complaints. The **RACI** matrix (responsible, accountable, consulted, informed) for each key process (registration, payment, monitoring, complaints) makes these separations visible and verifiable.

Finally, the programme manager organises a stable **management rhythm**: weekly team meeting on the work plan; monthly review of indicators, budget and risks; quarterly review with country management, which can decide on reallocations. A programme without a management rhythm discovers its problems in the final report.

Exercise: critical path and burn rate

Situation. Tasks: A, assessment (2 weeks); B, team recruitment (5 weeks, after A); C, purchase of kits (8 weeks, after A); D, team training (1 week, after B); E, distribution (2 weeks, after C and D). Total budget: 600,000 USD over 10 months. At the end of month 5, cumulative planned expenditure: 260,000 USD; actual expenditure: 182,000 USD.

Task. 1) Calculate the minimum duration and the critical path. 2) What is the float on recruitment? 3) Calculate the implementation rate and the monthly rate needed to spend the remaining budget.

Answer. A-C-E: 2 + 8 + 2 = 12 weeks. A-B-D-E: 2 + 5 + 1 + 2 = 10 weeks. Critical path: A-C-E, 12 weeks; recruitment has 2 weeks of float. To bring the distribution forward, act on the purchase of kits (launch the tender during the assessment, or use a framework agreement). Implementation: 182,000 / 260,000 = 70%. Remaining: 418,000 USD in 5 months, i.e. 83,600 USD per month against an average of 36,400 so far. The late lines must be identified and a revision or no-cost extension discussed with the donor now.

Key takeaways

- Only the critical path determines the project's duration; speeding up a task that has float is useless.
- A budget is built by activity: quantity × unit cost × duration, with a justification for each line.
- Cost per result compares approaches better than cost per person reached.
- The burn rate is compared with the cash-flow plan, not a straight line, and is monitored every month.
- Budget revision or no-cost extension: request it early, never resort to end-of-project purchases.

Quiz

1. A (2 wks), then B (4 wks) and C (3 wks) in parallel after A, then D (1 wk) after B and C. Minimum project duration:

   A. 10 weeks

   B. 7 weeks

   C. 6 weeks

   D. 9 weeks

2. Project of 300,000 USD, 1,500 children admitted, cure rate 80%. Cost per child cured:
   A. 200 USD
   B. 250 USD
   C. 240 USD
   D. 160 USD

3. To shorten a project, you act first on:
   A. The tasks with the most float
   B. The tasks on the critical path
   C. The indirect cost budget
   D. The final report

4. A cash programme transfers 720,000 USD for a total cost of 900,000 USD. The transfer ratio is:
   A. 0.80
   B. 1.25
   C. 0.72
   D. 0.20

5. Four months before the end, 40% of the budget is unspent because of access delays. The best response is:
   A. Buy equipment to use up the budget
   B. Analyse the causes line by line and ask the donor early for a revision or a no-cost extension
   C. Say nothing to the donor
   D. Return the money without explanation

Quiz answers : 1. B · 2. B · 3. B · 4. A · 5. B

Week 5 : The monitoring system: MEAL plan, data quality and remote monitoring

Objectives

- Build a complete and proportionate MEAL plan, from the data collection form to the management decision.
- Assess and improve data quality along five dimensions, and organise regular quality checks.
- Organise monitoring in inaccessible areas: remote monitoring, third-party monitoring and triangulation.

The MEAL plan: the backbone of monitoring

The **MEAL plan** describes how each indicator will be measured, by whom, when, and for what decision. It takes the form of a table: indicator, definition, baseline, target, source and method, frequency, disaggregation, person responsible for collection, person responsible for analysis, use (monthly review, donor report, targeting adjustment).

It covers four types of monitoring:

- **Activity and output monitoring**: what has been done, against the plan (distribution registers, attendance lists, worksite sheets).
- **Outcome monitoring**: changes among people (baseline, midline and endline surveys; post-distribution monitoring).
- **Context monitoring**: prices, access, security, population movements, and indicators for the logframe's critical assumptions.
- **Accountability monitoring**: complaints and feedback received, response time, satisfaction, participation (week 8).

The cost of the MEAL plan must be in the budget: staff, tools, transport, surveys, evaluations. A range of 3 to 10% of the project budget is usual; a pilot project whose effect is to be demonstrated can justify more.

> The best test of a MEAL plan: for each indicator, can you name the meeting where it will be discussed and the decision it can trigger? If not, ask yourself whether it needs to be collected.

From collection tools to the dashboard

The data chain has five links: **collection**, **entry and storage**, **cleaning**, **analysis**, **dissemination**. Mobile data collection (KoboToolbox, ODK, CommCare) has replaced paper in most contexts: automatic checks at entry (value ranges, required fields, skip logic), geolocation and timestamps, direct upload to the server.

A good form is designed from the analysis plan: first write the tables you want to produce, then the questions needed and only those. It is tested in the field with real respondents before deployment, then a version is frozen; any later change is documented, because it makes data before and after incomparable.

Cleaning follows written rules: duplicates, outliers (a household of 45 people, food expenditure of 2,000 USD per week), inconsistencies (a married 3-year-old), interviews that are too short (under 10 minutes for a 40-minute questionnaire). A value is never corrected without a source; it is flagged, checked with the enumerator or removed, and a cleaning log is kept.

The **dashboard** shows few indicators, with the target, the trend and a colour code according to thresholds set in advance. It is updated at the pace of decisions: weekly for distributions, monthly for intermediate outcome indicators.

Data quality: five dimensions

A **data quality assessment** (DQA) checks five dimensions:

Dimension	Question	Example of a problem
Validity	Do the data measure what the indicator intends?	Counting people trained instead of those applying the practice.
Reliability	Are methods stable across places, people and periods?	Two offices calculate the cure rate differently.
Precision	Is the margin of error small enough for the decision?	A sample of 50 households to track a 5-point change.
Timeliness	Do the data arrive in time to decide?	The monthly report arrives six weeks after the end of the month.
Integrity	Are the data protected against errors and manipulation?	The same officer enters, approves and corrects payment lists.

The most concrete method is **verification by recount**: choose a sample of sites, recount from primary sources (registers, forms) and compare with the reported figures. A **verification ratio** of 1.00 is perfect; a tolerance is set, for example between 0.95 and 1.05. Above 1.05, the reported figures are inflated; below 0.95, activities are not being reported.

Example. A health centre reports 1,240 consultations; the recount of the registers finds 1,085. Ratio: 1,240 / 1,085 = 1.14. The centre over-reports by 14%: the cause must be understood (double counting of follow-up visits, pressure on targets) before any sanction.

Quality checks during data collection

Quality is decided during collection, not afterwards. The standard checks:

- **Training and testing** of enumerators, with a written test and mock interviews; enumerators who fail do not collect data.
- **Direct supervision** of part of the interviews (for example 5%).

- **Back-checks**: 10% of households are re-interviewed on a few stable questions (household size, type of shelter). A low agreement rate for one enumerator signals errors or fabricated interviews.
- **Daily remote checks**: interview duration, GPS coordinates (interviews all in the same place), time (interviews at 3 a.m.), distribution of answers by enumerator (an enumerator whose households all have exactly 5 members).

Fabricated data ("curbstoning") are more frequent than people think when enumerators are paid per interview, under quota pressure, or sent to dangerous areas. Preventing them requires realistic targets, daily pay rather than piece rates, and checks announced in advance.

> A MEAL manager who never finds an error is not looking hard enough. Publishing internally the errors found and corrected strengthens confidence in the figures.

Remote monitoring and third-party monitoring

In areas inaccessible to staff (insecurity, restrictions), **remote monitoring** combines several sources to make up for the lack of visits:

- **Local partners** or local staff who have access, with standardised tools and evidence (geotagged photos, signed lists).
- **Third-party monitoring** (TPM): an independent company or organisation, often required by donors, which verifies a sample of activities.
- **Phone calls** to a random sample of beneficiaries (post-distribution monitoring by phone).
- **Remote data**: satellite imagery (camps, crops, destruction), mobile operators' transaction data, price data.
- **Community feedback** via a toll-free line and messaging.

Remote monitoring has **limits**: phone samples biased towards households with phones, often represented by men; risks of pressure on partners; transfer of security risks to local staff. This must be stated in reports, partners must be properly paid for this work, and they must not be asked to take risks the organisation would refuse for its own staff.

Triangulation is the rule: a distribution is considered verified when at least two independent sources agree (for example a signed list and calls to a sample of beneficiaries).

Making monitoring useful: from data to decision

Monitoring rarely fails for lack of data; it fails because nobody uses them. Three mechanisms make monitoring useful:

- **Alert thresholds** set in advance: if waiting time at distributions exceeds 3 hours, if the rate of unresolved complaints exceeds 20%, if the price of the basket rises by more than 10%, a predefined action is triggered.
- **Regular reviews** where the programme, MEAL, logistics and finance teams look at the same figures together and decide. The minutes record the decisions, who is responsible and the deadline.

- **Feedback to people and teams**: present to communities what has been learned and changed; show enumerators what their data were used for.

The independence of MEAL matters. If the MEAL manager reports only to the project manager whose results they measure, the temptation to smooth over bad figures is strong. A dual reporting line (operational to the project, technical to a country or headquarters MEAL manager) protects the quality of information.

