

GrantWriter360

What the Results Can Tell Us

Planning Evidence
Checking Progress and
Making Better Decisions

Aneel Shahzad

Institutional Growth & Proposal Strategy Advisor

GrantWriter360

Pakistan

A note to the reader

An evidence plan should help a programme make decisions it can explain. That requires more than a set of attractive indicators. People need to know what a record means, whose experience is absent, how far a finding can be trusted and what action would be reasonable while uncertainty remains.

This book is for people who plan, deliver, fund or advise programmes in nonprofits, social enterprises and other institutions. It follows the work from an initial question through collection, interpretation and use. The emphasis is practical. A small organisation should be able to improve its evidence without pretending to operate a research institute. A larger organisation should be able to see where its reporting machinery is concealing an unresolved question.

All programme records, numerical illustrations and operational situations in the teaching examples are hypothetical. The continuing example concerns adults receiving employment support. It is a teaching record, not a claim about a GrantWriter360 client or the author's results. Unnamed roles are used so that attention stays on the work and the decisions.

The examples adopt explicit definitions. Actual programmes should follow the applicable agreement, technical standard and local requirements, then explain how their measures are constructed. The source notes distinguish international terminology from UK government guidance and other bounded materials. This book does not establish a universal donor rule, a legal standard or an official employment measure.

The central discipline is to keep the claim close to the evidence. A completed activity can be reported as completed. An observed change can be described with its population and period. A causal explanation needs a suitable enquiry. An uncertain result can still support a decision if the decision acknowledges what is unknown and provides a way to learn more.

Aneel Shahzad · GrantWriter360 · Pakistan
First edition 2026

© 2026 Aneel Shahzad. All rights reserved.

Contents

1 Begin with the decision..... 1

2 Describe how change is expected to happen.....5

3 Define exactly what is being counted.....9

4 Establish a starting point that can be compared.....12

5 Set defensible targets.....15

6 Collect only what the decision needs..... 18

7 Check the record before interpreting it.....22

8 Understand what monitoring can establish.....25

9 Read mixed and missing results..... 29

10 Report what the evidence supports.....33

11 Use findings to choose the next action.....36

12 Bring learning into the next proposal.....40

Worked tools.....43

Source notes and further reading.....47

About the author.....49

1 Begin with the decision

A programme can collect accurate information and still leave its most important question unanswered. Workshop attendance may be recorded carefully while no one knows whether participants can use the material. A follow-up survey may produce a satisfaction score while the delivery team needs to know which part of the service people could not access. The problem begins before analysis, when the information requested is separated from the decision it is supposed to inform.

Start by writing down a decision in ordinary language. Should the next round use the same delivery arrangement? Does a service need to be adjusted during implementation? Is there enough evidence to continue a pilot at its present scale? These questions lead to different collection plans. They also have different deadlines. An answer received after a venue has been booked cannot change that booking, although it may still inform a later round.

A useful decision statement names the choice, the person or group responsible and the latest date on which evidence can help. It also identifies the consequence of being wrong. Changing the wording of an instruction sheet can usually be reversed cheaply. Expanding a service across several districts exposes many more people and resources to the underlying assumptions. The second decision deserves stronger evidence and a more careful examination of uncertainty.

Questions that can change an action

Consider a programme planning practical workshops and employment support for 120 adults. Asking whether participants like the service is reasonable, but that answer alone cannot determine whether to extend it. The programme needs to distinguish several questions. Did people receive the intended support? Could they complete the practical work? What happened to their paid-work status? What else changed in the places where they live? Which delivery changes are feasible before another cohort begins?

The OECD evaluation criteria offer several perspectives on performance, including relevance, effectiveness and sustainability. They help frame questions; they do not supply a research method or require every

evaluation to address every criterion equally. [9] These questions should not all be collapsed into a single success score. A high satisfaction rating can coexist with weak completion. Paid work can rise during a period of increased local recruitment even if the workshops are poorly delivered. A participant can find useful work before completing every activity. Each record answers part of the enquiry, and each has a different implication for action.

The planning team might identify one immediate decision: whether to offer an alternative workshop schedule in the second delivery area. The relevant evidence would include attendance by session, reasons people could not attend, the availability of alternative times and whether the missed sessions are necessary for later practice. An end-of-programme employment rate would arrive too late and be too distant from the scheduling problem. The team can decide to examine the attendance pattern early without claiming that a schedule change will itself create employment.

A separate decision concerns expanding the programme. That requires a longer view of delivery, outcomes, demand, external conditions and the credibility of the proposed explanation. The evidence needed for an early operational adjustment may be entirely adequate for that adjustment while remaining insufficient for expansion. Proportionate evidence means matching the enquiry to the consequence of the decision.

What would count against the preferred explanation

People usually find it easy to state what success would look like. They are less comfortable specifying evidence that would weaken their preferred explanation. Doing so before results arrive reduces the temptation to reinterpret every result as confirmation.

For the employment-support programme, the proposed explanation might be that practical preparation helps participants pursue suitable vacancies more effectively. Evidence that participants attend but cannot complete the practice exercise would raise a question about teaching or accessibility. Evidence that they can complete it but find no relevant vacancies would raise a different question about the opportunity available. Evidence of improved paid-work status among people who barely used the service would require care before attributing the change to participation.

These observations do not automatically settle the explanation. People who use fewer sessions may differ in prior experience or availability. A practice exercise may assess only part of what matters. The purpose of identifying contrary evidence is to create a genuine enquiry rather than a search for favourable material.

The decision owner should also consider what an inconclusive result would mean. If follow-up is incomplete, will the next cohort be paused, offered at a smaller scale or continued with a specific improvement to the evidence plan? A programme should not discover at the reporting deadline that uncertainty has no place in its decision process.

Put the evidence on a calendar

An evidence plan needs at least two calendars. One shows when the programme expects change to become observable. The other shows when decisions will be made. Comparing them exposes requests that are impossible or unhelpful.

A six-month programme cannot honestly offer a twelve-month retention outcome at closure. It can report a shorter observation and arrange an appropriately authorised later follow-up. An early report can explain that sustained employment has not yet been observed. Moving a future outcome into the present tense does not solve the timing problem.

Similarly, a baseline taken after several workshops may contain effects of the support already delivered. If it is the earliest feasible measurement, retain it and describe its timing. Avoid calling it an untouched starting point. An evidence calendar should therefore include the intervention start, each measurement window, the expected analysis date and the decision date. This gives the delivery team a chance to correct the plan before the evidence is needed.

At this stage, a short record is sufficient. For the scheduling decision, it might say that the delivery lead will review attendance after the third workshop, speak with participants who could not attend through an agreed contact process and decide before the next venue booking. The record should say what information will be considered and how the decision will be documented. It need not prescribe the answer in advance.

A small set of questions that belongs to someone

Evidence requests accumulate easily. A partner asks for an extra breakdown. A manager adds a question in case it becomes useful. A funder requires a measure that the programme already tracks in a slightly different form. Without an owner, the collection grows while the use remains uncertain.

For every proposed item, ask who will read it and what that person could do differently. Some information is required for accountability even when it does not change an immediate decision. Record that purpose explicitly. Other information may be valuable for understanding the programme over time. It should still have an intended use, a reasonable collection method and a review date.

The programme can then separate essential records from optional enquiry. Attendance and the agreed outcome measure may be essential. A detailed account of household circumstances may be unnecessary, intrusive or beyond the team's ability to protect. Removing an unjustified question improves the plan. It reduces the participant's burden and gives staff more time to obtain the information that matters.

Before building a form, write a one-paragraph evidence brief: the decision, the questions, the population, the period, the likely limitation and the person responsible for use. If the paragraph cannot explain why the information is needed, the form is premature.

2 Describe how change is expected to happen

A programme needs an account of how its work could lead to the change it seeks. The account should be detailed enough to examine and simple enough for the delivery team to use. A chain of broad nouns such as training, empowerment and prosperity leaves too much unexplained. It does not show what participants are expected to do differently or which conditions must be present.

In the continuing example, eight practical workshops take place from April to June 2025. Wider employment support runs from 1 April to 30 September. Participants enrol because they are seeking paid work or more stable paid work. Some already have work at the starting point. This eligibility choice matters. A programme open only to people without paid work would have a different population and a different interpretation of later change.

The proposed chain begins with support that participants can actually access. They attend enough of the workshops to practise the relevant tasks and receive useful feedback. They then use some of that preparation when seeking suitable work. Employers or customers must have opportunities, and participants must be able to take them. Paid-work status may subsequently change. Each step involves conditions that the programme should make visible.