Exercise: check the data quality of three sites

Situation. A nutrition project reports the admissions of three sites for one month. Site 1: reported 320, recounted 312. Site 2: reported 410, recounted 342. Site 3: reported 198, recounted 205. Tolerance: ratio between 0.95 and 1.05. Back-checks on the coverage survey show an agreement rate of 96% for all enumerators except one (61%), whose interviews last on average 9 minutes against 35 for the others.

Task. 1) Calculate the verification ratios and interpret them. 2) What do you do with the suspect enumerator's data? 3) What preventive measures do you propose?

Answer. Site 1: 320 / 312 = 1.03, within tolerance. Site 2: 410 / 342 = 1.20: over-reporting of 20%, to be investigated (double counting, pressure on targets, fraud). Site 3: 198 / 205 = 0.97, within tolerance (slight under-reporting). Total reported 928 against 859 recounted: the monthly report must be corrected and the donor informed. The suspect enumerator's data are set aside and some of their households re-interviewed; if falsification is confirmed, their interviews are redone or removed, and the cause is addressed (quotas, pay per interview). Prevention: daily pay, daily checks of duration and GPS, announced back-checks, realistic targets.

Key takeaways

- The MEAL plan links each indicator to a method, a person responsible, a frequency and a decision.
- Five dimensions of data quality: validity, reliability, precision, timeliness, integrity.
- Verification ratio = reported figure / recounted figure; beyond the tolerance, investigate before sanctioning.
- Back-checks, interview duration and GPS detect fabricated data.
- In remote monitoring, an activity is verified when two independent sources agree.

Quiz

1. A site reports 500 distributions; the recount finds 400. The verification ratio is:
   A. 0.80
   B. 1.25
   C. 1.20
   D. 0.90

2. Two offices calculate the same indicator differently. Which dimension of quality is at stake?
   A. Timeliness
   B. Reliability
   C. Integrity
   D. Precision

3. An enumerator conducts 8-minute interviews for a 35-minute questionnaire, all at the same GPS point. Most likely:
   A. They are very efficient
   B. Fabricated or rushed interviews
   C. A server failure
   D. A sample that is too large

4. Why avoid paying enumerators per interview?
   A. It is more expensive
   B. It encourages rushed or fabricated interviews
   C. Donors always forbid it
   D. It reduces the sample size

5. What is the main limitation of post-distribution monitoring by phone?
   A. It is too expensive
   B. The sample is biased towards households with a phone, often represented by men
   C. It is forbidden
   D. It allows no questions

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 6 : Surveys and sampling: getting the numbers right

Objectives

- Choose a sampling design (simple random, systematic, stratified, cluster) suited to the question and the field.
- Calculate a sample size with margin of error, finite population correction, design effect and non-response.
- Design a reliable questionnaire and correctly interpret a proportion and its confidence interval.

Population, sample and representativeness

A survey seeks to know a **population** (all displaced households in a district) from a **sample**. Results can be generalised only if each unit of the population has a **known, non-zero probability** of being selected: this is probability sampling. A "convenience" sample (households near the road, those present at the distribution) may be large, but it remains biased.

A **sampling frame** is needed first: a list of households, sites or villages, with their population. Without a reliable list, work in stages: select villages or blocks of a camp, then households within them, enumerating them on the spot.

The main designs:

- **Simple random**: random draw from the complete list. The simplest statistically, often impractical in the field.
- **Systematic**: calculate an interval (population / sample size), draw a random starting point, then take one household every "interval". For 2,400 shelters and 240 households to interview, interval = 10; start drawn between 1 and 10, for example 7: shelters 7, 17, 27…
- **Stratified**: the population is divided into groups (displaced and host, urban and rural areas) and a sample is drawn in each group. Essential if reliable results are wanted for each group.
- **Cluster**: first draw clusters (villages), then households in each cluster. Less costly in travel, but less precise: a larger sample is needed.

Calculating sample size

To estimate a proportion, the basic formula is: $\mathbf{n = z^2 \times p \times (1 - p) / e^2}$, where z is 1.96 for a 95% confidence level, p is the expected proportion (0.5 when unknown, because it is the most demanding case) and e the margin of error.

- 5% margin: $1.96^2 \times 0.25 / 0.0025 = 384$ households.
- 7% margin: 196 households.
- 10% margin: 96 households.

Halving the margin multiplies the sample by four. The choice of margin depends on the decision: to track a 10-point change between two surveys, a 10% margin is not enough.

Finite population correction. When the population is small, the required sample decreases: $n = n_0 / (1 + (n_0 - 1) / N)$. For N = 2,000 households: 384 / (1 + 383 / 2,000) = 384 / 1.19 = 323 households. For N = 100,000, the correction is negligible.

Design effect. In cluster sampling, households in the same village resemble each other, which reduces the information each provides. The sample is multiplied by the design effect: **DEFF = 1 + (m − 1) × ICC**, where m is the number of households per cluster and ICC the intra-cluster correlation coefficient. With 10 households per cluster and an ICC of 0.05: DEFF = 1 + 9 × 0.05 = 1.45. Without data, 1.5 to 2 is often used.

Non-response. Divide by the expected response rate: with 10% non-response, n / 0.9.

> Full calculation: 5% margin, p = 0.5, DEFF 1.5, 10% non-response: 384 × 1.5 = 576; 576 / 0.9 = 640 households, for example 64 clusters of 10 households.

Stratification and comparisons

If results are wanted for each stratum with the same precision, the sample size is needed **in each stratum**. For three districts with a 5% margin each, the sample is three times larger than for an overall estimate. This is often the main cause of overly expensive surveys: results are promised by district, by sex of the head of household and by status, without calculating the required sample.

When strata have different sizes and an overall result is wanted, **weighting** is applied: each stratum counts according to its share of the population, not its share of the sample. **Example.** Displaced: 20,000 households, sample 400, 60% food insecure. Host: 80,000 households, sample 400, 30% food insecure. Unweighted average: 45%. Weighted average: 0.2 × 60 + 0.8 × 30 = 36%. Presenting 45% would greatly overstate the situation.

To **compare two surveys** (baseline and endline) or two groups, the sample must be able to detect the expected difference: detecting a change from 40% to 50% with 95% confidence and 80% power requires about 385 households per group in simple random sampling, more with clusters. A small change requires a large sample.

Designing a reliable questionnaire

A questionnaire is built from the indicators and the analysis plan. Essential rules:

- **One idea per question**: "Do you have access to water and soap?" is two questions.
- **Defined recall period**: "In the last 7 days" for food, "in the last 30 days" for expenditure. A period that is too long produces memory errors.
- **Standard modules** for standard indicators (food consumption score, rCSI, Washington Group), unmodified, to remain comparable.
- **Neutral wording**: "What do you think of the distribution?" rather than "Are you satisfied with the distribution?".

- **Order**: simple questions at the beginning, sensitive ones at the end, and never a sensitive question without enumerator training and a referral option.
- **Translation and back-translation** into local languages, then field testing.
- **Length**: beyond 45 minutes, the quality of answers declines; remove what is not essential.

> Informed consent is read to every respondent: purpose of the survey, duration, voluntary nature, no link with aid, confidentiality. A refusal is respected and replaced according to the planned rule, never by the most convenient neighbour.

Interpreting a proportion and its confidence interval

A survey result is an **estimate**, surrounded by uncertainty. The 95% confidence interval of a proportion is approximately: $p \pm 1.96 \times \sqrt{(p \times (1 - p) / n)}$, multiplied by $\sqrt{DEFF}$ with clusters.

Example. 384 households in simple random sampling, 45% without sufficient access to water. Standard error: $\sqrt{(0.45 \times 0.55 / 384)} = 0.0254$. Margin: $1.96 \times 0.0254 = 5.0$ points. Interval: 40% to 50%.

Practical consequences:

- If the baseline was 45% (interval 40-50%) and the endline 49% (interval 44-54%), it cannot be claimed that the situation has worsened.
- Results by subgroup have wider intervals, because subgroups are smaller: 45% among 80 female-headed households has a margin of about 11 points.
- Always present the sample size, and the interval for key indicators.

Statistical significance does not mean practical importance: with a very large sample, a 1-point difference can be significant without changing any decision. Conversely, a large difference may not be significant with a small sample: then the data do not allow a conclusion, which is not the same as saying there is no effect.

Organising a field survey

A household survey is prepared in weeks, not days. The typical schedule:

- Week 1: objectives, indicators, analysis plan, sampling plan, authorities' approval.
- Week 2: questionnaire, translation, programming on tablets, tests.
- Week 3: recruitment and training of enumerators (3 to 5 days, with a test and practice), pilot test, corrections.
- Weeks 4 and 5: data collection, with daily quality checks and debriefing meetings.
- Week 6: cleaning, weighting, analysis, validation with the team and partners.
- Week 7: report and feedback, including to communities.

Logistics. Number of enumerators = sample size / (interviews per enumerator per day × days of collection). For 640 households, 4 interviews per enumerator per day, 10 days: 640 / 40 = 16 enumerators, i.e. 4 teams of 4 with one supervisor each. Allow 10 to 20% reserve days for contingencies (rain, security, access).