Activities and the records of delivery

An activity describes something the programme does. A workshop is an activity. A record that eight workshops were delivered tells the reader that those events occurred. It does not establish that all intended participants attended, understood the material or could use it later.

An output describes a product or service delivered, or another immediate result under the framework being used. For this book, workshop completion is treated as a delivery output. The completion rule requires attendance at at least six of the eight workshops and submission of the specified practice exercise by 30 June 2025. The exercise is submitted; its

submission alone does not demonstrate proficiency. If competence matters to a decision, the team needs an assessment with an explicit standard.

Terminology varies between institutions. The OECD glossary allows outputs to include some activity-produced changes in knowledge, skills or abilities. A different programme framework may place a similar change under an immediate outcome. The practical task is to agree the convention, describe the actual change and use the terms consistently. Moving an item to a more impressive part of a diagram does not make the evidence stronger. [2]

The programme's delivery record ultimately shows 96 completers out of 120 enrolled. That is 80 per cent under the stated rule. It is a useful account of receipt of the planned service. It leaves open whether the service was well suited to participants, whether the practice was useful and what happened afterwards.

Outcomes and the conditions around them

In this book, an outcome is a change in a participant's behaviour or circumstances that the programme seeks to influence. Paid-work status is one outcome in the example. It is deliberately narrow. It does not describe earnings, safety, working hours, stability or whether the work matches a participant's aims.

The team could choose additional measures if those matters are central to the programme. Each addition would require an explanation of purpose and a proportionate method. A programme concerned with stable livelihoods should be cautious about relying entirely on a single week of any paid work. Such a measure can show entry into work while missing whether that work continues or provides adequate income.

Longer-term changes may involve household security, organisational capacity or a wider labour market. They can be important to the programme's purpose even when the programme cannot measure or attribute them within six months. Stating an ambition is legitimate. Reporting that ambition as an achieved result requires evidence for the actual claim.

A change map should therefore distinguish what the programme can observe now from what it expects later. It should also identify conditions outside its direct control. Relevant vacancies, transport, caring

responsibilities and local recruitment practices may affect whether preparation leads to paid work. These are possible influences to investigate, not explanations to assert whenever the target is missed.

Examine the links as separate questions

A useful change map turns each connection into a question. Can participants reach the sessions? Do the exercises prepare them for the tasks they face? Do they use that preparation? Are suitable opportunities available? Does the paid-work measure capture a change that participants regard as worthwhile?

This approach prevents a common analytical shortcut. If employment does not increase as hoped, the team may conclude that the training failed. The evidence may instead show that people learned the material but faced a shortage of suitable opportunities. Conversely, an increase in employment does not prove that the teaching was effective. Recruitment conditions might have improved independently.

The distinction changes the next enquiry. If access is the immediate problem, the team examines scheduling and delivery. If practice is weak, it examines the instruction and assessment. If opportunities are scarce, it considers whether its service design fits the situation. The evidence plan can collect a small amount of information at each critical link rather than measuring only the first activity and the final outcome.

The map should include the possibility of unintended effects. Participants may spend time on activities that are less useful than other options available to them. A referral process might exclude people who lack a particular document. Staff incentives might encourage attendance counting without attention to the quality of participation. These possibilities deserve a safe way to be reported even when they sit outside the headline indicator.

Keep the map open to revision

A change map is a provisional explanation. Its value comes from being specific enough to learn from. When findings challenge a link, record what changed in the explanation and why. Keep the earlier version so that a reader can distinguish the original expectation from a later interpretation.

Suppose participants can complete the practice task but describe difficulty finding relevant vacancies. The team might add a question about how

people locate opportunities and examine the relevance of its referral information. It should not erase the original expectation that improved preparation would be sufficient. That expectation is part of what the programme has learned to examine more carefully.

A revision also affects evidence needs. If the programme adds a referral service, a record of workshop completion no longer describes the whole intervention. The team should define the new service, its start date, the intended recipients and the evidence needed to understand its use. Otherwise a later report may attribute results to a programme whose contents changed without a clear record.

3 Define exactly what is being counted

An indicator is a specified observation used to describe a result or condition. Its short label is only the beginning. Two teams can both report an employment rate while counting different people, asking about different periods and applying different criteria. Their percentages may look comparable even when the underlying observations are not.

The continuing example fixes its cohort at 120 adults enrolled by 31 March 2025. Every enrolled participant remains part of that cohort for the outcome measure, including people who do not complete the workshops. Late applicants are recorded separately. A programme may choose a different enrolment rule, but changing the population after seeing results would change the question being answered.

Each participant has a unique programme identifier. The identifier links the enrolment record, workshop record and follow-up response. It allows the programme to count people rather than form submissions and to recognise a participant who attends several sessions. Contact details can be kept separately with access limited to those who need them. An identifier is useful for linking records; it does not by itself make a dataset anonymous.

Specify the numerator and denominator together

The numerator is the count that meets the criterion. The denominator is the population against which that count is expressed. They must refer to compatible units and periods.

For the illustrative paid-work measure, a positive response means that the participant reports at least one hour of employment or self-employment for pay or profit during the fixed reference week. Unpaid volunteering and unpaid household work are excluded. This is an operational choice for the teaching programme. It is not presented as a complete national or international labour-statistics definition. The measure does not establish job quality or sustained employment.

The baseline reference week is 24 to 30 March 2025. Baseline interviews take place on 31 March. The follow-up reference week is 22 to 28

September 2025, with interviews conducted from 29 September to 10 October. Every follow-up interview asks about that same September week. Asking each participant about the seven days before their own interview would introduce a moving observation period.

At the data freeze on 15 October, the programme has valid follow-up responses for 90 participants. Of those, 42 report paid work under the definition. The respondent proportion is therefore 42 divided by 90, or 46.7 per cent. The known-positive share of the enrolled cohort is 42 divided by 120, or 35 per cent. These statements are both arithmetically correct. They answer different questions, and neither supplies the missing status of the other 30 participants.

A report should place the denominator close to the percentage. “Paid work was reported by 42 of 90 respondents” is easier to interpret than “employment reached 46.7 per cent” with the population hidden in an appendix. For small groups, the count also reminds readers that one response can materially affect the percentage.

A definition that can survive a change of staff

An indicator specification should let another competent person reproduce the calculation. It needs the purpose, population, unit, inclusion rule, observation period, source, calculation and treatment of missing or conflicting records. It should say who maintains the definition and when it was last approved.

For completion, the rule has two parts: sufficient attendance and exercise submission by the cutoff. Someone who attends six workshops but submits no exercise does not meet the rule. Someone who submits the exercise but attends five workshops does not meet it either. If the team later wants a measure of partial participation, it can report that separately. Quietly relaxing the completion rule would destroy comparability with the original target.

For paid-work status, the response categories should include yes, no and unknown. A refusal or an unsuccessful contact attempt belongs in the unknown category. A blank cell should not automatically become no. If the survey permits “prefer not to answer”, retain that response distinctly in the restricted working data even if the public table combines missing outcomes for simplicity.

The specification should also define the counting date. An interview received after the data freeze may be valid evidence about the same reference week. Adding it later requires a revised dataset and a revised report, not an unexplained change to the old total. Version control matters because two correct calculations made from different snapshots can otherwise appear to contradict one another.

Preserve the meaning when a measure changes

Some changes to an indicator are necessary. A question may be misunderstood, a service may be redesigned or an agreement may require a different definition. The programme should record the reason, the date and the effect on comparison. Where feasible, a short period of parallel measurement can show how the old and new measures relate.

Consider two separate reporting rounds that each show 80 per cent completion. One might be 96 of 120 enrolled under the six-session-plus-exercise rule. Another might be 80 of 100 people who attended at least one workshop, using attendance at four sessions as the threshold. The percentages match, but the populations and standards differ. Reporting them as stable performance would conceal the change.

A similar problem occurs when an outcome measure switches from all enrolled participants to completers only. Completers are a selected group. They may have more time, fewer access barriers or different prior experience. The programme can report their results if the subgroup is relevant, but it should retain the all-enrolled view and label the subgroup clearly.

A definition also needs a rule for disaggregation. If participants attend in both areas, which area owns their record? The teaching case uses the delivery area assigned at enrolment, with 60 participants in each. A transfer can be recorded as additional information without making the person appear twice in the cohort. If a different assignment is needed for a delivery question, give that table a different label.

4 Establish a starting point that can be compared

A baseline describes the relevant condition before the intervention or at another explicitly stated starting point. Its value depends on whether it can be compared with later evidence. A number collected early is not automatically a useful baseline if the population, question or measurement method changes afterwards.