The **replacement** of absent or refusing households follows a written rule (return once, then replace with the next household in the drawn list); otherwise enumerators replace distant households with nearby ones, which biases the sample.

Exercise: size a baseline survey

Situation. A project wants to know, with a 5% margin and 95% confidence, the proportion of food-insecure households, separately for displaced people (3,000 households) and hosts (40,000 households). Cluster sampling with 12 households per cluster, ICC estimated at 0.06. Expected non-response: 8%.

Task. 1) Calculate the basic sample with the finite population correction for each stratum. 2) Calculate the design effect and the final sample per stratum. 3) How many clusters are needed in total?

Answer. Basic: 384. Displaced: 384 / (1 + 383 / 3,000) = 384 / 1.128 = 340. Hosts: 384 / (1 + 383 / 40,000) = 384 / 1.0096 = 380. Design effect: 1 + 11 × 0.06 = 1.66. Displaced: 340 × 1.66 = 565; 565 / 0.92 = 614. Hosts: 380 × 1.66 = 631; 631 / 0.92 = 686. Total: 1,300 households. Clusters: 614 / 12 = 51.2, i.e. 52 clusters; 686 / 12 = 57.2, i.e. 58 clusters; 110 clusters in total. For an overall result, the two strata will be weighted according to their share of the population (7% and 93%).

Key takeaways

- $n = z^2 \times p \times (1 - p) / e^2$: 384 for a 5% margin at 95% confidence with p = 0.5.
- Correction for a small population: $n = n_0 / (1 + (n_0 - 1) / N)$.
- Cluster design effect: $1 + (m - 1) \times ICC$; then divide by the response rate.
- A result per stratum requires the sample in each stratum; an overall result requires weighting.
- Always present the sample size and the confidence interval; not significant does not mean no effect.

Quiz

1. With p = 0.5, 95% confidence and a 10% margin, the simple random sample size is about:
   A. 38
   B. 96
   C. 196
   D. 384

2. Clusters of 20 households, ICC of 0.04. The design effect is:
   A. 1.04
   B. 1.76
   C. 1.80
   D. 0.80
3. Stratum A: 10,000 households, 50% affected. Stratum B: 30,000 households, 20% affected. The weighted overall proportion is:
   A. 35%
   B. 27.5%
   C. 30%
   D. 25%
4. Baseline: 52% (interval 47-57%). Endline: 55% (interval 50-60%). The most accurate conclusion:
   A. The situation has clearly improved
   B. The data do not allow a change to be concluded
   C. The situation has worsened
   D. The survey is wrong
5. For 2,000 shelters and 200 households to interview with systematic sampling, the interval is:
   A. 5
   B. 10
   C. 20
   D. 200

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 7 : Analysing data: useful statistics, pitfalls and visualisation

Objectives

- Choose the right descriptive statistics and disaggregate to reveal the inequalities hidden by averages.
- Avoid the classic pitfalls: correlation and causation, Simpson's paradox, misleading denominators.
- Analyse qualitative data rigorously and present results so that they are understood and used.

Descriptive statistics: mean, median, dispersion

The **mean** is sensitive to extreme values; the **median** (the value that splits the sample into two halves) is not. For income, expenditure, debt or waiting times, which are highly skewed distributions, the median better describes the typical household.

Example. Debts of ten households (USD): 0, 0, 20, 30, 40, 50, 60, 80, 100, 1,620. Mean: 2,000 / 10 = 200 USD. Median: (40 + 50) / 2 = 45 USD. Saying "households have an average debt of 200 USD" describes a single household, not the situation.

Dispersion matters as much as the centre: standard deviation, or interquartile range (between the 25th and 75th percentiles) for skewed distributions. Two programmes with an average waiting time of 2 hours are very different if one has waits between 1 h 30 and 2 h 30, and the other between 10 minutes and 7 hours.

For proportion indicators, present the numerator and the denominator: "62% (248 out of 400)". Percentages calculated on very small numbers ("50% of older people", i.e. 2 out of 4) should be avoided or flagged.

> Change in points and in percent: going from 20% to 30% is an increase of 10 percentage points, i.e. a relative increase of 50%. Confusing the two is the most frequent error in reports.

Disaggregating to see inequalities

A satisfactory overall average can hide a group left behind. **Disaggregation** (by sex and age, disability, status, area, household type) is the way to check that the programme reaches everyone.

Example. Overall satisfaction rate: 82%. By group: households without a person with a disability 86%; households with a person with a disability 54%. The programme does not work for an entire group, and the overall figure was masking it.

The **cross-tabulation** is the basic tool: an indicator in rows, a group in columns. To know whether the difference observed between two groups exceeds what sampling chance can produce, a test is used (chi-square test for proportions, t-test or non-parametric test for means) or confidence intervals are compared. Software calculates it; the manager's role is to ask the right question and to interpret.

Beware of **multiple comparisons**: if 20 differences are tested at the 5% threshold, one is expected to be "significant" by pure chance. The main comparisons are therefore defined in advance in the analysis plan, and the others are presented as exploratory.

Correlation, causation and Simpson's paradox

Two variables that move together do not prove that one causes the other. Three alternative explanations always exist: **chance**, a **common factor** (confounder) and **reverse causation**.

Example. Households that took part in agricultural training have better yields. But they are also those with more land and tools, and who were more motivated to sign up: the training may have an effect, but the simple comparison cannot separate it from these initial differences. Measuring an effect requires a counterfactual (week 10).

Simpson's paradox occurs when a trend present in each group reverses when the data are pooled, because of an unequal distribution between groups.

	Centre A	**Centre B**
Severe cases	60 cured out of 100 (60%)	110 cured out of 200 (55%)
Moderate cases	270 cured out of 300 (90%)	170 cured out of 200 (85%)
Total	330 out of 400 (82.5%)	280 out of 400 (70%)

Here, A is better in each group and overall: no paradox. But if B mainly received moderate cases and A mainly severe cases, B could appear better overall while being worse in each group. Rule: compare performance only with **comparable composition**, or by disaggregating by the factor that changes the difficulty.

Analysing qualitative data

Interviews and focus group discussions explain the "why" behind the numbers. Their analysis follows a method, without which one retains only the quotes that confirm what one already thought.

- **Transcription or detailed notes** on the same day, in people's own words.
- **Coding**: codes (themes) are assigned to passages. Start with codes linked to the evaluation questions, and add those that emerge.
- **Matrix**: themes in rows, groups (women, men, young people, displaced, hosts) in columns. This shows which themes recur in which groups.
- **Active search for counter-examples**: cases that contradict the general trend are often the most instructive.

- **Saturation**: data collection stops when new interviews no longer bring new themes.

Qualitative results are not presented as percentages, but frequency can be indicated: "this theme came up in 9 of the 12 focus groups, including all the women's groups". A quote illustrates a finding; it does not prove it.

> Mixed methods: the number says "30% of households do not collect their transfer", the qualitative data say "the cash-out point is 3 hours' walk away and agents often run out of cash". Together, they lead to a decision.

Visualising to make things understood

A chart must answer a question in a few seconds. Practical rules:

- **Horizontal bars** to compare categories, sorted by value; **lines** for trends over time; **maps** for differences between areas; avoid pie charts beyond three categories and 3D effects.
- The bar axis starts at **zero**; otherwise a 2-point difference looks huge.
- A **title that states the message** ("Coverage has fallen in the northern areas since March") rather than a neutral title ("Coverage by area").
- Few colours, one accent colour for what matters, and a marker for the target.
- The source, date and sample size at the bottom of the chart.

For a management dashboard, limit it to 6 to 10 key indicators, each with the current value, the target, the trend and a colour code with defined thresholds. A dashboard with 40 indicators is read by no one.

On a map, show **rates** rather than absolute numbers when areas have different populations: a large city will always have more cases than a village.

From finding to recommendation

A useful analysis follows a three-step progression: **finding** (what the data say), **interpretation** (why, using qualitative data and context) and **recommendation** (what should be done, by whom, when). Recommendations are few, precise and realistic, and each one flows from a finding.

The **level of confidence** must also be stated: strong (several converging sources, sufficient sample), medium, weak (a single source, small sample). Management can decide on medium evidence in an emergency, but it must know that it is doing so.

Finally, results are **presented back** before the final report: a session with the team and partners makes it possible to correct errors of interpretation and to build buy-in for the recommendations. Results are also shared with communities, in an appropriate form (meeting, poster, voice message).

Exercise: analyse a post-distribution monitoring survey

Situation. Post-distribution monitoring of 400 households. Waiting time: mean 2 h 40, median 1 h 20. Overall satisfaction: 78%. Female-headed households (120): 60% satisfied; male-headed households (280): 85.7% satisfied. The average food consumption score rose from 31 to 38; the proportion of households with acceptable consumption rose from 30% to 45%. In the focus group discussions, 8 out of 10 women's groups mention the distribution time (after 3 pm) and the journey home.

Task. 1) What does the gap between the mean and the median waiting time tell you? 2) Express the change in acceptable consumption in points and in percent. 3) What conclusion and what recommendation do you draw on satisfaction?

Answer. A mean much higher than the median indicates that most people wait only a short time, but that one group waits a very long time: you need to identify which sites or which days. Acceptable consumption: +15 percentage points, i.e. +50% in relative terms (this change cannot be attributed to the programme alone without a counterfactual). Satisfaction: check, $0.3 \times 60 + 0.7 \times 85.7 = 78\%$. The 26-point gap between female-headed and male-headed households is large and explained by the focus groups: late distribution times and a dangerous journey home. Recommendation: morning distributions, a priority queue, closer distribution points, then verification at the next monitoring round.

Key takeaways

- Median rather than mean for skewed distributions (income, debt, waiting times).
- A change from 20% to 30% is +10 points, i.e. +50% in relative terms.
- Disaggregation reveals the groups left behind that the average hides.
- Correlation is not causation; compare performance with comparable composition.
- Qualitative analysis: coding, matrix by group, search for counter-examples; a quote illustrates, it does not prove.

Quiz

1. Expenditure of 5 households: 10, 15, 20, 25, 430 USD. Mean and median:
   A. Mean 100, median 20
   B. Mean 20, median 100
   C. Mean 25, median 20
   D. Mean 100, median 25

2. Coverage goes from 40% to 50%. This is:
   A. +10% in relative terms
   B. +10 percentage points, i.e. +25% in relative terms
   C. +50 points
   D. +25 points

3. 40 differences between groups are tested at the 5% threshold. How many "significant" results can be expected by chance alone?
   A. None
   B. About 2
   C. About 20
   D. Exactly 5

4. On a map comparing districts with very different populations, you should show:
   A. The absolute numbers of cases
   B. Rates (for example per 1,000 inhabitants)
   C. Only the names of the districts
   D. NGO budgets

5. A theme comes up in 9 of the 12 focus groups. The correct presentation is:
   A. 75% of the population thinks this
   B. This theme came up in 9 of the 12 groups, specifying which ones
   C. Everyone thinks this
   D. This result is statistically significant

Quiz answers : 1. A · 2. B · 3. B · 4. B · 5. B

Week 8 : Accountability to affected people: participation, complaints and feedback

Objectives

- Translate the Core Humanitarian Standard (CHS) into concrete, measurable practices within a programme.
- Design and manage a complaints and feedback mechanism that is accessible, safe and closes the loop.
- Organise genuine participation of people and transparent information, including on difficult decisions.

What being accountable to people means

Accountability to affected people (AAP) is the commitment to use one's power responsibly: taking people's views into account, reporting to them and answering to them for one's actions. It does not replace accountability to donors; it complements it, and the two are often in tension (a donor wants numbers, people want to be heard).

The **Core Humanitarian Standard on Quality and Accountability** (CHS), revised in 2024, sets out nine commitments to people and communities. In essence: aid that respects their rights and their participation, is accessible in time, relevant, strengthens their capacities and does no harm, with complaints handled safely, coordinated aid, an organisation that learns, competent and well-managed staff, and resources used responsibly.

For each commitment, an organisation can ask itself three questions: what do we actually do? how do we know? what do people themselves say? The most telling indicators come from people: "Do you think the aid meets your most important needs?", "Do you know how to make a complaint?", "Do you feel treated with respect?".

> Accountability is judged by what changes. A programme that consults without ever changing anything teaches people that their opinion does not count, and the next round of participation will be weaker.

Designing a complaints and feedback mechanism

A **complaints and feedback mechanism** (CFM or FCRM) receives, records, processes and responds to all communications from people: questions, suggestions, simple complaints, sensitive complaints. Its design starts with consultation: which channels do people use and whom do they trust? Women, children, older people and people with disabilities do not give the same answers.

Several channels complement each other: a free hotline, messaging, suggestion boxes (for anonymity, if they are opened by someone independent of the distribution team), help desks

and reception desks at distributions, home visits for people who cannot travel, community committees, and feedback collected during monitoring.

The mechanism defines written **categories** and **timeframes**. For example: request for information, response within 3 days; programme complaint (not registered, wrong amount), response within 7 days; sensitive complaint (fraud, abuse, sexual exploitation, security), referral within 24 hours to the designated person, through a separate confidential channel.

The non-negotiable principles: **confidentiality**, **safety** of people who complain (no retaliation), **accessibility** (languages, illiteracy, disability), **free of charge**, and **separation** between those who handle complaints and those who are the subject of them.

Closing the loop and measuring performance

"Closing the loop" means that every person receives a response, even when the answer is no, and that feedback trends inform decisions. Most mechanisms collect, but many do not respond.

Useful performance indicators:

- Number of feedback items received per month and per 1,000 people assisted, by channel, by sex and age.
- Share of feedback closed within the planned timeframe.
- Share of people satisfied with the response (checked by a call-back on a sample).
- Share of people who know at least one channel (monitoring survey).
- Number of programme changes resulting from feedback during the quarter.

Example. 340 feedback items in one month for 17,000 people: 20 per 1,000. 289 closed within the timeframe: 85%. Compared with a previous month of 2 feedback items per 1,000, this is not a deterioration: it is often a sign that the mechanism is finally known and that people trust it.

> A mechanism that receives very few complaints is rarely the sign of a perfect programme: it is almost always the sign of a mechanism that is unknown, inaccessible or not trusted.

Sensitive complaints and protection from sexual exploitation and abuse

Complaints of **sexual exploitation and abuse** (SEA) by humanitarian staff follow a strict channel, based on the IASC six core principles: SEA constitutes gross misconduct and grounds for termination; any sexual activity with a child (under 18) is prohibited regardless of the local age of consent; exchanging money, employment, goods or services for sex is prohibited; sexual relationships with beneficiaries are strongly discouraged because they are based on unequal power dynamics; staff must report any concern; and managers must create an environment that prevents SEA.

In the mechanism, this means: anyone who receives an allegation passes it on **immediately** to the designated focal point, without investigating it themselves; the **survivor** is referred to assistance (medical care, psychosocial support, protection, legal aid) according to their wishes;

the person implicated is removed from contact with beneficiaries during the investigation; information is shared strictly on a need-to-know basis; the donor is informed in accordance with the contract.

A **survivor-centred** approach respects four principles: safety, confidentiality, respect for choices, non-discrimination. The person is never confronted with the alleged perpetrator, and is not asked to repeat their testimony.

Participation and information

Participation ranges from information to shared decision-making. The classic ladder distinguishes: informing, consulting, involving, collaborating, and transferring the decision. Most humanitarian programmes remain at the level of consultation; aiming higher on certain decisions (choice of sites, targeting criteria, kit composition, schedule) improves relevance and acceptance.

Community committees are useful but fragile: they often reproduce existing power relations. Attention is paid to their composition (women, young people, people with disabilities, minorities), a written mandate is set, and they are not left alone to select beneficiaries without oversight.

Information is a form of aid: people have the right to know who the organisation is, what it does, for whom, according to which criteria, for how long, how to complain, and that aid is free. Appropriate languages and formats are used (radio, voice messages, illustrated posters, separate meetings for women).

Difficult decisions (ration cuts, end of a programme, change of targeting) must be announced early, explained honestly and accompanied: referral to other services, a transition period. Announcing a cut on the day of the distribution is an accountability failure that also creates security risks for staff.

Data protection and integration into the programme cycle

Feedback and complaints contain **personal data**, sometimes highly sensitive. Minimum rules: access limited to those who handle cases; separation of identifying data from content for analysis; anonymisation in reports; a defined retention period; no sharing with authorities or third parties without a clear basis and without an assessment of the risks to the people concerned.

Accountability must be integrated into every stage of the programme cycle, and budgeted:

- **Assessment**: ask people about their priorities and preferred channels.
- **Design**: targeting criteria explained, complaints mechanism planned, budget line (staff, hotline, information materials).
- **Implementation**: information, feedback handled, adjustments.
- **Monitoring**: accountability indicators in the dashboard.
- **Evaluation and learning**: people themselves judge the relevance and quality of the aid.

At collective level, common mechanisms (inter-agency information lines) reduce confusion for people, who often do not know the names of organisations. They require clear agreements on data sharing and case referral.

Exercise: fixing a complaints mechanism

Situation. A programme for 24,000 people. In six months, 31 feedback items were received, all through the suggestion box installed in the organisation's office and opened by the site manager; 12 received a response. A monitoring survey shows that 22% of people know a way to complain (13% of women, 31% of men). An enumerator reports a rumour that a distributor is asking women for sexual favours in exchange for registration.

Task. 1) Calculate the feedback rate per 1,000 people and the response rate. 2) Identify the four main problems. 3) What do you do within 24 hours about the rumour? 4) Propose three monitoring indicators.

Answer. 31 feedback items for 24,000 people: 1.3 per 1,000 in six months, extremely low. Response rate: 12 / 31 = 39%. Problems: a single channel, not very accessible and located in the office; box opened by a person who may be implicated (no independence); low awareness, especially among women; loop not closed. Within 24 hours: pass the rumour on to the PSEA focal point without investigating yourself, arrange the precautionary removal of the distributor from registration activities, plan assistance for any survivor identified, inform management and the donor according to procedures. Indicators: feedback per 1,000 people by sex and channel; share closed within the timeframe; share of people, by sex, who know at least one channel.