In the teaching case, all 120 enrolled participants answer the paid-work question on 31 March, using the reference week of 24 to 30 March. Thirty-six report paid work, giving a baseline proportion of 30 per cent. The programme begins on 1 April. The dates keep the starting observation separate from the support period and make the sequence easy to inspect.

The baseline also records which delivery area each participant belongs to. Area A has 24 participants reporting paid work out of 60. Area B has 12 out of 60. Those starting proportions are 40 and 20 per cent. Any later comparison between the areas needs to retain that difference. A higher final proportion in Area A would not by itself show that its delivery was more effective.

Match the people when the question concerns change

Follow-up reaches 90 participants. Among those same 90 people, 24 reported paid work at baseline. Their baseline proportion is 26.7 per cent. At follow-up, 42 of the 90 report paid work, or 46.7 per cent. The observed change for this matched respondent group is 20 percentage points.

Comparing 46.7 per cent among follow-up respondents with 30 per cent among all enrolled participants would mix populations. The difference of 16.7 percentage points would be a difference between two summaries with different membership. It would not describe the change for a fixed group. The report should show the full baseline and the matched comparison for their separate purposes.

The matched records also reveal movement that a net percentage conceals. Eighteen participants report paid work in both weeks. Six report paid work at baseline and none at follow-up. Twenty-four move from no paid work to

paid work. Forty-two report no paid work in either week. These four counts sum to 90.

That pattern changes the questions the team asks. The net increase is 18 people, but 24 entered the measured state and six left it. A programme concerned with sustaining work needs to understand the latter group as well as the new entrants. It should avoid describing all 42 follow-up positives as new placements. Eighteen already met the measure at baseline, and the programme has not established that any job was obtained because of its support.

Decide what else is needed at the start

Baseline collection can become excessive. Teams sometimes ask about every characteristic that might later explain a result. Much of that information is never used, and some of it creates privacy or safeguarding risks. The starting record should be shaped by the actual questions and the capacity to manage the information responsibly.

For the teaching programme, area and baseline paid-work status are sufficient to demonstrate several analytical problems. A real programme might need to understand accessibility requirements, prior experience or barriers relevant to service design. It should explain why each item is necessary, how it will be used and who can see it. Sensitive questions require particular care and the applicable safeguards.

The team should distinguish information needed to provide a service from information needed to evaluate it. A contact number may be needed to arrange a session. That does not automatically justify including the number in an analysis file or sharing it with a funding partner. Data can often be linked through an identifier while the contact record remains separately controlled.

A baseline also needs a collection protocol. The interviewer should use the same question and reference period for everyone. If someone does not understand “pay or profit”, an agreed explanation should be available. Unrecorded interviewer improvisation can create apparent differences that come from the collection method. A small pilot of the question can expose this problem before the full baseline is collected.

Work honestly with an imperfect starting point

Some programmes begin before an evidence plan is in place. Others inherit records collected for a different purpose. The team should first inspect what exists rather than assume that a retrospective question can recreate the past accurately.

An administrative record may provide a useful starting observation if its population, definition and timing are suitable. A retrospective self-report may offer relevant context but carry recall limitations. A first measurement taken during delivery can establish a reference point for later change, provided the report states that some support had already occurred. None of these options should be renamed a pre-programme baseline without explanation.

If no credible baseline exists, the programme can still describe current status and delivery. It can examine mechanisms, compare appropriately selected groups where the design supports that comparison or commission a suitable evaluation. What it cannot do is calculate a defensible before-and-after change by treating missing historical information as zero.

A baseline of zero is sometimes correct for a genuinely new output. No workshops may have been delivered before a new programme starts. It is much less plausible as an automatic starting value for knowledge, organisational capacity or paid work. People have lives and capabilities before enrolment. The absence of a programme record is not evidence that the condition was absent.

Check comparability before calculating change

Before calculating a difference, place the two measurement specifications side by side. Check the people included, the wording, the reference period, the collection method and the coding. Identify any changes and decide whether a direct comparison remains meaningful.

Seasonality may also matter. The teaching case compares a March week with a September week. Those observations tell us about those weeks. They do not control for seasonal recruitment, wider economic changes or other support participants may have received. Matching people improves the description of observed change, but it does not create a counterfactual account of what would have happened without the programme.

5 Set defensible targets

A target becomes useful when everyone can tell what would count as meeting it. That requires a population, a measure, a period and a threshold, together with a plan for dealing with incomplete information. A percentage written beside a broad ambition leaves too much unsettled. People may agree that the programme should improve employment while holding different ideas about whose employment counts, how long it must last and what evidence is sufficient.

The employment-support programme has three targets: 96 participants completing the specified activities, 108 valid follow-up responses and 60 of the 120 enrolled adults in paid work during the follow-up reference week. These targets concern different things. Completion describes participation in a defined part of the programme. Follow-up coverage describes the completeness of the evidence. Paid-work status concerns a participant outcome. Each needs its own judgement.

Give the threshold a clear meaning

The completion threshold is interpretable because completion has a rule: attendance at at least six of eight practical workshops and submission of the specified practice exercise by 30 June 2025. General engagement, enthusiastic participation and attendance at a later event cannot silently substitute for that rule. They may matter for understanding the programme, but they answer different questions.

The paid-work target also needs its full wording. It concerns at least one hour of employment or self-employment for pay or profit during 22 to 28 September 2025. Its denominator is the 120 adults enrolled by 31 March, including people who did not complete the workshops. The target is 60 people meeting that definition, equivalent to half the cohort. The programme seeks paid work or more stable paid work, but this particular indicator establishes neither stability nor quality. That limitation belongs beside the target when it is chosen, while there is still time to decide whether another measure is needed.

The baseline helps explain the ambition. Thirty-six participants reported paid work in the baseline week, so reaching 60 would put the full-cohort count 24 above its starting level. This difference would be a net change in

status across the cohort. It would not establish that 24 jobs had been created or that precisely 24 participants had moved into work. Some people could enter paid work while others left it. The target should therefore be described in the language of the measure, without allowing an attractive interpretation to become part of its definition.

Record why the target is plausible

A defensible threshold needs a reason beyond being easy to remember. Relevant starting points include previous delivery records, evidence from sufficiently comparable programmes, the starting circumstances of the intended participants and the opportunities available during the programme period. Each source needs examination. An earlier programme serving people already close to employment may offer little guidance for a cohort facing different barriers. A strong labour market in a previous year may not continue.

The planning record should say which evidence influenced the threshold, where comparability is weak and which assumptions remain uncertain. Where the evidence is thin, a target can still express an ambition, but it should not be presented as a well-supported forecast. Recording the distinction gives later readers a fair basis for interpreting the result. It also allows the next programme to learn from how the original expectation was formed.

Targets may need different degrees of stretch. Staff can directly organise workshops and establish practical arrangements for follow-up, although participant circumstances still affect both. Paid work also depends on employers, available opportunities and events outside the programme. The desired outcome should remain important, while the plan recognises how much influence the programme can reasonably exercise. Otherwise pressure to reach the number may encourage convenient reporting or selection of participants who already have fewer obstacles.

Keep the evidence target separate

The follow-up target of 108 valid responses represents 90 per cent of the cohort. It is intended to make the outcome evidence more complete. Reaching it would still leave some uncertainty about the full-cohort result if the remaining outcomes could change the target judgement. Missing it

does not by itself establish poor employment outcomes. These are separate questions even though the quality of one answer depends on the other.

At the data freeze, the programme has 96 completions, 90 valid follow-up responses and 42 reported positive paid-work outcomes. The resulting judgements are:

Target	Recorded position	Judgement at the freeze
96 completions	96 meet the completion rule	Achieved
108 valid follow-up responses	90 valid responses	Missed
60 of all 120 participants in paid work	42 known positive and 30 unknown	Unresolved

The last judgement requires care. If the accepted self-reports are accurate, the 42 positive responses establish a minimum of 42 people meeting the outcome definition. If all 30 unknown outcomes were positive, the count would be 72. The feasible count therefore runs from 42 to 72, or 35 to 60 per cent of the full cohort. The target of 60 lies within that range. These endpoints express what the missing records permit; they are not a confidence interval or a probability statement about the true count.

Reporting the target as missed because only 42 positives have been recorded would replace the original outcome target with a documented-positive target. Applying the respondent rate to everyone would introduce an unsupported assumption about the missing group. Here it would produce an estimated count of 56, which still would not establish whether the target was met. Neither replacing the target nor substituting an unsupported estimate resolves the missing outcomes. The programme can report that attainment remains unknown and identify incomplete follow-up as the immediate evidence problem.

Preserve the original commitment

Targets sometimes need revision when the programme itself changes. A revised target should have a date, a reason and a clear statement of which participants or future period it covers. Retaining the original expectation makes the change visible. Replacing it after seeing the results removes useful information about the quality of the original plan.