Key takeaways

- The CHS sets out nine commitments; the best indicators come from what people say.
- A feedback mechanism has several channels chosen with people, written timeframes and a separate sensitive channel.
- Closing the loop: every feedback item receives a response and trends change the programme.
- Few complaints often signal a mechanism that is unknown or not trusted, not a perfect programme.
- SEA allegation: immediate referral to the focal point, survivor-centred assistance, no improvised investigation.

Quiz

1. A mechanism receives 3 complaints in six months for 20,000 people assisted. The most likely interpretation:
   A. The programme is excellent
   B. The mechanism is unknown, inaccessible or not trusted
   C. People have nothing to say
   D. The mechanism should be scrapped
2. A staff member receives an allegation of sexual exploitation against a colleague. They must:
   A. Question the colleague themselves
   B. Pass it on immediately to the designated focal point, without investigating themselves
   C. Wait until they have evidence
   D. Raise it at a team meeting
3. According to the IASC principles, any sexual activity with a person under 18 is prohibited:
   A. Only if the local age of consent is 18
   B. Regardless of the local age of consent
   C. Only for international staff
   D. Only during working hours
4. Who should not open a complaints box?
   A. An independent accountability officer
   B. The head of the distribution team whose activities are the subject of complaints
   C. A pair including a woman
   D. A member of the MEAL team
5. A ration cut has been decided. The practice consistent with accountability is:
   A. Announce it on the day of the distribution
   B. Announce it early, explain it honestly, and refer people to other services during a transition
   C. Not announce it
   D. Let the authorities announce it

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 9 : Evaluating: criteria, terms of reference, matrix and use

Objectives

- Choose the type of evaluation and the criteria suited to the decision it must inform.
- Write precise terms of reference and build a rigorous evaluation matrix.
- Manage an evaluation independently and turn its recommendations into tracked decisions.

Why and when to evaluate

An **evaluation** is a systematic and objective assessment of an intervention, its design, its implementation and its results. It differs from monitoring: monitoring is continuous and internal, and asks "are we doing what we planned?"; evaluation is periodic, often external, and asks "are we doing the right thing, in the right way, and what difference does it make?".

It has two functions, sometimes in tension: **accountability** (reporting to people, donors and the board) and **learning** (improving this programme and future ones). An evaluation designed only for accountability produces scores; an evaluation designed for learning produces decisions.

Type	Timing	Main question
Real-time evaluation	First weeks of an emergency	Should the response be corrected now?
Mid-term evaluation	Middle of the project	What needs adjusting for the second half?
Final evaluation	End of the project	What results, what lessons for the future?
Impact evaluation	During and after	What change is attributable to the intervention?
Ex-post evaluation	1 to 3 years after the end	Have the effects lasted?

Before commissioning an evaluation, write down who will use it and for what decision. If no one can answer, the evaluation will be one more report on a shelf.

The OECD-DAC criteria

The evaluation criteria of the OECD Development Assistance Committee, revised in 2019, are six in number:

- **Relevance**: does the intervention respond to people's needs and priorities, and does it remain relevant if the context changes?
- **Coherence** (added in 2019): is it compatible with other interventions in the same context, both internal and external to the organisation?

- **Effectiveness**: does it achieve its objectives, including for different groups?
- **Efficiency**: does it deliver its results economically and on time?
- **Impact**: what longer-term effects, positive or negative, intended or not?
- **Sustainability**: will the benefits continue after funding ends?

In humanitarian work, the criteria adapted by ALNAP are also used: **coverage** (have all people in need been reached, without exclusion?), **connectedness** (link between the emergency and the longer term) and **coordination**.

The criteria do not all carry the same weight. A real-time evaluation of a three-month response focuses on relevance, coverage and effectiveness; talking about sustainability at that stage makes little sense. Choosing three or four criteria and going into them in depth is better than skimming over six.

Terms of reference that get a good evaluation

The **terms of reference** (ToR) are the intellectual contract of the evaluation. Vague ToR produce vague evaluations. They contain:

- **Context** and description of the intervention (with the theory of change and the logframe in annex).
- **Purpose and users**: why now, for whom, for which decisions.
- **Scope**: period, areas, components, what is excluded.
- **Evaluation questions**: five to eight main questions, not thirty. Each question can be evaluated with the means and time available.
- **Expected approach and methods**, leaving the team to make proposals, and the existing data available.
- **Ethics**: consent, data protection, protocols for sensitive topics, participation of people.
- **Deliverables and timeline**: inception report, debriefing, draft report, final report, summary.
- **Team profile**, budget, governance (reference group) and independence rules.

Order of magnitude. A good-quality external final evaluation of a multisectoral project often costs between 25,000 and 80,000 USD depending on the field setting and methods. Too small a budget leads to a few days of fieldwork and a report without evidence; fewer evaluations, better done, is preferable.

The evaluation matrix and the inception report

The **evaluation matrix** is the central tool of the inception report: for each question, it specifies the sub-questions, the indicators or judgement criteria, the data sources, the data collection and analysis methods, and the limitations.

Example. Question: "To what extent have cash transfers improved the food security of the most vulnerable households?" Judgement criteria: change in the food consumption score and rCSI between baseline and endline, compared with a comparison area; differences according to

the sex of the head of household and the presence of a person with a disability. Sources: baseline and endline surveys, post-distribution monitoring, focus group discussions. Methods: difference-in-differences, qualitative analysis by group. Limitations: comparison area partially assisted by other actors from month 6.

The **inception report** is the moment to correct course: the team confirms what is feasible, flags questions that cannot be addressed, and proposes the sampling. The evaluation manager validates it with the reference group before fieldwork. It is the best investment of time in the whole evaluation.

The **quality** of an evaluation is judged on the clarity of the link between **findings** (supported by data), **conclusions** (judgements based on the findings) and **recommendations** (actions that flow from the conclusions). A recommendation without a corresponding finding is the evaluator's opinion.

Independence, ethics and evaluation management

Independence protects credibility: the evaluation team took no part in the design or implementation; the evaluation manager is not the head of the project being evaluated; the team has free access to data and people; the programme may comment on the facts in the draft report, but may not impose changes to the conclusions.

The United Nations Evaluation Group (UNEG) norms and standards for evaluation and the ALNAP principles insist on **ethics**: do no harm to the people interviewed, informed consent, confidentiality, attention to people in distress and to risks of retaliation, feedback to communities.

The **evaluation manager** has an active role: preparing data, organising field access, convening the reference group, checking quality at key stages, protecting the team's independence, and organising the debriefing. The main risks are well known: start-up delays that shorten fieldwork, non-existent baseline data, limited access, pressure to soften findings.

> An evaluation that finds only positive points is not credible. A programme that accepts well-supported negative findings shows that it is serious.

Management response and use

The **management response** is the document that turns an evaluation into decisions. For each recommendation: accepted, partially accepted or rejected, with the justification; planned actions; person responsible; deadline. It is written within a month of the final report and published with it.

Follow-up of actions takes place at quarterly management reviews: where does each action stand? Mature organisations publish an annual implementation rate for accepted recommendations.

A recommendation can legitimately be rejected: too costly, outside the mandate, already under way. But the justification must be written down. What is not acceptable is the absence of a response.

Finally, **dissemination** adapts the format to the audience: a two-page summary for management and donors, a presentation for teams, a simple oral version for communities, and publication of the full report (for example in the ALNAP library) for the sector. Publishing evaluations, including critical ones, is a Grand Bargain commitment on transparency.

Exercise: writing the questions and matrix of an evaluation

Situation. An 18-month project (water, hygiene and protection) in three displacement camps reaches its midpoint. Country management must decide within four months whether and how to extend the project to a fourth camp. The donor has requested an evaluation but has not set any questions. Budget: 30,000 USD. Baseline data exist for water, not for protection.

Task. 1) What type of evaluation and which main criteria? 2) Write three evaluation questions. 3) For one of them, fill in one row of the matrix (judgement criteria, sources, methods, limitations).

Answer. A formative mid-term evaluation, with country management and the donor as users; main criteria: relevance, effectiveness, coverage, efficiency (sustainability is secondary at this stage). Possible questions: 1) To what extent do the water and hygiene services meet the planned standards for all groups, including people with disabilities? 2) Do the protection activities respond to the priority risks identified by women and children? 3) What conditions must be met to extend the project to a fourth camp with the same quality and a comparable cost per person? Matrix row for 2): criteria: match between reported priority risks and activities, use of services, perception of safety; sources: focus group discussions separated by sex and age, interviews with protection actors, service data (anonymised); methods: coded qualitative analysis, participatory risk mapping; limitations: no protection baseline, hence a contribution analysis rather than a measurement of change.

Key takeaways

- Monitoring asks "are we doing what we planned?"; evaluation asks "is it the right thing, done well, and what difference does it make?".
- Six DAC criteria (2019): relevance, coherence, effectiveness, efficiency, impact, sustainability; plus coverage, connectedness and coordination in humanitarian work.
- Precise ToR: users, decisions, five to eight evaluable questions, ethics, deliverables.
- The evaluation matrix links each question to criteria, sources, methods and limitations.
- Management response: each recommendation accepted or not, with action, person responsible and deadline.