6 Collect only what the decision needs

Every question places a small demand on the person answering it. The demand grows when a form repeats information, uses unfamiliar language or asks for details whose purpose is unclear. Before adding a field, consider the decision it could change and how the answer would be used. A collection plan becomes more manageable when each item has a defined purpose and someone knows what to do with it. The CDC framework similarly connects questions, evidence and use, while asking teams to consider language, context and collection burden. Its public-health guidance offers useful principles for other settings. [3]

For this programme, the immediate outcome question is whether each enrolled adult met the paid-work definition during the same specified week. Answering it requires an eligible participant record, an interpretable response and a reliable link to that participant's baseline. It does not require a complete employment history. Further information may be valuable, but its value should be connected to a separate question, such as whether working hours became more predictable.

Build the question around the reference period

The follow-up interviews take place from 29 September to 10 October 2025, while the reference week is 22 to 28 September. Keeping these dates distinct is essential. Asking whether someone is working “now” would allow an interview conducted on 10 October to describe a different period from one conducted on 29 September. The response could change even if both interviewers followed the same script.

The question should name the reference week and explain what counts. Employment and self-employment for pay or profit are included when they involve at least one hour during that week. Unpaid volunteering and unpaid household work are excluded. An interviewer needs guidance for an uncertain response, including when to seek clarification and when to record that no valid answer was obtained. Guessing from a person's occupation or an earlier conversation would introduce information that the measure does not authorise.

A useful collection script separates the factual question from praise or encouragement. Asking whether the programme “helped you get back into work” combines employment status with an explanation for it. Someone might have worked during the week while attributing that work entirely to an existing contact. Another person might value the programme but remain without paid work. Neutral wording lets both give an accurate answer without feeling that they are evaluating the staff member conducting the interview.

Test the wording before the main collection begins. Ask people to explain what they understood, rather than merely whether the question seemed clear. Ambiguity about casual work, self-employment or the dates can then be addressed in the instructions. Keep a record of wording changes so that the final question used at baseline and follow-up can be compared.

Design the record for linking and checking

A stable participant identifier allows the baseline and follow-up records to be joined without relying on a name typed consistently each time. The analysis file can use that identifier while contact details are held separately with appropriately limited access. This arrangement also reduces the temptation to place personal information into every working copy of the dataset.

The minimum record needs enough context to distinguish a valid response from an unsuccessful contact attempt. These fields serve different purposes:

Field	What it allows the programme to check
Participant identifier	Membership of the enrolled cohort and linkage over time
Reference period	Whether the answer concerns the intended week
Paid-work response and validity	Whether an outcome can be classified
Collection date and method	When and how the evidence was obtained
Contact outcome	Why an attempt did not produce a valid response

The exact structure can vary, but missing values need a consistent meaning. A person who declines the question, someone who cannot be

reached and an interview that has not yet been attempted should not all disappear into an unexplained blank. These distinctions guide collection work and later help readers understand its limitations. They do not establish the person's employment status.

Free-text notes should remain purposeful. A short explanation of an ambiguous response may be useful for resolving it. Detailed personal narratives can introduce information that the analysis does not need and that the participant did not expect to circulate. The collection plan should explain the purpose of the questions, how information will be used and the practical choices available to participants, in language suitable for the setting. UNEG guidance connects understandable, voluntary participation with responsible data handling. [6] Before collection, specify who can access identifiable records, what will be shared, how long records are needed and how that period will be reviewed under applicable requirements. Do not promise anonymity when identifiers can reconnect a response to a person. Explain any practical limits on withdrawing information after it has been aggregated or published.

Make follow-up accessible to the whole cohort

The contact list must begin with all 120 enrolled adults. Workshop completion is a characteristic to analyse, not an eligibility test for being contacted. If collection takes place only at a final workshop or celebration event, it will tend to reach people who remain connected to the programme. The method would then shape the evidence before any calculation began.

Plan how to reach people whose circumstances make a standard call difficult. Suitable contact times, understandable wording and accessible methods can make a difference. Where more than one method is used, preserve the same reference period and outcome definition. Record the method so that a suspicious pattern can be checked later. A short telephone interview and an online response can contribute to the same measure if the questions retain their meaning.

A contact protocol also needs boundaries. Repeated approaches should be proportionate and respect a person's stated preferences. Staff should know when another attempt is appropriate, how a refusal is recorded and when collection closes. This protects participants from persistent requests while

preventing arbitrary differences between staff members' follow-up practices.

Leave room for useful explanations

A compact outcome questionnaire need not exclude thoughtful enquiry. If the programme needs to understand why participants stopped attending, a separate, carefully chosen conversation may provide more useful evidence than adding a long compulsory section to every follow-up. Its purpose, selection approach and analytical use should be clear. Views from people willing to talk can illuminate difficulties without automatically representing everyone who left.

Similarly, a paid-work response cannot carry the full burden of assessing an employment programme. If the decision concerns job stability, the evidence plan needs an appropriate way to examine stability. If it concerns accessibility, it needs evidence about access. These additions should be made because the programme faces those decisions, with enough time to collect meaningful information.

7 Check the record before interpreting it

A result can be calculated correctly from a file that represents the programme incorrectly. Duplicate records, mismatched identifiers and misunderstood dates may produce plausible totals. Begin by establishing what the file contains and how it was assembled. Interpretation becomes safer when another person could reproduce the eligible population and trace each reported count to its underlying records. The UK Government Data Quality Framework distinguishes such matters as completeness, accuracy and timeliness. These are separate qualities: a fully populated file can still be wrong, and an accurate record can arrive too late for its intended decision. [4]

The programme's enrolment cut-off is 31 March 2025. Its cohort contains 120 adults, and the data freeze is 15 October. These dates define the analysis version. A late correction may be legitimate, but it should create a documented change rather than an invisible replacement of the evidence used in an earlier report. Preserve the source records and record what was corrected, why and when.

Reconcile the cohort before the outcomes

The first check concerns people. Each enrolled adult should have one cohort identifier, and each analysis record should link to an eligible person. A repeat contact does not create another participant. A response from someone outside the agreed cohort should not enter the outcome total simply because it is complete. Where a name or identifier differs across files, resolve the match through an agreed process rather than choosing the version that makes the totals align.

Next reconcile the overall counts with the programme's component areas. Areas A and B each enrolled 60 adults. Their completions, 54 and 42, sum to 96. Their valid follow-ups, 54 and 36, sum to 90. The missing follow-ups, 6 and 24, sum to 30. Agreement among these totals is a useful check, although it does not prove that every individual record is correct.

Record checks should also test logical relationships. A person cannot have a valid binary paid-work response and simultaneously be classified as

having no valid outcome in the same analysis version. Completion requires both attendance and the practice exercise by the specified deadline. A total that happens to match the target offers no reason to skip checking those conditions.

Check the dates and meanings

The baseline interview date is 31 March, but its paid-work reference period is 24 to 30 March. The follow-up interview window extends into October, but its reference period remains 22 to 28 September. A record containing only an interview date leaves room for confusion when someone later tries to reconstruct the measure. The dataset or its accompanying specification must preserve both concepts.

Validity also depends on meaning. A positive outcome requires a response that meets the agreed definition. A note saying that a participant “has a placement” may describe paid employment, unpaid experience or an arrangement that starts after the reference week. It needs clarification before classification. A checked record should not appear more certain than its source.

Where clarification is no longer possible, retain the uncertainty. Coding an ambiguous answer as negative may make the file easier to analyse, but it converts a data problem into a claim about the participant. Coding it as positive makes the opposite unsupported claim. A visible unknown is more useful than a tidy classification that cannot be defended.

Reconstruct the paired records

All 90 follow-up respondents have a valid baseline and follow-up work response. Their transitions provide an important reconciliation:

Baseline status	Follow-up paid work	Follow-up no paid work	Total
Paid work	18	6	24
No paid work	24	42	66
Total	42	48	90

The row totals reproduce the baseline position of the followed-up group, and the columns reproduce its follow-up position. The table also prevents a

misleading interpretation of the 42 positive outcomes. Eighteen respondents reported paid work at both points, while 24 moved from no paid work to paid work. Six moved in the other direction. The net increase is therefore 18 within this observed group.

A frequent mistake would be to compare the baseline count of 36 for the full cohort with the follow-up count of 42 for respondents and call the difference the programme's change. Those counts concern different sets of people. For the 90 paired records, the correct baseline count is 24. The other 12 baseline positive responses belong to people whose follow-up outcomes are missing.

Calculations must follow the population named in the sentence. Among respondents, paid-work status rose from 24 of 90 to 42 of 90, equivalent to 26.7 to 46.7 per cent after rounding. That is a 20 percentage-point rise among the people observed at both times. A statement about the full cohort needs to account for the 30 missing follow-up outcomes instead.

Examine where the gaps sit

An overall response rate of 75 per cent can conceal substantial differences. Area A has valid follow-up for 54 of 60 participants, or 90 per cent. Area B has 36 of 60, or 60 per cent. The weaker coverage in Area B matters when its outcome rate is placed beside Area A's. A comparison of observed rates should retain this information rather than giving both estimates an appearance of equal completeness.

Completion shows another difference in coverage. Eighty of the 96 completers have valid follow-up, compared with 10 of the 24 non-completers. The corresponding coverage rates are approximately 83.3 and 41.7 per cent. These records do not explain why contact was less successful among non-completers, but they establish that the observed group contains a larger share of completers than the original cohort.

The check should lead to a usable account of the file: who is included, which measures are valid, what has been corrected and what remains unknown. Maintain calculations at adequate precision and round only for presentation. Label denominators wherever their meaning could be confused. Keep a record of the final rules used so that the next report can be compared with this one.

8 Understand what monitoring can establish

Monitoring allows a programme to describe delivery and follow developments in the people it serves. In this case it establishes that 96 participants met the completion rule, 90 provided valid follow-up responses and 42 respondents reported paid work during the specified week. Those statements are useful. They describe different aspects of performance and reveal where further enquiry is needed.

The records also show change among the 90 people observed twice. Their paid-work count rose from 24 to 42, while the transition table shows movement in both directions. This is stronger descriptive evidence than a single end-of-programme total because it preserves the starting position of the same people. The remaining question is what conclusions can reasonably follow from that observed change.

Keep the descriptive claim intact

The clearest statement gives the population, measure and period together: among the 90 participants with valid responses at both points, the proportion reporting paid work increased by 20 percentage points between the two reference weeks. A reader can then distinguish this finding from a claim about everyone enrolled or about employment sustained throughout the programme.

The phrase “at both points” also requires care. Eighteen people reported paid work in both reference weeks. The records do not establish continuous employment in the months between them. Someone could have changed jobs, lost work and returned, or worked intermittently. The binary observations provide two positions in time, while the path between them remains unrecorded.

Likewise, the 24 transitions into paid work do not reveal whether the work offered sufficient hours, reliable pay or improved conditions. These may be important concerns for the programme's wider purpose. They require evidence suited to those concerns. Expanding the meaning of a positive response after collection would make the indicator promise more than the question asked.

Ask what else could explain the pattern

The programme took place during the same months in which participants continued their lives. Employers could recruit or reduce staff. Participants could receive other support, use existing networks or make independent changes to their search. Such possibilities need evidence if they are to become explanations, but their plausibility is enough to prevent a simple before-and-after change from establishing the programme's causal effect.

The baseline is a description of the earlier period. It does not show what the same people would have experienced by September without the programme. Treating the baseline as that alternative would assume that nothing else affecting paid work changed. The monitoring records provide no basis for making that assumption.

A comparison between completers and non-completers would not automatically solve the problem. Completion may be connected to circumstances that also influence employment and willingness to respond. Within the observed records, 38 of 80 completers and 4 of 10 non-completers reported paid work. The small observed non-completer group and the unequal follow-up coverage deserve attention before anyone interprets these figures. The groups were not created as equivalent alternatives for estimating programme effects.

Area comparisons need similar care. Differences in starting conditions, available opportunities, delivery and follow-up can all matter. The record can reveal a question worth investigating, such as why coverage differs so much between the areas. It cannot decide among these explanations merely because the areas belong to the same programme.

Match causal questions to an enquiry designed for them

The Magenta Book recognises experimental, quasi-experimental and theory-based approaches to impact evaluation. The appropriate choice depends on the question and circumstances; a randomised trial is not always required. Theory-based work can examine causal contribution by testing the proposed mechanisms against evidence and considering alternative explanations. Its strength depends on the design and quality of those tests. A favourable trend or testimonial on its own does not perform that work. [1]

A quantitative estimate of causal impact needs a credible account of what would have happened without the programme. With a suitably authorised random allocation, comparing assigned groups can estimate the effect of the offer. Follow-up loss, take-up and effects spreading between groups still need attention. Comparing only completers would discard the original basis for comparability. The setting, ethics and service obligations also determine whether allocation is appropriate. [8]

Non-randomised designs need explicit assumptions. Comparing changes in participating and non-participating groups requires a basis for expecting comparable trends without the programme. Earlier observations can help examine that expectation. Matching on recorded characteristics cannot ensure that unrecorded differences have disappeared. Estimates also remain tied to a population and period; results under intensive pilot support may not transfer directly to ordinary delivery. These questions warrant specialist advice proportionate to the intended claim.

For this programme, a mechanism worth examining might concern how practice activities affect the way participants approach employment opportunities. A useful enquiry would specify what should be observable if that explanation were correct, when the expected changes should occur and what findings would weaken it. It would also ask how other support or changing opportunities could produce a similar pattern.

This changes the purpose of collection. A participant's account can help describe a sequence of events, but the enquiry would seek enough detail to examine that sequence rather than collecting praise. Did the relevant activity occur before the reported change? What did the person do differently? Is there evidence of that action beyond a general statement that the programme was helpful? What happened in cases where the activity occurred but the expected change did not follow?

The resulting account may support a carefully bounded contribution claim. It would still need to state which explanation was tested, what evidence supported it and what uncertainty remained. A precise numerical effect would require an appropriate basis for that additional claim. The scale of the conclusion should follow the evidence assembled.

Use limited findings for proportionate decisions

A programme does not need to resolve every causal question before improving an operational process. The follow-up records already show that

the evidence is substantially less complete in Area B and among non-completers. That is a reason to examine contact arrangements and accessibility. It is not evidence that either group experienced worse employment outcomes than the records reveal.

Similarly, the completion target has been achieved under its stated definition. The organisation can recognise that achievement while asking whether completion captures enough about the learning experience. A future collection plan might examine the usefulness of the practice exercise if that would inform a real delivery decision. Such an addition should not retrospectively alter the completion result already reported.

Larger decisions require broader support. Expanding a programme because employment improved among respondents involves assumptions about the missing participants, causal contribution and whether delivery would work similarly at a larger scale. Those assumptions should be made visible, with further evidence sought where the decision depends on them. Urgency may limit what can be learned before acting, but it does not remove the need to describe the uncertainty.

9 Read mixed and missing results

The employment-support case presents a mixed account: completion reached its target, follow-up fell short, and the paid-work target remains unresolved. The interpretation should explain this combination rather than force it into one overall verdict. A manager can acknowledge delivery while taking the evidence gap seriously and examining what the observed outcomes mean.

The task becomes harder when results differ between areas or groups. A difference may identify a genuine problem, a different starting position or a consequence of who was reached. Interpretation needs to work through these possibilities with the records available.

Keep the different results in view

A single overall rating would conceal where the uncertainty lies. Calling the programme successful because it achieved 96 completions would substitute a delivery result for the employment result. Calling it unsuccessful because only 42 paid-work responses are recorded would treat the unknown outcomes as negative. Calling the evidence inconclusive in every respect would discard the completion finding and the information supplied by respondents.

The more useful reading preserves these distinctions. Delivery achieved its specified completion count. Outcome collection fell short of its coverage target. The employment target remains unresolved. This gives a programme manager several different matters to address rather than a broad judgement that explains little.

Area comparisons need similar care. Area A recorded 54 completions among 60 enrolled participants; Area B recorded 42 among 60. Follow-up coverage was also different: 54 participants in Area A and 36 in Area B. The areas started from different employment positions, with 24 and 12 participants respectively reporting paid work at baseline. A higher observed outcome percentage in one area could reflect several features of these differing groups. It is a reason to investigate delivery and context, rather than sufficient grounds for copying one area's approach wholesale.

Read qualitative evidence with the same care

A small set of feedback records can help explain a pattern without establishing how common an experience is. Consider a separate coding exercise in which six records have been selected to explore contrasting experiences of access to a service. Selection is deliberate: the team wants variation, including accounts that challenge its expectation that session timing is the main problem.

ID	Account retained in the note	Initial topic
F01	Session ends after the last available bus	Travel and timing
F02	An early work shift cannot be changed	Timing
F03	Timing is suitable but instructions are difficult to follow	Instructions
F04	A later start would ease a caring responsibility	Timing
F05	In-person help is useful and online delivery would be difficult	Delivery method
F06	Reports being able to attend at the offered times without difficulty	Contrary account

The table contains short paraphrases for the exercise. In real work, keep the authorised source account and enough context to avoid changing its meaning. Code F01 under both travel and timing because the difficulty depends on their interaction. Do not force every record into one category simply to make the totals add to six. A second reader can examine the coding and record disagreements. The purpose is to expose interpretive choices, not to claim that agreement between readers makes the account true.