Quiz

1. Which criterion was added by the OECD-DAC in 2019?
   A. Relevance
   B. Coherence
   C. Efficiency
   D. Impact

2. An evaluation that asks whether all people in need were reached, without exclusion, addresses the criterion of:
   A. Sustainability
   B. Coverage
   C. Efficiency
   D. Connectedness

3. The most useful moment to correct a poorly focused evaluation is:
   A. Reading the final report
   B. Validating the inception report and the evaluation matrix
   C. Publication
   D. The management response

4. A recommendation that is not based on any finding in the report is:
   A. Good practice
   B. The evaluator's opinion, to be challenged
   C. Mandatory
   D. A conclusion

5. Who should not manage the evaluation of a project?
   A. A country MEAL manager
   B. The head of the project being evaluated
   C. A headquarters adviser
   D. A steering committee

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 10 : Measuring impact: counterfactual, methods and contribution

Objectives

- Understand the counterfactual problem and why a before-after comparison does not measure an effect.
- Choose and apply a quasi-experimental or experimental method suited to the humanitarian context, with its validity conditions.
- Use contribution approaches when attribution is impossible, and decide what it is reasonable to prove.

The fundamental problem: what would have happened without the programme?

The **impact** of a programme is the difference between what happened with the programme and what would have happened without it: the **counterfactual**. It is never possible to observe the same household directly both with and without the programme; all impact evaluation methods are ways of building a credible estimate of this counterfactual.

Two naive comparisons are misleading:

- **Before-after** among beneficiaries: the observed change mixes the effect of the programme with everything else that changed at the same time (harvest, prices, security, other aid). If the harvest was good, food security improves even without the programme; if it was poor, an effective programme can appear to have no effect.
- **Beneficiaries versus non-beneficiaries** after the programme: the two groups were different at the outset, precisely because targeting selected the most vulnerable. The comparison then underestimates the effect, or overestimates it if it was the most motivated who signed up (selection bias).

> The question to ask of any impact claim: "compared with what?". If the answer is "compared with before", it is not a measure of impact.

The randomised controlled trial and its use in humanitarian settings

In a **randomised controlled trial** (RCT), among the eligible people, those who receive the programme (treatment group) and those who do not receive it or receive it later (control group) are selected by lottery. Random assignment makes the two groups comparable on average on all characteristics, observed or not. The effect is simply the difference in average outcomes at the end.

In humanitarian settings, denying life-saving aid to eligible people in order to form a control group is **unethical**. But RCTs remain possible and have produced major knowledge (on cash, nutrition, mental health) in specific situations:

- **Insufficient resources** for everyone in any case: random selection is then a fair way of allocating, and the group not served acts as the control.
- **Phased roll-out**: everyone will receive the programme, but the order is drawn by lot (stepped wedge design).
- **Comparison of two variants** of a programme, both considered acceptable: monthly or lump-sum cash, with or without accompanying support.

RCTs are expensive, take time and answer specific questions; they are justified when the answer will serve many programmes. For most projects, quasi-experimental methods are sufficient.

Difference-in-differences

Difference-in-differences compares the change in an assisted group with the change in a non-assisted comparison group: effect = (after − before) in the assisted group − (after − before) in the comparison group. It eliminates stable initial differences between the groups and common trends (harvest, prices).

Example. Proportion of households with acceptable food consumption. Assisted group: 28% before, 52% after (+24 points). Comparison group: 35% before, 44% after (+9 points). Estimated effect: 24 − 9 = 15 points. A before-after comparison would have attributed 24 points to the programme; a comparison of final results alone would have attributed 8.

Validity condition: the **parallel trends** assumption. Without the programme, the two groups would have changed in the same way. It is checked where possible with several measurements before the programme, and one makes sure that the comparison group does not receive other significant aid during the period (contamination) and does not suffer a specific shock.

In humanitarian settings, the comparison group is often a neighbouring area not covered for lack of funding, or eligible households on a waiting list. A comparison group is never created by denying aid for the purposes of the study.

Matching and regression discontinuity

Matching builds a comparison group by pairing each assisted household with one or more non-assisted households that resemble it on characteristics measured before the programme (size, assets, income, status, area). The **propensity score** method summarises these characteristics as a probability of being assisted, and households with close scores are compared. Its limitation: it corrects only observed differences. If motivation or social networks influenced registration and are not measured, the bias remains.

Regression discontinuity applies when eligibility depends on a threshold on a score: for example, households with a vulnerability score of 60 or more are assisted, those below are not. Households just above and just below the threshold (57-59 and 60-62) are almost identical; the difference in outcomes between them estimates the effect of the programme around the

threshold. Conditions: the threshold is strictly applied, and households cannot manipulate their score to get above it.

Example. Households with a score of 57-59: 38% with acceptable consumption. Households with a score of 60-62 (assisted): 51%. Estimated effect around the threshold: about 13 points. This result holds for households close to the threshold, not necessarily for the most vulnerable.

The **interrupted time series** compares the trend of an indicator before and after the start of a programme, over many measurement points (for example weekly nutrition admissions). It is useful when no comparison group exists, provided there are enough points before the intervention.

Contribution approaches: when attribution is impossible

Many important effects do not lend themselves to statistical methods: advocacy that changes a law, capacity strengthening of local organisations, improved coordination. They cannot be compared with a quantified counterfactual. The aim is then to establish credibly that the programme **contributed** to the change.

- **Contribution analysis** (Mayne): set out the theory of change, gather evidence for each link, examine alternative explanations, and build a "contribution story" that is then critiqued by stakeholders.
- **Process tracing**: causal hypotheses are tested by looking for evidence whose presence or absence makes it possible to confirm or rule out an explanation (for example, the word-for-word reuse of an advocacy note in a decree).
- **Outcome harvesting**: first identify the changes observed among actors, then work backwards to determine whether and how the programme contributed to them.
- **Most significant change**: stories of change are collected from people, and panels select the most significant ones, explaining why. Useful for discovering unintended effects.

> Rigour does not mean quantitative. A well-conducted contribution analysis that seriously examines alternative explanations is more credible than a quantified before-after comparison.

Choosing the method and deciding what needs proving

The choice of method depends on four questions:

- **What decision?** Scaling up a pilot model nationally justifies strong evidence; adjusting an ongoing programme can make do with sufficient evidence.
- **Does the evidence already exist?** The effect of cash on food consumption is very well documented: there is no need to demonstrate it again with an RCT in every project. It is often enough to check that the programme is faithful to the model and that the intermediate effects are present.
- **What is feasible and ethical?** Is there a comparison group without denying aid? A clear eligibility threshold? A phased roll-out?

- **What budget and what timeframe?** A serious impact evaluation requires a baseline before the start: it is designed with the programme, not at the end.

As regards **sample size**, an impact evaluation must have sufficient power to detect the expected effect. Detecting a 15-point effect requires far fewer households than a 5-point effect. If the expected effect is small and the sample is small, the study will almost certainly conclude "no significant effect", which will be misinterpreted as "no effect".

A final point: present the **level of evidence** honestly. "Assisted households improved their consumption by 24 points; compared with a comparison area, the attributable effect is estimated at 15 points, under the parallel trends assumption supported by data from two previous rounds" is a sentence that inspires confidence.

Exercise: estimating an effect and critiquing the method

Situation. A livelihoods programme supported 1,200 households. Median monthly income: assisted households 42 USD before, 71 USD after. Neighbouring area not covered (for lack of funding): 47 USD before, 58 USD after. During the programme, another NGO distributed seeds in half of the villages in the neighbouring area. The final report states: "the programme increased incomes by 69%".

Task. 1) Where does the 69% figure come from and why is it misleading? 2) Calculate the effect by difference-in-differences. 3) What is the likely effect of the seed distribution on your estimate, and what do you propose?

Answer. 69%: (71 − 42) / 42, a before-after comparison that attributes the entire increase to the programme, including the increase due to the context. Difference-in-differences: (71 − 42) − (58 − 47) = 29 − 11 = 18 USD per month, i.e. about 43% of the initial income of assisted households. The seeds probably increased incomes in the comparison area, which inflates its trend and underestimates the programme's effect (contamination violates the parallel trends assumption). Proposal: analyse separately the comparison villages without seeds, or control for this variable; present the estimate as a lower bound and state the limitation clearly.

Key takeaways

- Impact is measured against a counterfactual; before-after is not a measure of impact.
- Difference-in-differences = change in the assisted group − change in the comparison group, under the parallel trends assumption.
- An RCT in humanitarian settings is ethical only if it deprives no one of available aid: limited resources, phased roll-out, or comparison of variants.
- Regression discontinuity: compare households just above and just below a strict eligibility threshold.
- When attribution is impossible: contribution analysis, process tracing, outcome harvesting, most significant change.