The team can then write a modest theme: some selected accounts connect access to the relationship between session times and other obligations. F03 and F06 prevent that observation from becoming a claim that everyone has a timing problem. F05 also warns that changing the delivery method could create a different difficulty. These contrary accounts deserve space in the analysis even if they complicate the preferred recommendation.

It would be misleading to report that half of all participants face scheduling barriers. The six records were selected for explanation, and the selection does not support a population percentage. A useful report states

how the records were chosen, describes the theme and its exceptions, and explains what further enquiry the theme suggests. Here, the next step could be to test the feasibility of an alternative session time while preserving access to in-person help. The feedback supports that enquiry; it does not establish that the adjustment will improve employment outcomes.

Examine who remains unknown

Missingness is partly visible because baseline information exists for all 120 participants. Of the 36 reporting paid work at baseline, 12 lack follow-up responses, a missing rate of 33.3 per cent. Among the 84 without baseline paid work, 18 are missing, or 21.4 per cent. This difference flags a response pattern worth examining. It does not establish whether missing participants were more or less likely to report paid work at follow-up. J-PAL discusses how selective attrition can bias an effect estimate; the wider lesson here is to examine who is missing before treating respondents as representative of the cohort. [5]

One tempting explanation is that participants already working became too busy to answer. Another is that some experienced disrupted work and disengaged. Both fit possibilities in the records; neither is demonstrated by them. Choosing the reassuring explanation would turn an untested story into a correction to the data.

Completion status provides another useful view. Follow-up reached 80 of 96 completers and ten of 24 non-completers. The remaining 16 completers and 14 non-completers have unknown outcomes. Restricting the report to completers would therefore remove a group with particularly limited coverage. It would also change the original follow-up population, which includes everyone enrolled regardless of completion.

A practical investigation would check contact records, permitted contact methods and reasons already documented for unsuccessful contact. Any renewed enquiry should ask about the original reference week, 22 to 28 September 2025. A response about work today cannot silently replace the missing historical response. Where recall has become unreliable, that limitation should remain visible rather than being resolved through confident coding.

Use bounds to understand the decision

On the same assumption that accepted self-reports are accurate, the 42 known positive outcomes establish a lower bound of 42 participants meeting the paid-work definition, or 35 per cent of the enrolled cohort. If all 30 unknown outcomes met the paid-work definition, the upper bound would be 72, or 60 per cent. These are logical bounds based on the missing records. They are neither forecasts nor confidence intervals, and values within them are not presented as equally likely.

The target of 60 lies within those bounds. Reaching it would require at least 18 of the 30 unknown outcomes to be positive. Conversely, if 13 of those unknown outcomes were established as negative, the maximum possible total would fall to 59. These thresholds explain what additional information could settle the target question without predicting what that information will show.

Resolving the target would still leave other purposes for follow-up. Stopping contact as soon as the employment target could be classified as reached would weaken understanding of the cohort and leave the coverage target unaddressed. A sound collection plan follows its agreed population and stopping rules. The bounds help explain what remains uncertain while that plan is completed; they do not provide a reason to seek only the answers that improve the headline.

10 Report what the evidence supports

A useful report lets a reader move from a short statement to the records and reasoning behind it. This remains important when the statement is favourable. Readers should be able to tell whose experience is described, when it was observed, how much information is missing and which conclusions require further evidence. A report that provides these details clearly can support action without making the reader reconstruct the analysis.

The pressure to simplify often appears during editing. A sentence about reported paid work becomes a sentence about jobs created. A result among respondents becomes a result for all participants. A short reference period becomes sustained employment. These changes can occur without altering a single number. Accurate reporting therefore requires attention to verbs, subjects and time as well as arithmetic.

Give the headline a defensible meaning

Consider the sentence, “The programme placed 42 participants in employment and achieved a 46.7 per cent success rate.” The records do not establish placement by the programme. The percentage refers to 90 respondents, while “success rate” supplies an undefined judgement. It also leaves out the target, the unknown outcomes and the limited paid-work measure.

A corrected executive-style paragraph could read:

“The employment-support programme ran from 1 April to 30 September 2025 for 120 adults enrolled by 31 March 2025. At the 15 October 2025 data freeze, interviews conducted from 29 September to 10 October had established that 42 of 90 respondents reported at least one hour of paid work or self-employment for pay or profit during 22 to 28 September 2025. This represents 46.7 per cent of respondents and a known-positive share of 35 per cent of all enrolled participants. Outcomes remain unknown for 30 participants, so achievement of the full-cohort target of 60 remains unresolved. Follow-up included non-completers. These self-reported observations do not establish sustained employment, job quality

or a causal programme effect. The recommended next action is for the evaluation lead to examine the missing-response pattern and assess whether authorised additional contact can recover reliable responses for the same reference week, then issue a versioned update.”

The paragraph is longer than the original claim because the original hid several questions. In a shorter summary, the population, missingness and key limitation should remain beside the result, with the fuller account readily available. Brevity should come from removing repetition and secondary detail rather than removing facts that change the meaning.

Make comparisons intelligible

A percentage carries its denominator even when the sentence leaves it out. The same apparent percentage can describe very different evidence. In this case, 24 of 60 Area A participants reported paid work at baseline, giving 40 per cent. Four of ten follow-up respondents who did not complete the programme also reported paid work, again giving 40 per cent. The matching percentages say nothing about continuity, equivalence or improvement between these groups.

Less obvious changes can be equally misleading. Among responding completers, 38 of 80 reported paid work, or 47.5 per cent. Among all respondents, the figure was 42 of 90, or 46.7 per cent. A report could move between those percentages with little apparent change while quietly excluding non-completers from the later headline. Readers comparing editions might miss the altered question because the numbers remain close.

When a comparison changes population or coverage, explain the change in the sentence that makes it. If the purpose is to describe change among people observed twice, use their paired records. If the purpose is to judge a target for all enrolled participants, retain the full cohort and its unknown outcomes. Neither percentage should be chosen because it creates the most attractive account.

Rounding also deserves restraint. Counts should normally accompany percentages in small groups, where a single response changes the percentage substantially. Excessive decimal precision can suggest a level of certainty the design does not possess. Conversely, rounding a value towards a target should never determine target achievement; use the underlying counts and the agreed target rule.

Put limitations where they affect interpretation

A limitations section is useful, but it cannot repair an unsupported claim earlier in the report. If a headline describes respondents, identify them there. If employment status is self-reported for one reference week, the outcome description should carry that meaning. If missing records leave a target unresolved, show that status in the target account itself.

This does not require inserting every qualification into every sentence. A clear opening account can establish the measure, cohort and dates, followed by consistent shorthand. The test is whether a reader who sees the summary alone would leave with a materially false impression. Appendices can carry detailed coding rules and tabulations; essential limits belong with the conclusion.

Separate an observation from the interpretation placed upon it. “Follow-up coverage was lower in Area B” is an observed comparison. “The contact approach may have contributed to that difference” is a hypothesis requiring supporting evidence. “The contact approach caused the difference” is a causal claim. Keeping these statements distinct allows a report to contain useful questions without presenting them as established explanations.

Make the report usable after publication

A report should identify its data freeze and provide a route to the underlying analysis appropriate to the reader's access. Versioning matters when late responses or corrections arrive. A revised report should explain which records changed, whether the change affects a conclusion and whether a previously recommended action remains appropriate. Quietly replacing a file can leave different readers acting on incompatible accounts.

The closing recommendation also needs an evidential basis. “Improve follow-up” is too broad to show what would be learned. Examining the missing-response pattern and testing whether reliable historical responses can be recovered addresses the specific uncertainty shown here. It gives a responsible role a defined question without pretending that renewed contact is certain to succeed.

11 Use findings to choose the next action

Evidence rarely removes all uncertainty before a decision is needed. The practical task is to determine which actions the available findings justify, what uncertainty matters for each action and when the decision should be revisited. A weak outcome study may still reveal a concrete collection problem. A strong descriptive result may still provide insufficient grounds for expanding a programme. The required evidence depends partly on the consequence of being wrong.

In the employment-support case, three possible decisions need different treatment: recover missing outcome information, change how support is delivered and expand the existing model. They should not move together merely because they appear in the same programme report. Each has a different purpose and requires a different explanation.

Match the decision to the evidence

The shortfall in follow-up coverage directly supports examining the collection process. Thirty outcomes are unknown, coverage differs by area and non-completers are less well represented. These facts provide a reasonable basis for a limited investigation of contact procedures and participant burden. They do not tell the investigator which contact method will work better, so any change should be checked rather than assumed effective.

Changing the support model requires more information. Area B's lower completion and follow-up counts could reflect delivery difficulties, participant circumstances or several interacting conditions. The areas also had different baseline employment counts. Removing an element of support solely because one area's observed paid-work percentage is lower would skip the work of establishing what happened and why.

Expansion has a further requirement: evidence relevant to expected outcomes under the proposed conditions. The 20 percentage point increase among paired respondents describes an observed change. It does not establish the additional employment attributable to the programme or predict what a larger programme would achieve. A decision to expand

could be made for other stated reasons, but the present records should not be offered as a causal estimate supporting it.

Reversibility helps determine a proportionate response. A small, authorised change to appointment reminders can be observed and revised relatively easily. Removing a form of support that participants rely on may have consequences that are harder to undo. The latter decision warrants more careful investigation of affected participants and plausible harms before implementation.

Specify an action that can be checked

A decision becomes accountable when a reader can identify a responsible role, the evidence supporting the choice, the uncertainty accepted and a condition for reconsidering it. These elements need not become a lengthy form. A short record can make the reasoning clear enough to revisit later.

For the case, an appropriate proposed decision would be to ask the evaluation lead to review the 30 missing follow-up records before making any renewed contact. The basis is the incomplete outcome picture and uneven coverage. The review would establish whether contact is permitted, whether previous attempts followed the agreed procedure and whether the original reference week can still be recalled reliably. Its uncertainty concerns recoverability and possible response differences. The decision would be revisited when that review establishes which records can be pursued and which must remain unknown.

A separate proposed delivery decision could assign the programme manager responsibility for examining why completion differed between Areas A and B. Its basis would be the recorded completion counts, rather than an assumption about employment impact. The enquiry would compare documented access difficulties, actual delivery and participants' accounts. A change in support would be reconsidered when the evidence identifies a specific barrier that the proposed change could reasonably address.

Before an assignment is made, the responsible role needs the authority and capacity to carry out the action. Access to the relevant information matters as much as a title in the record. A proposed enquiry should remain visibly pending until its scope and responsibility are agreed.

Decide what further information is worth collecting

Additional evidence has value when it could change an important choice or improve how that choice is carried out. Recovering missing paid-work responses could settle the target question and improve description of the cohort. Asking every participant a long retrospective questionnaire may add less value if recall is poor and the questions are weakly connected to a decision.

Before collecting more, specify the information that would alter the recommendation. If the question concerns whether participants could attend workshops, attendance records and focused accounts of access barriers are relevant. If it concerns sustained employment, another isolated question about one week's paid work is insufficient. The collection design would need to match the intended meaning of sustained employment and the period over which it is assessed.

The six recorded transitions from paid work to no paid work also warrant care. They suggest that an account focused only on gains would be incomplete. They do not identify six people who lost work because of the programme. An appropriately permitted enquiry could explore changes in work and support needs without leading participants towards that explanation. Participants should be able to decline, and responses should not determine whether they retain access to support.

No outcome collection process should reward staff for obtaining positive answers. Such an incentive would weaken the evidence being sought. The useful objective is a valid response obtained through an appropriate process, including responses that make the programme's results look less favourable.

Keep decisions open to later evidence

A revisit trigger should be specific enough to distinguish new information from routine repetition. For the missing outcomes, a trigger might be completion of the record review or receipt of additional valid responses under the agreed collection procedure. For a delivery change, it could be evidence that the proposed adjustment creates an access problem or fails to address the documented barrier. "Review regularly" leaves the reason for reconsideration unclear.

New information may strengthen one decision while leaving another unchanged. Additional responses could establish whether the target was reached without showing that the programme caused the outcomes. Better accounts of access difficulties could support a workshop adjustment without resolving employment impact. Treating these as separate decisions prevents progress in one area from being used to claim certainty in another.

The decision record should also preserve alternatives considered and the main reason for the chosen course. This need not record every discussion. It should retain enough context to explain why the action was proportionate at the time, including the limitations accepted. When the next cycle begins, that account helps distinguish a reasonable decision made under uncertainty from a conclusion that was never supported by the evidence available.

12 Bring learning into the next proposal

The next proposal is an opportunity to improve the programme's reasoning as well as its description. Evidence from earlier delivery can show what was achieved, what remains uncertain and which assumptions require a better test. Its value is reduced when a writer extracts only the strongest-looking percentage and leaves the conditions behind.

Reuse findings with their conditions attached

Evidence reuse begins with the original question and population. Forty-two participants reported paid work at follow-up; this count includes 18 who also reported paid work at baseline. It cannot be reused as “42 people helped into jobs”. The net increase of 18 among paired respondents cannot become “18 jobs created” either. Both statements would add attribution and change the meaning of the underlying observations.

A concise account for a future proposal might say: “Earlier delivery achieved the completion target of 96 among 120 enrolled adults. Among 90 participants with valid baseline and follow-up responses, the number reporting paid work increased from 24 to 42. Thirty enrolled participants had unknown follow-up outcomes, leaving achievement of the full-cohort employment target unresolved. The next phase would strengthen follow-up across the whole enrolled group and investigate the mechanisms through which support is expected to influence employment.”

That account offers an evidential starting point. It shows the reader why stronger follow-up and a more deliberate enquiry belong in the next design. It does not imply that the proposed changes have been implemented or that their effectiveness is already known.

Reuse also requires checking the fit between past and proposed settings. Eligibility, local employment conditions, support intensity and the period available for outcomes may differ. A finding from one setting can remain informative while being less directly applicable to another. The proposal should identify the differences most likely to affect the expected pathway and explain how they will be examined.

Turn unresolved questions into design choices

Learning becomes visible through a change in what will be done or tested. “We will improve monitoring” says little about the lesson. The case supports a more specific design question: how can follow-up remain feasible for participants who do not complete the workshops? The original follow-up population already included them, but coverage among this group was limited. The next design could examine permitted contact preferences, the burden of the interview and whether outcome contact has been made unnecessarily dependent on workshop attendance.

Any revised process should preserve the purpose of the measure. Making interviews shorter may improve response while omitting information needed to interpret paid work. Contacting participants earlier may improve recall while changing the observation period. A design improvement therefore needs a clear account of what will change, what question it will answer and whether the new results can be compared with the earlier ones.

The transition records suggest further questions about job entry and retention. Those questions may require different evidence. A participant's account of how an application changed can inform enquiry into the support process. A later account of employment continuity can inform understanding of retention. Neither should be collected merely because it would create a persuasive story. Each should have a defined place in the programme's reasoning and an appropriate method of assessment.

Design a stronger test of the explanation

A proposal should explain how planned evidence could challenge the programme's expected pathway. If practical workshops are expected to improve applications, which observable changes would support that mechanism? What evidence would weaken it? What other influences could produce employment changes during the same period? Questions about other support, changing vacancies and participants' prior experience help avoid attributing every favourable movement to the intervention.

For example, the next phase could examine whether participants use the practice exercises when preparing real applications. The proposal would need to say how cases are selected, what evidence of use will be examined and how accounts that challenge the expected mechanism will be included.

Permission to examine application material would need to be obtained, with unnecessary personal information excluded.

The appropriate design follows the question, the setting and the information that can credibly be obtained. If a numerical estimate of additional employment is essential, the proposal needs an approach capable of supporting that estimate and a clear account of its assumptions. If the immediate purpose is to understand whether a specific support mechanism operates under particular conditions, a carefully designed theory-based enquiry may be more suitable. Specialist advice can help establish whether the proposed design supports the strength of claim intended.

A learning plan should allow findings to weaken the original explanation. Evidence that participants did not use the workshop material, that access barriers prevented attendance or that work was obtained through unrelated routes deserves attention. Otherwise, the next phase may produce more information without improving the organisation's ability to judge its own account.

Preserve a useful record for the next cycle

The evidence carried into a proposal should have a traceable source, an observation period and a statement of scope. Keep the original count alongside the percentage and preserve important limitations when findings are shortened. A later reader should be able to recover why the claim was considered reasonable, including any revisions after the original data freeze.

Targets for the next phase require their own rationale. The earlier target of 60 paid-work outcomes remains an aspiration whose achievement is unresolved in the available records. It cannot be treated as an achieved benchmark for forecasting. A new target may draw on earlier observations, relevant external evidence and the proposed delivery conditions, but the assumptions connecting those sources to the target should remain explicit.