Quiz

1. Assisted group: 30% before, 50% after. Comparison: 32% before, 40% after. Effect by difference-in-differences:
   A. 20 points
   B. 12 points
   C. 10 points
   D. 8 points

2. Which assumption is essential to difference-in-differences?
   A. The groups are identical at the outset
   B. Without the programme, the two groups would have changed in the same way
   C. The sample exceeds 1,000 households
   D. The programme lasted more than a year

3. In which situation is random selection ethically acceptable in humanitarian settings?
   A. To deny available life-saving aid to half of the eligible people
   B. When resources are in any case insufficient for all eligible people, or for the order of a phased roll-out
   C. Never
   D. Only if the donor requires it

4. The main limitation of propensity score matching is:
   A. It only works with RCTs
   B. It corrects only measured differences, not unobserved differences such as motivation
   C. It requires an eligibility threshold
   D. It is prohibited by the OECD

5. To show that advocacy contributed to the adoption of a law, the most suitable method is:
   A. A randomised trial
   B. A difference-in-differences
   C. A contribution analysis or process tracing
   D. A household survey

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. C

Week 11 : Value for money: cost-efficiency, cost-effectiveness and trade-offs

Objectives

- Apply the 4E framework (economy, efficiency, effectiveness, equity) to judge a programme's value for money.
- Carry out a rigorous cost-efficiency and cost-effectiveness analysis, with allocation of shared costs and sensitivity analysis.
- Use these analyses to choose between options, without sacrificing equity or quality.

Value for money: the 4E framework

Value for money (VfM) does not mean "the cheapest". It is the best use of resources to achieve the intended results. The most widely used framework, which comes from UK development cooperation, has four dimensions:

- **Economy**: buying inputs of the required quality at the best price (prices of kits, salaries, rents compared with the market).
- **Efficiency**: converting inputs into outputs at the lowest cost (cost per kit distributed, per household registered, transfer ratio).
- **Effectiveness**: converting outputs into outcomes (share of households improving their consumption, cure rate).
- **Equity**: reaching the most vulnerable and hardest to reach, which often costs more per person.

Cost-effectiveness is sometimes added as a summary: cost per outcome achieved. The four dimensions hang together: a very economical programme that buys poor-quality kits, or a very efficient one that avoids remote areas, does not deliver value for money.

> Reaching the most isolated people or people with disabilities costs more per person. An honest VfM report explains this instead of hiding it, and shows that this extra cost is an equity choice.

Cost-efficiency analysis: the method

A **cost-efficiency analysis** calculates the cost of producing an output (for example the cost per household receiving a transfer for six months). It requires attributing to the activity **all** the costs it consumes, not just direct costs: programme staff, a share of support staff (logistics, finance, HR), offices, vehicles, and a share of headquarters costs.

The steps, inspired by the systematic cost analysis approach developed by NGOs and taken up by the Dioptra tool:

- **Define** the output and the period.

- **Gather** the actual expenditure (not the budget) for all the lines concerned.
- **Allocate** each line to the activity according to a key: 100% for an exclusively related cost; share of time for shared staff; share of direct expenditure for support costs.
- **Divide** the total allocated cost by the number of outputs.
- **Compare** with other programmes or other periods, taking the context into account.

Example. A 6-month cash programme for 2,000 households: transfers 1,080,000 USD; programme staff 96,000; payment service provider fees (2%) 21,600; share of country support 54,000; share of headquarters 42,400. Total cost: 1,294,000 USD. Cost per household: 647 USD. Transfer ratio: 1,080,000 / 1,294,000 = 0.83.

Cost-effectiveness analysis

Cost-effectiveness analysis relates cost to an outcome: cost per child cured, per household lifted out of food insecurity, per score point gained, per cholera case averted. It makes it possible to compare approaches aimed at the same outcome.

Example of a comparison. Two options to improve the food consumption of 1,000 households over six months. Option A, in-kind food: total cost 520,000 USD, 380 households move to acceptable consumption (compared with the counterfactual). Option B, cash: total cost 450,000 USD, 410 households. Cost per household that improved: A = 1,368 USD; B = 1,098 USD. B is more cost-effective in this context.

Three rules of interpretation:

- The effect must be **attributable** (week 10): dividing the cost by a before-after improvement grossly overestimates or underestimates cost-effectiveness.
- Comparisons are valid only between options aimed at the **same outcome**, in comparable contexts.
- **Multiple** effects complicate the comparison: cash can also reduce debt and support markets. In health, composite measures are used (disability-adjusted life years, DALYs), which are rarely available in other sectors.

Social return on investment (SROI), which converts all effects into money, is attractive but fragile: the monetary values attributed to social effects are highly debatable. These ratios must be read with caution and the assumptions always examined.

Economies of scale, fixed costs and sensitivity analysis

Unit cost depends heavily on **scale**. Fixed costs (coordination team, payment system, office) are spread over more beneficiaries as the programme grows.

Example. Fixed costs: 150,000 USD; variable cost per household: 400 USD. For 500 households: (150,000 + 200,000) / 500 = 700 USD per household. For 2,000 households: (150,000 + 800,000) / 2,000 = 475 USD. Comparing the unit cost of a pilot project with that of a large-scale programme is therefore unfair; compare at comparable scale, or model the cost at the target scale.

Sensitivity analysis tests the robustness of conclusions when assumptions change: what happens to the comparison if the effect of option B is 20% lower than expected? if payment fees double? If the conclusion changes for a realistic variation, this must be said: the decision then rests on an assumption that needs checking.

The **break-even point** of an investment is another useful tool: a mobile payment system that costs 60,000 USD to set up and saves 12 USD per household per month pays for itself from 5,000 household-months (60,000 / 12), for example 1,000 households for 5 months.

Presenting value for money to donors

Donors increasingly ask for a VfM section in proposals and reports. A solid section:

- Presents **indicators for each of the 4Es**, with their value and a comparison (previous period, other area, external benchmark).
- Shows the **concrete measures** taken: joint procurement, framework agreements, sharing offices and vehicles with other projects, electronic payment, loss reduction.
- Explains **justified extra costs**: hard-to-reach areas, accessibility, security, quality, localisation (support to partners).
- Acknowledges the **limitations**: non-attributable effects, incomplete cost data.

Pitfalls to avoid: presenting short-term savings that degrade quality (less monitoring, less durable materials); comparing unit costs from very different contexts; displaying a high transfer ratio obtained by cutting monitoring and accountability.

> Value for money is also an argument for support costs: a programme without solid logistics, finance and MEAL is not efficient, it is risky.

Trade-offs: efficiency, equity and quality

Cost analyses inform decisions; they do not make them. Common trade-offs:

- **Coverage versus depth**: assist 10,000 households with a transfer covering 50% of the basket, or 6,000 with a transfer covering 85%? If a partial transfer does not make it possible to achieve the intended outcome (for example avoiding negative coping strategies), helping more people less well is not more effective.
- **Easy areas versus hard-to-reach areas**: the cost per person can double in remote areas, where needs are often the most severe.
- **Short term versus sustainability**: repairing a water point costs less than training a management committee and setting up a cost-recovery system, but the water point is likely to break down again.

One way to decide is to combine cost-effectiveness analysis with **explicit criteria**: a minimum quality threshold (Sphere standards), priority to the most severe, a ceiling on cost per person beyond which another approach is sought. The decision and its reasons are written down, so that they can be explained to people, teams and donors.

Exercise: comparing two water supply options

Situation. To supply 15 litres of water per person per day to 12,000 people for 12 months, there are two options. Option A, water trucking: fixed costs 30,000 USD, variable cost 0.012 USD per litre. Option B, rehabilitation of two boreholes and a network: investment 210,000 USD, operation 2,500 USD per month. The lifespan of the rehabilitated boreholes is at least five years.

Task. 1) Calculate the cost of each option over 12 months. 2) Calculate the cost per person per year. 3) After how many months does option B become cheaper? 4) What recommendation, and under what conditions?

Answer. Volume: 12,000 × 15 × 365 = 65,700,000 litres per year. Option A: 30,000 + 0.012 × 65,700,000 = 30,000 + 788,400 = 818,400 USD, i.e. 68.2 USD per person per year. Option B: 210,000 + 12 × 2,500 = 240,000 USD, i.e. 20 USD per person in the first year. Break-even: monthly trucking cost ≈ 788,400 / 12 = 65,700 USD; B is cheaper as soon as 30,000 + 65,700 × m > 210,000 + 2,500 × m, i.e. m > 180,000 / 63,200 = 2.85 months. Recommendation: B, with trucking during the works (2 to 3 months), provided that hydrogeology confirms the yield and water quality, and that a management committee and funding for operation are planned beyond the project.

Key takeaways

- 4E: economy, efficiency, effectiveness, equity; the cheapest is not the optimum.
- Cost-efficiency = total allocated cost (including support and headquarters) / outputs.
- Cost-effectiveness = cost / attributable outcome; compare only options aimed at the same outcome.
- Fixed costs make small programmes more expensive per person: compare at comparable scale.
- Sensitivity analysis: if the conclusion changes for a realistic variation, the decision rests on an assumption that needs checking.