The next proposal should show which uncertainty has been reduced, whose outcomes remain poorly represented and what the proposed work will make possible to decide. That account gives the next programme a useful starting point and makes its remaining assumptions available for examination.

Worked tools

Adapt these records to the actual programme and leave unsupported fields unresolved. Do not copy the teaching figures into a real report.

Tool 1 A short evidence brief

Decision: Whether the next cohort should use the same workshop schedule in Area B, offer an alternative schedule or test a smaller adjustment first.

Decision owner: The role responsible for programme delivery. The final record should identify the actual person with authority to make this choice.

Question: Which attendance difficulties are connected to timing, and which arise from other access barriers? The current completion difference is a reason to enquire, not proof that scheduling caused it.

Evidence to inspect: Session attendance linked to enrolment identifiers; the actual dates and times offered; accessible feedback from participants who attended fewer sessions; delivery staff's account of changes or cancellations. Include people who did not complete. Check whether the people who respond to the enquiry differ from those who remain unreachable.

Decision rule: Select an adjustment only if the evidence identifies a plausible, addressable barrier and the proposed arrangement is feasible. If accounts conflict, test the smallest useful adjustment with a defined review rather than treating the preferred explanation as established.

Record of use: Document the chosen arrangement, evidence supporting it, unresolved questions, role responsible for implementation and the next point at which attendance and participant feedback will be reviewed. This is a delivery decision. It does not settle whether the programme caused the paid-work results.

Tool 2 The continuing case at a glance

All participants are adults. Enrolment closes on 31 March 2025. Support runs from 1 April to 30 September. Workshop completion is assessed by 30 June. Follow-up includes every enrolled person, irrespective of completion, and the dataset is frozen on 15 October.

Record	Area A	Area B	Total
Enrolled cohort	60	60	120
Workshop completers	54	42	96
Baseline paid work	24	12	36
Valid follow-up responses	54	36	90
Follow-up paid work	30	12	42
Unknown follow-up status	6	24	30
Baseline paid work among respondents	20	4	24

The 90 follow-up respondents include 80 completers and 10 non-completers. Among those groups, 38 and four respectively report paid work. Sixteen completers and 14 non-completers have unknown follow-up status. These descriptive groups are not equivalent comparison groups for estimating a programme effect.

The baseline week is 24 to 30 March, observed through interviews on 31 March. The follow-up week is 22 to 28 September, observed through interviews from 29 September to 10 October. Both use the same operational paid-work definition from Chapter 3.

Tool 3 A compact data dictionary

Field	Meaning and allowed treatment
Programme identifier	One stable value per enrolled adult; no duplicate enrolments in the analytical cohort
Enrolled cohort	Included by the 31 March cutoff; denominator fixed at 120
Delivery area	Area assigned at enrolment, A or B; later transfers recorded separately
Completion	Yes only if attendance and exercise-submission requirements are both met by 30 June
Baseline paid work	Yes or no for the fixed March week; all 120 have a valid response in this case

Field	Meaning and allowed treatment
Follow-up paid work	Yes, no or unknown for the fixed September week; never infer no from a failed contact
Response date	Date of interview, kept separate from the reference week
Record version	Data-freeze date and revision identifier; preserve a correction history

A real dictionary should add the question wording, coding instructions, data source, responsible role and permitted access. WHO and UNICEF provide a detailed example of this metadata approach in their health-indicator technical specifications; the fields are instructive without making those health measures applicable here. [7] Health, identity, contact and other sensitive or identifying fields require their own justification and safeguards. The presence of a field in an existing form does not establish a need to reuse it.

Tool 4 Reconcile a finding before reporting it

The paired records contain 18 people with paid work in both weeks, six moving from paid work to no paid work, 24 moving into paid work and 42 with no paid work in either week. Their sum is 90. Baseline positives are 18 plus six, or 24. Follow-up positives are 18 plus 24, or 42. The net difference is 18, giving a 20-percentage-point increase within the matched group.

For the full cohort, start with the 42 observed positives. If none of the 30 unknown outcomes is positive, the count remains 42. If all are positive, it becomes 72. The resulting 35 to 60 per cent range is a statement of simple possible limits under the binary definition and accepted observed responses. It is not a confidence interval or a probability statement.

The actual-status target is 60 of 120. At least 18 unknown outcomes would have to be positive for it to be met. The current record cannot settle that question. A report should carry the target as unresolved, alongside the known-positive count and missing count.

Tool 5 Record the action and its limits

Finding: Valid follow-up is available for 90 of 120 participants, against the planned 108. Area B accounts for 24 of the 30 unknown outcomes.

Decision: Examine the follow-up process before drawing a conclusion about relative area performance. Review contact attempts and the accessibility of the agreed contact method. Any further contact must respect the participant information, consent arrangements and applicable requirements.

Reason: The missing information is concentrated, and the respondent groups begin from different paid-work levels. A simple ranking would overstate what the records establish.

Responsible role and review: Assign the evidence lead to document the process review and return with a recommendation before the next cohort's collection arrangements are fixed. Set a real date when adapting this record. If new valid responses are obtained, preserve the original snapshot and issue a dated revision. If no further information is available, retain the uncertainty in subsequent reports.

Source notes and further reading

The numerical case and worked decisions are original teaching material. The sources below support selected terminology and methodological distinctions. They do not endorse the book or establish a universal requirement for every programme. Publication details and linked official materials were checked on 5 October 2026.

[1] HM Treasury and Evaluation Task Force. Magenta Book Central Government Guidance on Evaluation. Updated 15 May 2026. Sections 1.9, 2.2, 3.4 and 4.4, and Chapter 6 support proportionate evaluation, explicit causal reasoning, quality and use. This is UK central-government guidance. Its administrative provisions should not be treated as global donor rules. [Official guidance](#).

[2] OECD. Glossary of Key Terms in Evaluation and Results Based Management for Sustainable Development. Second edition, 2023. See the entries for baseline, outputs, outcomes, impacts, target and theory of change. The glossary helps readers compare terminology while retaining the definitions used by their own programme. [Official publication](#).

[3] Kidder and colleagues. CDC Program Evaluation Framework 2024. MMWR Recommendations and Reports 73, number 6, pages 1 to 37. Published 26 September 2024. Steps 4 to 6 address credible evidence, conclusions and action. Its public-health setting should remain visible when applying the wider principles elsewhere. [Official report](#).

[4] UK Government. The Government Data Quality Framework. Published 3 December 2020. Useful for distinguishing completeness, uniqueness, consistency, timeliness, validity and accuracy, and for assessing fitness for a specified use. It is UK public-sector guidance rather than a universal certification scheme. [Official framework](#).

[5] J-PAL. Data Analysis. Updated August 2023. The attrition discussion explains why missing follow-up and differences in who remains observed matter. Its formal analytical methods require the relevant research design and assumptions. The book's simple missing-outcome bounds are arithmetic, not an estimate supplied by J-PAL. [Official resource](#).

[6] United Nations Evaluation Group. Ethical Guidelines for Evaluation. Revised 2020 edition. Pages 9, 18 and 32 address meaningful participation, responsible data and consent. The guidance allows adaptation outside the UN, but does not replace applicable local law or organisational responsibilities. [Official PDF](#).

[7] WHO and UNICEF. Web Annex Technical Specifications for Primary Health Care Measurement Framework and Indicators Monitoring Health Systems through a Primary Health Care Lens. Published 28 February 2022. The annex provides a concrete example of indicator definitions, data sources, disaggregation and, where relevant, numerators and denominators. Its health indicators are not requirements for an employment programme. [Official publication](#).

[8] Gibson and Sautmann. Introduction to Randomized Evaluations. J-PAL, updated April 2023. Useful for counterfactual reasoning, selection differences and the assumptions of evaluation designs. Its emphasis on quantitative impact evaluation should not be read as excluding other rigorous approaches to causal enquiry. [Official resource](#).

[9] OECD. Applying Evaluation Criteria Thoughtfully. Published 15 March 2021. Explains how relevance, coherence, effectiveness, efficiency, impact and sustainability can inform evaluative questions. The criteria are perspectives for judgement; they do not themselves provide an evaluation method or a requirement to use every criterion equally. [Official publication](#).

About the author

Aneel Shahzad is an Institutional Growth & Proposal Strategy Advisor based in Punjab, Pakistan.

Through GrantWriter360, his practice covers opportunity qualification and funding strategy, grant and donor proposals, government RFP and tender response strategy, and monitoring and evaluation frameworks. Related work includes compliance review, budget alignment, submission management and donor frameworks.

grantwriter360.com

Professional enquiries: shahzad_aneel@yahoo.com

WhatsApp: +92 332 6383691