Quiz

1. Reaching people with disabilities and remote areas increases the cost per person. In the 4E framework, this extra cost falls under:
   A. Economy
   B. Equity
   C. Efficiency
   D. A management error

2. Fixed costs 100,000 USD, variable cost 300 USD per household, 1,000 households. Cost per household:
   A. 300 USD
   B. 400 USD
   C. 1,300 USD
   D. 100 USD

3. Option A: 600,000 USD for 400 children cured. Option B: 500,000 USD for 250 children cured. Which is more cost-effective?
   A. B, because it costs less in total
   B. A: 1,500 USD per child cured against 2,000 for B
   C. They are equivalent
   D. They cannot be compared

4. Why must a share of support and headquarters costs be included in a cost-efficiency analysis?
   A. To inflate the figures
   B. Because the activity actually consumes these resources; ignoring them underestimates its cost
   C. Because the donor pays for everything
   D. It is not necessary

5. An investment of 40,000 USD saves 8 USD per household per month. It pays for itself from:
   A. 500 household-months
   B. 5,000 household-months
   C. 320,000 household-months
   D. 50,000 household-months

Quiz answers : 1. B · 2. B · 3. B · 4. B · 5. B

Week 12 : Adaptive management, learning and reporting: leading with evidence

Objectives

- Establish adaptive management: regular reviews, decision rules and documented revision of the programme.
- Build an organisational learning system that captures and reuses lessons, beyond reports.
- Write honest and convincing donor reports, and present a complete MEAL package as a final project.

Adaptive management

Humanitarian contexts change faster than logframes. **Adaptive management** consists of deliberately and transparently adjusting activities, methods or targets according to data and context, while keeping the objective. It is not improvisation: it rests on explicit assumptions, monitoring that tests them, and planned decision points.

Its components:

- **Working assumptions** clearly stated in the theory of change ("women will be able to get to the cash-out points").
- **Context and early warning indicators** that test them (reported travel time, security incidents on the roads).
- **Decision rules** written in advance: "if more than 20% of transfers are not collected within 10 days, mobile cash-out points are opened".
- **Regular reviews** (quarterly, or monthly in an acute crisis) with the authority to decide.
- **An agreement with the donor** on what can change without a formal request (activities, methods) and what requires an amendment (objectives, overall budget, areas).

Donors are often more open than people think to well-argued adaptation. What they refuse are unannounced changes discovered in the final report.

> A project that has changed nothing in 18 months in a volatile context has probably not looked at its data, or has not dared to talk about it.

Programme reviews and after action reviews

The **quarterly review** of a programme brings together the programme team, MEAL, logistics, finance and, where possible, representatives of partners and communities. The standard agenda:

- Progress on key indicators against targets and trends.

- Budget execution and burn rate.
- Feedback from people and partners: main themes, responses provided.
- Context and assumptions: what has changed.
- Decisions: what, who, when. And follow-up of decisions from the previous review.

The **after action review**, which originated in the military and was then widely adopted, is a short tool (one to two hours) to be used after a significant event: a distribution, an evacuation, a first wave of response. Four questions: what did we plan? what actually happened? why is there a gap? what will we do differently next time? It works if people talk about facts, without looking for culprits, and if the last question produces actions with a person responsible.

Lessons must be **specific** and **actionable**: "communication needs to be improved" is not a lesson; "announce distribution dates by radio and by voice message at least 5 days in advance, because 40% of households had not received the information" is one.

Organisational learning

Humanitarian organisations produce many reports and learn little: the same mistakes are repeated from one crisis to the next, often because staff turn over quickly and lessons remain in documents that no one reads. An effective learning system:

- **Defines priority learning questions** for the organisation (for example: "how can monitoring quality be maintained when access is intermittent?") and directs the MEAL of several projects towards answering them.
- **Captures** lessons in usable forms: two-page briefs, checklists, document templates, rather than long reports.
- **Circulates** lessons at the right time: at the start of a new project, the lessons from similar projects are read (lessons review at the design stage).
- **Values** the reporting of failures, for example through regular sessions where teams present what did not work and what they learned from it.
- **Shares** with the sector: published evaluations, contributions to cluster working groups, case studies.

The role of the MEAL manager then changes: from a producer of data for donors, they become a **facilitator of reflection** who helps teams ask good questions, read the data and decide.

Writing donor reports

A good-quality **narrative report** follows the donor's structure and meets three requirements: **accuracy** (figures consistent with the financial report and previous reports), **analysis** (not just what was done, but what changed and why) and **honesty** (gaps and difficulties explained).

For each result, a simple structure works: **target, achieved, gap, explanation, consequence, action**. For example: "Target: 3,000 households. Achieved: 2,450 (82%). Gap due to the closure of the road to X for 7 weeks. Consequence: two transfer cycles delayed in 4 villages. Action: mobile payment introduced in May; catch-up planned by August, at no additional cost."

Mistakes to avoid:

- Output figures presented as outcomes ("12,000 people reached by awareness-raising have improved their practices").
- Double counting between activities (the same person counted in each activity); unique beneficiaries must be distinguished from participations.
- Individual success stories presented as representative; a story illustrates, it does not prove.
- No mention of complaints received and responses provided.

> The donor will find out about problems sooner or later. It is better that they hear about them from you, with the analysis and the measures taken, than through an evaluation or an audit.

Programme closure and sustainability

Closure is prepared from the design stage (exit strategy) and implemented at least three months before the end. It includes:

- **Informing** communities and authorities about the end date and arrangements, and referral to other services.
- **Transfer** of assets (water points, equipment) according to the donor's rules, with written agreements with the recipients of these assets.
- **Handover** to partners, technical services or committees, with capacity strengthening that is genuinely planned and budgeted.
- **Endline survey** and evaluation, then feedback to communities.
- **Archiving** of data and documents, and secure destruction of personal data in line with the retention policy.

Sustainability is judged by the conditions in place: an institution or community that takes responsibility, resources for operation (public budget, contributions, cost recovery), skills, and genuine demand from users. Without these conditions, it is better to acknowledge it and look for a successor (other funding, other actor) rather than announce a sustainability that will not exist.

Final project: the complete MEAL package

The final project of this Mini MBA consists of presenting a **complete MEAL package** for a real or realistic programme of your choice. It brings together everything you have learned in twelve weeks:

- A two-page **needs analysis note**: sources, severity, priority groups, coverage gaps, response options compared.
- A **theory of change** and a **logframe** with indicators, baselines, calculated targets and assumptions.
- A **targeting strategy** and, where applicable, the calculation of the transfer value.

- A **MEAL plan** with two indicator reference sheets, the sampling plan and the sample size calculation for the baseline survey.
- A **complaints and feedback mechanism**: channels, timeframes, sensitive channel, indicators.
- An **evaluation plan**: type, questions, method for estimating the effect and its validity conditions.
- A **VfM analysis**: 4E indicators and a comparison of options.
- A **review schedule** and three adaptive decision rules.

Quality criteria: consistency between the parts, correct and verifiable calculations, realism (MEAL budget of 3 to 10%), a genuine place given to people, and honesty about limitations. The case studies in the diploma exam draw on exactly these skills.

Exercise: writing a decision rule and a report paragraph

Situation. A cash programme for 4,000 households, mobile payment. In the third month: 18% of transfers are not collected after 10 days (6% in the first month); in 3 villages out of 20, this rate exceeds 50%. Feedback reports agents without cash and a road cut off by armed groups. Target for the quarter: 12,000 transfers collected; achieved: 9,840.

Task. 1) Write an adaptive decision rule for this problem. 2) Write the paragraph of the quarterly report on this result, following the structure target, achieved, gap, explanation, consequence, action.

Answer. Rule: "If more than 15% of a village's transfers are not collected 10 days after sending, the team checks agents' liquidity and access within 72 hours; if the cause is confirmed, a mobile cash-out point or an additional agent is organised before the next cycle, and the donor is informed in the monthly report." Paragraph: "Target: 12,000 transfers collected. Achieved: 9,840 (82%). The gap comes from 3 villages where more than half of the transfers were not collected, because of agents' lack of cash and the closure of the road to Y. Consequence: about 2,160 transfers pending, and a risk of households concerned resorting to debt. Actions: two mobile cash-out points since week 12, agreement with the operator to supply agents with cash, information to households by voice message; the transfers remain available and will be monitored until fully collected."

Key takeaways

- Adaptive management: explicit assumptions, early warning indicators, written decision rules, regular reviews, agreement with the donor.
- After action review: planned, actual, why the gap, what to do differently; without looking for culprits.
- A lesson is specific and actionable; otherwise it is just an impression.
- Report: target, achieved, gap, explanation, consequence, action; unique beneficiaries distinguished from participations.
- Sustainability is judged by the conditions in place (responsible party, resources, skills, demand), not by a statement.

Quiz

1. Which wording is a useful lesson learned?
   A. Coordination must be strengthened
   B. Announce distributions by radio and voice message 5 days in advance, because 40% of households were not informed
   C. We need to communicate better
   D. The project was a success

2. In an after action review, the main rule is to:
   A. Identify the person responsible for the error
   B. Talk about facts without looking for culprits, and end with actions with a person responsible
   C. Write a long report
   D. Invite the donor

3. A report counts 5,000 participants in training and 5,000 in the distribution, i.e. "10,000 beneficiaries". The error is:
   A. Possible double counting: unique beneficiaries must be distinguished from participations
   B. No error
   C. A percentage calculation error
   D. An exchange rate error

4. Which change generally requires a formal amendment with the donor?
   A. Changing the awareness-raising method
   B. Modifying the specific objective or the areas of intervention
   C. Adjusting the schedule of an activity by two weeks
   D. Changing the agenda of a meeting

5. Programme closure is prepared:
   A. In the last month
   B. From the design stage with an exit strategy, and implemented at least three months before the end
   C. After the final report
   D. Only if the donor requests it

Quiz answers : 1. B · 2. B · 3. A · 4. B · 5. B